AROMATIC AMINO ACIDS IN PEPTIDES AND PROTEINS: NOVEL SYNTHESES, INFLUENCES ON STRUCTURE, AND THE NATURE OF C–H/π AND S–H/π INTERACTIONS

by

Christina R. Forbes

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Biochemistry

Fall 2016

© 2016 Christina R. Forbes
All Rights Reserved

ProQuest Number: 10241137

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10241137

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

AROMATIC AMINO ACIDS IN PEPTIDES AND PROTEINS:

NOVEL SYNTHESES, INFLUENCES ON STRUCTURE,

AND THE NATURE OF C–H/π AND S–H/π INTERACTIONS

by

Christina R. Forbes

Approved: _____

Murray Johnston, Ph.D.
Chair of the Department of Chemistry & Biochemistry

Approved: _____

George H. Watson, Ph.D.
Dean for the College of Arts and Sciences

Approved: _____

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Neal J. Zondlo, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Colin Thorpe, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

John T. Koh, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Steven J. Metallo, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

It is said that “it takes a village to raise a child,” and doctoral training as a scientist is no exception. Training in scientific techniques and methodology, critical thinking, creativity, entrepreneurism, networking, emotional and mental balance, and physical endurance are all taught from many people over duration of a Ph.D. I express my deepest gratitude to all of those who have been a part of this experience, including those not mentioned here. First and foremost, I thank my adviser, Neal J. Zondlo, for all of his support and inspiration. With his drive and mentorship, and his recognition of my potential, I’ve surpassed my own expectations and achieved more than I realized I was capable of. I also thank all of those faculty members on my committee, who provided guidance for both my career and my research. I thank my labmates, past and present, including Caitlin Tressler, Andrew Urmey, Dr. Himal Ganguly, and Dr. Anil Pandey, who provided intellectual and emotional support throughout my studies. I thank our collaborators for their time and contributions to the work described in the next few pages, specficially Dr. Shi Bai, Dr. Glenn P. A. Yap, Dr. Sudipta K. Sinha, and Dr. David J. Craik. I thank Dr. Anne K. Jones for her advice, patience, understanding, and support. I thank my parents, who instilled and cultivated my desire for learning; I never would have achieved this without them. I thank all of the staff in the department (Susan Cheadle, Gary, and Jesse), who shared their support and prayers with me. Finally, I thank my husband, Joseph Holland, for his patience, concern, and understanding; he kept me grounded, and carried me through the hardest days. Finally, it is “someday.”

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In biology, the function of a protein or an enzyme is implicitly defined by its structure. The efficiency of a catalyst is dictated by the defined structure and chemical nature of the molecule. Specific functional groups within proteins and small molecules empart specialized reactivity. Functional groups can also tune and modulate noncovalent interactions in proteins and small molecules, such as salt bridges, hydrogen bonding, and aromatic interactions. Understanding how a functional group influences chemical structure and reactivity is critical for designing de novo peptide-based inhibitors, therapeutics, rationally designed proteins, catalysts, and supramolecular materials.

While Nature has evolved a plethora of biological molecules with specific and reactive functional groups, the 20 natural amino acid side-chains have been limited by natural abundance and development of biosynthesis pathways. We sought to expand the "tool-box" of amino acid side-chains in proteins beyond the canonical amino acids, which provides access to side-chain groups with unique functions and reactivity. A practical methodology was developed for synthesizing peptides containing 4-thiophenylalanine, the sulfur analogue of tyrosine. 4-Thiophenylalanine was synthesized within peptides on solid-phase using a copper-mediated cross-coupling reaction. 4-Thiophenylalanine was subjected to alkylation reactions and oxidation reactions to generate a variety of different functionalized derivatives. 4-Thiophenylalanine or related derivatives can potentially be used for site-specific

labeling in proteins, as spectroscopic probes for sulfur oxidation state, for studying chalogen effects in proteins and peptides, and for modulating peptide structure.

In addition, 4-thiophenylalanine was synthesized as an amino acid monomer, which provided detailed insights into the fundamental nature of S–H/π aromatic interactions. The strength and geometry of the intermolecular S–H/π aromatic interaction in 4-thiophenylalanine was characterized by solution and solid-state NMR, x-ray crystallography, and IR analysis. S–H/π aromatic interactions exhibited a distinctive geometry compared to cation/π interactions in crystal structures from the Cambridge Structural Database: the S–H bond was oriented towards the edge of the face of the aromatic ring rather than the centroid. In addition, the S–H/π interaction distances were often less than the sum of the van der Waals radii for carbon and hydrogen. *ab initio* calculations indicated that the interaction was stabilized by a molecular orbital interaction between the aromatic π and the S–H σ* orbitals.

Noncovalent interactions were explored within model peptides to gain insights into the fundamental nature of C–H/π aromatic interactions. In proteins and peptides, prolines in the cis amide bond conformation are likely to be preceded by an aromatic amino acid. It has been suggested that the aromatic-cis-proline motif is stabilized by a C–H/π aromatic interaction between proline and the aromatic ring. With practical access to a variety of non-natural aromatic amino acids, the nature of the aromatic-cis-proline C–H/π interaction was explored in more than 50 peptides of the model sequence Ac-TXPN-NH_2 (where X = a substituted aromatic amino acid). Aromatic amino acids with electron-donating substituents increased the cis populations of the model peptides, with a significant upfield chemical shift in the proline Hα, suggesting a C–H/π interaction. Selected peptides Ac-TXPN-NH_2 were further examined in

organic solvents and characterized via van′t Hoff analysis in order to understand the enthalpic and entropic contributions to aromatic-cis-proline conformations. Dipeptide motifs were further examined via solution NMR and x-ray crystallography. Extensive, C–H/π aromatic interactions with close contact distances were observed in the crystal structures of the dipeptides. The geometry and contact distances observed in the crystal structures of the aromatic-proline motifs suggested an orbital interaction between the aromatic π and the C–H σ* orbitals. The similar interaction geometry and contact distances between S–H/π and C–H/π aromatic interactions suggests that the $\pi \rightarrow \sigma^*_{X-H}$ interaction is general in weakly polar X–H interactions.

With the unique reactivity in 4-thiophenylalanine, and the increased acidity compared to cysteine or tyrosine, we sought to utilize this aromatic amino acid in the context of native chemical ligation reactions. Native chemical ligation forms a native amide bond with peptides containing N-terminal cysteine and peptides containing C-terminal thioesters. Peptides containing 2-thiophenylalanine at the N-terminus were synthesized in a practical manner using a solid-phase cross-coupling reaction. The peptides containing 2-thiophenylalanine at the N-terminus reacted rapidly with a variety of peptides containing C-terminal thioesters. Native chemical ligation reactions mediated by 2-thiophenylalanine and were further applied for the rapid synthesis of modified proteins. Practical access to this novel amino acid provided unique insights into the mechanism of ligation reactions using aryl thiolated amino acids.

With the development of practical methodologies for synthesizing a variety of aromatic amino acids within peptides, we have synthesized peptides and proteins with novel functionality and reactivity, and have explored the fundamental nature of noncovalent aromatic interactions in peptides and proteins.

Chapter 1

SYNTHESIS OF PEPTIDES CONTAINING 4-THIOPHENYLALANINE AND RELATED SULFUR DERIVATIVES VIA A COPPER-MEDIATED CROSS-COUPLING REACTION ON PEPTIDES ON SOLID-PHASE

1.1 Introduction

Biological processes involve complex pathways that include a concert of enzymes, reactive substrates, and conformational dynamics of proteins. Protein function is driven in part by local interactions and folding motifs that are dictated by the amino acid sequence. By examining the underlying chemical principles that influence local interactions in amino acids, and how these interactions impact the tertiary structure of proteins, we can gain greater insight into protein function and dysfunction. Furthermore, these fundamental principles of protein folding and dynamics enhances our ability to rationally design novel peptides and proteins for synthetic applications. Designed peptides and proteins with specific, defined structural preferences and reactivity have potential applications in therapeutics, materials science, and nanotechnology.

The sulfur-containing amino acids, cysteine and methionine, are relatively uncommon in proteins, but can be crucial for protein structure and function.[1-3] Many biological functions of proteins rely on the diverse and reactive roles of cysteine, which has been known to perform nucleophilic roles in catalysis, undergo post-translational modifications, coordinate metal active sites, form structural disulfides for protein folding, and orchestrate cellular signaling via multiple oxidation states of sulfur.[4-7] Much progress has been made in understanding how cysteine is modified

naturally in proteins, and how these modifications influence cellular signaling,[4, 8] but there is still much to be understood about the more short-lived intermediates of cysteine.[8-10] For example, it is known that cysteine sulfenic acids have dramatically different reactivity compared to cysteine, but how this oxidative modification influences its local interactions in proteins, how it alters cellular signaling events, and how its misregulation can cause diseases have yet to be fully established.[8, 11, 12] In addition, cysteine is also widely used synthetically in biochemistry and materials science for tagging and modifying proteins with fluorophores, small molecules, synthetic protein domains, nanoparticles, and electrode surfaces.[3, 13-15] With growing applications in therapeutics, nanotechnology, and materials science, protein modification and conjugation techniques must be continually improved, including the development of more novel chemical tools for site-selective modification of proteins.

4-Thiophenylalanine, the sulfur analogue of tyrosine, was originally synthesized in 1912 with the anticipation that it might be found naturally in proteins.[16] It was proposed that since serine and amide bonds have natural sulfur-containing isosteres (cysteine and thioamides, respectively), then tyrosine might also have a similar sulfur isostere.[16] Although 4-thiophenylalanine was never found as a naturally occurring amino acid, it introduced the concept that natural biomolecules can be synthesized, mimicked, altered, and engineered. In the way that cysteine is used for tagging and modifying proteins, 4-thiophenylalanine can potentially be incorporated into proteins and peptides for bioconjugation strategies. 4-Thiophenylalanine is the sulfur analogue of tyrosine, and it can therefore be utilized to probe the roles of tyrosine in proteins.[17] Tyrosine is known to participate in non-covalent aromatic interactions in proteins, which can dictate protein folding, substrate recognition, or

enzyme activity,[18, 19] and replacement of tyrosine with 4-thiophenylalanine can potentially strengthen or modulate these roles. However, previous reported syntheses of 4-thiophenylalanine require 6-9 synthetic steps, with approaches that utilize dangerous reagents (i.e. thionyl chloride), or other approaches that generate racemic mixtures of the amino acid product.[16, 20-23] These difficult synthetic strategies have potentially hindered the broader application and characterization of this unique amino acid. Given the recent advancements in transition-metal-mediated cross-coupling methodologies, particularly for C–S bond formation,[24-26] a more practical synthesis of 4-thiophenylalanine was sought via a cross-coupling reaction on commercially available halogenated phenylalanine derivatives.

1.1.1 Context of sulfur in biological systems

The natural sulfur-containing amino acids, methionine and cysteine, were late additions into the genetic code: these amino acids were essentially non-existant in ancient proteins, but have increased in proteomic frequency during the course of evolution.[1, 27, 28] Mutations involving cysteine are more commonly associated with genetic diseases in humans over other amino acids, implicating the biological significance of cysteine in chemical biology, physiology, and evolution.[1, 29] The pK_a of cysteine thiols can vary from 8.5 to 2.0, depending on solvent exposure or stabilizing effects from other amino acid side chains.[1, 5] Cysteine thiolates are important biological nucleophiles and are closely involved in enzyme active-site mechanisms, even with its relatively low abundance in proteins. Cysteine residues have the second highest propensity to be involved in catalytic roles in enzymes after histidine,[30] and catalytic cysteine residues tend to be highly conserved throughout enzyme classes.[1, 31] The specific functions of cysteine in catalytic roles are varied, including proteolysis,

protein ligation, metabolic pathways, and intein-mediated protein splicing.[7, 32, 33] Sulfur readily binds to metal ions such as Fe^{2+}, $Zn^{2+,}$ Cd^{2+}, or Cu^{+}, and cysteine is often associated with coordinating metals within protein active sites.[1] For example, cysteine sulfur atoms are integral components of iron-sulfur clusters, and the positioning of these clusters is crucial for efficient electron-transfer in redox-active enzymes. Cysteine also has the unique ability to undergo many post-translational modifications, including oxidation (sulfenate, sulfinate, sulfonate, and S-nitrosylated), disulfide formation (S-glutathionylation, symmetric cysteine disulfides), acylation (palmitoylation), alkylation (prenylation) as well as some less common enzymatic modifications such as phosphorylation, methylation, and ubquitination (Figure 1.1).[5, 34] These post-translational modifications of cysteine can be both functional and regulatory.

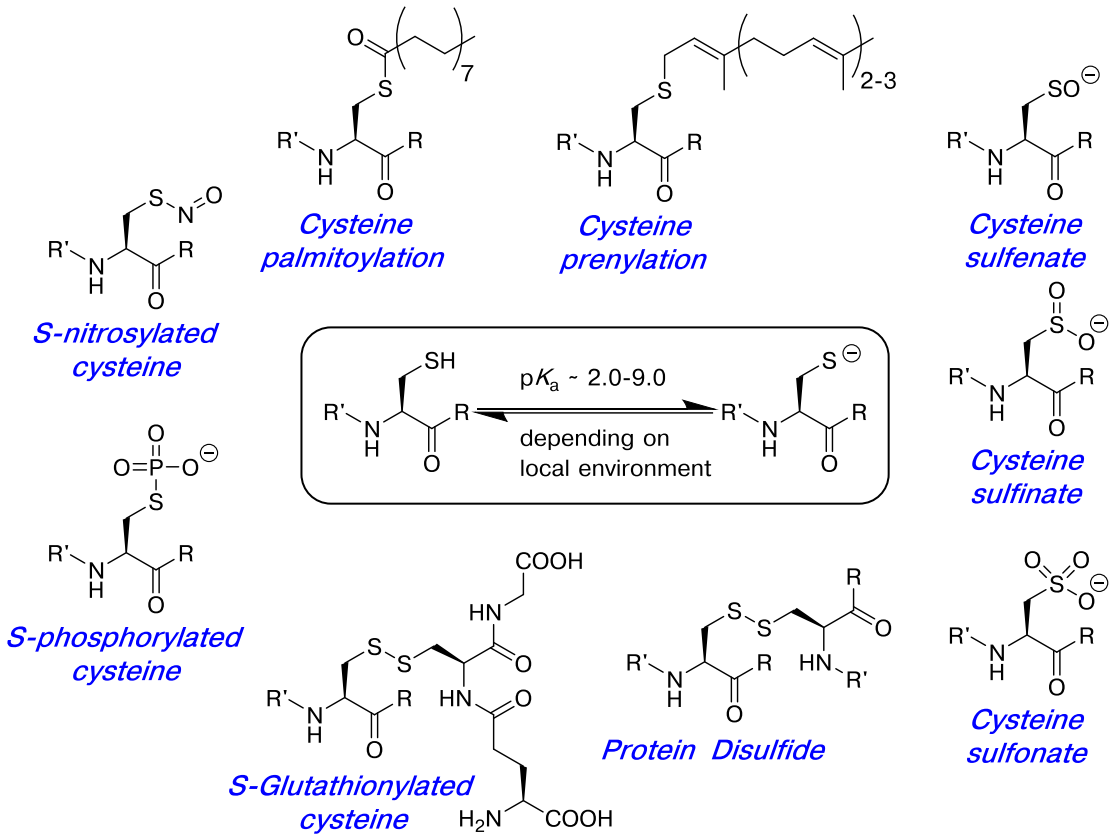


Figure 1.1 Cysteine and some of its biological post-translational modifications Cysteine functions in catalytic and regulatory roles in proteins. Cysteine residues are highly conserved, due to their biological importance in protein structure and function, even though they are relatively uncommon amino acids.

Acyl transfer reactions involving cysteine occur ubiquitously throughout chemical biology (Figure 1.2): cysteine peptidases degrade proteins via acyl transfer through a catalytic cysteine residue;[1, 2, 35] post-translational modifications on cysteine, including palmitoylation and ubquitination, proceed via nucleophilic attack of cysteine to form a thioester;[1, 34] protein splicing and precursor peptide processing rely on acyl transfer reactions, with many examples through cysteine-thioester intermediates;[36-38] the primary function of Coenzyme A, a central cofactor in fatty acid metabolism and

numerous biosynthetic pathways, involves an acyl transfer reaction at cysteine.[39, 40] The diversity of these acyl transfer reactions, and the central role of cysteine in many of these mechanisms, underscores the biological significance of this amino acid. The mechanism of protein splicing through cysteine thiolates inspired a new strategic paradigm for the chemical synthesis of proteins, which will be discussed in greater detail in Chapter 4.[36-38, 41]

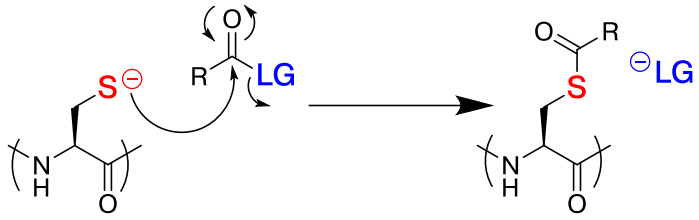


Figure 1.2 Acyl transfer reaction involving nucleophilic cysteine
Acyl transfer reactions occur throughout chemical biology, involving proteins, lipids, or cofactors. Cysteine catalyzes many of these reactions. "LG" refers to "leaving group."

The sulfur atom in cysteine can be oxidized to multiple oxidation states (Figure 1.3), either enzymatically or as a result of oxidative stress.[4] Cysteine sulfenates were initially viewed solely as reactive intermediates, but the importance of this post-translational modification as a regulatory mechanism in proteins has recently been realized.[5, 8] For example, oxidation of Cys797 to the sulfenate (but not to the disulfide) is known to increase kinase activity activity in epidermal growth factor receptor (EGFR),[1, 42] and therefore cysteine oxidation to the sulfenate is crucial for regulating cellular growth and division through this tyrosine kinase receptor. Since sulfenic acids exhibit both nucleophilic and electrophilic roles, oxidation of cysteine to the sulfenate can dramatically alter its catalytic ability.[5] Furthermore, cysteine sulfenates can be

used for site-selective tagging in proteins.[43, 44] Oxidation of cysteine to the sulfinate or sulfonic acid will also change reactivity by substantially increasing its acidity, and this has been observed to affect metal-binding ability.[5] Cysteine sulfinates can also potentially affect local conformation in proteins through hydrogen-bonding interactions, both with amino acid side-chains or in the peptide backbone. For proteins containing structural disulfides, oxidation of cysteine can be a rate-determining step in protein folding pathways.[45, 46] As ubiquitous as protein phosphorylation, S-glutathionylation of cysteine plays a major role in regulating redox-active proteins and associated cell-signaling cascades.[5, 8] S-nitrosylation of cysteine is also an important post-translational modification that signals cellular responses to nitrosative (reaction with NO) and oxidative stress, which is particularly pertinent to NO signaling in cardiac physiology and heart disease.[1, 47, 48]

In addition to the natural modifications in biological contexts, cysteine is also important in synthetic, non-natural modifications of peptides and proteins. Selective modification of proteins is possible with cysteine residues under mild conditions, even in the presence of other biologically reactive groups, such as amines or phenols.[15] A plethora of biological reagents are commercially available for synthetic modification of cysteine, such as alkyl halides (e.g. iodoacetamides) or Michael acceptors (e.g. maleimides). Modifications to cysteine can also be performed using radical or photochemical "thiol-yne" or "thiol-ene" chemistries with alkynes or alkenes.[49, 50] Sequences of cysteines can also be used for site-selective tagging in proteins, such as the CCXXCC hairpin motif and FlAsH tags.[51, 52] The types of modifications at cysteine are varied, and provide the ability to tag proteins with fluorophores, imaging agents, drug conjugates, or to design multi-domain protein conjugates or hybrid

protein-materials.[13] Peptide "stapling," or covalent attachment of synthetic protein domains can also be performed via oxidation of cysteine pairs to form disulfides (reversible) or through bis-alkylation of cysteine disulfides (irreversible).[1, 15, 53]

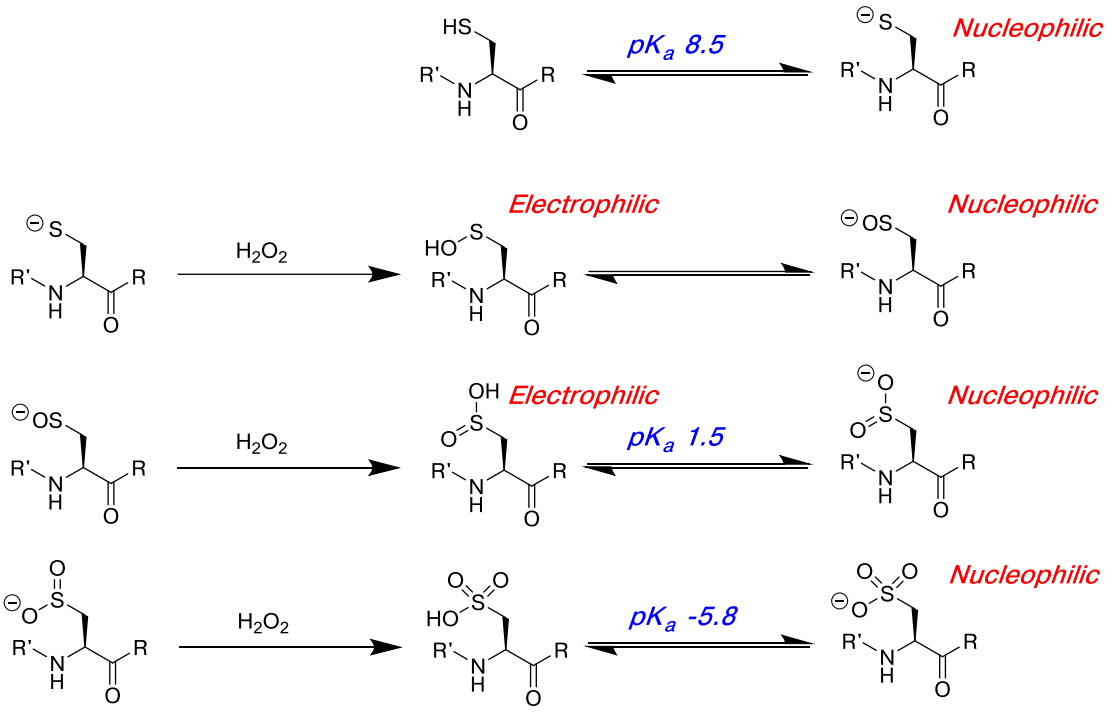


Figure 1.3 Unique reactivity in different sulfur oxidation states of cysteine
Oxidation at the sulfur atom in cysteine generates cysteine sulfenic acid, sulfinic acid, and sulfonic acid (indicated). The protonation state for each of these derivatives of cysteine has distinctive reactivity,[54, 55] allowing for detection of specific sulfur oxidation states in proteins.[10, 44, 55, 56]

The reactive, structural, and regulatory roles of sulfur in chemical biology have been widely studied for more than 100 years. Oxidation of sulfur-containing amino acids in proteins is both structural and regulatory: both nonspecifically in stress responses, and specifically as a concerted post-translational modification.[1] Although much has been discovered about sulfur oxidation in biology, there is a need for new

tools to probe the dynamic roles of the less stable oxidation states of sulfur. Cysteine has been widely used as a versatile functional handle for protein modification, but there remains the possibility of cross-reactivity with other residues under mild conditions, and synthetic modification of cysteine is only site-selective if there is a single cysteine residue.[51]

1.1.2 Synthesis of an aryl thiolated amino acid: 4-thiophenylalanine

In 1912, with only early knowledge of protein composition and structure, researchers sought to synthesize and characterize the reactive properties of 4-thiophenylalanine. Since serine and amide bonds were found to have natural sulfur isosteres (cysteine and thioamide bonds, respectively), it was speculated that tyrosine could potentially have a sulfur-containing analogue in Nature, and characterization of so-called "thiotyrosine" (Figure 1.4) would be necessary in the event it was ever found. This first reported synthesis of 4-thiophenylanine involved a 10-step synthesis via hydrolysis of a modified hydantoin (Figure 1.5),[16] and suggested the possibility of designing and synthesizing non-natural amino acids for modulation of protein function, drawing on inspirations from naturally occurring biomolecules. This synthesis includes a number of potentially problematic steps, including aromatic nitration, tin reduction, diazotization, and thermal rearrangement steps.

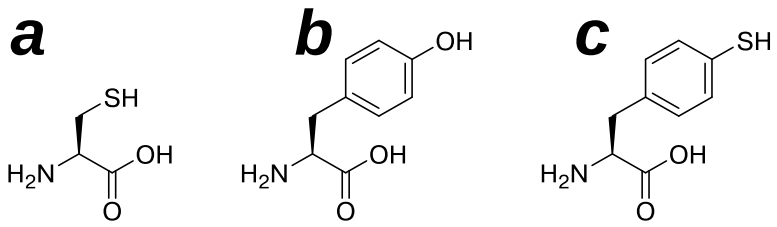


Figure 1.4 Structure of cysteine, tyrosine, and 4-thiophenylalanine
In the early 20th century, having found that serine had a natural sulfur-analogue, cysteine (a), researchers speculated that tyrosine (b) might have a natural sulfur analogue, 4-thiophenylalanine (c).

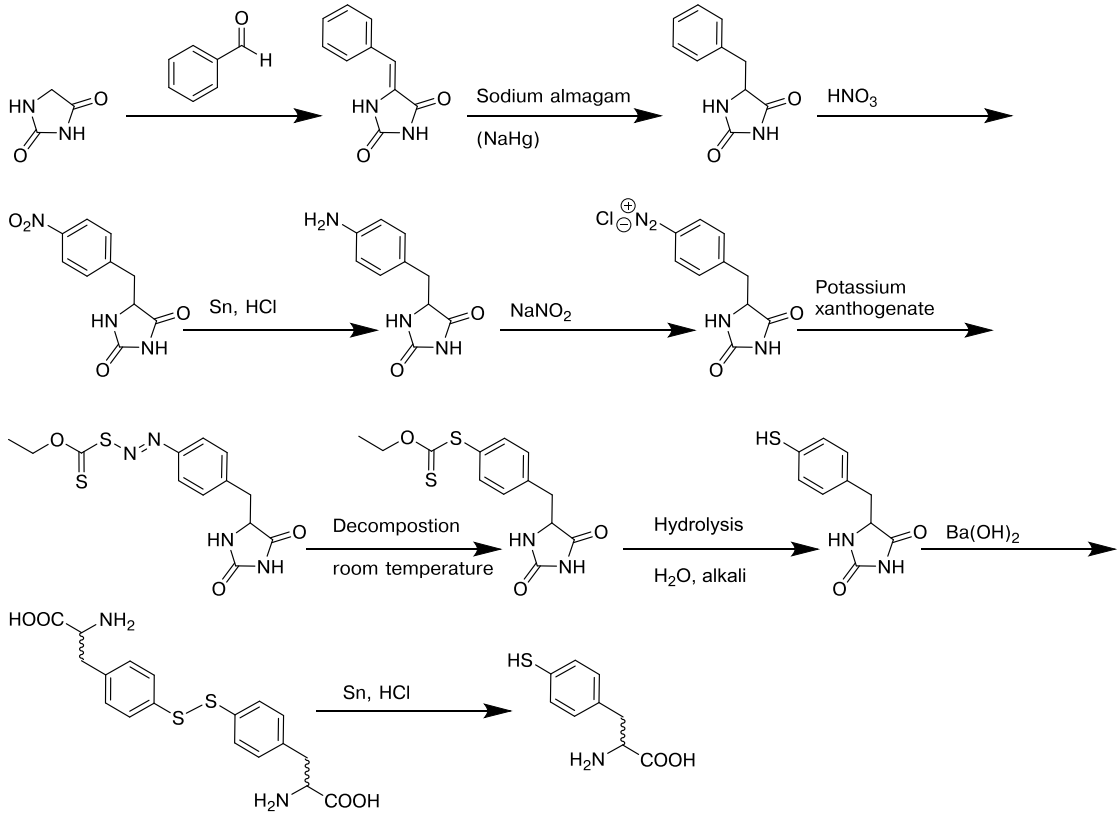


Figure 1.5 Earliest reported synthesis of 4-thiophenylalanine in 1912[16]
"Thiotyrosine," or 4-thiophenylalanine, was synthesized in order to characterize its chemical properties, in an effort to determine if it was naturally found in proteins. The researchers reported that it can be distinguished from tyrosine on reaction with sulfuric acid, where 4-thiophenylalanine and its disulfides produce a brilliant purple color.

4-Thiophenylalanine has never been found in native proteins, and this synthesis was never subsequently utilized. In 1949, Eliot & Haringtон[20] reported a new synthesis of 4-thiophenylalanine, in an attempt to discover novel bacteriostatic molecules based on amino acid analogues (Figure 1.6). This 10-step synthetic approach used somewhat milder reagents (though still requiring synthetic steps with red phosphorus or sodium), but still generated a racemic mixture of the 4-thiophenylalanine product.

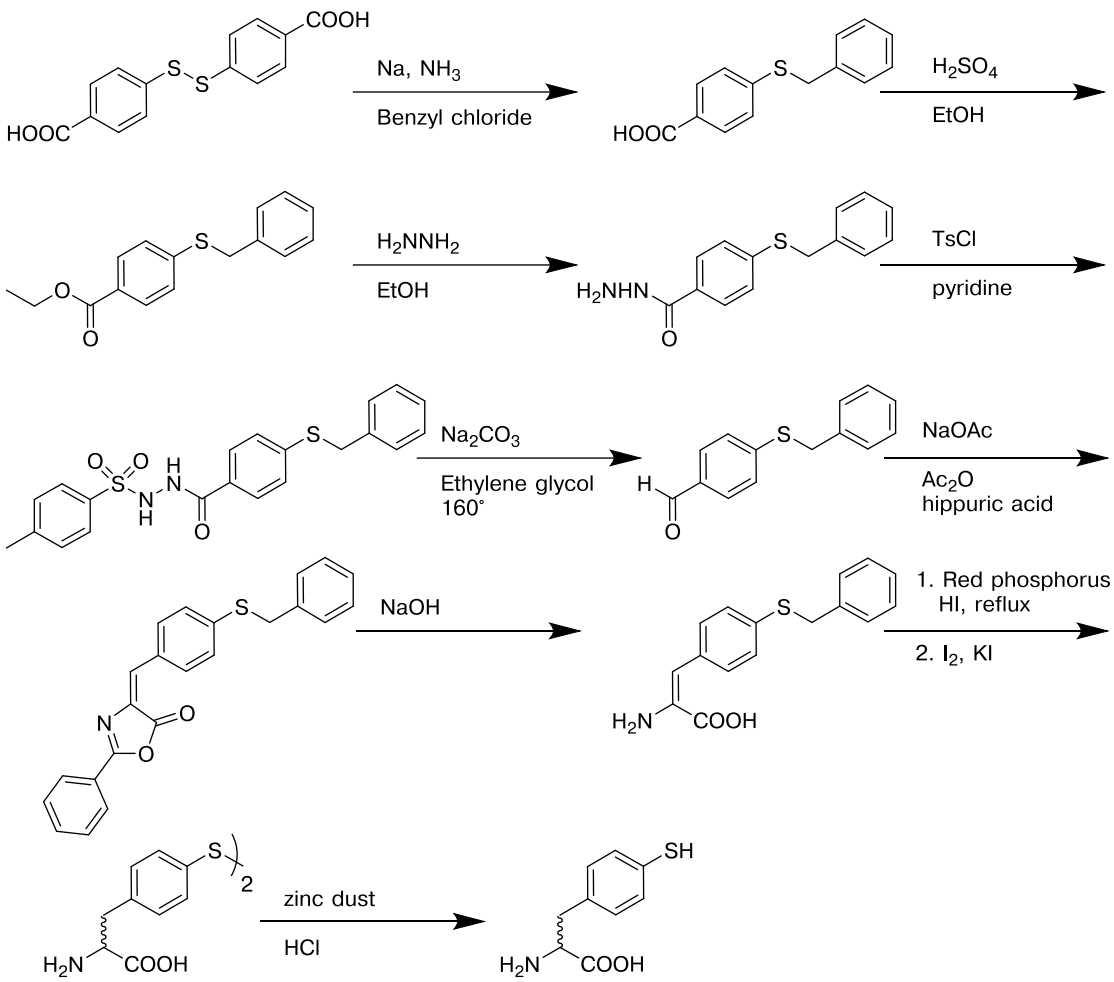


Figure 1.6 1949 synthesis of 4-thiophenylalanine[20]
Analogues of amino acids were sought for bacteriostatic properties, and the sulfur analogue of tyrosine was resynthesized through a slightly shorter route.

The synthesis was later improved by Colescott *et al.*[21, 57] to a 3-step synthesis from potassium ethyl xanthate and *p*-diazobenzylacetamidomalonate, in order to further explore the potential for thiophenylalanine and related derivatives in bacterial growth inhibition. While this improved synthesis was much shorter and resulted in a higher product yield, this approach still generated racemic mixture of 4-thiophenylalanine.[21]

The first enantiopure synthesis of L-4-thiophenylalanine, and its first application in peptides, was reported in 1983 by Escher and coworkers (Figure 1.7),[22] in order to examine structure-activity relationships with tyrosine analogues at residue 4 in angiotensin II (DRVYVHPF). Based on these studies, the authors proposed that Tyr4 is involved in an aromatic interaction with angiotensin II receptor, since the electron-donating effects of aromatic substituents modulated activity.[58] This synthetic route generated 4-thiophenylalanine as a a free L-amino acid in only 2 steps from commercially available L-phenylalanine, but required chlorosulfonic acid and tin reduction. The synthesized L-4-thiophenylalanine was readily protected for incorporation into peptides via Boc- solid-phase peptide synthesis (SPPS). The peptide containing 4-S(benzyl)-thio-L-phenylalanine was subjected to deprotection conditions in HF.[58] HF is a dangerous and corrosive reagent that is necessary for cleavage and deprotection reactions in Boc- SPPS, and so the safer approach using Fmoc- SPPS is more widely utilized.[59] Escher also suggested the utility of 4-thiophenylalanine for covalent attachment to solid supports via Michael addition or metal-complexation, and described its versatility in synthesizing a number of 4-thiophenylalanine derivatives for SAR studies of angiotensin II inhibitors.[22]

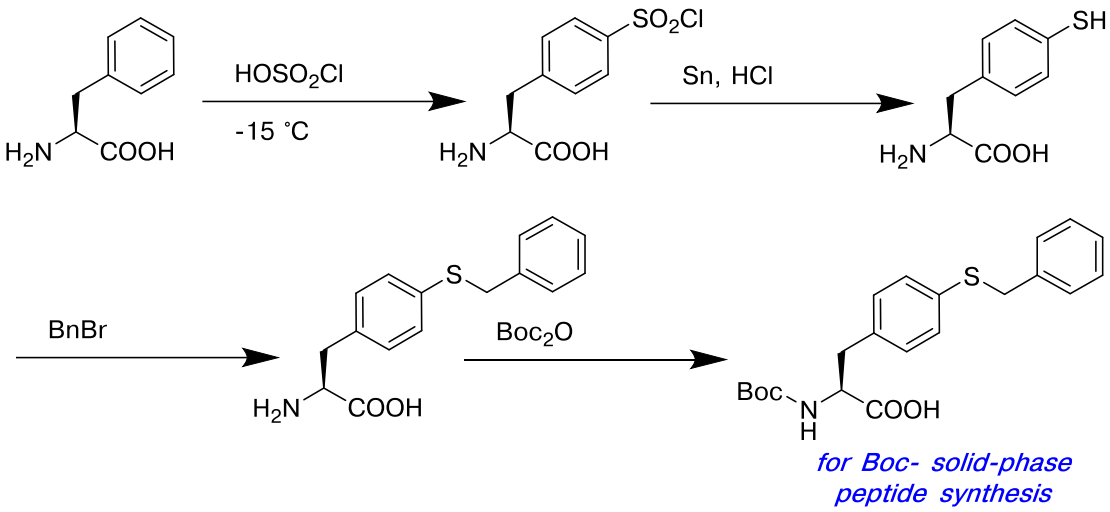


Figure 1.7 First reported synthesis of 4-thiol-L-phenylalanine[22]
Prior syntheses of 4-thiophenylalanine would produce racemic mixtures of 4-thiol-D,L-phenylalanine. Escher reported this synthesis to generate 4-thiol-L-phenylalanine, and incorporated it into peptides for SAR studies of angiotensin II analogues. The synthesized Boc-4-S(benzyl)-thio-L-phenylalanine was incorporated into peptides via Boc- solid-phase peptide synthesis, and the peptide was subjected to deprotection conditions in HF.

Hobbs & Still[60] reported a synthetic strategy for peptides containing 4-thiophenylalanine via saponification of the xanthate species, which was obtained through reaction of the 4-diazonium-phenylalanine with potassium O-ethyl xanthate (Figure 1.8), similar to a synthesis reported by Colescott *et al.*[21, 57] In this work, peptides containing 4-thiophenylalanine were used to generate biaryl-thioether analogues of a peptide macrocycle via a photo-initiated coupling reaction with an aryl iodide.[60, 61] While this 3-step synthetic route is more practical than prior methods, it does not produce 4-thiophenylalanine in a form that is readily compatible with Fmoc-SPPS (where HF is not required).

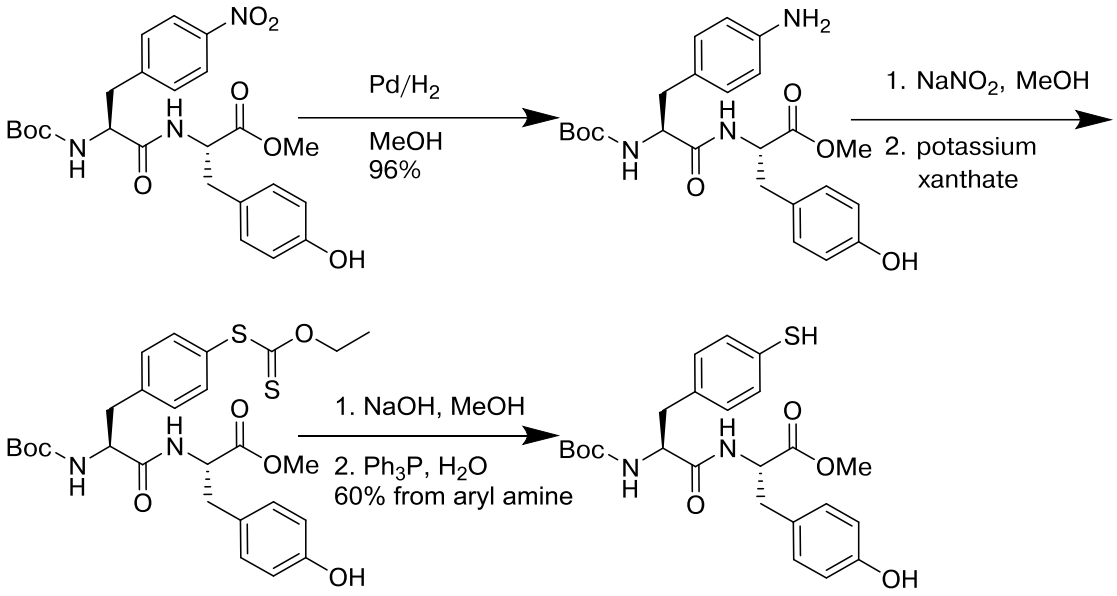


Figure 1.8 Synthesis of 4-thiophenylalanine within a dipeptide
A synthesis to generate thiophenylalanine initially demonstrated by Colescott *et al.*[21, 57] and later improved upon and utilized by Hobbs & Still.[60] This work demonstrated a photo-initiated coupling reaction to form biaryl thioethers from iodo-tyrosine derivatives and 4-thiophenylalanine.

Rajagopalan *et al.*[23] noted the difficulty in prior synthetic approaches to 4-thiophenylalanine, which "has undoubtedly hindered its preparation, inclusion into bioactive peptides and characterization."[23] Rajagopalan *et al.*[23] reported a convenient synthesis of 4-thiophenylalanine using a palladium-catalyzed cross-coupling reaction with Boc-4-iodophenylalanine with *tert*-butyl mercaptan to generate the *tert*-butyl thioether of 4-thiophenylalanine (Figure 1.9). This method conveniently generated protected 4-thiophenylalanine in 1 step from a commercially available amino acid, and can be utilized directly for incorporation into peptides via Boc-SPPS. However, for incorporation into peptides using Fmoc- SPPS, the *tert*-butyl thioether can not be readily deprotected in solution to form 4-thiophenylalanine within peptides.

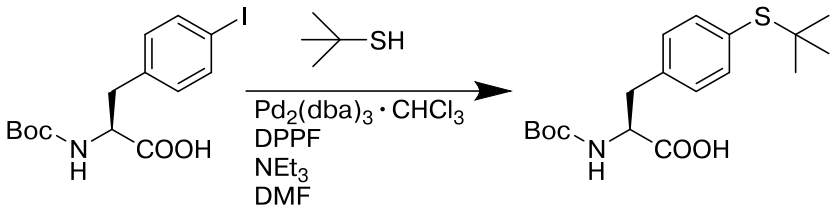


Figure 1.9 First reported synthesis of 4-thiol-L-phenylalanine via cross-coupling methodology[23]
Advances in metal-catalyzed cross-coupling reactions inspired this one-step synthesis of a 4-thiophenylalanine derivative. The product was isolated in 67% yield.

Despite the synthetic limitations, 4-thiophenylalanine has been employed in the context of proteins and peptides. In 1997, 4-thiophenylalanine and similar amino acids were synthesized by DeGrado and coworkers[62] to examine peptide and protein folding dynamics, where 4-thiophenylalanine disulfide was employed as an optical trigger release within a peptide. DeGrado and coworkers[62] developed a protecting-group strategy in order to incorporate 4-thiophenylalanine into peptides via Fmoc-SPPS (Figure 1.10), utilizing an extension of the synthesis developed by Escher and coworkers.[22, 58] In this work, a helical peptide was designed with 4-thiophenylalanine incorporated at the N- and C-termini, which acted as a disulfide linker that prevented helix formation.[62] The disulfide was photolyzed to form the thiyl radicals of 4-thiophenylalanine, and early nucleation steps of helix formation could be studied. However, it was found that the unstable thiyl radical of 4-thiophenylalanine rapidly reformed the disulfide linkage, and helix-formation could not be studied using 4-thiophenylalanine, and a different amino acid was utilized for this study. In 2011, Hecht and coworkers[17, 63, 64] explored the nucleophilic role of tyrosine in DNA topoisomerase I by expressing different aromatic amino acids in place of Tyr274. Similar to DeGrado and coworkers,[62] the Escher synthesis[22, 58] was utilized to generate

4-thiophenylalanine as a protected form for expression into DNA topoisomerase I via modified aminoacyl tRNA.[17, 63, 64] Compared to native tyrosine, the rate of DNA backbone cleavage was 30 times lower with the topoisomerase mutant containing 4-thiophenylalanine, but the rate of DNA ligation was 3 times faster.[17] In 2013, Rudolf & Poulter[65] examined the substrate scope and enzyme promiscuity of tyrosine O-prenyltransferase. Several canonical and non-natural aromatic amino acids, including 4-thiophenylalanine, were subjected to prenylation by SirD. In this work, 4-thiophenylalanine was synthesized via Boc-4-S(*tert*-butyl)-thiophenylalanine, described by Rajagopalan *et al.*,[23, 65] and subsequent deprotection using concentrated HCl. In these approaches to 4-thiophenylalanine, harsh reaction conditions (i.e. chlorosulfonic acid, hydrofluoric acid, concentrated HCl) were necessary to generate the usable form of 4-thiophenylalanine, either for Fmoc- SPPS or for expression in proteins.

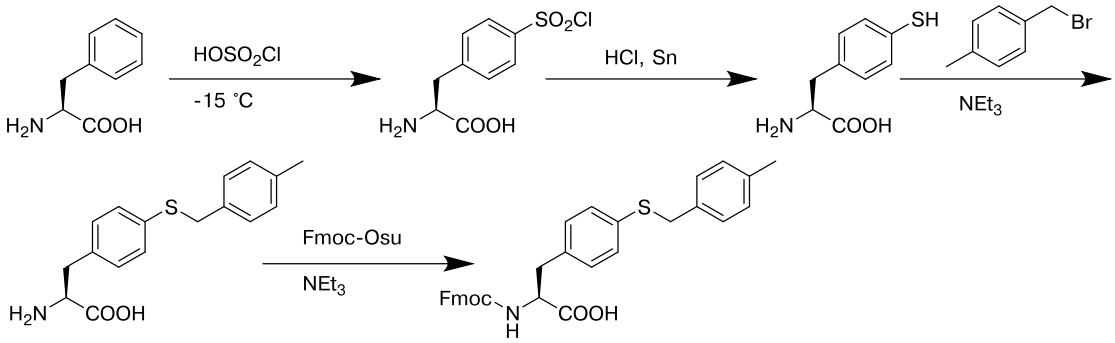


Figure 1.10 Synthesis of protected 4-thiophenylalanine for use in solid-phase peptide synthesis[62]
Protecting groups to 4-thiophenylalanine for solid-phase peptide synthesis adds synthetic steps and decreases overall yield (30% from L-phenylalanine). Protected 4-thiophenylalanine was incorporated into peptides via Fmoc-SPPS, and the thiol functional group was deprotected in the presence of HF.

These synthetic strategies of 4-thiophenylalanine encompass a wide variety of applications across chemical biology and engineering. 4-Thiophenylalanine has been explored for potential antibiotic activity, for synthesizing analogues of peptide hormones for exploring roles of aromatic amino acids, for elucidation of nucleophilic roles of tyrosine in proteins, as a photolyzable release to examine protein folding kinetics, and was suggested for future use as a chemical "handle" for bioconjugation reactions in materials science. However, only two synthetic routes have been applied in recent years, which either require hazardous reagents or generate a thioether derivative that must be deprotected using corrosive reagents (i.e. hydrofluoric acid). This might be generally acceptable for expressing 4-thiophenylalanine in proteins, but use of this sulfur analogue of tyrosine in peptides and proteins has been relatively limited. A more practical, inexpensive, and safer approach to synthesize 4-thiophenylalanine is required, ideally in a form that can be utilized for Fmoc- SPPS or for expression in proteins.

1.1.3 Transition-metal-mediated cross-coupling methodologies for the synthesis of aryl thiols

Transition-metal-catalyzed cross-coupling methodologies have been well developed over the past 100 years for carbon–carbon and carbon–heteroatom bond formation.[24, 25, 66-68] These reactions have been especially useful for coupling aryl halide and pseudo-halide substrates with nucleophilic coupling partners. This area of research has grown dramatically since the discovery and success of palladium-catalyzed approaches.[24, 69] In comparison to other heteroatoms, cross-coupling methodologies for C–S or C–Se bond formation have been developed only much more recently.[26] In addition to their sensitivity to oxidation, sulfur and selenium nucleophiles are often avoided in transition-metal catalysis since they are prone to "poisoning" the catalysts.[26] Given the versatility of aryl thiols in organic synthesis, numerous examples of cross-coupling reaction methodologies for generating aryl thiols and thioethers have emerged. These approaches and have been extensively reviewed.[26, 70, 71]

The first palladium-catalyzed cross-coupling reaction for C–S bond formation was reported by Migita *et al.*[72, 73] in 1978, which utilized aryl and alkyl thiolates as coupling partners for substituted aryl halides (Figure 1.11). During this time period, palladium-catalyzed cross-coupling methodologies had gained much attention among synthetic chemists, and researchers were working to explore the substrate scope of these reactions. Progress in exploring and optimizing carbon–carbon bond formation using these palladium-catalyzed reactions had already been underway since the ground-breaking work of Heck,[74] Mizoroki,[75] Julia,[76] and others in the late 1960s and early 1970s.[24, 69] However, carbon–heteroatom bond formation via palladium-catalysis was still in early stages of development. Prior to the report by Migita *et al.*,[72, 73] Oka

and coworkers[77] had already demonstrated that aryl nitriles could be synthesized via palladium-catalyzed cross-coupling with aryl halides and potassium cyanide. This work by Oka and coworkers[77] had shown that nucleophilic anions could participate as coupling partners in palladium-catalyzed cross-coupling reactions. Migita *et al.*[72, 73] briefly explored the reaction mechanism of synthesizing aryl sulfides via palladium-catalysis, and found that this reaction did not proceed in the absence of palladium(II); that the thiolate species reacted with an $Ar–Pd^{II}–X$ intermediate, consistent with other palladium-catalyzed cross-coupling reactions; and that the reaction was not inhibited by the presence of *p*-dinitrobenzene, which suggested that the reaction did not proceed through an $S_{RN}1$ mechanism.[72, 73, 78] Since this work, the scope of palladium-catalyzed cross-coupling reactions for C–S bond formation has expanded to include other aryl substrates and sulfur-containing coupling partners.[23, 79, 80] In 1996, Boc-protected 4-iodophenylalanine was utilized used in a palladium-catalyzed cross-coupling reaction for generating aryl thioethers.[23, 80] Typical for palladium-catalyzed approaches, individual reactions can be tailored and optimized for specific substrates through screening different palladium sources, ligands, bases, reaction temperatures, and solvents.[26]

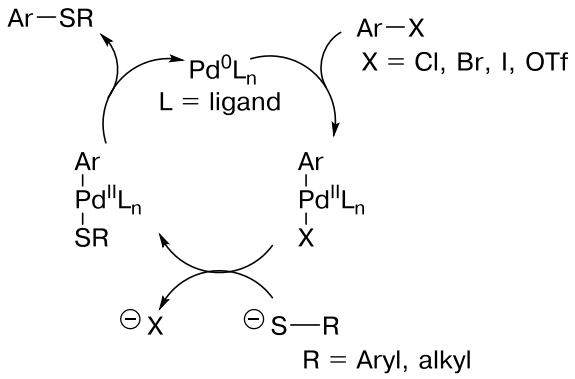


Figure 1.11 General mechanism for palladium-catalyzed cross-coupling reactions for aryl C–S bond formation

Although palladium-catalyzed cross-coupling reactions have drawn much attention in recent years, copper-mediated strategies have been extensively utilized in synthetic organic chemistry since the early 1900s.[25, 26] Much of the inspiration for development of palladium catalysis originated from the copper-mediated approaches that were initially described by Ullmann in a series of papers at the beginning of the 20th century.[25, 66-68] Initially, Ullmann demonstrated aryl C–C bond formation in his synthesis of a biphenyl derivative by refluxing 2-nitro-bromobenzene in the presence of copper.[68] In 1903, he synthesized a biaryl amine (C–N bond formation) by refluxing an aryl chloride in the presence of aniline and copper, and in 1904 he reported synthesis of a biphenyl ether derivative (C–O bond formation) from an aryl chloride and phenol using catalytic copper at high temperatures.[66, 67] With the expense and air-sensitivity of palladium sources and their ligands, and cases of substrates that are incompatible with palladium, many researchers have revisited these early reports in copper-mediated catalysis in search of alternatives to palladium for carbon–carbon and carbon–heteroatom bond formation.[25, 26, 71, 81] The number of reports in copper-mediated aryl C–S bond formation has grown considerably in the past few decades, in

part of a so-called "Renaissance" of Ullman-coupling, with major advancements in lower reaction temperatures and lower catalyst loading.[82-88]

In light of these recent advancements in transition-metal-mediated cross-coupling strategies in aryl C–S bond formation, and with the potential utility of 4-thiophenylalanine in chemical biology, we sought to develop a practical approach to synthesize this amino acid via cross-coupling methodology. Palladium-catalyzed cross-coupling has been previously employed for synthesis of the Boc-protected 4-S(*tert*-butyl)-thiophenylalanine,[23] but this strategy was never applied to peptides. Boc-4-S(*tert*-butyl)-thiophenylalanine[23] is not practically compatible for incorporation into peptides, since deprotection of the *tert*-butyl thioether requires additional reagents that are not necessarily compatible with peptides. Copper-catalyzed approaches to aryl halide cross-coupling reactions generally have greater functional group tolerance, are less sensitive to the nature of the metal source, less sensitive to oxygen, and are less expensive and more readily available than palladium sources.[25, 26] We describe development of a novel synthetic strategy, utilizing copper catalysis, in order to synthesize 4-thiophenylalanine and related derivatives for practical application in peptides.

1.2 Results

1.2.1 Initial screening of cross-coupling conditions with iodobenzene

We sought to develop a transition-metal mediated cross-coupling approach to generate 4-thiophenylalanine in a form that can be readily utilized for Fmoc- SPPS. The work by Rajagolpalan *et al.*[23] demonstrated that metal-mediated cross-coupling approaches can be used to form aryl-thiolated amino acids from Boc-4-

iodophenylalanine. Rudolf & Poulter[65] subjected Boc-4-S(*tert*-butyl)-thiophenyalanine to concentrated HCl to generate 4-thiophenylalanine as a free amino acid. In Fmoc-SPPS, the protected peptide is subjected to TFA to cleave the peptide from resin and removed side-chain protecting groups, but an S(*tert*-butyl) aryl thioether can not be readily deprotected (to form the aryl thiol) using standard TFA cleavage conditions. In contrast, by protecting 4-thiophenylalanine as a thioester, mild reductants can be used to deprotect the thiol following standard TFA deprotection reaction conditions. Combined, we sought to utilize a transition-metal-mediated cross-coupling reaction using 4-iodophenylalanine and a thioacid, such as thiolacetic acid, to generate 4-thiophenylalanine with a thioester protecting group.

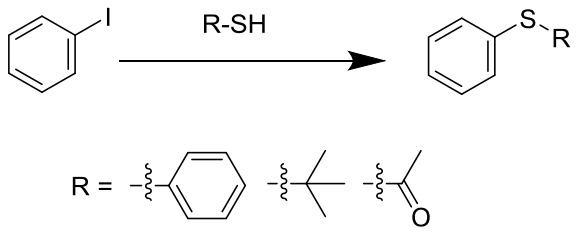
In the synthesis described by Colescott *et al.*,[21] 4-thiophenylalanine was generated within a dipeptide, rather than as a protected amino acid. This synthetic strategy essentially utilized the peptide backbone as protecting groups on the amine and carboxylic acid. In a similar manner, we sought to develop reaction conditions to generate 4-thiophenylalanine *within peptides*, which would avoid reaction steps that are required to incorporate and remove protecting groups. In essence, a peptide containing 4-iodophenylalanine would act as the aryl halide reactant for cross-coupling reactions with thioacids.

Rajagolpalan's[23] work demonstrated the utility of palladium-catalyzed cross-coupling approaches, but copper-catalyzed approaches for C–S bond formation have the advantage of being relatively air- and moisture-stable.[82, 88, 89] An initial screen was conducted to examine different thiol reaction partners in copper-mediated cross-coupling reactions. Thiols examined in this screen included *tert*-butyl mercaptan, thiophenol, and thiolacetic acid, which exhibit different acidities and different

potential reactivities. Iodobenzene was used to in place of 4-iodophenylalanine for the initial screen, and the resultant products of the cross-coupling reactions were analyzed via GC-MS to rapidly identify reaction products. The results of these initial screens are reported in Table 1.1.

Table 1.1. Initial screening of the copper-catalyzed cross-coupling conditions using iodobenzene and different thiols

Crude reaction products were evaluated via GC-MS. Peak identity was established via MS data. Product conversions were normalized relative to total peak integrations, and were calculated based on product formed (as PhSR).
[a]refers to total cross-coupled products as PhSR, PhSPh, and PhSSPh.



Entry	Thiol	Catalyst	Base	Solvent	Temperature (°C)	Reaction Duration	Product Conversion	Ref.
1	*t*-BuSH	20 mol% CuI 2 eq. ethylene glycol	2 eq. K_2CO_3	*tert*-amyl alcohol	100	23 h	12%	[82]
2	PhSH	20 mol% CuI 2 eq. ethylene glycol	2 eq. K_2CO_3	isopropanol	80	23 h	77%	[82]
3	AcSH	20 mol% CuI 2 eq. ethylene glycol	2 eq. K_2CO_3	*tert*-amyl alcohol	100	21 h	10%	[82]
4	*t*-BuSH	10 mol% CuI 20 mol% cis-1,2-cyclohexanediol	2 eq. K_3PO_4	DMF	50	23 h	<1%	[89]
5	AcSH	10 mol% CuI 20 mol% cis-1,2-cyclohexanediol	2 eq. K_3PO_4	DMF	50	23 h	<1%	[89]
6	*t*-BuSH	10 mol% CuI 20 mol% 1,10-phenanthroline	2 eq. DIPEA	toluene	110	23 h	<1%	[88]
7	AcSH	10 mol% CuI 20 mol% 1,10-phenanthroline	2 eq. DIPEA	toluene	110	23 h	47% (67%)[a]	[88]

The data in Table 1.1 demonstrate that a copper-mediated cross-coupling approach between aryl iodides and thiolacetic acid is viable for to generate thioesters after appropriate reaction optimization. The ligand-free reaction conditions, similar to those described by Buchwald and coworkers,[82] reacted efficiently with thiophenol (Table 1.1, entry 2), but were less successful for *tert*-butyl mercaptan and thiolacetic acid, even when utilizing higher reaction temperatures (Table 1.1, entries 1 and 3). The approach at lower reaction temperature described by Cook *et al.*,[89] using cyclochexanediol as the ligand, was not successful for either *tert*-butyl mercaptan or thiolacetic acid (Table 1.1, entries 4 and 5). Using thiolacetic acid as the reaction substrate, the conditions described by Sawada *et al.*[88] were the most successful (Table 1.1, entry 7). However, using these reaction conditions (Table 1.1, entry 7), multiple reaction products were observed: S-phenyl thioacetate was generated as the major product, as well as the biaryl thioether and symmetric disulfides of thiophenol (Figure 1.12). These products potentially resulted from deacetylation of the S-phenyl thioacetate product, and either subsequent oxidation of the resultant thiophenolate, or the resultant thiophenolate reacting with iodobenzene as a substrate in the cross-coupling reaction (Figure 1.12).

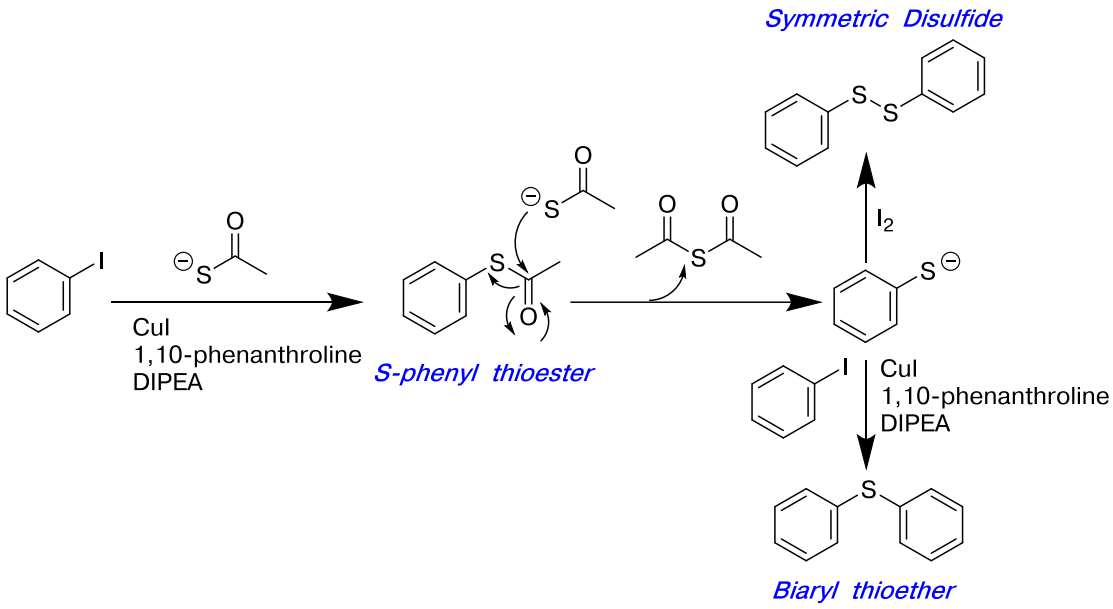


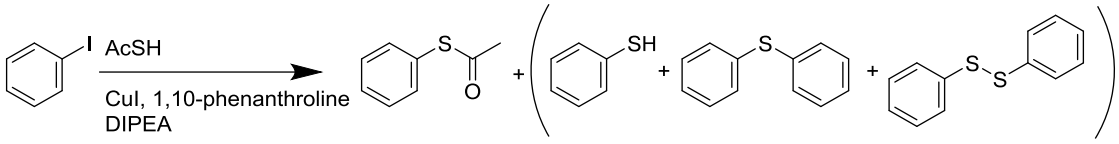
Figure 1.12 Potential mechanisms for formation of the disulfide and thioether products during the cross-coupling reaction conditions
GC-MS data of the crude reaction products (Table 1.1, entry 7) showed a mixture of aryl thiolated species. Thiolacetate can potentially react with the S-phenyl thioester product to generate thiophenolate. Thiophenolate can compete with thioacetate as the substrate for the copper-catalyzed cross-coupling conditions, generating the biaryl thioether product. Alternatively, thiophenolate can form disulfides in solution to generate the symmetric disulfide. The S-phenyl thioester and thiophenol disulfide products can be converted into the aryl thiol with reductive work-up conditions.

Although the copper-catalyzed cross-coupling reaction conditions successfully formed S-phenyl thioacetate from iodobenzene and thiolacetic acid, the S-phenyl thioacetate product was prone to thiolysis during the reaction. The thiophenol product that resulted from thiolysis of S-phenyl thioacetate can oxidize to form disulfides, or compete with thioacetate as the thiol substrate in the cross-coupling reaction. In consideration of all products from the reaction (S-phenyl thioester, biaryl thioether, thiophenol disulfides), the overall reaction conversion with iodobenzene and thiolacetic acid was 67% using the conditions described by Sawada *et al.*[88] (Table 1.1,

entry 7). These additional reaction products are important to consider in the overall conversion, because the disulfide reaction products can be converted to the aryl thiol after reductive work-up conditions. However, thiolysis of the S-phenyl thioacetate product must be suppressed during the reaction conditions, as to reduce potential formation of the biaryl thioether (which can not be converted to the aryl thiol). Having established initial cross-coupling conditions, a secondary screen was conducted to optimize the reaction conditions.[88] The reaction conditions (Table 1.1, entry 7) were screened using different solvents and reaction temperatures, shown in Table 1.2. In anticipation of further optimization of the reaction conditions for peptides containing 4-iodophenylalanine, solvents that can solubilize peptides were emphasized in this reaction screen.

Table 1.2. Optimization of the copper-mediated cross-coupling reaction conditions using iodobenzene and thiolacetic acid

Crude reaction products were evaluated via GC-MS. Peak identity was established via MS data. Product conversions were normalized relative to total peak integrations. For formation of the biaryl ether product, it is assumed a reaction occurred between iodobenzene and thiophenol (resulting from thiolysis of S-phenyl thioacetate). For formation of the disulfide product, it is assumed an oxidation reaction occurred between two thiophenolates (resulting from thiolysis of S-phenyl thioacetate). Overall conversion refers to all cross-coupling reaction products.
PhSAc = S-phenyl thioacetate
PhSPh = biaryl thioether product



Entry	Solvent	Temperature (°C)	Reaction Duration	PhSAc Yield	PhSPh Yield	Overall Conversion
1	*tert*-amyl alcohol	80	21 h	0%	0%	0%
2	*tert*-amyl alcohol	100	22 h	37%	0%	37%
3	*tert*-amyl alcohol	110	21 h	37%	26%	84%
4	*tert*-amyl alcohol	120	21 h	29%	38%	85%
5	10% H_2O/*tert*-amyl alcohol	90	24 h	3%	0%	70%
6	20% H_2O/*tert*-amyl alcohol	90	24 h	0%	0%	76%
7	DMF	100	22 h	6%	18%	32%
8	toluene	110	26 h	32%	10%	64%

The copper-catalyzed cross-coupling reaction with iodobenzene and thiolacetic acid successfully generated the S-phenyl thioacetate product. Additional products were observed as a result of the copper-catalyzed cross-coupling reaction, which potentially resulted from the thiolysis of the S-phenyl thioacetate product. These

additional products included the biaryl thioether (PhSPh) and thiophenol disulfide (PhSSPh). In the context of generating peptides containing 4-thiophenylalanine, the corresponding biaryl thioether product could not be readily converted to 4-thiophenylalanine. In addition, if the S-phenyl thioacetate product is prone to thiolysis during the reaction conditions, then the resultant thiophenol product can react as a potent nucleophile in the reaction. In the context of peptides as reaction substrates, a thiophenolate nucleophile can potentially react with side-chain functional groups. Therefore, conditions were identified with the highest reaction conversion to form the S-phenyl thioacetate with minimal side-products.

In order to suppress potential thiolysis of the resultant S-phenyl thioacetate reaction product, reaction conditions were screened, including different reaction solvents and different reaction temperatures. The cross-coupling reaction conducted at 80 °C generated no products, and only iodobenzene and reagents were observed (Table 1.2, entry 1). Conducting the cross-coupling reactions with iodobenzene in *tert*-amyl alcohol at the highest reaction temperatures (110-120 °C) resulted in high reaction conversion (Table 1.2, entries 3 and 4), but also resulted in higher conversion to the biaryl thioether product (26-38% conversion). In reactions where water was used as a cosolvent in *tert*-amyl alcohol, most of the reaction conversion was in the form of thiophenol disulfides, indicating significant thiolysis in the presence of water. The cross-coupling reaction conducted in DMF resulted in overall reaction conversion that was comparable to a similar reaction conducted in *tert*-amyl alcohol (Table 1.2, entries 2 and 7). However, when DMF was used as a solvent for the cross-coupling reaction, a substantial component of the reaction conversion was in the form of the biaryl thioether (Table 1.2, entry 7). In contrast, when toluene was used as a reaction

solvent at higher reaction temperature, the conversion to the S-phenyl thioacetate product was comparable to a similar reaction conducted in *tert*-amyl alcohol, with lower conversion to the biaryl thioether (Table 1.2, entries 3 and 8).

Combined, these data suggest that the optimal copper-mediated cross-coupling reaction conditions on iodobenzene utilize a non-polar solvent at higher reaction temperatures. No product formation was observed as a result of the reaction conducted at 80 °C, and significant disulfide formation was observed resulting from the cross-coupling reaction conducted at 120 °C. Reaction conversion to generate the S-phenyl thioacetate product, rather than the disulfide or biaryl thioether, was improved on utilizing a non-polar solvent, rather than *tert*-amyl alcohol or DMF. In addition, anhydrous reaction conditions should be utilized, due to the fact that significant thiolysis of the S-phenyl thioacetate reaction product was observed in the presence of water (Table 1.2, entries 6 and 7).

Having established reaction conditions for generating a thioester from iodobenzene and thiolacetic acid via a copper-catalyzed cross-coupling reaction, the potential for synthesizing 4-S(acetyl)-thiophenylalanine from 4-iodophenylalanine was explored. The optimal reaction conditions identified in this reaction screen on iodobenzene were further optimized for a peptide containing 4-iodophenylalanine.

1.2.2 Optimization of copper-mediated cross-coupling reaction conditions for peptides containing 4-iodophenylalanine in solution phase

Copper-catalyzed cross-coupling reaction conditions were established using iodobenzene for synthesizing S-phenyl thioacetate. This reaction is the first reported copper-mediated cross-coupling reaction between and aryl halide and thiolacetic acid. The conditions were further optimized for reaction with 4-iodophenylalanine within

peptides. Solution phase synthesis of amino acids inherently involves consideration of protecting group strategies on the amine and carboxylic acid functional groups. Rather than adding synthetic steps for incorporating and removing protecting groups, we examined direct reaction on a peptide containing 4-iodophenylalanine, effectively using the peptide backbone to protect the amine and carboxylic acid. Using this synthetic strategy within peptides, synthesis of peptides containing 4-thiophenylalanine will require less time and reagents than if the amino acid was independently synthesized for Fmoc-SPPS (Figure 1.13).

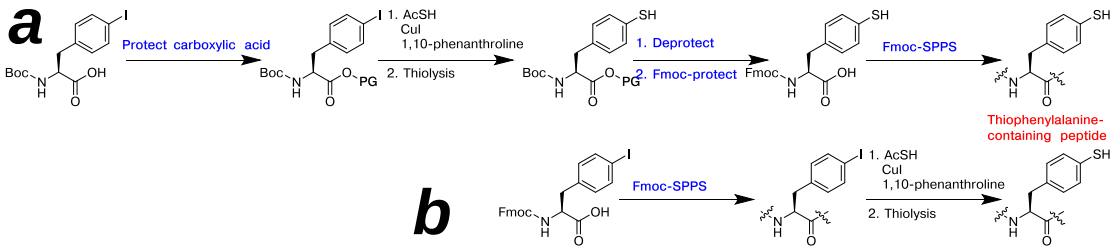


Figure 1.13 Cross-coupling reaction on peptides containing 4-iodophenylalanine
(a) Synthesizing 4-thiophenylalanine for use in Fmoc- SPPS can require several protection/deprotection steps on the amine and the carboxylic acid, which requires additional time and reagents. (b) By incorporating the aryl halide functional group within a synthesized peptide, the peptide backbone acts as the "protecting groups," and peptides containing 4-thiophenylalanine can be synthesized with fewer overall reaction purification steps.

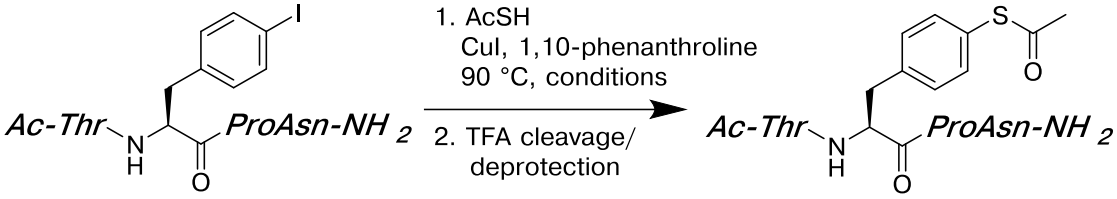
In order to evaluate the potential for a cross-coupling reaction strategy for the synthesis of peptides containing 4-thiophenylalanine, we synthesized a model peptide, Ac-TXPN-NH_2, where X = 4-iodophenylalanine (4-I-Phe). This model peptide was chosen for three reasons: (1) Ac-TXPN-NH_2 contains a variety of functional groups

(Thr hydroxyl, Asn amide) and can be used to test the compatibility of an un-protected peptide substrate for the copper-mediated cross-coupling reaction conditions; (b) the peptide context Ac-TXPN-NH_2 (where X = aromatic amino acid) has been well characterized by NMR,[90-92] and comparison of the peptide products from the cross-coupling reaction can be compared with published NMR data in order to determine the compatibility of the peptide with the copper-mediated cross-coupling reaction conditions; (c) substituents on aromatic amino acids are known to influence peptide structure (i.e. tyrosine v.s. phenylalanine),[91, 93] and characterization of the Ac-TXPN-NH_2 containing 4-thiophenylalanine allows for the exploration of the functional effects of thiol oxidation and alkylation to control to peptide structure.[5] Thus, the products of the cross-coupling reaction can be used directly to characterize the functional effects of sulfur oxidation on peptide structure.

The copper-mediated cross-coupling reaction conditions were further optimized using the purified peptide Ac-T(4-I-Phe)PN-NH_2, in place of iodobenzene. Solubility of the peptide, thiolacetic acid, and the copper catalyst is crucial for efficient reaction conversion. However, the reaction temperature must be 90-110 °C, as determined based on optimization of the reaction conditions with iodobenzene (Table 1.2). Solvents such as DMF or *tert*-amyl alcohol, polar solvents with high-boiling points, were examined for the initial reaction screen with the peptide Ac-T(4-I-Phe)PN-NH_2. The reaction progress was monitored via RP-HPLC. The results of the initial reaction condition screen are shown in Table 1.3, and a representative reaction scheme and resultant chromatogram of the products of the reactions are shown in Figure 1.14. This solution-phase approach, if successful, would provide direct access to peptides containing 4-S(acetyl)-thiophenylalanine (4-SAc-Phe).

Table 1.3. Optimization of the copper-mediated cross-coupling reaction with the peptide Ac-T(4-I-Phe)PN-NH_2 in solution

Crude reaction products were evaluated via HPLC. Peak identity was established via ESI-MS. Product conversions were normalized relative to total peak integrations, and were calculated based on product conversion for desired product (as Ac-T(4-SAc-Phe)PN-NH_2) and as total products resulting from the cross-coupling reaction (including additional side-products) shown in parentheses.
n.d. refers to "not determined" due to unidentifiable products of the reaction.



Entry	Base	Solvent	Reaction Duration	Product Conversion
1	1 M DIPEA	DMF	18 h	n.d.
2	1 M DIPEA	20% *tert*-Amyl alcohol in H_2O	9 h	n.d.
3	0.5 M Na_2CO_3	*tert*-Amyl alcohol	6 h	44% (59%)
4	100 mM DIPEA	*tert*-Amyl alcohol	6 h	15% (26%)
5	0.5 M DIPEA	*tert*-Amyl alcohol	6 h	11% (60%)
6	1 M DIPEA	*tert*-Amyl alcohol	6 h	22% (58%)
7	1 M DIPEA	*tert*-Amyl alcohol	3 h	36% (48%)
8	1 M DIPEA	*tert*-Amyl alcohol	9 h	14% (49%)
9	1 M DIPEA	*tert*-Amyl alcohol	11 h	29% (65%)
10	1 M DIPEA	*tert*-Amyl alcohol	24 h	0% (18%)

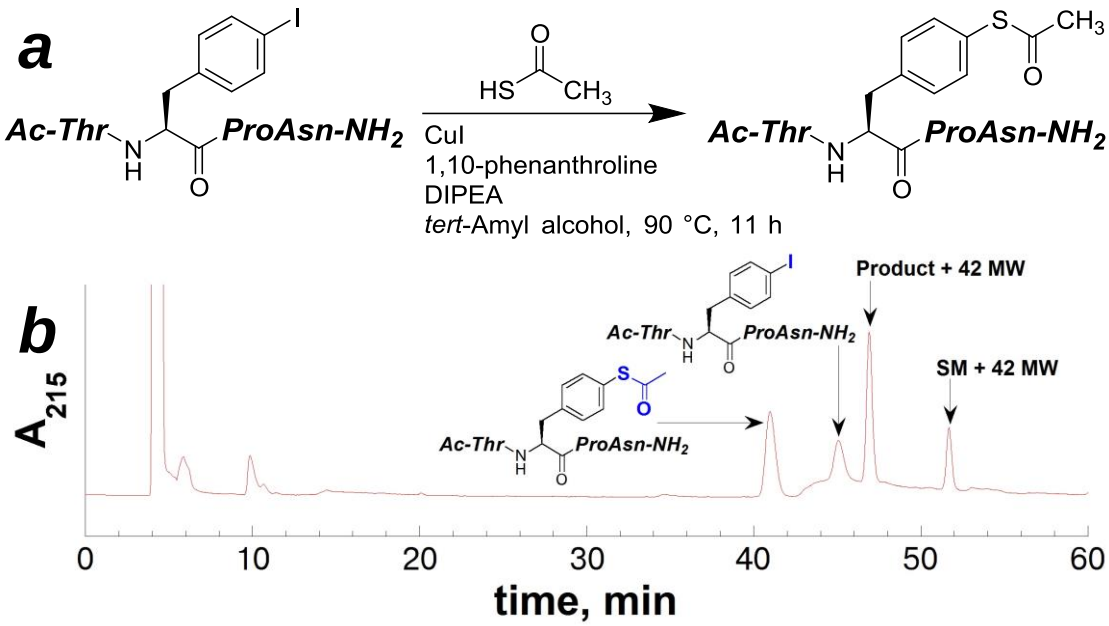


Figure 1.14 Representative HPLC chromatogram of the solution phase copper-mediated cross-coupling on the peptide Ac-T(4-I-Phe)PN-NH_2

(a) Reaction scheme of solution phase copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2 peptide (Table 1.3, entry 11); (b) analytical HPLC chromatogram (UV detection at 215 nm) of crude peptide products resulting from optimized copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2 (Table 1.3, entry 11), using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.

"SM + 42 MW" refers to a product from the reaction that was consistent with the peptide Ac-T(4-I-Phe)PN-NH_2 with an additional functional group of 42 mass units (identified via ESI-MS). This additional 42 mass units is consistent with an additional acetyl group on the peptide.

"Product + 42 MW" refers to a product from the reaction that was consistent with the peptide Ac-T(4-SAc-Phe)PN-NH_2 with an additional functional group of 42 mass units (identified via ESI-MS). This additional 42 mass units is consistent with an additional acetyl group on the peptide.

This copper-mediated cross-coupling reaction approach was successfully utilized for synthesizing the model peptide containing 4-S(acetyl)-thiophenylalanine from the peptide Ac-T(4-I-Phe)PN-NH_2 in solution. The resultant peptide from the reaction, Ac-T(4-SAc-Phe)PN-NH_2, was subjected to thiolysis reaction in solution using dithiothreitol (DTT) to generate the peptide Ac-T(4-SH-Phe)PN-NH_2 (described in the next section, Chapter 1.2.3). However, additional side-products were identified as a result of this reaction (Figure 1.14b, 46.9 minutes and 51.7 minutes, respectively). Both of the reactant and product peptides, containing Ac-T(4-I-Phe)PN-NH_2 or Ac-T(4-SAc-Phe)PN-NH_2, respectively, had associated side products that were consistent with an undesired product. The product masses (each respective mass +42 mass units) were suggestive of acetylation under the reaction conditions, potentially of the threonine hydroxyl, the asparagine carboxyamide, or the C-terminal carboxyamide.

Two possibilities were considered: first, that the reaction conditions generated a reactive acylating agent, such as thiolacetic anhydride, which could react with the peptide to effect acetylation; second, acetylation could result from direct reaction of a peptide with the thioacetyl peptide product. In order to address these possibilities, the peptide Ac-T(4-I-Phe)PN-NH_2 was subjected to the reaction conditions (Table 1.3, entry 6) but without any copper(I) iodide (Figure 1.15). These conditions generated a side-product that was consistent with acetylation on the peptide Ac-T(4-I-Phe)PN-NH_2 (determined via ESI-MS, Ac-T(4-I-Phe)PN-NH_2 +42 mass units), and, as expected, did not generate any of the peptide Ac-T(4-SAc-Phe)PN-NH_2 as a reaction product. These reaction products indicated that the reaction conditions caused these additional acetylated side-products, and did not result from inherent reaction with the product peptide Ac-T(4-SAc-Phe)PN-NH_2. In addition, the peptide Ac-TYPN-NH_2, which

contains no aryl halide as a reactant functional group, was also subjected to the copper-mediated cross-coupling conditions in the presence of copper(I) iodide (Figure 1.16). Side-products were observed as a result of this reaction, consistent with acetylation on the peptide Ac-TYPN-NH_2 at one or more functional groups (Figure 1.16c). These side-products that resulted from the reaction conditions on the peptide Ac-TYPN-NH_2, which had no aryl halide functional group, further indicated that the reaction conditions caused acetylation on the side-chain functional groups of the peptide. The reaction conditions potentially generate thiolacetic anhydride, which is a reactive acylating reagent (Figure 1.12).

The reaction conditions were screened for the copper-mediated cross-coupling reactions on the peptide Ac-T(4-I-Phe)PN-NH_2 as shown in Table 1.3. Multiple products were observed, and the results are shown as conversion to the desired peptide Ac-T(4-SAc-Phe)PN-NH_2, and as overall conversion to products resulting from the cross-coupling reaction (shown in parentheses, Table 1.3). In reactions where DMF or 20% H_2O in *tert*-amyl alcohol were used as a solvent, the products did not result in significant reaction conversion to desired products (Table 1.3, entries 1 and 2), consistent with comparable reactions in the screens with iodobenzene (Table 1.2). Potentially, the presence of water in the solvents increased the potential for thiolysis of the desired peptide Ac-T(4-SAc-Phe)PN-NH_2 and subsequent side reactions, similar to observations from the reaction screen with iodobenzene (Table 1.2, entries 6 and 7). The base was also screened during these reaction conditions, since the concentration or steric hindrance of the base can potentially play a role in the potential thiolysis and subsequent side-reactions. Products resulting from the side-reactions were suppressed by using 0.5 M Na_2CO_3 rather than DIPEA (Table 1.3, entry 3 vs. entry 5), but the risk

of epimerization of the peptide products makes this approach less ideal than using DIPEA. Interestingly, higher concentrations of DIPEA (Table 1.3, entries 4-6) resulted in a slightly higher percent conversion to the desired peptide Ac-T(4-SAc-Phe)PN-NH_2 over the thiolysis side products. Using cross-coupling reaction conditions with 1 M DIPEA in *tert*-amyl alcohol, the highest conversion was reached after 9-11 hours of reaction time, but little product was recovered after 24 hours, and only thiolysis side-products were observed. Based on this reaction screen, the optimal reaction conditions to generate the peptide Ac-T(4-SAc-Phe)PN-NH_2 utilize 1 M DIPEA in *tert*-amyl alcohol for 9-11 hours.

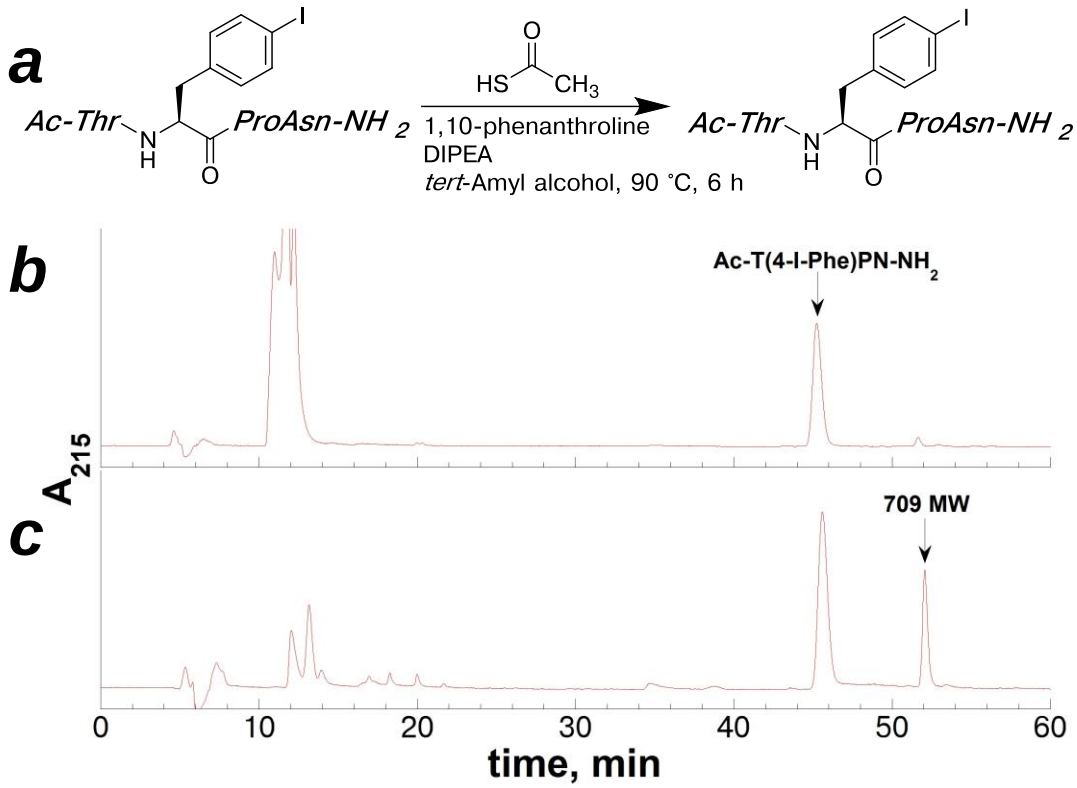


Figure 1.15 Solution phase reaction on the peptide Ac-T(4-I-Phe)PN-NH_2 in the absence of copper(I) iodide

(a) Reaction scheme of solution phase reaction on the peptide Ac-T(4-I-Phe)PN-NH_2 in the absence of copper(I) iodide; (b) HPLC chromatogram of the resultant peptide products from the reaction conditions on the peptide Ac-T(4-I-Phe)PN-NH_2 after 1 hour; (c) HPLC chromatogram of the resultant peptide products from the reaction conditions on the peptide Ac-T(4-I-Phe)PN-NH_2 after 6 hours. HPLC chromatograms where conducted using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.

"709 MW" refers to a product from the reaction that was consistent with the peptide Ac-T(4-I-Phe)PN-NH_2 with an additional functional group of 65 mass units (identified via ESI-MS). This additional 65 mass units is consistent with an additional acetyl group on the peptide (as a sodium ion: 65 mass units = acetyl (42 mass units) + Na^+ (23 mass units).

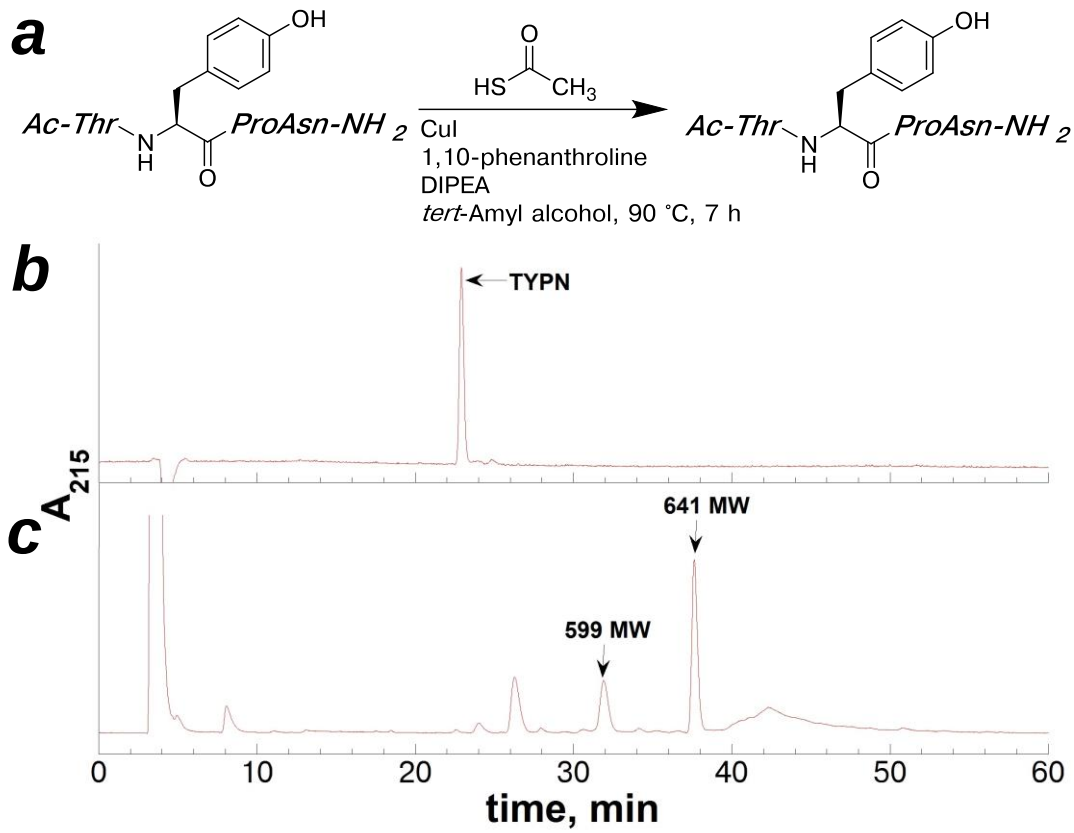


Figure 1.16 Solution phase reaction on the peptide Ac-TYPN-NH_2
(a) Reaction scheme of the solution phase reaction conditions on the peptide Ac-TYPN-NH_2; (b) analytical HPLC chromatogram (UV detection at 215 nm) of the purified peptide Ac-TYPN-NH_2; (c) HPLC chromatogram of the peptide products resulting from the reaction conditions on the peptide Ac-TYPN-NH_2. HPLC chromatograms where conducted using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.
"599 MW" refers to a product from the reaction that was consistent with the peptide Ac-TYPN-NH_2 with an additional functional group of 65 mass units. This additional 65 mass units is consistent with an additional acetyl group on the peptide (as a sodium ion: 65 mass units = acetyl (42 mass units) + Na^+ (23 mass units).
"641 MW" refers to a product from the reaction that was consistent with the peptide Ac-TYPN-NH_2 with an additional functional group of 107 mass units. This additional 107 mass units is consistent with two additional acetyl groups on the peptide (as a sodium ion: 107 mass units = 2 × acetyl (42 mass units) + Na^+ (23 mass units).

The results above indicate that cross-coupling reaction to generate the peptide containing thiophenylalanine can be achieved. However, due to the observation of peptide acetylation under the reaction conditions, we next examined the possibility of utilizing the cross-coupling reaction approach on the *protected* peptide, by subjecting the synthesized peptide on solid phase to cross-coupling reaction conditions (Figure 1.17). Use of this approach would protect reactive functional groups from potential side reaction with thiolacetic anhydride, potentially generating the desired peptide Ac-T(4-SAc-Phe)PN-NH_2 in high yield without side-reaction.

1.2.3 Optimization of copper-mediated cross-coupling conditions for peptides containing 4-iodophenylalanine on solid phase

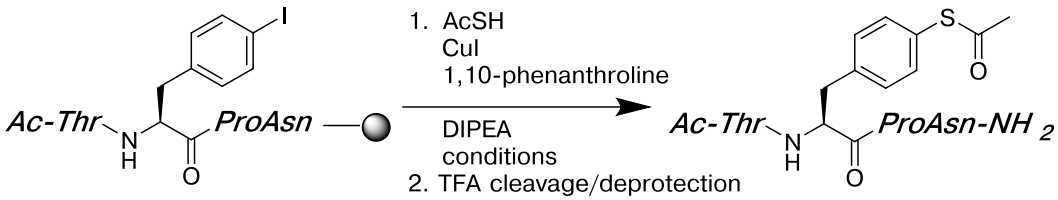
The copper-mediated cross-coupling reaction on the peptide Ac-T(4-I-Phe)PN-NH_2 successfully synthesized the peptide Ac-T(4-SAc-Phe)PN-NH_2. Reactions on peptides suggested that the reaction conditions caused acetylation reactions on the amino acid side-chains, unrelated to the cross-coupling reaction with 4-iodophenylalanine and thiolacetic acid. Therefore, we examined the possibility of utilizing the *protected* peptide containing 4-iodophenylalanine on solid-phase as the cross-coupling reaction substrate. With this approach, the side-chain functional groups would be protected from potential side-reactions.

In order to establish the synthesis of 4-thiophenylalanine within protected peptides solid-phase, the model peptide Ac-T(4-I-Phe)PN was synthesized on Rink amide resin. The threonine hydroxyl was protected as a *tert*-butyl ether, and the asparagine was protected as a trityl amide. The copper-mediated cross-coupling reaction conditions were screened using the resin-bound peptide containing 4-iodophenylalanine, based on the reaction conditions that were optimized using the

peptide Ac-T(4-I-Phe)PN-NH_2. In addition to *tert*-amyl alcohol, toluene was also examined as a solvent in the reaction screen, since it was the solvent utilized by Sawada in similar copper-catalyzed cross-coupling reactions with iodobenzene and thiobenzoic acid.[88] Using the protected peptide, where the side chains were protected with protecting groups, and with the peptide bound to a polystyrene support, the requirement for a polar solvent for fully solubilizing the peptide was eliminated, and less polar solvents such as toluene could be utilized. Following the copper-mediated cross-coupling reaction on solid phase, the peptide products were subjected to cleavage and deprotection reaction using TFA, which removed all side-chain protecting groups. A representative reaction scheme and resultant chromatogram of reaction products is shown in Figure 1.17.

Table 1.4. Optimization of the copper-mediated cross-coupling reaction on the peptide Ac-T(4-I-Phe)PN-NH_2 on solid phase

Crude reaction products were evaluated via HPLC. Peak identity was established via ESI-MS. Product conversions were normalized relative to total peak integrations, and are calculated based on product conversion for desired product (as the peptide Ac-T(4-SAc-Phe)PN-NH_2) and as total products resulting from the cross-coupling reaction (including disulfide side-products) shown in parentheses.



Entry	Solvent	Temperature (°C)	Reaction Duration	Product Conversion
1	*tert*-Amyl alcohol	90	6 h	40% (45%)
2	*tert*-Amyl alcohol	100	6 h	59% (69%)
3	*tert*-Amyl alcohol	100	11 h	37% (85%)
4	*tert*-Amyl alcohol	100	24 h	53% (90%)
5	*tert*-Amyl alcohol	110	16 h	0% (99%)
6	Toluene	100	18 h	57% (67%)
7	Toluene	110	16 h	22% (99%)

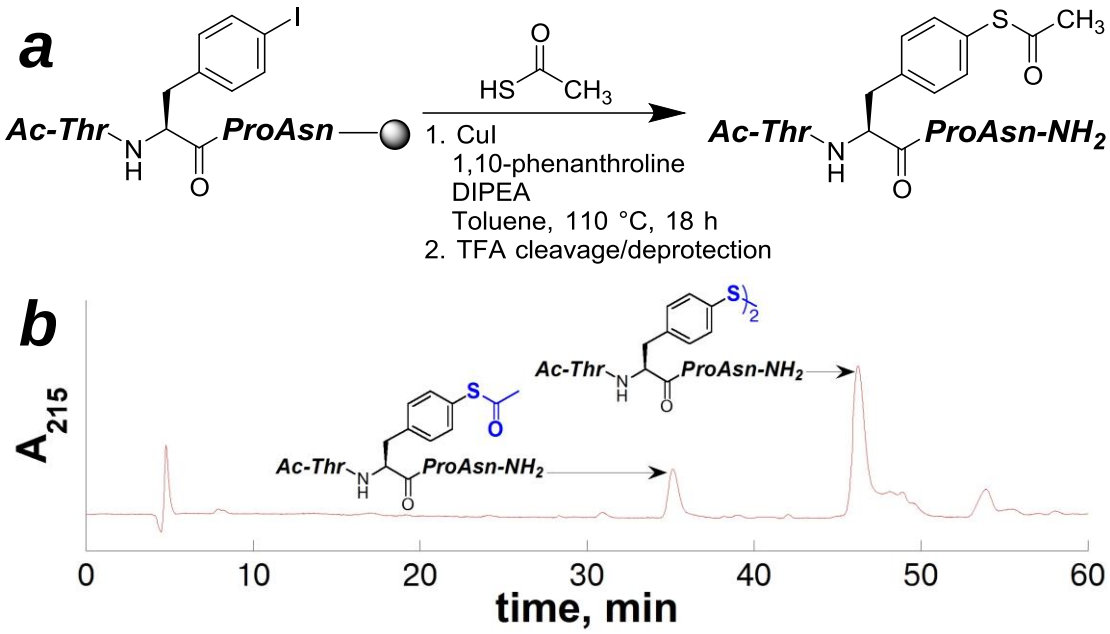


Figure 1.17 Optimized conditions for copper-mediated cross-coupling reaction on the peptide Ac-T(4-I-Phe)PN-NH_2 on solid phase
(a) Optimized reaction scheme of solid phase copper-mediated cross-coupling conditions on the peptide Ac-T(4-I-Phe)PN-NH_2; (b) analytical HPLC chromatogram (UV detection at 215 nm) of crude peptide products resulting from optimized copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2, and subsequent cleavage from resin, using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.

These copper-mediated cross-coupling reactions on the protected peptide Ac-T(4-I-Phe)PN on solid-phase successfully generated the peptides containing derivatives of 4-thiophenylalanine in high conversion. Two products were observed as a result of the cross-coupling reactions on solid phase: the desired peptide Ac-T(4-SAc-Phe)PN-NH_2 and the corresponding symmetric disulfide, which potentially resulted from thiolysis during the reaction. The reactions with iodobenzene indicated

that thiolysis of the S-phenyl thioacetate product during the reaction generated the thiophenolate, which could compete with the thiolacetic acid for the cross-coupling reaction, generating biaryl thioether products. Therefore, the ideal reaction conditions would generate high conversion of peptides containing 4-thiophenylalanine derivatives, with minimal conversion of the product to the disulfide by product.

For reactions that were conducted in *tert*-amyl alcohol, increasing the reaction temperature (from 90 °C to 100 °C and 110 °C) and reaction time (from 6 hours to 24 hours) significantly improved the overall conversion to cross-coupled reaction products. Consistent with reactions with iodobenzene, the cross-coupling reaction on peptides on solid phase required high reaction temperatures, as product conversion was reduced if the reaction was conducted at 90 °C (Table 1.4, entry 1). In *tert*-amyl alcohol, the highest reaction conversion was obtained when the reaction was conducted at 110 °C (Table 1.4, entry 5). However, at high reaction temperature, the product conversion was mostly in the form of disulfides rather than the peptide containing 4-S(acetyl)-thiophenylalanine (Table 1.4, entries 5 and 7). These products indicate that thiolysis of the reaction product, the peptide Ac-T(4-SAc-Phe)PN-NH_2, is more likely at higher reaction temperature. With toluene as the reaction solvent at 110 °C (Table 1.4, entry 7), there was reduced potential for thiolysis of the peptide Ac-T(4-SAc-Phe)PN-NH_2, with overall reaction conversion nearly complete. No additional acetylation reactions on the peptide side-chains were observed as a result of the cross-coupling reactions with the peptide containing 4-iodophenylalanine on solid phase.

The copper-mediated cross-coupling reaction conditions can be used to efficiently generate peptides containing 4-S(acetyl)-thiophenylalanine and the corresponding disulfides on solid phase. Aryl thioethers can be converted into thiols, and disulfides can be reduced to thiols. Thus, both of the peptide products that resulted from the copper-mediated cross-coupling reaction on the peptide Ac-T(4-I-Phe)PN can be converted to the desired peptide containing 4-thiophenylalanine. In order to generate the peptide Ac-T(4-SH-Phe)PN-NH_2, the cleaved, deprotected, crude peptides resulting from the cross-coupling reaction were subjected to solution-phase thiolysis and reduction using DTT under neutral or basic conditions (Figure 1.18). Alternatively, TCEP or sodium borohydride can be used to produce similar results, without requiring neutral or basic conditions.

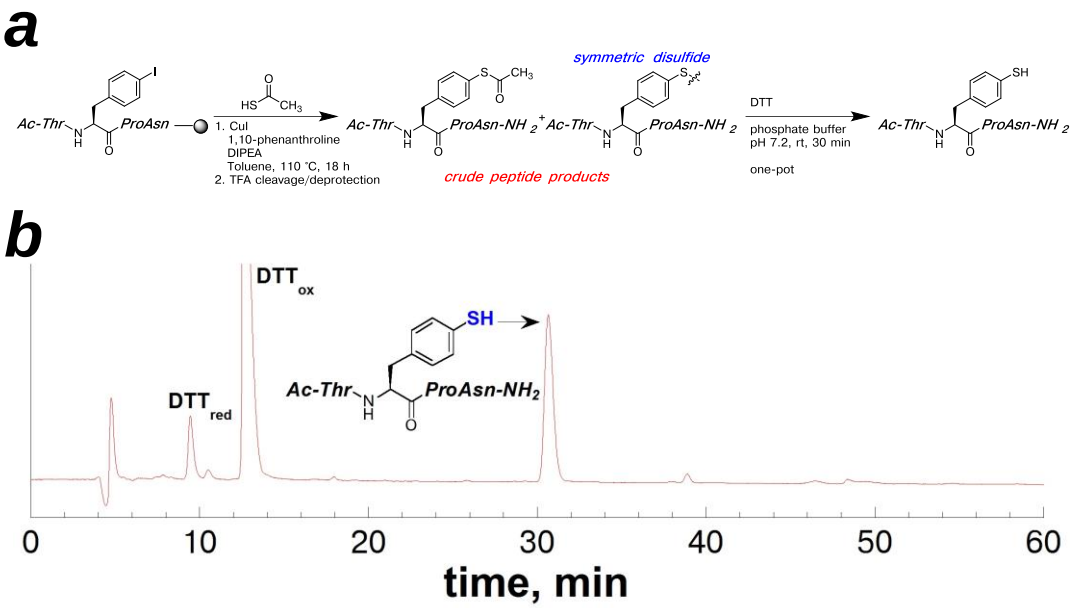


Figure 1.18 Synthesis of the peptide Ac-T(4-SH-Phe)PN-NH_2 via cross-coupling reaction on solid phase and subsequent thiolysis in solution
(a) Reaction scheme of solid phase copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2 peptide followed by TFA cleavage from resin and solution phase reduction and thiolysis of the crude reaction products; (b) analytical HPLC chromatogram (UV detection at 215 nm) of crude peptide products resulting from optimized copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2, and subsequent reaction with DTT (pH 7.2) after cleavage from resin, using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.
DTT_{red} and DTT_{ox} indicates dithiothreitol in reduced and oxidized forms, respectively.

The peptide Ac-T(4-SH-Phe)PN-NH_2 was cleanly synthesized after solution-phase thiolysis and reduction reaction using the products that resulted from the copper-mediated cross-coupling reaction on the peptide containing 4-iodophenylalanine on solid phase. This synthetic approach to peptides containing 4-thiophenylalanine utilizes commercially available Fmoc-4-iodophenylalanine and inexpensive reagents (i.e. copper(I) iodide, thiolacetic acid, 1,10-phenanthroline). In

contrast to all other approaches to synthesize 4-thiophenylalanine,[16, 20, 22, 23, 58] this synthetic strategy did not require a multi-step solution synthesis to prepare the protected amino acid for use in SPPS. Using this approach, a peptide containing 4-thiophenylalanine can be synthesized in less than 24 hours with only one HPLC purification step, by using the protected peptide containing 4-iodophenylalanine and reaction on solid phase.

The aim of this work was to generate peptides containing 4-thiophenylalanine in peptides in a practical manner using a copper-mediated cross-coupling approach. The reaction conditions shown in Figure 1.17 generated the peptides Ac-T(4-SAc-Phe)PN-NH_2 and corresponding disulfides in 99% conversion, which were both cleanly converted to Ac-T(4-SH-Phe)PN-NH_2 in solution following the cross-coupling reaction. However, in some applications, it may be preferable to generate only the peptide containing 4-S(acetyl)-thiophenylalanine from the cross-coupling reaction on solid phase. The HPLC chromatogram representing the conditions that resulted in high conversion to the peptide Ac-T(4-SAc-Phe)PN-NH_2, with the lowest conversion to the corresponding disulfides, is shown in Figure 1.19.

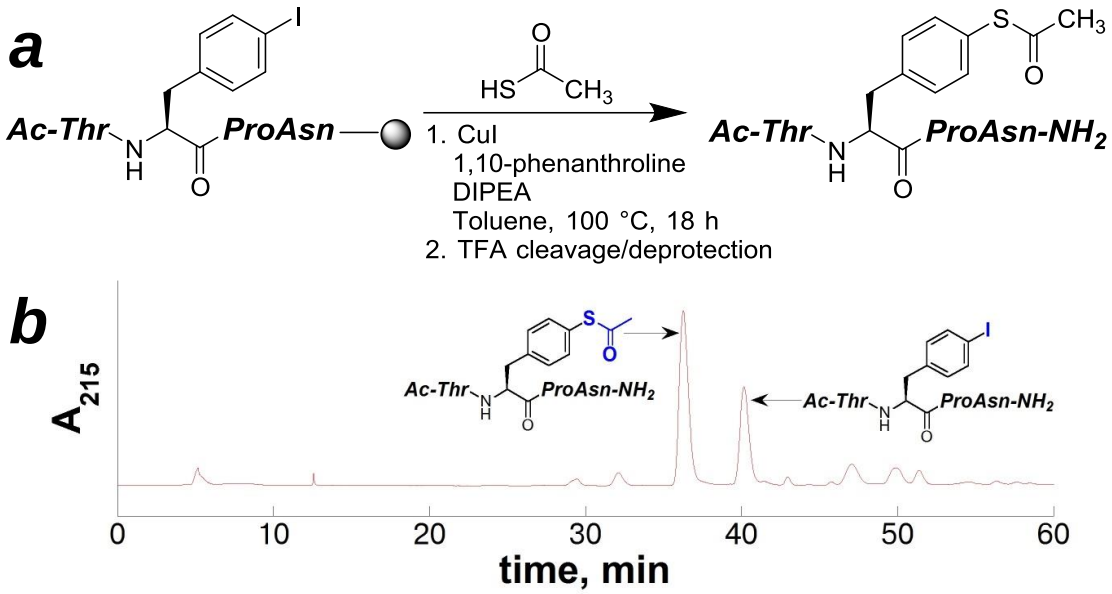


Figure 1.19 Optimized copper-mediated cross-coupling reaction to synthesize the peptide containing 4-S(acetyl)-thiophenylalanine with minimal disulfide formation

(a) Optimized reaction scheme of solid phase copper-mediated cross-coupling conditions on the peptide Ac-T(4-I-Phe)PN-NH_2; (b) analytical HPLC chromatogram (UV detection at 215 nm) of crude peptide products resulting from optimized copper-mediated cross-coupling conditions on Ac-T(4-I-Phe)PN-NH_2, and subsequent cleavage from resin, using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.

These reaction conditions generated the peptide containing 4-S(acetyl)-thiophenylalanine as the sole reaction product, and avoided thiolysis of the resultant peptide during the cross-coupling conditions. By conducting the cross-coupling reaction at a lower temperature, thiolysis of the peptide Ac-T(4-SAc-Phe)PN-NH_2, and the potential for side-chain reactivity, was suppressed during the cross-coupling reaction. However, the overall reaction conversion was only 57% from the peptide Ac-

T(4-I-Phe)PN, compared to 99% overall reaction conversion resulting from a comparable reaction that was conducted at only 10 °C higher temperature (Table 1.4, entries 6 and 7). These reaction conditions shown in Figure 1.19 can be utilized in applications where thiolysis of the peptide could lead to additional side reactions, although this approach results in significantly reduced reaction conversion.

The model peptide containing 4-thiophenylalanine was characterized via NMR at both protonation states (thiol, pH 4.0, and thiolate, pH 8.5, pK_a determination described in Chapter 1.2.7). The model peptide that was used for solid-phase reaction screening contains a proline residue, which is known to undergo cis-trans isomerism. Both the cis and trans amide bond conformations are observed by NMR (Figure 1.20). The structural details of 4-thiophenylalanine within peptides will be further discussed in Chapter 3.

The optimized copper-mediated cross-coupling reaction conditions generated the peptide Ac-T(4-SH-Phe)PN-NH_2 in 99% conversion from the peptide Ac-T(4-I-Phe)PN-NH_2 on solid phase, after reaction with DTT in solution. Ac-TXPN-NH_2 is a well-characterized tetrapeptide, and was a suitable model for optimizing the cross-coupling reaction conditions. Longer peptides with bulky side chains have the potential to decrease the reaction efficiency due to steric crowding. Furthermore, some side-chain protecting groups may not be stable to the high reaction temperatures that were required for efficient reaction conversion. Therefore, additional screening was conducted in order to determine the compatibility of the optimized cross-coupling reaction conditions with other amino acid functional groups and longer peptides.

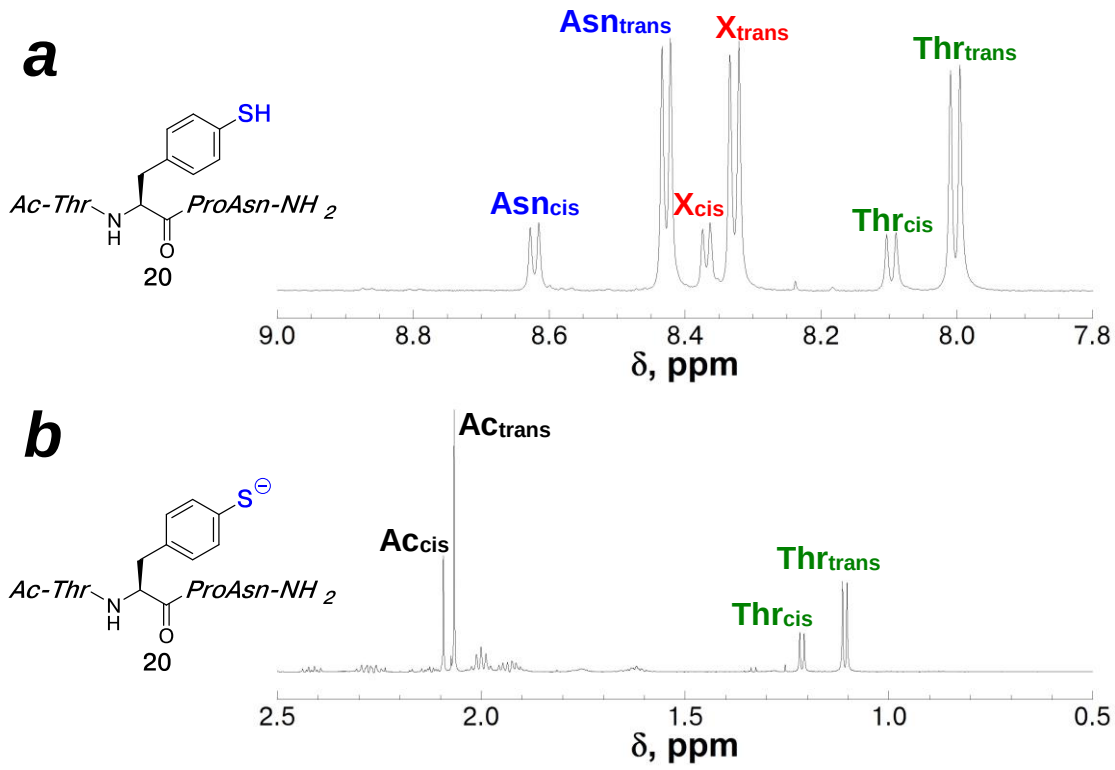


Figure 1.20 ^{1}H NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH$_2$ containing 4-thiophenylalanine
NMR spectrum for the peptide Ac-Thr(4-thiophenylalanine)ProAsn-NH$_2$ at (a) pH 4.0 and (b) pH 8.5. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by TOCSY spectrum. The aliphatic region is shown for pH 8.5, because rapid amide exchange at high pH obscures the amide peaks.

1.2.4 Peptide substrate scope of the optimized solid-phase, copper-mediated cross-coupling reaction

In order to determine the compatibility of the optimized cross-coupling reaction conditions with other side-chain functional groups, resins containing additional protected peptides with 4-iodophenylalanine were subjected to the copper-mediated cross-coupling reaction with thiolacetic acid. A series of peptides were synthesized, representing all canonical amino acid functional groups, and included peptides with diverse structures and sequence lengths.

In order to evaluate the broader compatibility of the copper-mediated cross-coupling reaction conditions on protected peptides on solid-phase, three additional peptides were synthesized on Rink amide resin and subjected to the optimized reaction conditions. The three peptides that were synthesized for this reaction scope included (where X = 4-iodophenylalanine) (1) the trp cage miniprotein (Ac-NLXIQWLKDGGPSSGRPPPS-NH_2), (2) a proline-rich peptide Ac-GPPXPPGY-NH_2, and (3) a peptide containing sensitive side-chain functional groups, Ac-KKHMCX-NH_2. The trp cage miniprotein was included in the peptide substrate scope because it contains many of the canonical amino acid functional groups. Furthermore, the trp cage miniprotein has been well characterized via NMR and CD as the subject of miniprotein stability and protein folding studies.[94-96] Therefore, the product of the cross-coupling reaction, the Y3(4-SH-Phe) variant of the trp cage miniprotein, can be used to characterize the structural consequences of tyrosine versus 4-thiophenylalanine within a peptide. The peptide Ac-GPPXPPGY-NH_2 was included in the reaction scope because a proline-rich sequence can potentially hinder the cross-coupling reaction efficiency, both due to the greater steric demand of proline and due to the possibility of formation of poly-proline helices in organic solvents.[97-99] In addition, aromatic amino acids have been shown to influence local conformation of proline-rich sequences.[100] The peptide Ac-GPP(4-SH-Phe)PPGY-NH_2, and related derivatives of 4-thiophenylalanine, can be used to evaluate the influence of this aryl thiolated amino acid on poly-proline structure. Peptides that contain amino acids that coordinate metals or are prone to oxidation, including cysteine, methionine, and histidine, may not be compatible with the cross-coupling conditions, given the use of stoichiometric copper(I) iodide in the reaction. Therefore, the peptide Ac-KKHMCX-NH_2 was

included in the substrate reaction scope in order to address the potential for metal-coordination, or oxidation, with these amino acids under the cross-coupling reaction conditions.

In order to determine the compatibility of the copper-mediated cross-coupling reaction with a broader scope of amino acid side chains, the peptides described above were synthesized on Rink amide resin containing 4-iodophenylalanine. The protected peptides were subjected to the optimized copper-mediated cross-coupling reaction on solid phase. The amino acid protecting groups in this screen included: trityl amide (Gln, Asn), *tert*-butyl ether (Thr, Ser, Tyr), *tert*-butyl ester (Glu, Asp), Boc-carbamate (Lys), trityl thioether (Cys), and trityl imidazole (His); the tryptophan indole was not protected. In addition to evaluating the resultant peptides containing 4-thiophenylalanine via HPLC and NMR, the Y3(4-SH-Phe) trp cage variant was characterized by CD (described in Chapter 1.2.6), in order to determine if the cross-coupling reaction affected the structure of the peptide substrates in a manner that could not be observed by HPLC, ESI-MS, or NMR. All of the peptides were subjected to reaction with DTT to effect thioester thiolysis and to reduce disulfides following the cleavage and deprotection reaction (Figure 1.21).

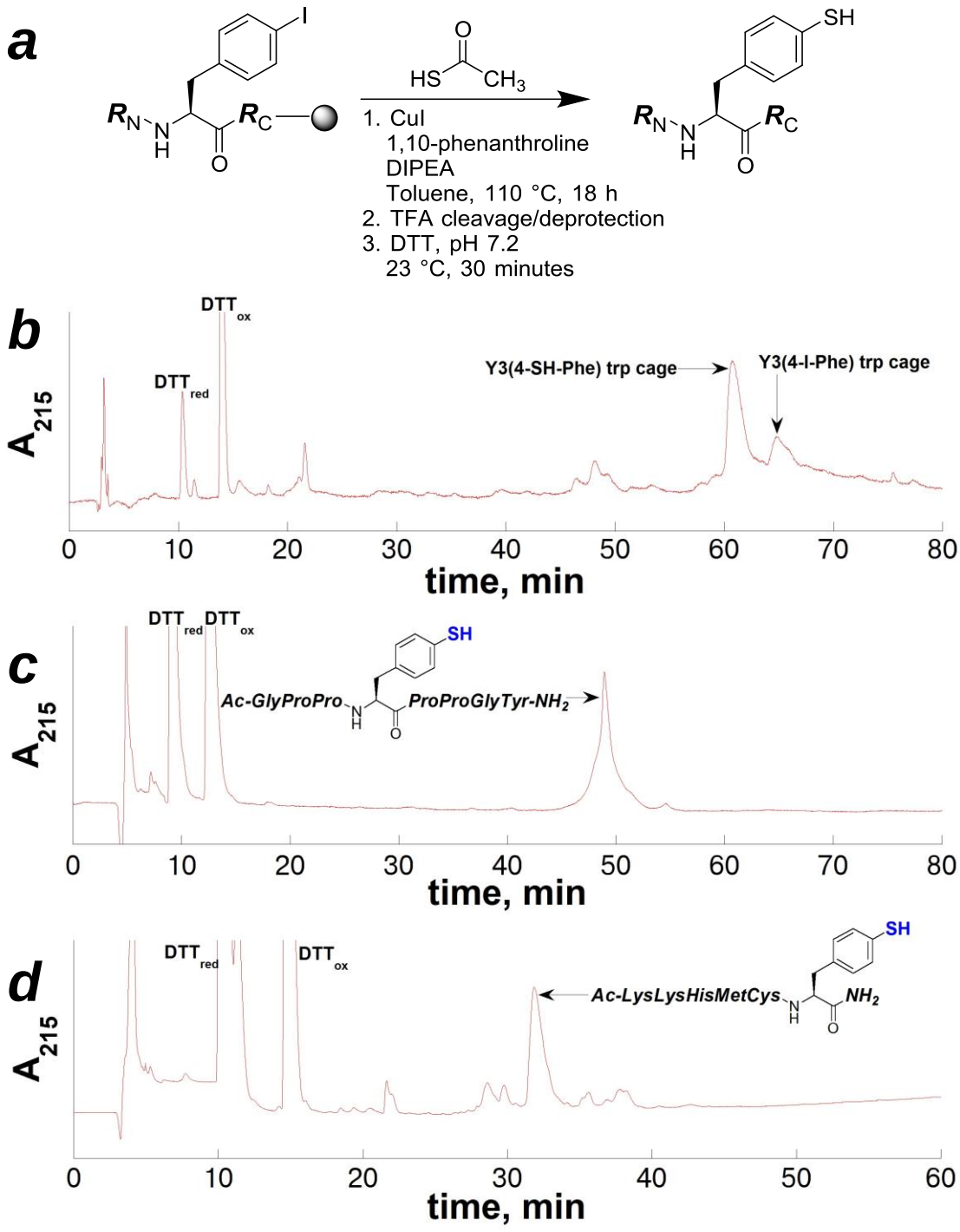


Figure 1.21 Peptide scope of solid phase, copper-mediated cross-coupling reaction to generate peptides containing 4-thiophenylalanine

(a) Reaction scheme for the copper-mediated cross-coupling reactions on peptides containing 4-iodophenylalanine on solid phase. Analytical HPLC chromatograms (UV detection at 215 nm) of the crude peptide products resulting from the optimized copper-mediated cross-coupling reaction, and subsequent reaction with DTT (pH 7.2) in solution. Peaks were identified via ESI-MS. (b) HPLC chromatogram of the resultant products from the cross-coupling reaction on the Y3(4-I-Phe) trp-cage miniprotein using a linear gradient of 0-55% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 80 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the resultant products from the cross-coupling reaction on the peptide Ac-GPP(4-I-Phe)PPGY-NH_2 using a linear gradient of 0-55% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (d) HPLC chromatogram of the resultant products from the cross-coupling reaction on the peptide Ac-KKHMC(4-I-Phe)-NH_2 using a linear gradient of 0-70% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

The optimized solid-phase, copper-mediated, cross-coupling reaction exhibited excellent peptide substrate scope, based on this set of peptide sequences. The Y3(4-SH-Phe) trp cage variant was generated via cross-coupling reaction in slightly lower conversion compared to Ac-TXPN-NH_2, 67% from the peptide containing 4-iodophenylalanine (Figure 1.21b). The peptide Ac-GPP(4-SH-Phe)PPGY-NH_2 was cleanly generated in 98% conversion (Figure 1.21c). The peptide containing oxidation-sensitive side-chains, Ac-KKHMC(4-SH-Phe)-NH_2, was generated in 90% conversion after solution phase reaction with DTT (Figure 1.21d). The cross-coupling reaction effectively generated 4-thiophenylalanine within peptides on a variety of peptide substrates.

In the peptide containing oxidation-sensitive amino acids, Ac-KKHMCX-NH_2, the cysteine thiol was protected as a trityl thioether, which is readily removed under the deprotection reaction conditions using TFA. However, it is not clear from the

resultant peptides shown in Figure 1.21d if the cysteine side-chain remained protected during the cross-coupling reaction. Potentially, the trityl protecting group on cysteine could be removed during the cross-coupling reaction, and the resultant thiol could react with thiolacetic anhydride, generating a cysteinyl thioacetate. Reaction with DTT in solution would remove the acetyl group and generate the cysteine thiol, rendering this potential side-reaction unobservable. In order to examine the potential for this side reaction on the trityl-protected cysteine side-chain, the peptide Ac-KKHMC(4-I-Phe) was subjected to cross-coupling reaction conditions in the absence of the copper catalyst (Figure 1.22). The resultant products of this reaction were not subjected to thiolysis after the cleavage and deprotection reaction with TFA. The expected product was Ac-KKHMC(4-I-Phe)-NH_2, provided there was no reaction on the cysteinyl side chain under the reaction conditions.

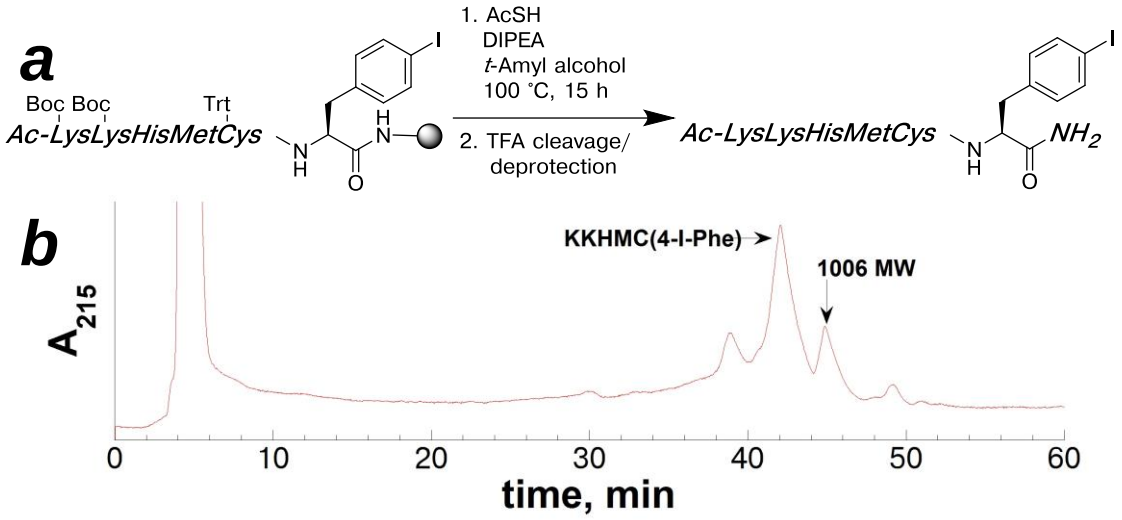


Figure 1.22 Potential side reaction on the peptide containing trityl-protected cysteine as a result of the cross-coupling reaction conditions
(a) Scheme of the reaction conditions on the peptide Ac-KKHMC(4-I-Phe)-NH_2; (b) analytical HPLC chromatogram (UV detection at 215 nm) of the crude peptide products resulting from the reaction conditions after the peptide was subjected to cleavage and deprotection reaction using a linear gradient of 0-60% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.
"1006 MW" refers to a product with a mass of 1006 (idenfited via ESI-MS). This mass is consistent with the peptide Ac-KKHMC(4-I-Phe)-NH_2 with an additional 43 mass units, suggesting an additional acetyl group (Ac-KKHMC(4-I-Phe)-NH_2 + Ac + H^+).

The peptide Ac-KKHMC(4-I-Phe)-NH_2 that was subjected to cross-coupling reaction conditions, without any copper catalyst, on solid-phase resulted in a product that was consistent with the peptide containing an additional acetyl group. Potentially, the acetyl group is attached to the cysteinyl thiol, as a result of trityl deprotection under the reaction conditions and subsequent reaction with thiolacetic anhydride. This product containing a cysteinyl thioacetyl group can be converted to the thiol via reaction with DTT in solution. However, deprotection of a cysteine thiol is undesirable

for some applications, including syntheses where the thiol-protecting groups on cysteine and 4-thiophenylalanine must be orthogonal. Therefore, the synthetic approach for generating peptides containing both 4-thiophenylalanine and cysteine would need to be revised.

In order to examined the potential for generating of peptides containing orthogonal protecting groups on cysteine and 4-thiophenylalanine, the peptide Ac-KKHMC(4-I-Phe) was synthesized with an alternative side-chain protecting group on cysteine. The cysteine thiol was protected as an S-*tert*-butyl thioether, and the peptide was subjected to the optimized copper-mediated cross-coupling reaction conditions (Figure 1.23).

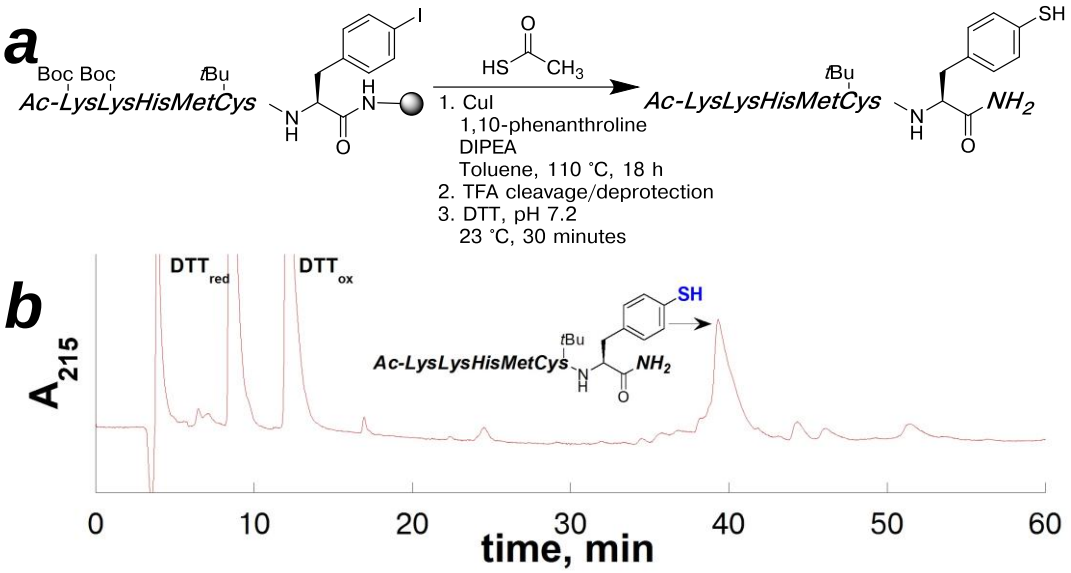


Figure 1.23 Copper-mediated cross-coupling reaction on the peptide Ac-KKHM(*tert*-butyl-Cys)(4-I-Phe) on solid phase
(a) Optimized reaction scheme for copper-mediated cross-coupling conditions on the peptide Ac-KKHMC(*t*Bu)(4-I-Phe) on solid phase; (b) analytical HPLC chromatogram (UV detection at 215 nm) of the crude peptide products resulting from the cross-coupling reaction, and subsequent reaction with DTT (pH 7.2) after cleavage from resin using a linear gradient of 0-75% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Peaks were identified via ESI-MS.

The peptide containing histidine, methionine, and S-*tert*-butyl-protected cysteine had no observed cross-reactivity under the copper-mediated cross-coupling reaction conditions, generating the peptide containing 4-thiophenylalanine in 90% conversion. Combined, these data indicate that the optimized copper-mediated cross-coupling reaction is compatible with all canonical amino acid functional groups. Peptides up to 20 residues in length were synthesized with 4-thiophenylalanine in good conversion, suggesting reaction compatibility with diverse peptide sequences. The lower yield for the reaction on the trp cage miniprotein variant could be due to

"steric crowding" on the resin, or due to the proximity of the 4-iodophenylalanine to bulky, protected amino acid side chains (Trp, Ile, trityl-Gln). Potentially, a lower loading-capacity on the resin might improve the conversion for longer peptides.

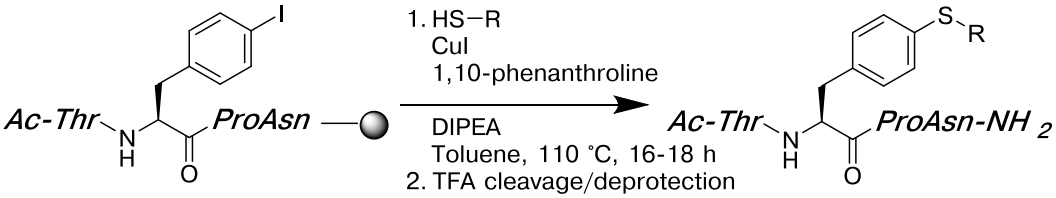
1.2.5 Thiol substrate scope of the optimized solid-phase, copper-mediated cross-coupling reaction

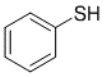
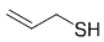
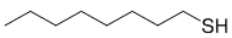
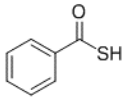
The synthesis of 4-S(acetyl)-thiophenylalanine within peptides from thiolacetic acid and peptides containing 4-iodophenylalanine is a practical approach for synthesizing 4-thiophenylalanine .However, a diverse range of thiol, thioester, and thioether products can also be generated using this solid-phase cross-coupling reaction approach. Biaryl ethers are found in some macrocyclic peptides as a result of oxidative coupling reactions with tyrosine.[61] In the work by Escher and coworkers,[58] 4-thiophenylalanine was alkylated with a variety of different thioether functional groups, and these products were utilized for structure-activity relationship analysis in angiotensin analogues. Thioether and thioester derivatives of 4-thiophenylalanine might be synthesized more directly by utilizing alternative thiols in the cross-coupling reaction with peptides containing 4-iodophenylalanine.

In the initial screening reactions on iodobenzene, acidic thiols demonstrated increased reactivity for the cross-coupling reaction conditions, including thiophenol (Table 1.1). In order to examine the role of the acidity of the thiol in the copper-mediated cross-coupling reaction, a series of mercaptans was screened using the optimized cross-coupling conditions on the peptide Ac-T(4-I-Phe)PN on solid phase. The products that resulted from the cross-coupling reactions were evaluated via analytical HPLC, and conversion was established via integration of the peaks in the HPLC chromatograms (Table 1.5).

Table 1.5. Thiol substrate scope of the cross-coupling reaction on resin-bound peptide Ac-T(4-I-Phe)PN-NH_2

Crude reaction products were evaluated via HPLC. Peak identity was established via ESI-MS. "Product conversion" refers strictly to the formation of the peptide containing 4-SR-Phe. Product conversions were normalized relative to total peak integrations.
Conversions to total cross-coupled products, including disulfides, trisulfides, and tetrasulfide peptide products, are shown in the in parentheses.



Entry	Thiol Substrate	Temperature (°C)	Ac-T(4-SR-Phe)PN-NH_2 product conversion
1		110	22% (99%)
2		110	90%
3		110	2%
4		110	0%
5		100	21% (45%)
6		100	57% (67%)

Alkyl thiols demonstrated minimal reactivity towards peptides containing 4-iodophenylalanine on solid-phase in comparison to thiolacetic acid (Table 1.5, entries 3 and 4). In contrast, thiophenol efficiently reacted with Ac-T(4-I-Phe)PN-NH_2 to

generate the biaryl thioether product under these conditions (Table 1.5, entry 2). Thiobenzoic acid also exhibited efficient reactivity in the cross-coupling reaction, although this reaction generated an unusual mixture of products. Interestingly, products consistent with trisulfides (–S–S–S–) and tetrasulfides (–S –S–S–S–) were observed in the reaction with thiobenzoic acid and the peptide Ac-T(4-I-Phe)PN-NH_2 (identified via ESI-MS). From a mechanistic perspective, it is unclear how the tri-sulfide and tetra-sulfide side-products form in the cross-coupling reaction utilizing thiobenzoic acid, or why these products were not observed for thiolacetic acid cross-coupling reactions. Trisulfides were previously observed by Davis & coworkers[101] in a reaction with sodium methanethiosulfonate and an alkyl halide, but the reaction mechanism for formation of this product was not explained. Further work on these reaction products would be necessary to fully characterize the trisulfide and tetrasulfide products and to establish a mechanism of formation. Collectively, these data indicate a requirement for a relatively acidic thiol for successful copper-mediated cross-coupling reaction under these conditions.

The model peptides synthesized in significant yield from these reactions using different thiol substrates were characterized via NMR (Figures 1.24-1.26). The model peptide has a high propensity for adopting the cis-proline conformation, and so the two proline amide-bond rotamers of the peptide, cis proline and trans proline, are both observable by NMR. Structural analysis of these peptides will be further discussed in Chapter 3.

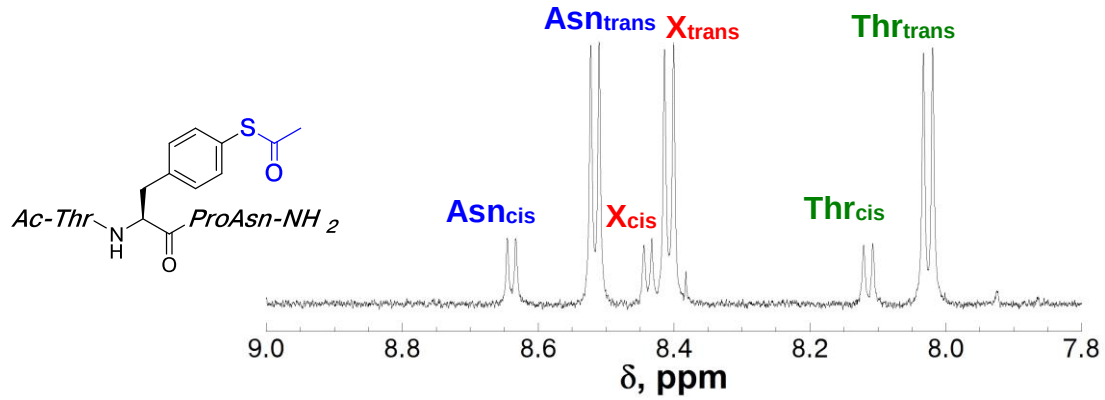


Figure 1.24 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S(acetyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH_2 (where X = 4-S(acetyl)-thiophenylalanine) at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by analogy to related peptides.

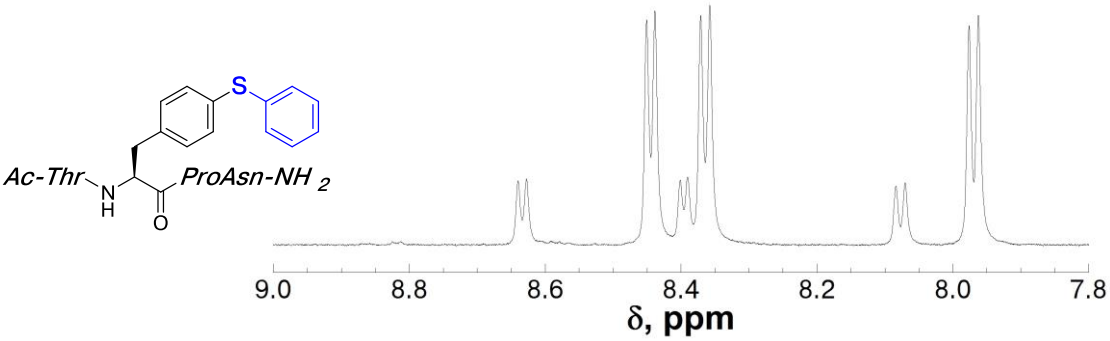


Figure 1.25 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S(phenyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH_2 (where X = 4-S(phenyl)-thiophenylalanine) at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by analogy to related peptides.

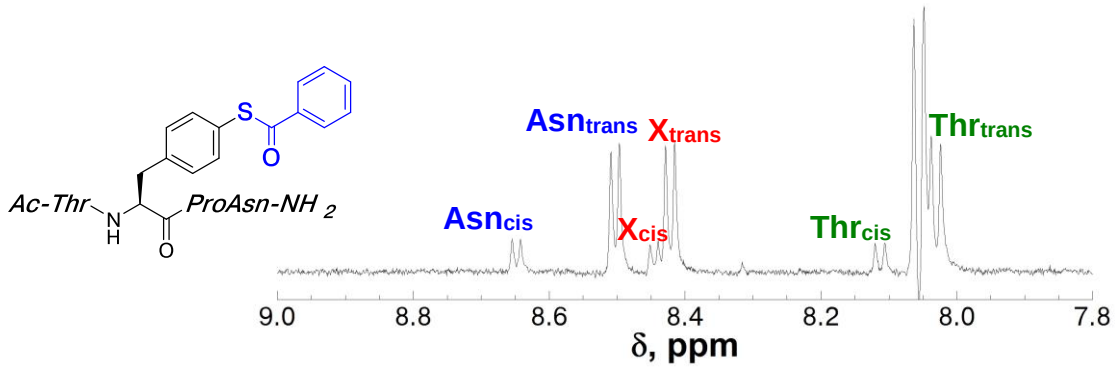


Figure 1.26 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S(benzoyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH_2 (where X = 4-S(benzoyl)-thiophenylalanine) at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were via TOCSY spectrum.

1.2.6 Structural characterization of the 4-thiophenylalanine-containing trp cage miniprotein

In Chapter 1.2.4, the Y3(4-SH-Phe) variant of the trp cage miniprotein was successfully generated via the copper-mediated cross-coupling reaction without side reactions. The trp cage miniprotein is a designed protein that is comprised of an α-helix and a polyproline helix linked by a short, flexible sequence of amino acids (Figure 1.27). The relatively short trp cage miniprotein exhibits remarkable structural stability, which is well characterized via circular dichroism (CD) and NMR.[94, 95] The trp cage miniprotein was derived from exendin-4 (also Exenatide or Byetta), an FDA approved glp-1 agonist that is used in treatment of type 2 diabetes, originally isolated from Gila monster saliva.[94, 102] Designed stabilized or modified variants of the trp cage miniprotein can potentially be used as templates or scaffolds for bioactive peptides.

The folding and stability of the trp cage miniprotein have been extensively examined by solution, solid-phase, and computational techniques.[94-96, 103] The small size of the trp cage miniprotein has led to its wide use in computational investigation

of protein folding pathways and mechanisms.[104] With the extensive structural characterization of the trp cage miniprotein, it represents an ideal model for examining potential effects of the copper-mediated cross-coupling reaction on protein folding and stability.

The stability of the trp cage is based, in part, upon non-covalent contacts surrounding Trp6 (Figure 1.27), which provides hydrophobic contacts for the proline residues on the adjacent helix (Pro18, Pro19) and for Tyr3 (highlighted in red, Figure 1.27).[105] The aromatic ring of Tyr3 participates in C–H/π interactions with Pro19.[105] Given the electronic nature of C–H/π interactions,[91, 93, 106] the overall stability of the trp cage is potentially tunable by modulating the electronic effects in the aromatic ring at position 3. This possibility was explored using 4-thiophenylalanine, the sulfur analogue of tyrosine.

In order to examine the structural implications of 4-thiophenylalanine and the copper-mediated cross-coupling reaction conditions on the trp cage miniprotein, the Y3(4-SH-Phe) trp cage variant was synthesized (Chapter 1.2.4) and characterized. Circular dichroism (CD) and thermal denaturation data were obtained for the Y3(4-SH-Phe) trp cage variant, and compared to the "native" trp cage miniprotein containing Tyr3 (Figure 1.28, Table 1.6).

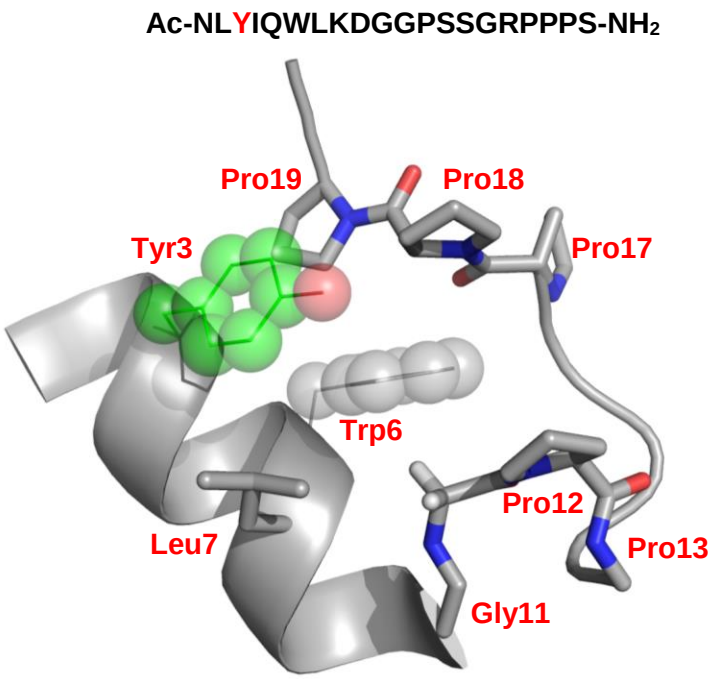


Figure 1.27 trp cage miniprotein
Cartoon representation of the trp cage miniprotein (PDB ID: 1L2Y). Key residues for stability are shown as sticks. Aromatic side chains are shown as spheres. The Y3(4-SH-Phe) trp cage miniprotein variant replaces Tyr3 with 4-thiophenylalanine (site shown in green).

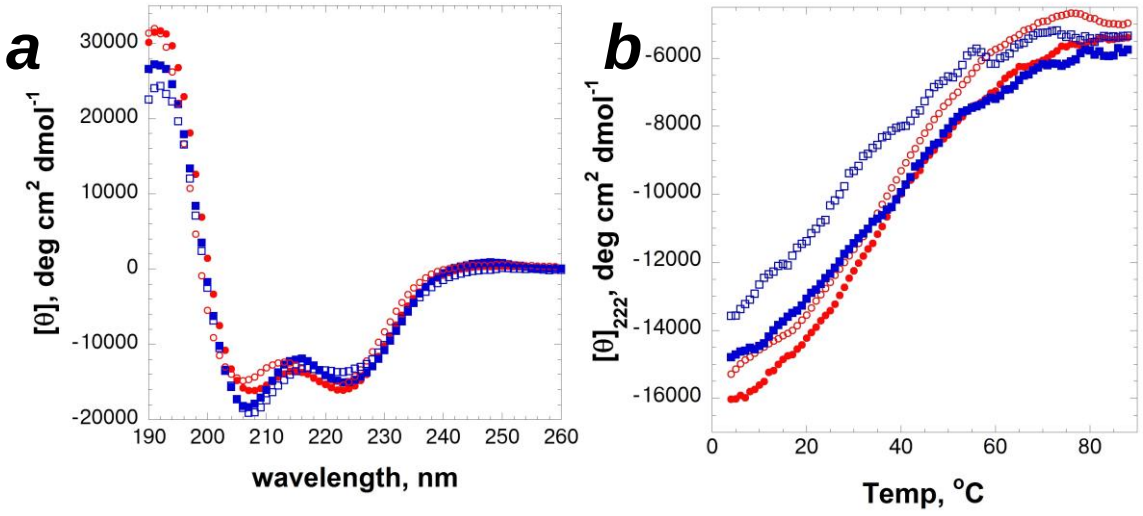


Figure 1.28 Circular dichroism (CD) and thermal denaturation data of the trp cage miniprotein

Circular dichroism of native trp cage miniprotein (red circles) and Y3(4-SH-Phe) trp cage miniprotein variant (blue squares). Closed figures indicate pH 4.0. Open figures indicate 7.0 or 8.5 (for native or modified trp cage, respectively). Samples were measured in 15 mM phosphate in the presence of TCEP (0.56 mM) to prevent disulfide formation. Concentrations of stock solutions of peptides containing 4-thiophenylalanine were measured via Ellman's test. All spectra were background subtracted. (a) CD spectra of peptides at 4 ˚C. A cartoon representation of the Y3(4-SH-Phe) modified trp cage miniprotein is shown in the inset; (b) thermal denaturation data of the native trp cage miniprotein and the Y3(4-SH-Phe) variant at the indicated pH.

Table 1.6. Derived thermal denaturation and CD data for modified and native trp cage miniproteins

Values were calculated from the plots shown in Figure 1.28. Samples were dissolved in 15 mM phosphate buffer at the indicated pH. Samples containing the Y3(4-SH-Phe) variant contained 0.56 mM TCEP to reduce disulfides. [θ] = mean residue ellipticity (deg cm^2 dmol^{-1}) at the indicated wavelength at 4 °C. T_m = melting temperature, determined using a slope for the 100% unfolded baseline (d[θ]$_{222}$/dT) of –8.83 deg cm^2 dmol^{-1}/°C; and a slope for the 100% folded baseline (d[θ]$_{222}$/dT) of 63.5 deg cm^2 dmol^{-1}/°C, based on prior data for trp cage miniprotein variants.[96]
[a]Data obtained at 2 °C.[96]

Ac-NLXIQWLKDGGPSSGRPPPS-NH$_2$

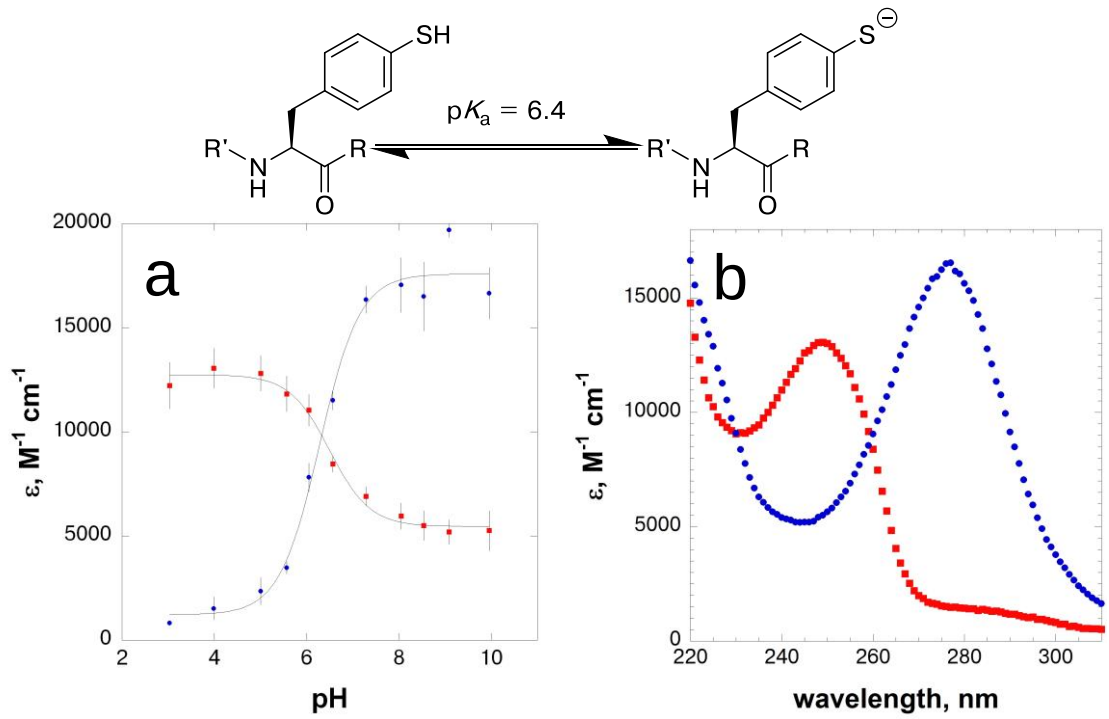
Residue at Position 3	pH	T_m, °C	[θ], 222 nm	[θ], 208 nm
Tyr	4.0	43	-16000	-16200
Tyr	7.0	39	-15600[a]	-14400[a]
4-SH-Phe	4.0	42	-14800	-17900
4-S$^-$-Phe	8.5	27	-13700	-19000

The overall α-helicity of the Y3(4-SH-Phe) trp cage variant is similar to the native trp cage under both acidic and neutral conditions. Notably, the pK_a of 4-thiophenylalanine was measured to be 6.4 (Chapter 1.2.7), and so it is mostly anionic under neutral and basic conditions. The trp cage and the Y3(4-SH-Phe) trp cage variant exhibit similar thermal stability at pH 4.0, where the thiol of 4-thiophenylalanine is neutral, as is the tyrosine. In contrast, the thiolate form of the Y3(4-SH-Phe) trp cage variant is significantly less stable than the neutral form, or the native trp cage, as indicated by the lower T_m for the peptide with the thiolate. The decreased stability of the Y3(4-S$^-$-Phe) trp cage variant is potentially due to a charge repulsion of the anionic sulfur atom and the π electrons of the tryptophan aromatic ring or the hydrophobic proline rings (Figure 1.27). The charge repulsion would

disrupt the hydrophobic core of the miniprotein and reduce the overall stability. These data suggest that replacement of the tyrosine residue at position 3 with its sulfur analogue introduces a structural "switch" for the stability of the trp cage miniprotein.

1.2.7 UV-Vis and fluorescence spectroscopy characterization of 4-thiophenylalanine within model peptides

The fundamental chemical and spectral properties of 4-thiophenylalanine had not been characterized during any of the prior syntheses of this amino acid, emphasizing the importance of developing a practical approach to generate this amino acid. Having optimized a practical approach for the synthesis of peptides containing 4-thiophenylalanine, the chemical and spectroscopic properties of this amino acid were examined. The acid dissociation constant of the thiol in 4-thiophenylalanine in aqueous conditions was measured by UV spectroscopy, with the UV spectra of both the neutral and anionic forms of 4-thiophenylalanine characterized (Figure 1.29).



$$e_{249} = \frac{e_{249,thiolate}10^{(pH-pKa)} + e_{249,thiol}}{1+10^{(pH-pKa)}} \qquad pK_a = 6.50 \pm 0.08$$

$$e_{276} = \frac{e_{276,thiolate}10^{(pH-pKa)} + e_{276,thiol}}{1+10^{(pH-pKa)}} \qquad pK_a = 6.28 \pm 0.09$$

Figure 1.29 UV spectroscopy and pK_a determination of Ac-T(4-SH-Phe)PN-NH_2
UV spectra were measured in 50 mM phosphate in the presence of 100 µM DTT (to reduce disulfides). Peptide concentrations were 45-65 µM. Concentrations of stock solutions of peptides containing 4-thiophenylalanine determined via Ellman's test. (a) Measured absorbance of the peptide Ac-T(4-SH-Phe)PN-NH_2 at 243 nm (red) and 276 nm (blue) under various pH conditions. The pK_a was calculated based on a curve fit of pH-dependence on molar absorbance at 249 nm and 276 nm, following the above equations. Results from both calculations are shown; (b) UV spectra of the peptide Ac-T(4-SH-Phe)PN-NH_2 at pH 4.0 (red, neutral, λ_{max} 243 nm) and 8.5 (blue, anionic, λ_{max} 276 nm). Error bars indicate standard error.

The UV spectra of the peptide Ac-T(4-SH-Phe)PN-NH_2 exhibited significant changes in the λ_{max} and absorbance intensity upon shifting from neutral to anionic

forms, with a bathochromic shift in the λ_{max} from 249 nm to 276 nm on increasing pH (Figure 1.29b). The observed increase in absorbance intensity and red shift in λ_{max} on shifting from acidic to basic pH is consistent with auxochromes such as phenolic –OH or thiol –SH groups. The lone pairs of electrons in the thiolate form of 4-thiophenylalanine stabilizes the π^* state, lowering the energy of the $\pi \rightarrow \pi^*$ transition, resulting in a bathochromic shift of the $\lambda_{max.}$[107] The increased area of electron delocalization upon deprotonaton of the aryl thiol also increases the molar extinction coefficient.[107] The significant changes in the UV spectrum depending on the protonation state of 4-thiophenylalanine allowed for measurement of the p*K*a.

The pK_a of 4-thiophenylalanine was measured to be 6.4 in water (Figure 1.29a), which is similar to the pK_a of thiophenol (pK_a 6.6).[45, 108] In contrast to cysteine (pK_a = 8.5[109]) or tyrosine (pK_a = 10.1[110]), 4-thiophenylalanine is primarily anionic at physiological pH. The nucleophilic thiolate form of 4-thiophenylalanine is readily accessible for further modification under mildly acidic conditions (pH 4 or 5), which can be utilized for site-specific labeling orthogonal to cysteine.

Given these significant changes in the UV spectrum upon deprotonation of 4-thiophenylalanine, and the associated changes in the aromatic electronic properties, the fluorescence properties of this amino acid were also characterized (Figure 1.30).

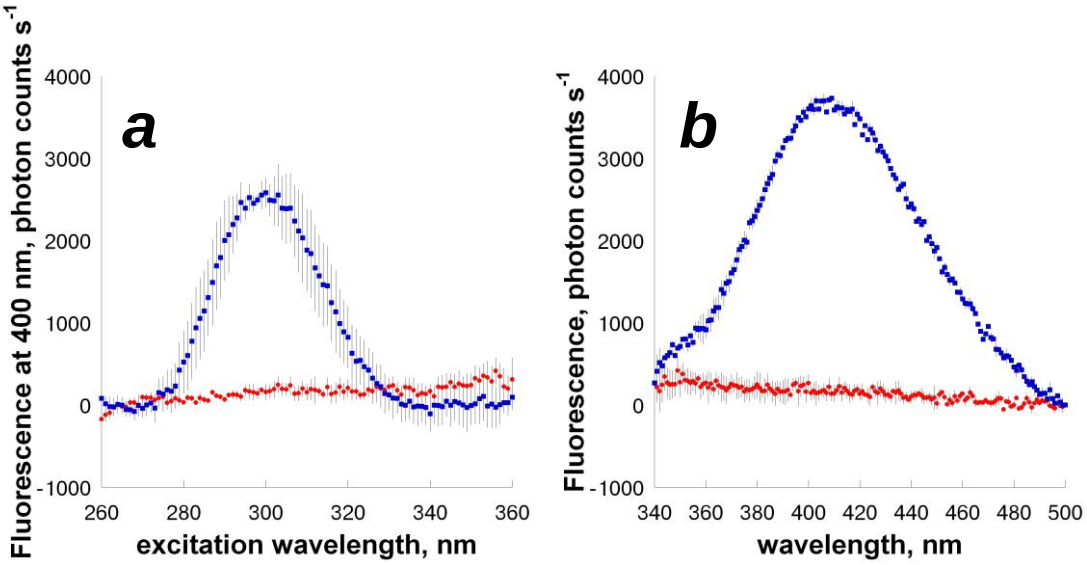


Figure 1.30 Fluorescence properties of Ac-T(4-SH-Phe)PN-NH_2
Fluorescence excitation and emission and spectra of peptides measured in 50 mM phosphate at pH 4.0 (red circles) and pH 8.5 (blue squares). Peptide concentrations were 100 µM with 125 µM DTT (to reduce disulfides). Concentrations of peptides containing 4-thiophenylalanine were determined via Ellman's test. Error bars indicate standard error. (a) Excitation spectra of Ac-T(4-SH-Phe)PN-NH_2, with detection at 400 nm; (b) emission spectra of Ac-T(4-SH-Phe)PN-NH_2, with excitation at 295 nm.

The anionic form of the peptide containing 4-thiophenylalanine exhibited fluorescence emission at 405 nm that was not observed in the neutral form. The fluorescence emission properties of 4-thiophenylalanine are not strong enough for utility in fluorescent labelling of proteins. Under similar conditions, tyrosine fluorescence emission at 300 nm was 500,000× stronger than the fluorescence emission for anionic 4-thiophenylalanine at 405 nm. However, the fluorescence emission and excitation observed in the anionic form of 4-thiophenylalanine indicate that the neutral and thiolate forms are electronically distinct, consistent with the observations in the UV absorbance spectra (Figure 1.29). The π orbitals of the aromatic ring in 4-thiophenylalanine are influenced by the protonation state of the

thiol, which can be utilized in applications that tune reactivity or structure.[91] The potential applications for 4-thiophenylalanine for electronic control of protein structure will be further discussed in Chapter 3.

1.2.8 Modification of peptides containing 4-thiophenylalanine via alkylation in solution phase

Cysteine thiolates are commonly alkylated for protein tagging or modification, because cysteine can be site-selectively modified in the presence of lysine, tyrosine phenolates, or other reactive amino acid functional groups.[49-52] 4-Thiophenylalanine is more acidic than cysteine (pK_a = 6.4 vs. 8.5, respectively), and could potentially be modified in a similar manner under more acidic conditions. Alkylation on the thiol of 4-thiophenylalanine might influence the electron distribution of the aromatic ring, and thus influence the spectroscopic properties of this amino acid. Thioether derivatives of 4-thiophenylalanine could be used to integrate novel functionality into proteins, including fluorophores, bioorthogonal handles, or photo-cleavable protecting groups.

In order to explore the potential thioether derivatives that can be readily generated from peptides containing 4-thiophenylalanine, the model peptide Ac-T(4-SH-Phe)PN-NH_2 was subjected to a series of alkylation reactions in solution. In this manner, thioether derivatives of 4-thiophenylalanine were generated in one step from the purified peptide containing 4-thiophenylalanine (Figure 1.31). Alkylation reactions on the peptide Ac-T(4-SH-Phe)PN-NH_2 included reaction with methyl iodide, allyl iodide, propargyl bromide, 2-nitrobenzyl bromide, and N-ethyl maleimide.

A simple alkylation reaction was demonstrated with methyl iodide, generating an aromatic analogue of the methionine thioether (Figure 1.31b). Functional groups that can be used in bioorthogonal reactions were also incorporated via alkylation of the

peptide Ac-T(4-SH-Phe)PN-NH_2, including an alkene functional group (substrate for olefin cross-metathesis reactions[111-113] or thiol-ene reactions[49, 114, 115]) and an alkyne functional group (substrate for azide-alkyne click-chemistry[116-118] and Sonogashira coupling reactions[119]). Reactive functional groups in proteins (i.e. tyrosine or cysteine) can be protected using a photo-cleavable mask, such as a 2-nitrobenzyl group, which can be inducibly "released" exposure to visible light.[120-122] 4-Thiophenylalanine can participate photo-initiated radical reactions,[62] and therefore "photocaging" the reactive thiol may be useful for some applications. 4-Thiophenylalanine was subjected to an alkylation reaction with 2-nitrobenzyl bromide to generate a thioether that is expected to be cleaved upon exposure to intense 365 nm light.[121] Thiols can also react as nucleophiles in Michael addition reactions with alkenes. 4-Thiophenylalanine was subjected to alkylation reaction with N(ethyl)maleimide (NEM) to demonstrate its utility as a nucleophile in Michael reactions (Figure 1.31f).

Each of these alkylation reactions on the peptide Ac-T(4-SH-Phe)PN-NH_2 generated a novel derivative of 4-thiophenylalanine. These resultant peptides were characterized via NMR, and the amide regions for all peptides are shown in Figures 1.32-1.36. The model peptide Ac-TXPN-NH_2 can undergo cis-trans isomerism at proline, and both rotamers are observed via NMR. The structural influence of these derivatives of 4-thiophenylalanine on peptide structure will be discussed in greater detail in Chapter 3.

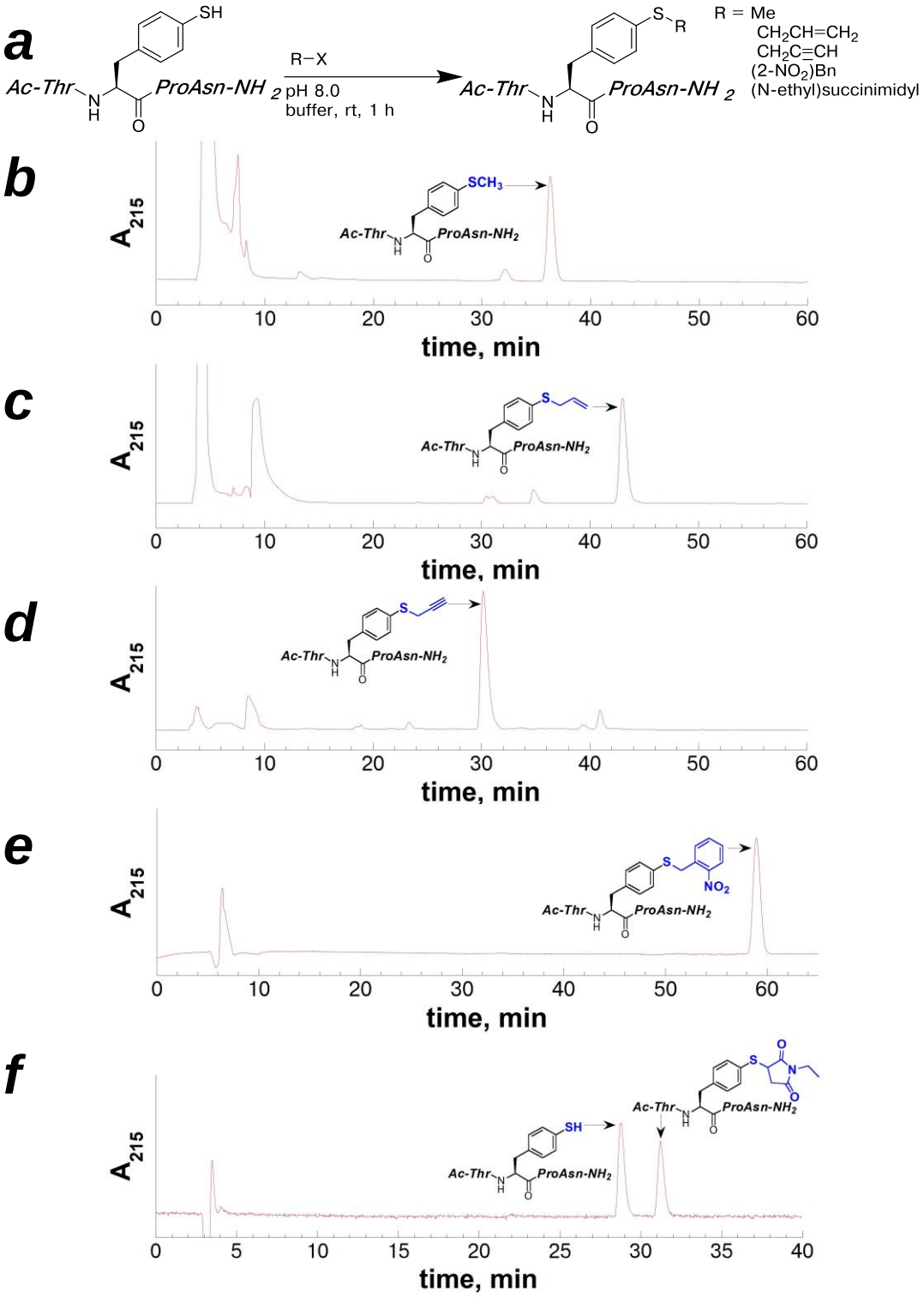

a
Ac-Thr
ProAsn-NH 2
R−X
pH 8.0
buffer, rt, 1 h
R = Me
CH2CH=CH2
CH2C≡CH
(2-NO2)Bn
(N-ethyl)succinimidyl
b
c
d
e
f
A215
time, min

Figure 1.31 Alkylation reactions on the peptide Ac-T(4-SH-Phe)PN-NH_2 in solution

(a) Generalized reaction scheme for solution phase alkylation reactions on the peptide Ac-T(4-SH-Phe)PN-NH_2. HPLC chromatograms of the products that resulted from the solution-phase alkylation reactions with Ac-T(4-SH-Phe)PN-NH_2: (b) HPLC chromatogram of the reaction with excess methyl iodide at pH 8 (100 mM phosphate buffer) at room temperature for 1 hour to produce Ac-T(4-SMe-Phe)PN-NH_2 using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the reaction with excess allyl iodide at pH 8 (100 mM phosphate buffer) at room temperature for 1 hour to produce Ac-T(4-S-allyl-Phe)PN-NH_2 using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (d) HPLC chromatogram of the reaction with excess propargyl bromide at pH 8 (100 mM phosphate buffer) at room temperature for 1 hour to produce Ac-T(4-S-propargyl-Phe)PN-NH_2 using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (e) HPLC chromatogram of the reaction with excess 2-nitrobenzyl bromide at pH 8 (100 mM phosphate buffer) at room temperature for 1 hour to produce Ac-T(4-S((2-NO_2)Bn)-Phe)PN-NH_2 using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (f) HPLC chromatogram of the reaction with excess (*N*-ethyl)maleimide at pH 8 (100 mM phosphate buffer) at room temperature for 1 hour to produce Ac-T(4-S((N-ethyl)succinimide)-Phe)PN-NH_2 using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

4-Thiophenylalanine was readily modified via alkylation reaction in one step using all alkylating reagents examined, including both alkyl halides and Michael acceptors, in solution under mildly basic conditions. These thioether derivatives of 4-thiophenylalanine represent a range of versatile modifications that can be used in therapeutic or materials applications, including bioorthogonal conjugation reactions, photorelease/photo-uncaging reactions, methionine mimics, or protein labeling.

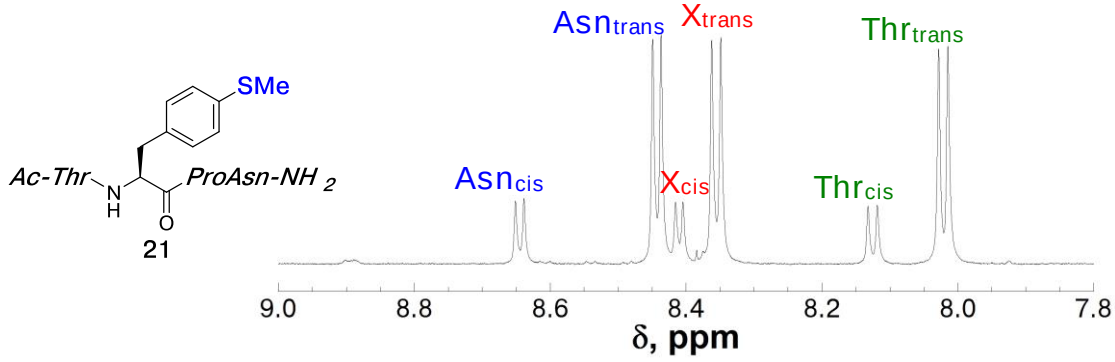


Figure 1.32 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S(methyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S(methyl)-thiophenylalanine)ProAsn-NH$_2$ at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K. Amide peaks were identified by TOCSY spectrum.

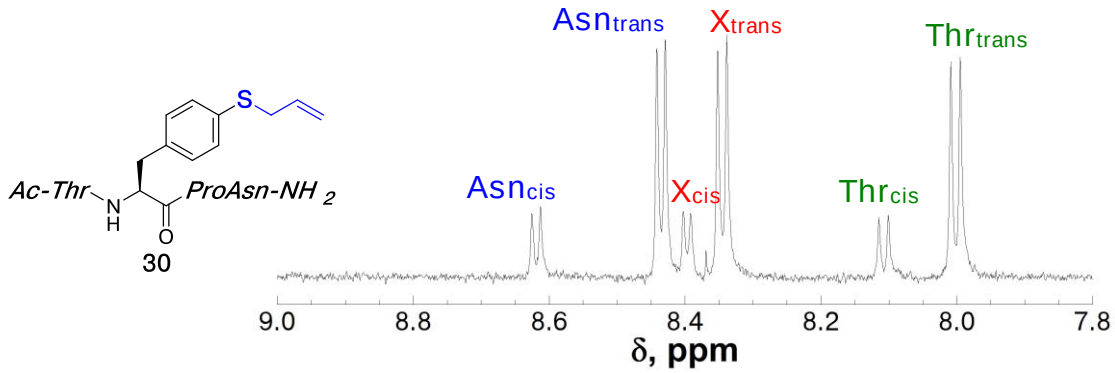


Figure 1.33 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-allyl-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S-allyl-thiophenylalanine)ProAsn-NH$_2$ at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K. Amide peaks were identified by TOCSY spectrum.

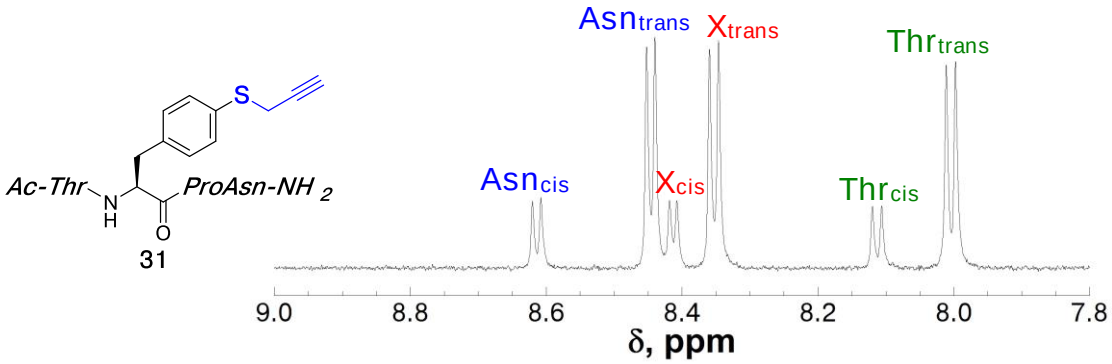


Figure 1.34 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S-propargyl-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S-propargyl-thiophenylalanine)ProAsn-NH_2 at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by TOCSY spectrum.

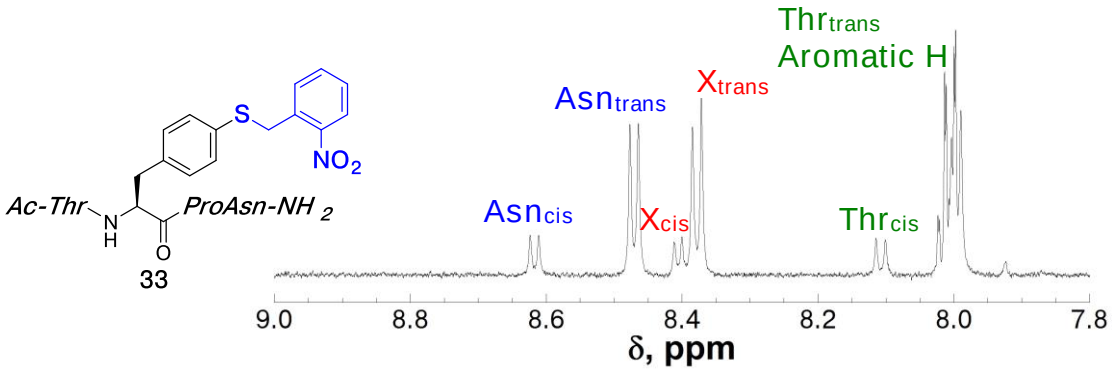


Figure 1.35 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S(2-nitrobenzyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S(2-nitrobenzyl)-thiophenylalanine)ProAsn-NH_2 at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by analogy to similar spectra.

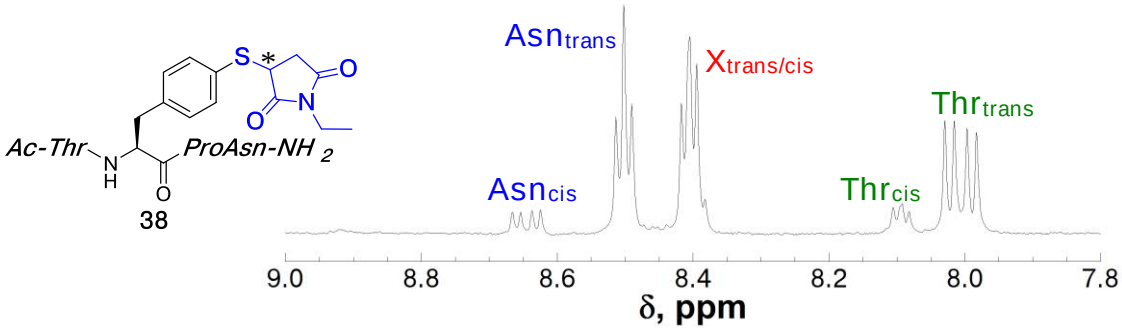


Figure 1.36 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S(*N*-ethyl succinimidyl)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S(*N*-ethyl-succinimidyl)-thiophenylalanine)ProAsn-NH_2 at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by TOCSY spectrum. The purified sample is a mixture of diastereomers at the succinimdyl thioether (indicated).

1.2.9 Comparative reactivity of peptides containing 4-thiophenylalanine for alkylation reaction in solution phase

4-Thiophenylalanine could also potentially be used in concert with cysteine, given the difference in acidity between these two thiolated amino acids. The sulfur atom in cysteine serves as a functional handle for modification in proteins. The aryl thiol in 4-thiophenylalanine could potentially serve a similar role. An advantage of using cysteine for protein labelling is that it can be readily alkylated at pH 8. However, at this pH, cysteine can be only modestly selective in the presence of lysine amines, tyrosine phenolates or other reactive amino acid functional groups.[51] Given that 4-thiophenylalanine is more acidic than cysteine, 4-thiophenylalanine can potentially be used in similar alkylation reactions at a lower pH, with reactivity orthogonal to cysteine. Site-specific alkylation of 4-thiophenylalanine in the presence of cysteine residues could be applied for multiple, orthogonal modifications in proteins and biomolecules.

Maleimides are common functional groups for modification of cysteine, and were demonstrated to react efficiently with the model peptide containing 4-thiophenylalanine in the previous section. In order to compare the relative reactivity of 4-thiophenylalanine and cysteine for alkylation with a Michael accepter in water, two model peptides Ac-TXPN-NH_2 (where X = 4-thiophenylalanine or cysteine) were synthesized. The reaction products of these competitive alkylation experiments were characterized via HPLC (Figure 1.37). Each peptide (100 μM) was allowed to react, separately, in the presence of 0.5 equivalents of *N*-ethyl maleimide (NEM, 50 μM) at pH 6.5, in order to evaluate reactivity of cysteine and 4-thiophenylalanine for this alkylating reagent (Figure 1.37a and 1.37b). The competitive alkylation reactions allowed for the reaction of both peptides in a single solution (200 μM total peptide concentration, 100 μM per peptide), with *N*-ethyl malemide (50 μM) at pH 6.5, pH 5.0, and pH 9.0 (Figure 1.37e, 1.37f, 1.37g, respectively). The competitive alkylation experiments determined whether 4-thiophenylalanine or cysteine was more reactive with the dilute alkylating reagent in solution, providing insight into relative reaction kinetics of alkylation (Figure 1.37). Approximate conversions of these alkylation reactions are shown in Table 1.7 (changes in molar extinction coefficient in NEM-labeled products were not considered, and these conversions are approximate).

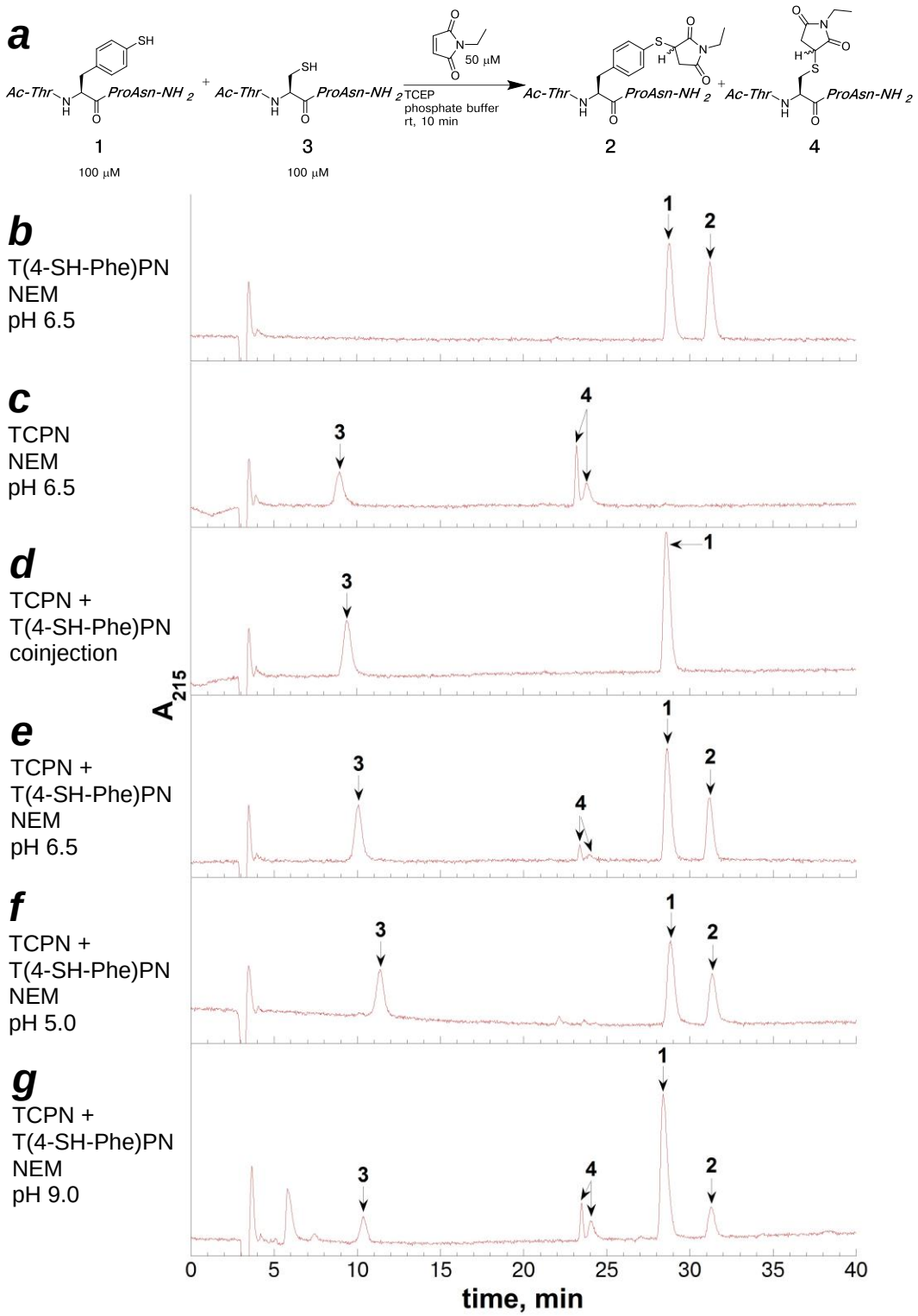


Figure 1.37 Competitive alkylation reactions of model peptides containing

cysteine or 4-thiophenylalanine with *N*-ethyl maleimide (NEM)

(a) General reaction scheme for competitive alkylation reaction between the peptides Ac-Thr(4-SH-Phe)ProAsn-NH_2 (**1**) and Ac-ThrCysProAsn-NH_2 (**3**) with NEM to yield the corresponding NEM-conjugated peptides (**2** and **4**, respectively). A buffered solution of the peptides (100 μM each) was incubated with NEM (50 μM) in the presence of TCEP (50 μM) for 10 minutes at room temperature at the indicated pH; b) HPLC chromatogram of the alkylation reaction of Ac-Thr(4-SH-Phe)ProAsn-NH_2 (100 μM) at pH 6.5; c) HPLC chromatogram of the alkylation reaction of Ac-ThrCysProAsn-NH_2 at pH 6.5; d) HPLC chromatogram of the coinjection of the peptides Ac-Thr(4-SH-Phe)ProAsn-NH_2 and Ac-ThrCysProAsn-NH_2; e) HPLC chromatogram of the alkylation reaction with the peptides Ac-ThrCysProAsn-NH_2 and Ac-Thr(4-SH-Phe)ProAsn-NH_2 at pH 6.5; f) HPLC chromatogram of the alkylation reaction with the peptides Ac-ThrCysProAsn-NH_2 and Ac-Thr(4-SH-Phe)ProAsn-NH_2 at pH 5.0; g) HPLC chromatogram of the alkylation reaction with the peptides Ac-ThrCysProAsn-NH_2 and Ac-Thr(4-SH-Phe)ProAsn-NH_2 at pH 9.0.

Table 1.7. Approximate conversion for competitive alkylation reactions of model peptides containing cysteine or 4-thiophenylalanine with *N*-ethyl maleimide (NEM)

Crude reaction products were evaluated via HPLC (Figure 1.37). Peak identity was established via ESI-MS. Product conversions were normalized relative to total peak integrations, and are calculated based on product conversion for NEM-modified product. Changes in molar extinction coefficient of the NEM-modified peptide products were not considered, and these conversions are only for relative comparison. Reaction conditions: 50 mM phosphate buffer in water, 50 μM TCEP, and 50 μM NEM. Reactions were allowed to incubate at room temperature for 10 minutes.

Entry	HPLC chromatogram	Peptide (concentration)	pH	Conversion
1	Figure 1.37b	Ac-T(4-SH-Phe)PN-NH_2 (100 μM)	6.5	43%
2	Figure 1.37c	Ac-TCPN-NH_2 (100 μM)	6.5	59%
3	Figure 1.37e	Ac-T(4-SH-Phe)PN-NH_2 (100 μM) + Ac-TCPN-NH_2 (100 μM)	6.5	25% 5%
4	Figure 1.37f	Ac-T(4-SH-Phe)PN-NH_2 (100 μM) + Ac-TCPN-NH_2 (100 μM)	5.0	28% 0%
5	Figure 1.37g	Ac-T(4-SH-Phe)PN-NH_2 (100 μM) + Ac-TCPN-NH_2 (100 μM)	9.0	13% 14%

Each peptide containing either 4-thiophenylalanine or cysteine cleanly reacted with NEM at pH 6.5 (Figure 1.37a-c). In Figure 1.37b and Figure 1.37c, the peptides containing 4-thiophenylalanine and cysteine were separately incubated (100 μM total peptide concentration) in a buffered solution with 0.5 equivalents of *N*-ethyl maleimide (NEM) for 10 minutes at pH 6.5 at room temperature. The diastereomers of the NEM-modified peptide Ac-TCPN-NH_2 were visible as separate peaks by HPLC (Figure 1.37c). The expected diastereomeric products did not resolve under these HPLC with the reaction of NEM and the peptide Ac-T(4-SH-Phe)PN-NH_2. Both reactions with 0.5 equivalents of NEM using the peptides containing either 4-thiophenylalanine or cysteine resulted, expectedly, in approximately 43-59% approximate conversion at pH 6.5 in 10 minutes (Table 1.7, entries 1 and 2).

For the competitive alkylation reactions, both peptides Ac-TCPN-NH_2 or Ac-T(4-SH-Phe)PN-NH_2 (100 μM each, 200 μM total peptide concentration) were allowed to react with NEM (50 μM), maintaining the same relative concentration of reactants as in the single peptide experiments. Through these experiments, the relative reactivity of cysteine and 4-thiophenylalanine for alkylation with NEM under different pH conditions was compared. At pH 6.5 and 5.0, the peptide containing 4-thiophenylalanine exhibited significantly greater reactivity with NEM over the peptide containing cysteine (Table 1.7, entries 3 and 4). At pH 9.0, under conditions where both cysteine and 4-thiophenylalanine are mostly anionic, both peptides exhibited similar reactivity with NEM, generating nearly equal amounts of NEM-modified Ac-TCPN-NH_2 and Ac-T(4-SH-Phe)PN-NH_2 (Table 1.7, entry 5). These competitive experiments demonstrate the potential utility of 4-thiophenylalanine for site-selective

alkylation reactions in the presence of cysteine, which can allow for orthogonal labelling of biomolecules in mildly acidic, aqueous conditions.

1.2.10 Modification of peptides containing 4-thiophenylalanine via alkylation on peptides on solid-phase

Solution alkylation reactions on 4-thiophenylalanine within peptides were demonstrated with a series of alkyl halides and Michael acceptors, demonstrating the versatility for modification of this amino acid. These alkylation reactions in solution were rapid, and generated the resultant thioether product in only one step with one purification from the peptide Ac-T(4-SH-Phe)PN-NH_2. In some applications, it may be necessary to modify 4-thiophenylalanine on protected peptides, particularly in applications involving orthogonal protecting groups or modifications on both 4-thiophenylalanine and cysteine. 4-Thiophenylalanine was demonstrated to react more efficiently with alkylating reagents than cysteine under mildly acidic conditions, but some thiol modifications may require greater site selectivity. Solid-phase alkylation reaction approaches were developed in order to alkylate 4-thiophenylalanine on protected peptides on solid phase (Figure 1.38). This approach specifically allows for the synthesis of S-alkylated peptides with a single HPLC purification step.

In order to effect alkylation reactions on peptides containing 4-thiophenylalanine on solid-phase, the protected peptide was subjected to solid-phase thiolysis reaction conditions following the copper-mediated cross-coupling reaction. Immediately following the thiolysis reaction conditions on the peptide on solid phase, the peptide was allowed to react with an alkylating reagent (allyl bromide) in the presence of base and reductants (to reduce disulfides). The resultant HPLC chromatograms are shown in Figure 1.38.

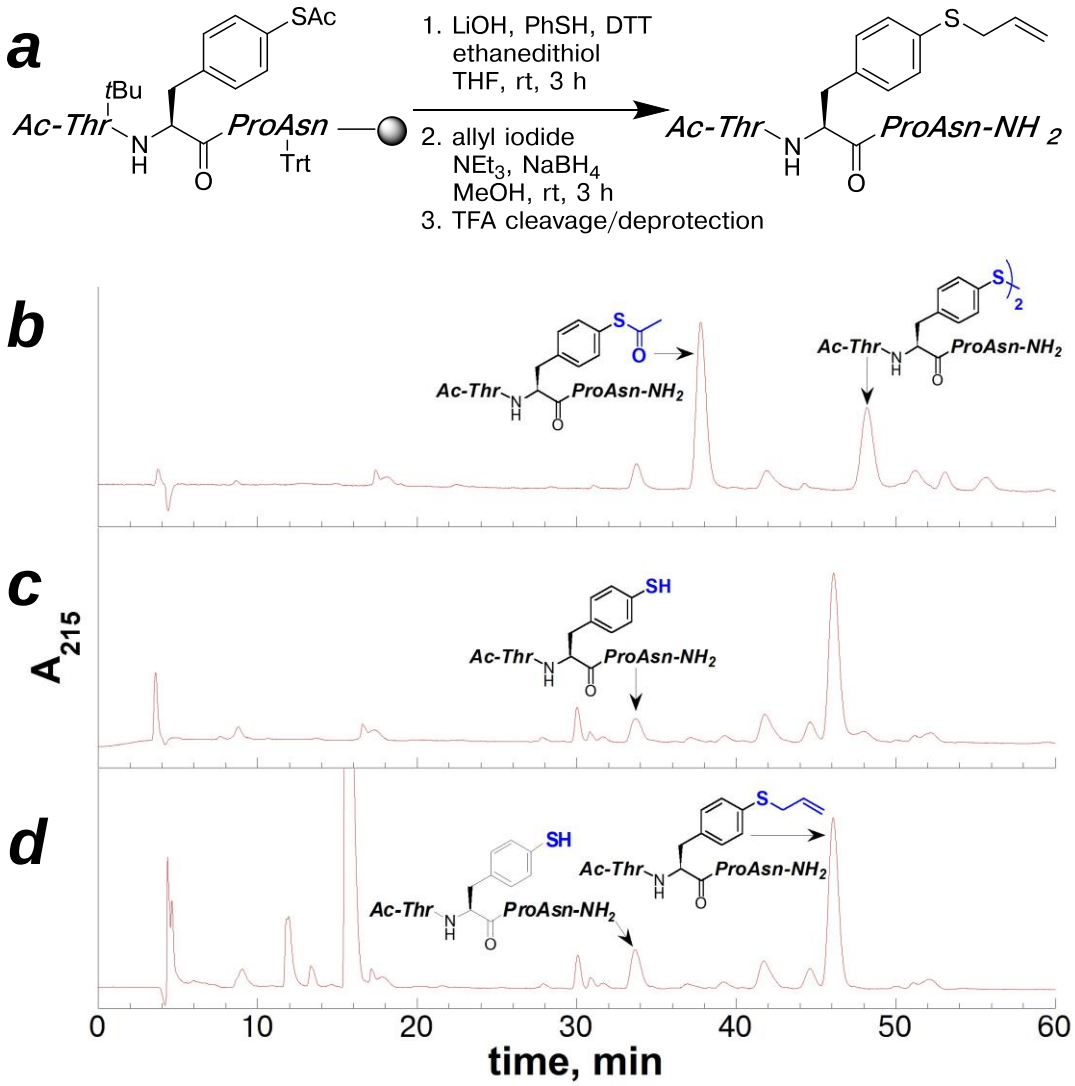


Figure 1.38 Optimized synthesis of the peptide containing 4-S-allyl-thiophenylalanine via alkylation reaction on peptides on solid phase

Crude reaction products were analyzed via HPLC using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were characterized via ESI-MS. (a) Scheme for one-pot thiolysis/reduction reaction, followed by alkylation reaction on the peptide Ac-Thr(4-SH-Phe)ProAsn-NH_2 on solid-phase. The resin containing the peptide was used immediately after the solid-phase copper-mediated cross-coupling reaction, and contained a mixture of peptides containing 4-S(acetyl)-thiophenylalanine and the corresponding disulfide products; (b) HPLC chromatogram of the crude reaction products following the solid-phase copper-mediated cross-coupling reaction, and TFA cleavage and deprotection from resin; (c) HPLC chromatogram of the crude reaction products following the solid-phase deacetylation and reduction reaction on solid phase (reaction 1 in scheme, Figure 1.38a). Resin containing the peptide products was subjected to cleavage and deprotection reaction using TFA; (d) HPLC chromatogram of the crude reaction products following the solid-phase alkylation reaction with allyl iodide, following the copper-mediated cross-coupling and thiolysis reactions on solid phase. Resin containing the peptide products was subjected to cleavage and deprotection reaction using TFA.

As proof-of-principle of solid-phase cross-coupling reaction and subsequent S-alkylation, we examined the solid-phase synthesis of peptides containing an allyl functional group. The allyl group would introduce a bioorthogonal alkene handle into peptides, which can potentially be used for modification via substrate for olefin cross-metathesis reactions[111-113] or thiol-ene reactions.[49, 114, 115] The optimized solid phase alkylation reaction conditions generated the peptide containing 4-S-allyl-thiophenylalanine in 82% conversion from the peptide containing 4-iodophenylalanine (Figure 1.38d). Curiously, the resultant peptide products from the solid-phase deacetylation reaction were primarily symmetric disulfides, rather than the peptide containing 4-thiophenylalanine in reduced form. Iodine that was generated during the cross-coupling reaction, potentially still retained in the resin even after multiple washing steps, could potentially react with 4-thiophenylalanine on solid phase to form disulfides. In order to reduce 4-thiophenylalanine to allow for reaction with the

alkylating reagent, allyl iodide and triethylamine were immediately added to the protected peptide on solid phase following thiolysis. Immediate addition of the alkylating reagent was crucial for efficient synthesis of the peptide containing 4-S-allyl-thiophenylalanine. With this one-pot reduction/alkylation method, 4-thiophenylalanine was readily alkylated on the peptide on solid-phase, selective modification of 4-thiophenylalanine without undesired cysteine alkylation is possible. This approach is expected to be able to be modified for other solid phase reactions on 4-thiophenylalanine.

1.2.11 Aqueous olefin cross-metathesis reaction on a model peptide containing 4-S-allyl-thiophenylalanine

Alkene functional groups have been used for bioorthogonal reactions in small molecules, peptides and proteins, including thiol-ene reactions, olefin cross-metathesis reactions, Heck reactions, and Diels-Alder reactions.[13, 51] The allyl functionality has been incorporated into proteins, both genetically and synthetically, for a variety of applications, including peptide stapling and synthesis of glycol-protein or protein-protein conjugates.[51, 123-129] Davis and coworkers[111] have described the unique reactivity of allyl sulfides as substrates for olefin cross-metathesis in small molecules, peptides, and proteins (Figure 1.39). In this work, cysteine was modified to S-allyl-cysteine in proteins, and a variety of terminal alkenes were conjugated to proteins under mild, aqueous conditions using the Hoveyda-Grubbs II ruthenium catalyst.[111] In a subsequent paper, allyl sulfides and allyl selenides were both found to have enhanced reactivity for olefin-cross metathesis reactions over comparable O-allyl substrates, potentially due to additional coordination between the sulfur atom and the ruthenium in the catalyst.[111, 112] S-allyl cysteine and Se-allyl selenocysteine have been

incorporated into proteins via alkylation reactions on expressed proteins,[111, 112] but this method is not site-selective in proteins containing more than one cysteine residue. Attempts to incorporate allyl sulfides and allyl selenides using methionine-auxotrophic *E. coli* were unsuccessful.[113] In his work demonstrating enhanced chalcogen reactivity in cross-metathesis reactions, Davis[113] speculated that 4-S-allyl-thiophenylalanine could provide an opportunity for enhanced reactivity over allyl-cysteine. Furthermore, 4-S-allyl-thiophenylalanine could be genetically incorporated into proteins via amber nonsense suppression, given the success with genetic incorporation of the oxygen-analogue by Noren *et al.*[130]

Having established a synthesis of peptides containing 4-S-allyl-thiophenylalanine, conditions for olefin cross-metathesis reactions were examined using allyl alcohol (Figure 1.40, Figure 1.41, Table 1.8). The model peptide containing 4-thiophenylalanine, Ac-T(4-SH-Phe)PN-NH_2, was synthesized and subjected to alkylation reaction with allyl iodide, using either solid-phase or solution approaches, as described in previous sections. The purified peptide containing 4-S-allyl-thiophenylalanine was subjected to cross-metathesis conditions similar to those described by Davis and coworkers.[112, 113] Davis noted that high concentrations of $MgCl_2$ were required for efficient reaction conversion.[111] Therefore, this additive was also examined in cross-metathesis reactions on the peptide containing 4-S-allyl-thiophenylalanine.

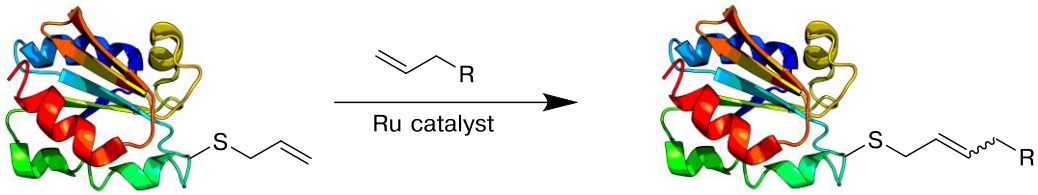


Figure 1.39 Generalized olefin cross-metathesis reactions for protein modification

Davis and coworkers demonstrated that allyl sulfides, obtained via modification of exposed cysteine side-chains, have enhanced reactivity for olefin cross-metathesis reactions.[15, 111-113, 126] Other allyl-functionalized amino acids, including allylglycine, O-allyl-tyrosine, and dehydroalanine, have also been used as substrates for olefin-cross metathesis reactions.

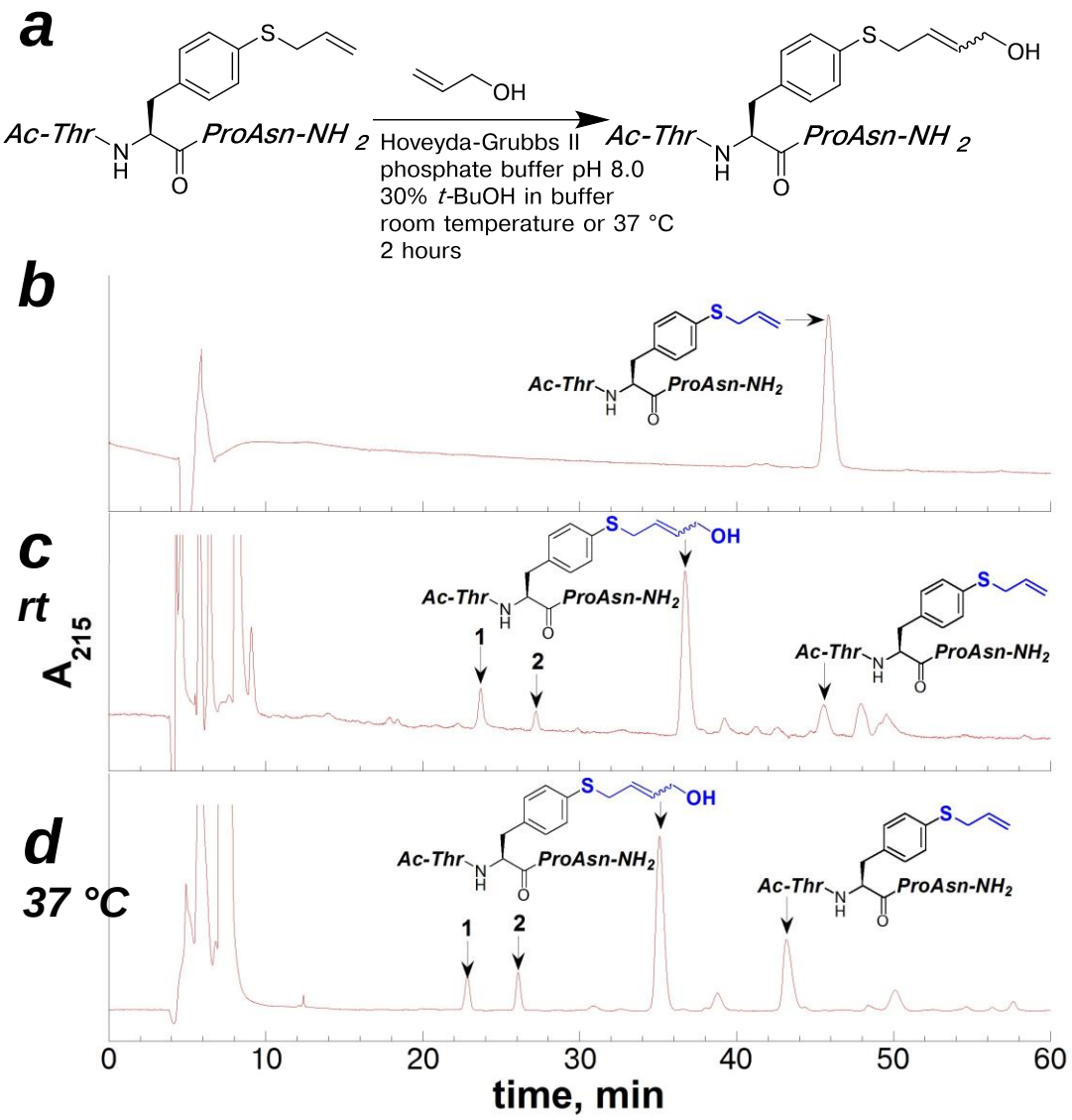


Figure 1.40 Olefin cross-metathesis reactions of allyl alcohol with the peptide containing 4-S-allyl-thiophenylalanine

(a) Scheme for the olefin cross-metathesis reaction in solution on the peptide Ac-Thr(4-S-allyl-Phe)ProAsn-NH_2 in solution; (b) HPLC chromatogram of the purified reactant peptide, Ac-T(4-S-allyl-Phe)PN-NH_2; (c) HPLC chromatogram of the crude reaction products following the cross-metathesis at room temperature; (d) HPLC chromatogram of the crude reaction products resulting from the cross-metathesis reaction at 37 °C. Crude reaction products were analyzed via HPLC using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS are indicated. The additional products noted in panel (c) and (d) are the corresponding sulfoxides of the product (**1**) and starting material (**2**), respectively.
Reaction conditions (see experimental section 1.4.9): Ac-Thr(4-S-allyl-Phe)ProAsn-NH_2 (approx. 1 mM) in phosphate buffer (50 mM, pH 8.0), with 30% *tert*-butanol, Hoveyda-Grubbs II catalyst (3.5 mM), allyl alcohol (177 mM). The resultant reaction was allowed to incubate for 2 h, either at room temperature or at 37 °C.

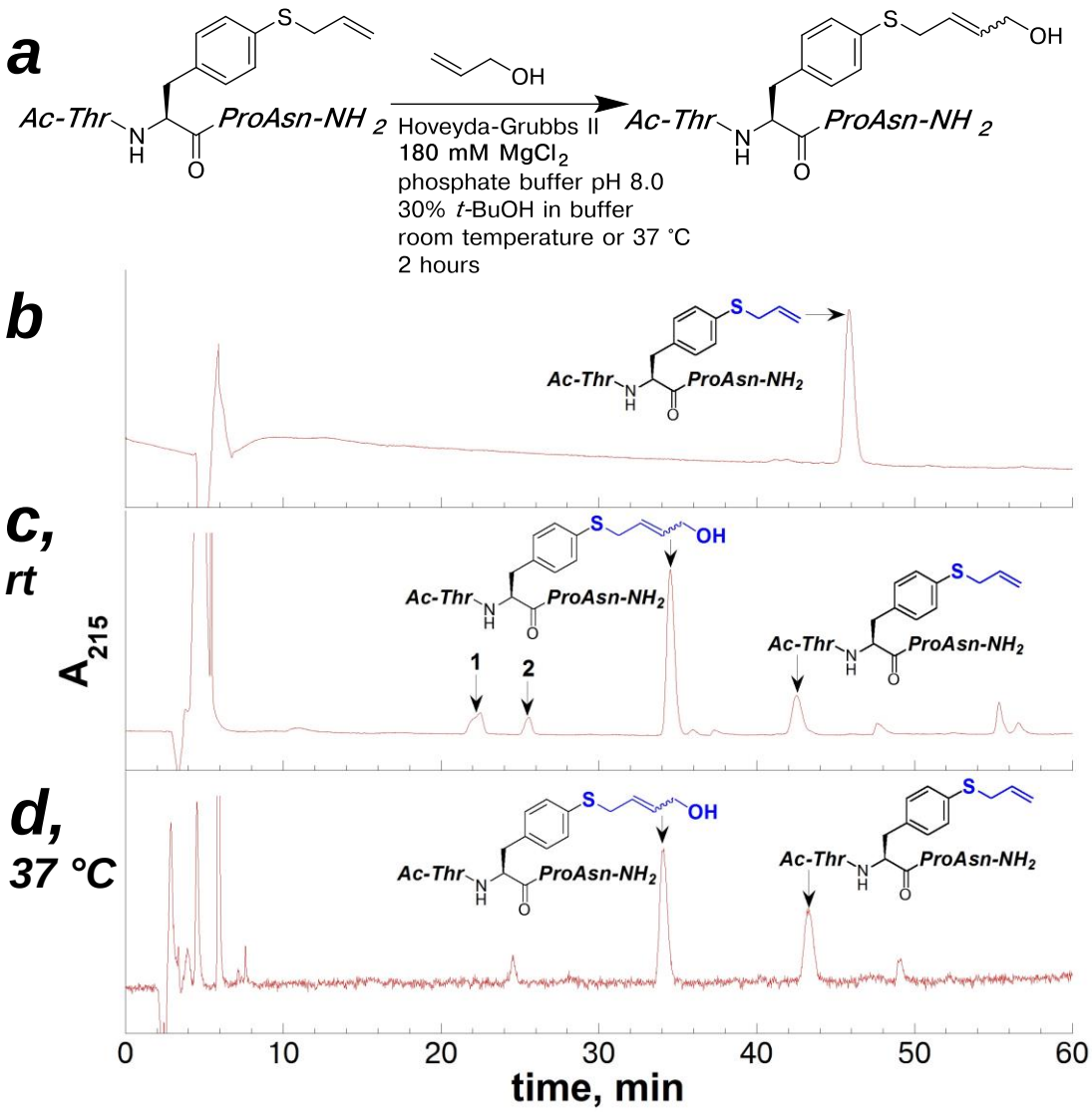
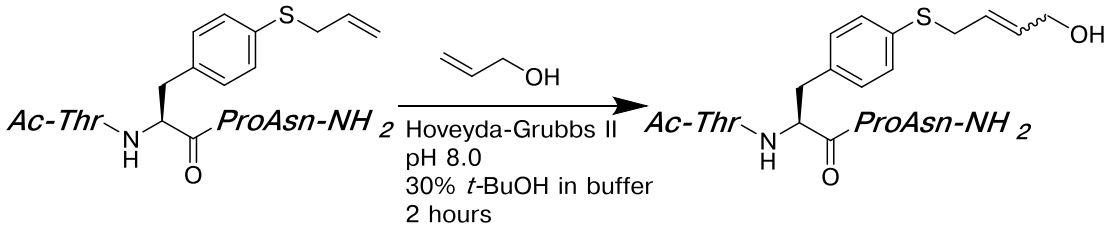


Figure 1.41 Olefin cross-metathesis reactions of allyl alcohol with the peptide containing 4-S-allyl-thiophenylalanine with $MgCl_2$ as an additive

(a) Scheme for the olefin cross-metathesis reaction in solution on the peptide Ac-Thr(4-S-allyl-Phe)ProAsn-NH_2 in solution; (b) HPLC chromatogram of the purified reactant peptide, Ac-T(4-S-allyl-Phe)PN-NH_2; (c) HPLC chromatogram of the crude reaction products following the cross-metathesis with 180 mM $MgCl_2$ at room temperature; (d) HPLC chromatogram of the crude reaction products resulting from the cross-metathesis reaction with 180 mM $MgCl_2$ at 37 °C. Crude reaction products were analyzed via HPLC using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS are indicated.
The additional products noted in panel (c) and (d) are the corresponding sulfoxides of the product (**1**) and starting material (**2**), respectively.
Reaction conditions (see experimental section 1.4.9): Ac-Thr(4-S-allyl-Phe)ProAsn-NH_2 (approx. 1 mM) in phosphate buffer (50 mM, pH 8.0) containing with 180 mM $MgCl_2$, with 30% *tert*-butanol, Hoveyda-Grubbs II catalyst (3.5 mM), allyl alcohol (177 mM). The resultant reaction was allowed to incubate for 2 h, either at room temperature or at 37 °C.

Table 1.8. Olefin cross-metathesis reactions of allyl alcohol on the peptide Ac-T(4-S-allyl-thiophenylalanine)PN-NH_2

Percent conversion from starting material was calculated via integration of the HPLC chromatograms shown in Figures 1.40 and 1.41.



Entry	Temperature	Additive	Product conversion
1	room temperature	none	69%
2	37 °C	none	61%
3	room temperature	180 mM $MgCl_2$	70%
4	37 °C	180 mM $MgCl_2$	64%

Peptides containing 4-S-allyl-thiophenylalanine reacted efficiently with allyl alcohol to generate the olefin cross-metathesis products, both in the presence and absence of $MgCl_2$, and at both room temperature and 37 °C. Side products were observed associated with the sulfoxide of the reactant peptide and product peptide (Ac-T(4-S-allyl-Phe)PN-NH_2 and Ac-T(4-S-(1-butenol)-Phe)PN-NH_2, respectively). Formation of these side-products was lowest in the reaction conducted at 37 °C in the presence of 180 mM $MgCl_2$ (Figure 1.41d). Davis and coworkers[111] noted in their work with S-allyl cysteine that high concentrations of $MgCl_2$ (>100 mM) were required for formation of the cross-metathesis product. However, for these cross-metathesis reactions with peptides containing 4-thiophenylalanine, $MgCl_2$ had little apparent effect on the reaction conversion (Table 1.8).

With 4-S-allyl-thiophenylalanine, room temperature reactions generated higher reaction yields than reactions conducted at 37 °C, and the reactivity was comparable in either the presence or absence of Mg^{2+}. In contrast, Davis and coworkers reported excellent yields using S-allyl-cysteine with allyl alcohol at room temperature, although other allyl substrates required reaction at 37 °C and high concentrations of Mg^{2+} (180 mM).[111, 112] Davis[111, 112] speculated that the high concentration of Mg^{2+} was necessary to prevent unproductive chelation of the catalyst with amino acid side-chains. It is possible that the side-chains in the model peptide Ac-TXPN-NH_2 are not especially prone to chelation of the catalyst, therefore eliminating the need for high concentrations of $MgCl_2$. Alternatively, the enhanced reactivity of 4-S-allyl-thiophenylalanine could abrogate the need for $MgCl_2$. Davis and coworkers[111-113] discovered that S-allyl substrates are more reactive than O-allyl substrates for cross-

metathesis reactions, and therefore 4-S-allyl-thiophenylalanine can be expected to have enhanced reactivity over 4-O-allyl-tyrosine in these bioconjugation reactions.

The resultant peptide from the olefin cross-metathesis reaction, Ac-T(4-S-(1-butenol)-Phe)PN-NH_2, was characterized via ^{1}H NMR (Figure 1.42). The peptide containing 4-S-(1-butenol)-thiophenylalanine can adopt the cis and trans conformations at proline, and both rotamers are observable via NMR. The structural implications of this amino acid will be further discussed in Chapter 3.

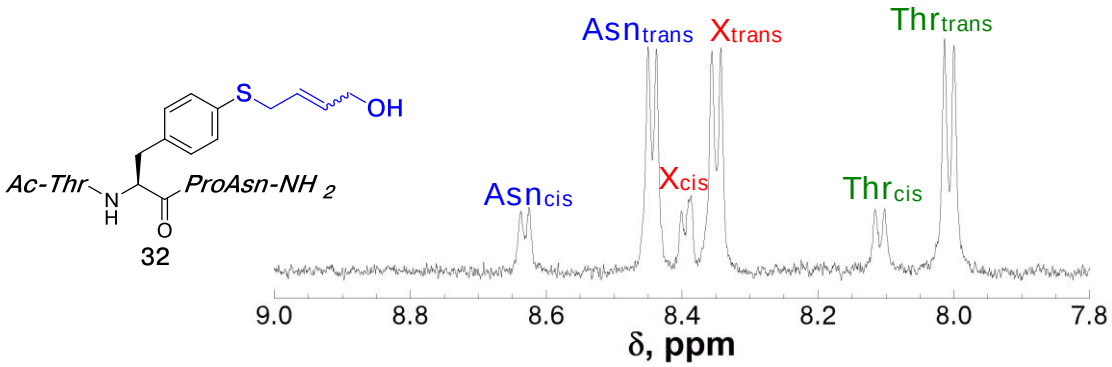


Figure 1.42 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-S-(1-butenol)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-Thr(4-S-(1-butenol)-thiophenylalanine)ProAsn-NH_2 at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by analogy to similar peptides.

1.2.12 Synthesis of peptides containing oxidized thioether derivatives of 4-thiophenylalanine in solution phase

Canonical amino acids are subject to posttranslational modification, including phosphorylation, methylation, and glycosylation. Sulfur oxidation, through multiple oxidation states, is a primary mechanism of posttranslational modification in cysteine and methionine.[10] For both of these amino acids, oxidation can occur either via

concerted enzymatic oxidation or as a result of non-specific oxidative stress.[55, 131, 132] Cysteine can be oxidized to form the sulfenate, sulfinate, or sulfonate, or can form symmetric or asymmetric disulfides (Figure 1.43).[133] Methionine posttranslational modifications include oxidation to the sulfoxide and sulfone; sulfoxide formation introduces a new stereocenter at the sulfur atom. Methionine oxidation can act both as a general biological antioxidant in proteins, as an enzyme activation mechanism, and it may also influence structure of peptides and proteins.[132] Chemical oxidation of methionine generates the diastereotopic sulfoxide, but enzymatic oxidation has been found to be stereospecific.[131, 134]

Thiols are considered "soft" hydrogen bond donors, but the roles of sulfur in other non-covalent interactions (both as thiols and as thioethers) have gained more recent appreciation.[7, 135-137] Methionine sulfur/π interactions were found to be a common motif in proteins that can potentially stabilize α-helical or tertiary structures.[135, 136] For example, a sulfur-aromatic interaction between Met56 and the FMN cofactor in flavodoxin has been suggested to modulate the redox potentials of this enzyme.[138] Non-covalent interactions involving thiols and thioethers have also been examined in drug design.[7] Potentially, the oxidation state of methionine can influence its non-covalent interactions, such as a sulfur-aromatic interaction between methionine and a tryptophan ring might be disrupted upon methionine oxidation.

In aryl thioethers, the torsion angle of the thioether linkage can be affected by oxidation, which can invoke structural preferences and orientation of the ether linkage with respect to the aromatic ring. In order to examine how oxidation of 4-thiophenylalanine thioethers could change the inherent structural preference of the associated alkylation, the torsion angles were obtained for aryl thioethers in crystal

structures in the Cambridge Structural Database (CSD) as a function of sulfur oxidation state. The torsion angle C_{aro}–C_{aro}–X–Csp^3 was measured (X = O, S, SO, or SO_2; Figure 1.44) with respect to the plane of the aromatic ring: 0° refers to an ether or thioether linkage that is coplanar with the aromatic ring; and 90° refers to an ether or thioether linkage that is orthogonal to the plane of the ring (Figure 1.44).

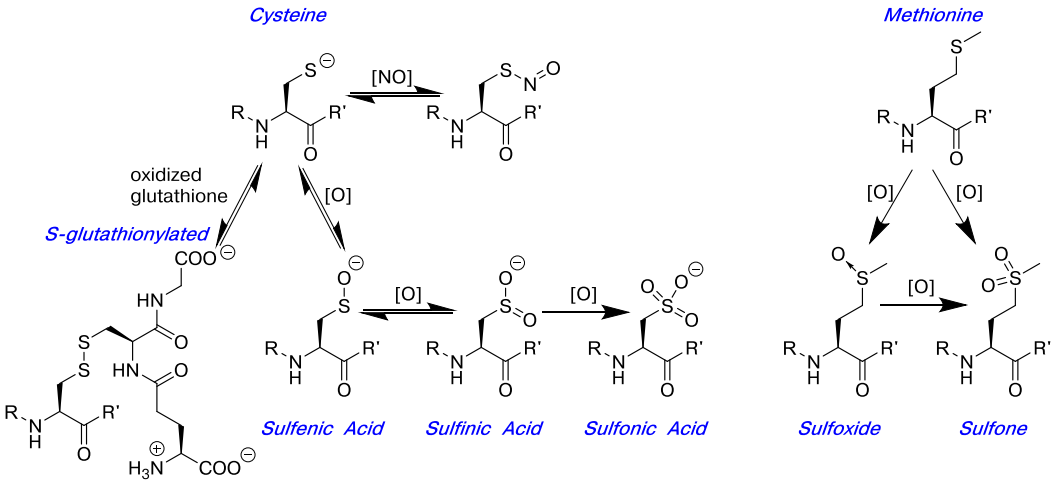


Figure 1.43 Oxidative modifications of cysteine and methionine
Oxidation of the sulfur-containing amino acids, either enzymatically or as a result of oxidative or nitrosative stress, have implications in cell signaling, enzyme activation, and reactivity of catalytic thiols.[5, 8] Oxidiation of cysteine via reactive-oxygen-species (ROS) or reactive-nitrogen-species (RNS) generates several reversible intermediates, with the sulfonic acid being the stable end product. Each of the multiple oxidation states of sulfur can have distinct implications, including electronic, structural, or reactivity changes.

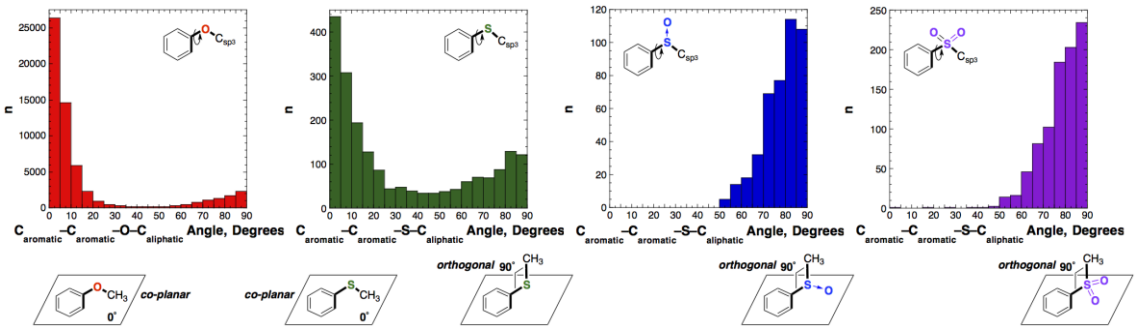


Figure 1.44 Conformational preferences about the aromatic C_{aro}–C_{aro}–X–Csp^3 torsion angle in crystal structures
Histograms reflecting the torsion angles of C_{aro}–C_{aro}–X–Csp^3 (X = O, S, S(O), or SO_2) bonds in crystal structures in the Cambridge Structural Database. In ethers and thioethers, the ether linkage tends to be coplanar with the aromatic ring. In oxidized thioethers, for both sulfoxides and sulfones, most ether linkages tend to orient perpendicular to the plane of the aromatic ring. Details regarding the CSD search parameters are included in the Experimental Section.[139]

Aryl thioethers and ethers exhibited coplanar orientation with the aromatic ring (0° < torsion < 20°), where aryl sulfoxides and sulfones were presented orthogonal to the plane of the ring (70° < torsion < 90°). The conformational preferences between thioethers and oxidized thioethers is striking: 82% and 84% of oxidized aryl thioethers crystal structures were orthogonal to the plane of the ring, while aryl ethers and thioethers were primarily coplanar with the aromatic ring (83% and 54%, respectively). Therefore, oxidation of an aryl thioether, such as 4-thiophenylalanine, can induce strong structural preferences on the presentation of the thioether linkage. Thioether derivatives of 4-thiophenylalanine could potentially be used in designed peptides and proteins to introduce new interactions or structural preferences via non-covalent interactions. Alkylation of the sulfur atom in 4-thiophenylalanine could be used to attach a functional group on 4-thiophenylalanine. Moreover, subsequent

oxidation to the sulfoxide or sulfone could control the presentation of this functional group.

In order to investigate this additional level of structural and functional control, protocols were established for oxidation of thioether derivatives of peptides containing 4-thiophenylalanine. The model peptide Ac-TXPN-NH_2 containing 4-S-methyl-thiophenylalanine was oxidized to form the sulfoxide and sulfone derivatives in solution phase (Figure 1.45 and Figure 1.46).[137] Both of these model Ac-TXPN-NH_2 peptides, containing either 4-S(methyl sulfoxide)-thiophenylalanine or 4-S(methyl sulfone)-thiophenylalanine, were characterized via NMR (Figures 1.47 and 1.48, respectively). The details of structural preferences in the peptide Ac-TXPN-NH_2 containing 4-S-methyl-thiophenylalanine sulfoxide or sulfone will be discussed in greater detail in Chapter 3.

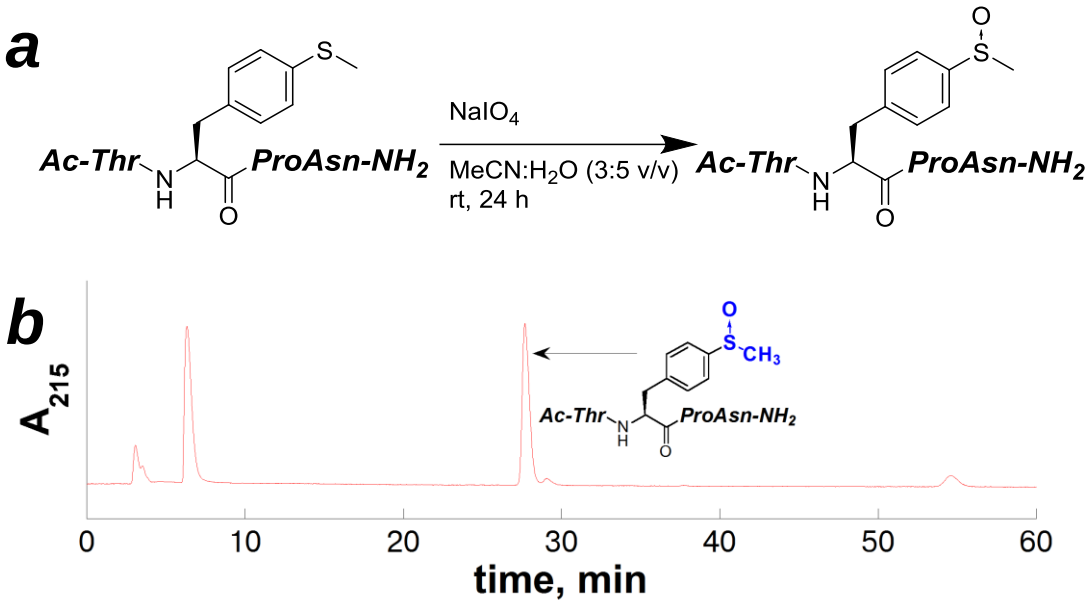


Figure 1.45 Synthesis of 4-S-methyl(sulfoxide)-thiophenylalanine in a model peptide in solution

(a) Scheme for oxidation reaction of the peptide Ac-Thr(4-SMe-Phe)ProAsn-NH_2 in solution to generate the sulfoxide; (b) HPLC chromatogram of the crude reaction products of the reaction with excess sodium periodate in a solution of acetonitrile and water (3:5 v/v solution) for 24 hours at room temperature and the peptide Ac-Thr(4-SMe-Phe)ProAsn-NH_2 generate the sulfoxide derivative of 4-S-methyl-thiophenylalanine using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated. This reaction generated a mixture of diastereomers that were inseparable by HPLC, as identified by NMR (Figure 1.47).

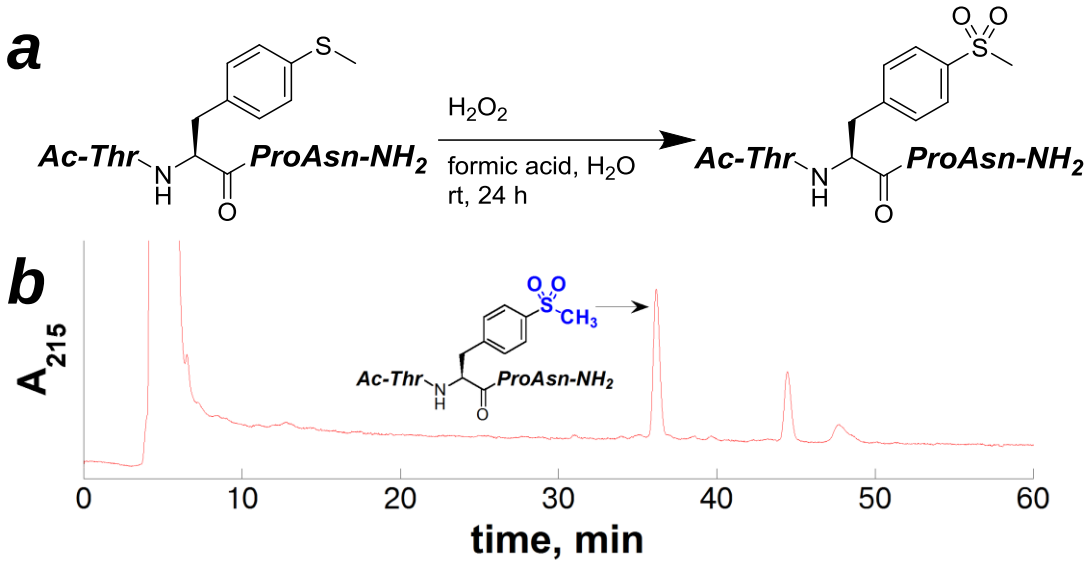


Figure 1.46 Synthesis of 4-S-methyl(sulfone)-thiophenylalanine in a model peptide in solution

(a) Scheme for oxidation reaction of the peptide Ac-Thr(4-SMe-Phe)ProAsn-NH_2 in solution to generate the sulfone; (b) HPLC chromatogram of the crude reaction products of the reaction with excess hydrogen peroxide in a solution of formic acid in water (86% solution) for 24 hours at room temperature and the peptide Ac-Thr(4-SMe-Phe)ProAsn-NH_2 generate the sulfone derivative of 4-S-methyl-thiophenylalanine using isocratic buffer A (98% water, 2% MeCN, 0.06% TFA) for 20 minutes, followed by a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

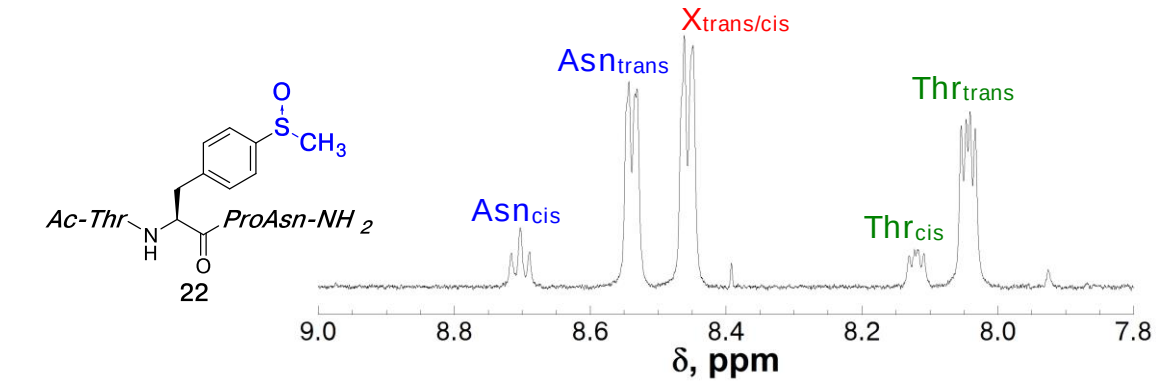


Figure 1.47 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-methyl(sulfoxide)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH$_2$ (where X = 4-S-methyl(sulfoxide)-thiophenylalanine) at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K. Amide peaks were identified by TOCSY spectrum. The NMR shows a mixture of diastereomeric mixture of sulfoxides.

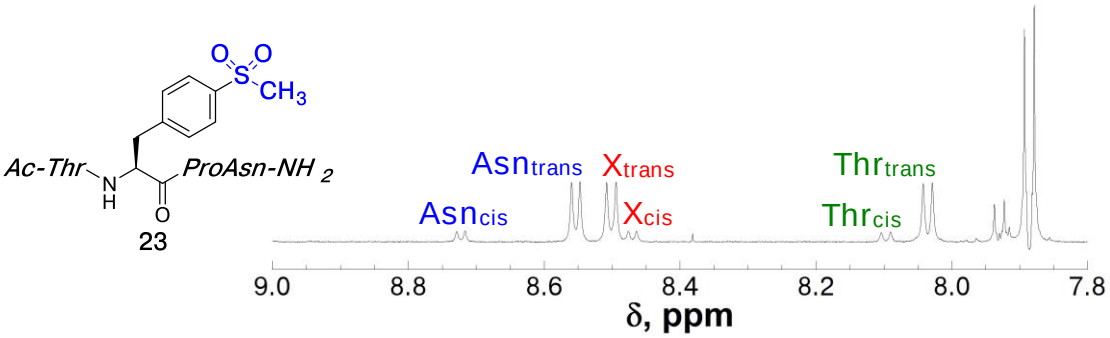


Figure 1.48 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-methyl(sulfone)-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH$_2$ (where X = 4-S-methyl(sulfone)-thiophenylalanine) at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K. Amide peaks were identified by analogy to related spectra.

Both the sulfoxide and sulfone derivatives of 4-S-methyl-thiophenylalanine were cleanly generated within the model peptide Ac-TXPN-NH_2 in solution phase. In the synthesis of the sulfoxide, the oxidation reaction with sodium periodate generated a mixture of two inseparable diastereomers, as identified via ^{1}H NMR (Figure 1.47). Oxidation of thioethers to form sulfoxides introduces a new chiral center on the sulfur atom. Chemical oxidation to the sulfoxide is not stereospecific in reaction at this site, in contrast to some enzymatic oxidations of methionine.[132, 134] From the synthesized peptide containing 4-iodophenylalanine, these two distinct derivatives were synthesized in 4 steps with only three HPLC purification steps. This rapid approach demonstrates the practicality and versatility of the solid-phase cross-coupling reaction for generating derivatives of peptides containing 4-thiophenylalanine.

1.2.13 Synthesis of peptides containing oxidized derivatives of 4-thiophenylalanine in solution phase

Oxidative modifications of cysteine are important posttranslational modifications in proteins and biomolecules which play significant roles in cellular signaling and activation of enzymes.[5, 8] Cysteine can be oxidized to different oxidation states, including symmetric or asymmetric disulfides (e.g. glutathione disulfide), sulfenic acid (–SOH), sulfinic acid (–SO_2H), and sulfonic acid (–SO_3H) (Figures 1.1, 1.3, and 1.43). Cysteine also reacts with nitric oxide to form the S-nitrosylated species (Figures 1.1 and 1.43). S-nitrosylated cysteine is involved in cellular signaling and oxidative/nitrosative stress responses.[8, 10, 140, 141] As the oxidation state of sulfur changes, the charge on the sulfur atom decreases, which ultimately affects the reactivity of cysteine and the specific biological or catalytic roles of the thiol.[4-6, 55] For example, oxidation of a thiolate to a sulfenic acid changes the sulfur atom from a

nucleophile to an electrophile (Figure 1.3).[5] Chemical probes, such as dimedone, have been developed that harness the unique reactivity of oxidized sulfur species, and can selectively alkylate and 'trap' specific oxidation states of cysteine.[10, 11, 142] As an aryl thiol, electronic perturbations on the thiol group of 4-thiophenylalanine would be reflected in changes in the electronic properties of the aromatic ring. These perturbations in electronic structure can be observed as a change in the UV spectrum or as a structural effect in a peptide. These properties can allow for 4-thiophenylalanine to potentially function as a direct probe of sulfur oxidation state. In contrast to trapping techniques, 4-thiophenylalanine could potentially be used to directly probe *dynamic, reversible* oxidative processes in cells. The direct spectroscopic identification of a sulfenic acid or sulfinic acid intermediate, via 4-thiophenylalanine, could provide insights into the consequences of oxidative posttranslational modifications of cysteine. To explore this possibility, several derivatives of 4-thiophenylalanine were synthesized with different oxidative modifications in the model peptide Ac-TXPN-NH_2.

The S-glutathionylated asymmetric disulfide of 4-thiophenylalanine can be generated via disulfide exchange with oxidized glutathione. However, this process is relatively slow without using large excesses of glutathione. Alternatively, an activated 2-thiopyridyl disulfide derivative of 4-thiophenylalanine can be used for directed disulfide exchange reactions, where the 2-thiopyridyl leaving-group promotes disulfide exchange thiols, including glutathione.[55, 143, 144] In order to characterize the electronic and structural properties of mixed disulfides of 4-thiophenylalanine, the model peptide Ac-TXPN-NH_2 containing 4-thiophenylalanine was subjected to

disulfide formation with excess 2,2′-thiopyridyl disulfide under basic conditions (Figure 1.49).

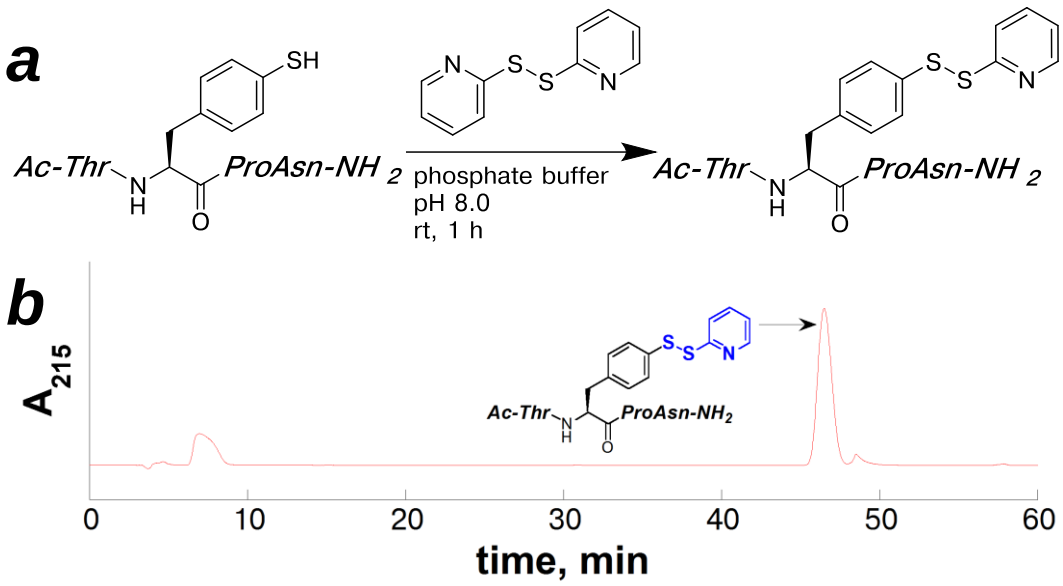


Figure 1.49 Synthesis of the model peptide containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine in solution
(a) Scheme for the reaction with 2,2′-thiopyridyl disulfide and the peptide Ac-Thr(4-SH-Phe)ProAsn-NH_2 in solution at pH 8; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with excess 2,2′-thiopyridyl disulfide in pH 8 phosphate buffer (100 mM) for 2 hours at room temperature and the peptide Ac-Thr(4-SH-Phe)ProAsn-NH_2 generate the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The peptide containing 4-thiophenylalanine-successfully formed the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine within 1 hour. The product was isolated, and characterized via ^{1}H NMR (shown at the end of this section, Figure

1.56). The details regarding the structural influence of the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine will be discussed in Chapter 3.

Next, the S-glutathionylated derivative of 4-thiophenylalanine was synthesized, in order to characterize the structural and spectroscopic properties that result from this post-translational modification. The isolated peptide Ac-TXPN-NH_2 containing the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine was allowed to react with reduced glutathione in solution under basic conditions to generate the S-glutathionylated disulfide derivative (Figure 1.50).

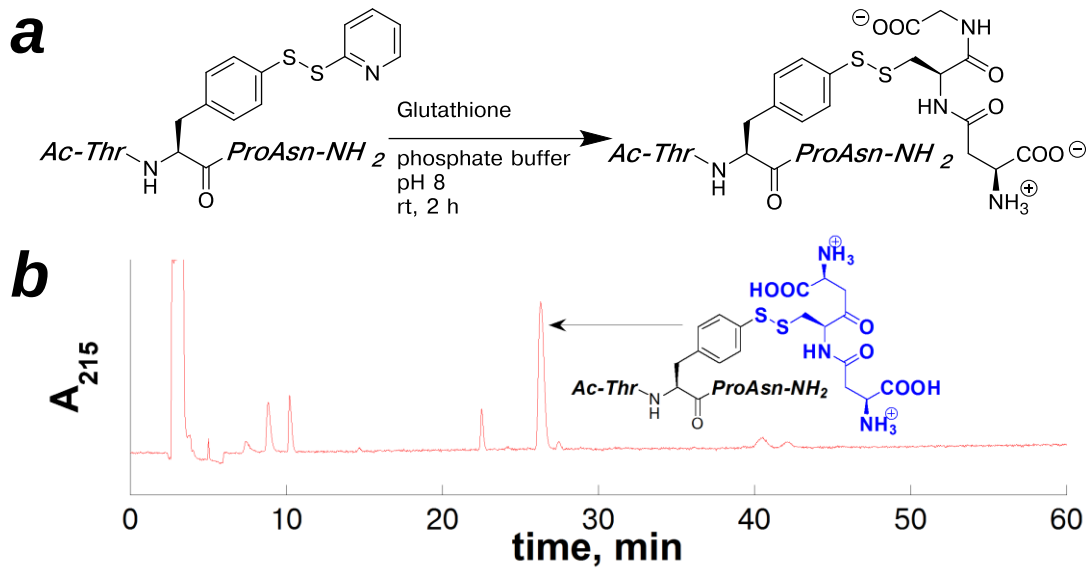


Figure 1.50 Synthesis of the model peptide containing 4-S-S-glutathione disulfide-thiophenylalanine in solution
(a) Scheme for disulfide exchange reaction with reduced glutathione and the peptide Ac-Thr(4-S-S-(2-thiopyridyl) disulfide-Phe)ProAsn-NH_2 in solution at pH 8; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with excess glutathione (reduced) in phosphate buffer (pH 8) for 2 hours at room temperature and the peptide Ac-Thr(4-S-S-(2-thiopyridyl) disulfide-Phe)ProAsn-NH_2 generate the S-glutathione disulfide derivative of 4-thiophenylalanine using a linear gradient of 0-50% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The peptide containing the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine successfully formed the glutathionylated derivative of 4-thiophenylalanine within 2 hours. The product was isolated, and characterized via ^{1}H NMR (shown at the end of this section, Figure 1.57). The details regarding the structural influence of the 2-thiopyridyl disulfide derivative of 4-thiophenylalanine will be discussed in Chapter 3.

In addition to forming asymmetric disulfides, cysteine can be oxidized to the sulfenic acid, sulfinic acid, and sulfonic acid derivatives. Each of these oxidation states on sulfur induce unique reactivity (Figure 1.3) and structural consequences (Figure 1.44). In order to examine the structural and spectroscopic properties of these sulfur oxidation states of sulfur in the context of 4-thiophenylalanine, the model peptide containing 4-thiophenylalanine was subjected to oxidation reactions to generate the sulfinic acid and sulfonic acid derivatives. Attempts to generate and isolate the sulfenic acid were unsuccessful, potentially due to rapid reaction of the sulfenate with unreacted thiolates to form the symmetric disulfide.

The sulfinic acid derivative of 4-thiophenylalanine was synthesized via oxidation reaction on the peptide Ac-T(4-SH-Phe)PN-NH_2 with hypochlorite (Figure 1.51). An alternative synthesis of the sulfinic acid was achieved via elimination of a sulfone derivative of 4-S(*N*-ethyl succinimide)-thiophenylalanine (Figure 1.52).

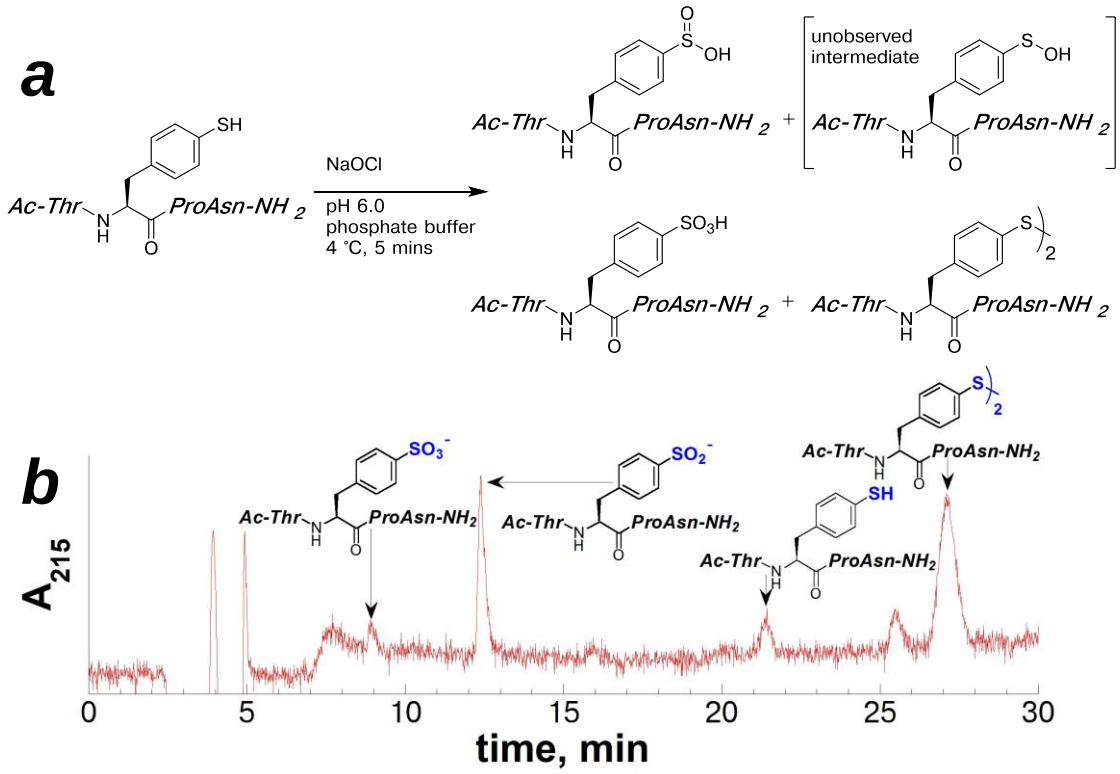


Figure 1.51 Synthesis of the model peptide containing 4-SO_2^--thiophenylalanine in solution using sodium hypochlorite
(a) Scheme for the general oxidation reaction of Ac-T(4-SH-Phe)PN-NH_2 using sodium hypochlorite; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with sodium hypochlorite (11 μM) in phosphate buffer (pH 6.0) for 5 minutes at 4 °C and the peptide Ac-Thr(4-SMe-Phe)ProAsn-NH_2 (27.5 μM) generate the with the peptide containing 4-thiophenylalanine with sodium hypochlorite using isocratic buffer A (98% water, 2% MeCN, 0.05% TFA) for 20 minutes, followed by a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.06% TFA) in buffer A over 30 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The model peptide containing 4-thiophenylalanine was successfully oxidized in solution phase to generate the sulfinic acid derivative. Oxidation of the peptide containing 4-thiophenylalanine with sodium hypochlorite at pH 6 at 4 °C for 5 minutes (Figure 1.51)[145] resulted in a mixture of products including the sulfinic acid,

sulfonic acid, and the symmetric disulfide. The model peptide containing 4-thiophenylalanine was allowed to react at pH 6 at 4 °C for 5 minutes in the absence of sodium hypochlorite, and no oxidation was observed. It is possible that the symmetric disulfides that resulted from the reaction with sodium hypochlorite and the peptide Ac-T(4-SH-Phe)PN-NH_2 were generated via rapid reaction of the sulfenic acid with unreacted thiolates.

In order to specifically synthesize the sulfinic acid of 4-thiophenylalanine, aa alternative synthetic route was developed where the peptide containing 4-SO_2(*N*-ethyl succinimide)-thiophenylalanine was subjected to base-mediated elimination (Figure 1.52).[146] The model peptide containing 4-SO_2(*N*-ethyl succinimde)-thiophenylalanine was generated via alkylation of the peptide Ac-T(4-SH-Phe)PN-NH_2 with *N*-ethyl malelimide, and subsequent oxidation reaction using hydrogen peroxide as described in Chapter 1.2.12.

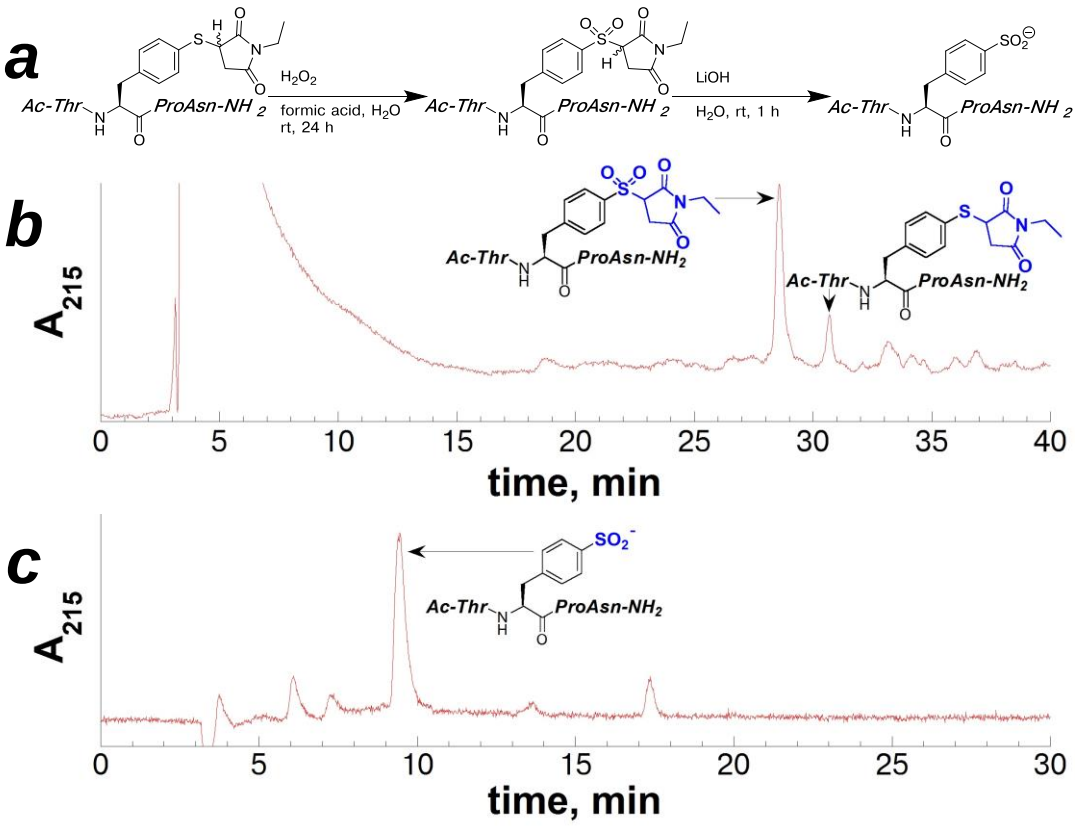


Figure 1.52 Synthesis of the model peptide containing 4-SO_2^--thiophenylalanine in solution via oxidation reaction of the S-succinimide derivative and subsequent elimination reaction

(a) Scheme for the two-step selective synthesis of the peptide Ac-T(4-SO_2^--Phe)PN-NH_2 via base-mediated elimination of the peptide Ac-T(4-SO_2(*N*-ethyl succinimide)-Phe)PN-NH_2, which was synthesized via oxidation reaction on the peptide Ac-T(4-S(*N*-ethyl succinimide)-Phe)PN-NH_2; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with the peptide containing 4-S-(*N*-ethyl succinimide sulfone)-thiophenylalanine with excess hydrogen peroxide in a solution of formic acid in water (86% solution) for 24 hours at room temperature using isocratic buffer A for 10 minutes, followed by a linear gradient of 0-40% buffer B in buffer A over 40 minutes; (c) crude reaction products resulting from the base-catalyzed elimination reaction on the peptide Ac-T(4-SO_2(*N*-ethyl succinimide)-Phe)PN-NH_2 with lithium hydroxide (10 mM solution in water) at room temperature for 1 h to generate the peptide Ac-T(4-SO_2^--Phe)PN-NH_2 using a linear gradient of 0-45% buffer B in buffer A over 30 minutes. Products identified via ESI-MS are indicated.

The sulfinic acid of 4-thiophenylalanine was synthesized selectively via base-mediated elimination reaction on the peptide containing a sulfone derivative of 4-S-(*N*-ethyl succinimide-thiophenylalanine. In contrast to the direct oxidation reaction using sodium hypochlorite on the peptide Ac-T(4-SH-Phe)PN-NH_2, which resulted in a mixture of products, the sulfinic acid derivative of 4-thiophenylalanine was cleanly generated as the sole product of the base-mediated elimination reaction. With sufficient quantities of the sulfinic acid derivative of 4-thiophenylalanine generated by this synthetic route, the peptide Ac-T(4-SO_2^--Phe)PN-NH_2 was characterized via NMR (shown at the end of this section, Figure 1.58). The details of structural preferences in the peptide Ac-T(4-SO_2^--Phe)PN-NH_2 will be discussed in greater detail in Chapter 3.

The highest oxidation state of cysteine is the sulfonic acid derivative (Figures 1.3 and 1.43). In order to characterize the potential spectroscopic and structural properties of the sulfonic acid derivative in the context of 4-thiophenylalanine, the peptide containing 4-thiophenylalanine was subjected to oxidation reaction with methyltrioxorhenium (MTO). In contrast to the oxidation reaction on the peptide Ac-T(4-SH-Phe)PN-NH_2 using hypochlorite, MTO can selectively form sulfonates from disulfides in the presence of acetonitrile an hydrogen peroxide (Figure 1.53).[147, 148]

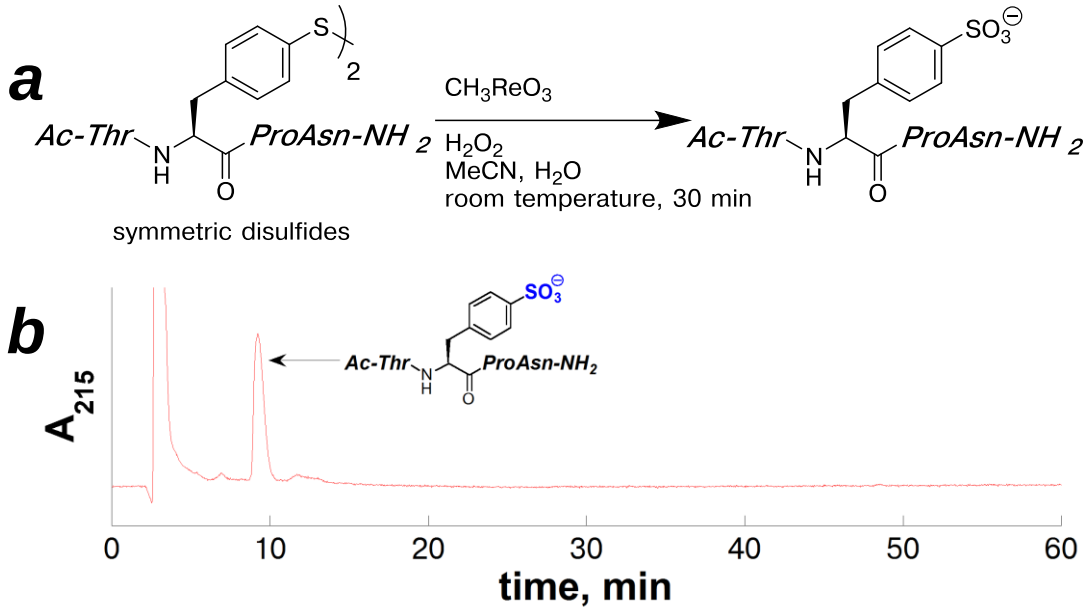


Figure 1.53 Synthesis of the model peptide containing 4-SO_3^--thiophenylalanine in solution with methyl trioxorhenium
(a) Scheme for the oxidation reaction to generate the peptide Ac-Thr(4-SO_3^--Phe)ProAsn-NH_2 in solution; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with the peptide containing 4-thiophenylalanine disulfide with methyltrioxorhenium (1.5 mg/mL solution in acetonitrile) in the presence of excess hydrogen peroxide (30% in water) in for 30 minutes at room temperature using isocratic buffer A (98% water, 2% MeCN, 0.06% TFA) for 20 minutes, followed by a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The reaction with MTO in the presence of acetonitrile and hydrogen peroxide on the peptide containing 4-thiophenylalanine disulfide selectively generated the sulfonic acid derivative. The sulfonic acid derivative represents the highest oxidation state of sulfur, and is the stable end-product of many oxidative pathways in post-translational modifications of cysteine. Isolation of sufficient quantities of this derivative of 4-thiophenylalanine allowed for the characterization of the spectroscopic and structural properties resulting from this oxidation state of sulfur. The peptide Ac-

T(4-SO_3^--Phe)PN-NH_2 was characterized via ^{1}H NMR (shown at the end of this section, Figure 1.59). The details of structural preferences in the peptide Ac-T(4-SO_3^--Phe)PN-NH_2 will be discussed in greater detail in Chapter 3.

S-nitrosylation is another important posttranslational modification of cysteine that results from reaction with reactive-nitrogen species, and is implicated in cell signaling and cardiac disease.[48] The S-nitrosylated derivative of the peptide containing 4-thiophenylalanine was synthesized using nitrosonium tetrafluoroborate (Figure 1.54).[149] The product was identified via UV spectra and by ESI-MS, where a mixture of the S-nitrosylated and symmetric disulfide derivatives of 4-thiophenylalanine were observed.

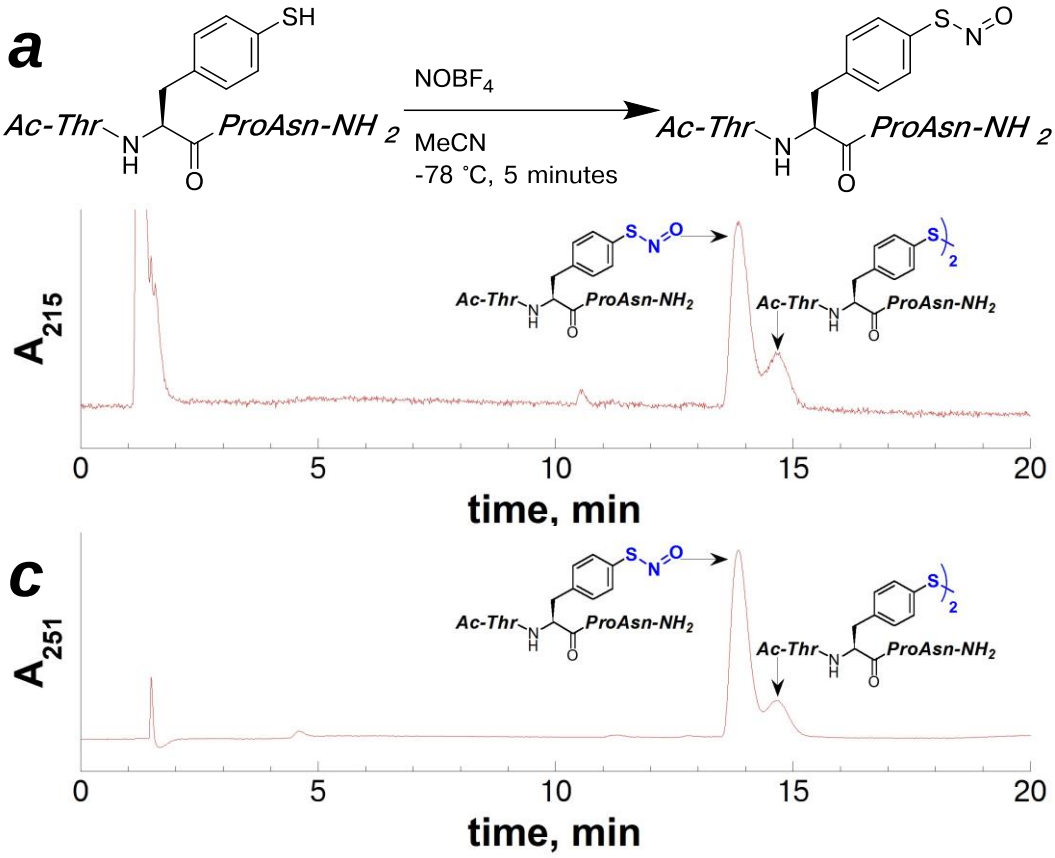


Figure 1.54 Synthesis of the model peptide containing 4-S-NO-thiophenylalanine in solution
(a) Scheme for the S-nitrosylation reaction on the peptide Ac-Thr(4-SH-Phe)ProAsn-NH_2 in solution; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with the peptide containing 4-thiophenylalanine with excess nitrosonium tetrafluoroborate in dry acetonitrile for 5 minutes at –78 °C using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.06% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 15 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

Synthesis of an S-nitrosylated derivative of 4-thiophenylalanine was successfully acheived within peptides in solution phase using nitrosonium tetrafluoroborate at –78 °C. The S-nitrosylated product was identified by flash-freezing (in liquid nitrogen) of the sample from the HPLC, and immediate analysis via

ESI-MS showed a mixture of the S-nitrosylated and symmetric disulfide derivatives of 4-thiophenylalanine within the model peptide. Due to the instability of the S-nitrosylated species at room temperature (and rapid disulfide formation), the product peptide Ac-T(4-SH-Phe)PN-NH_2 was not characterized via ^1H NMR. However, the UV spectrum of 4-S-NO-thiophenylalanine was detected via diode array on HPLC purification. The results from the diode array allow for qualitative comparisons of the λ_{max} between the different oxidized species of 4-thiophenylalanine within peptides.

In order to characterize the different spectroscopic properties of each sulfur oxidation state, the UV spectra for each of the described derivatives were obtained (Figure 1.55, Table 1.9). These derivatives of 4-thiophenylalanine represent the different oxidation states that occur as posttranslational modifications on cysteine. Due to the instability of the S-nitrosylated derivative under the conditions for obtaining a UV spectrum, only the qualitative spectrum from HPLC diode array is shown (Figure 1.55b). The UV spectrum of the sulfonic acid derivative of 4-thiophenylalanine was also only obtained qualitatively via detection on diode array (Figure 1.55b). The spectroscopic properties for these derivatives of 4-thiophenylalanine are summarized in Table 1.9.

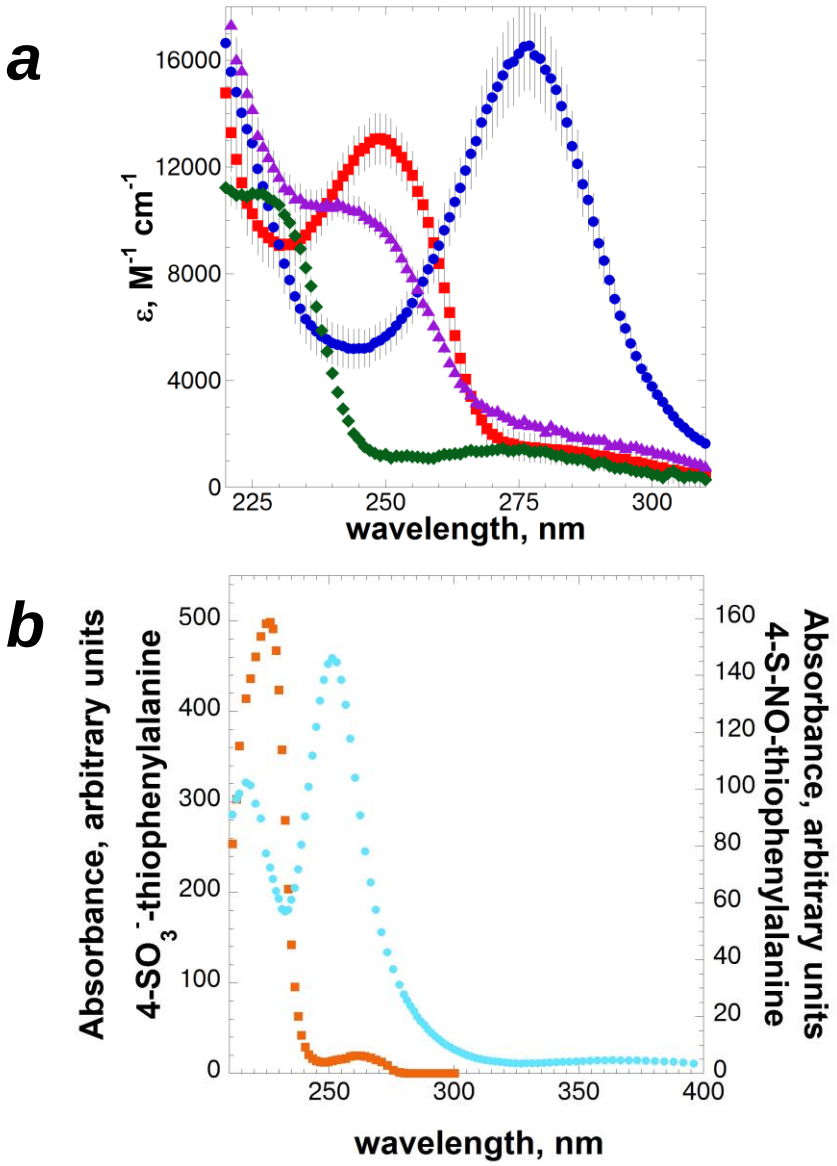


Figure 1.55 UV spectra for oxidized derivatives of 4-thiophenylalanine in the peptide Ac-TXPN-NH_2

Red squares: Ac-T(4-SH-Phe)PN-NH_2 at pH 4.0; blue circles: Ac-T(4-S^--Phe)PN-NH_2 at pH 8.5; green diamonds: Ac-T(4-SO_2^--Phe)PN-NH_2 at pH 4.0; purple triangles: Ac-T(4-S-S-glutathione disulfide-Phe)PN-NH_2 at pH 4.0; orange squares: Ac-T(4-SO_3^--Phe)PN-NH_2 at pH 2; cyan circles: Ac-T(4-S-NO-Phe)PN-NH_2.
(a) Quantitative UV spectra for the model peptide containing derivatives of 4-thiophenylalanine. Samples containing 4-thiophenylalanine were quantified using Ellman's test; all other samples were quantified via NMR. Data represent an average of at least three trials. Error bars indicate standard error.
(b) qualitative UV spectra for the model peptide containing derivatives of 4-thiophenylalanine obtained from HPLC equipped with a diode array.

Table 1.9. Summary of UV absorbance spectra for oxidized derivatives of 4-thiophenylalanine in the peptide Ac-TXPN-NH_2

λ_{max} of the absorbance and molar extinction coefficients (ε) were derived from data shown in Figure 1.55a.
[a]Qualitative data obtained via HPLC chromatogram detected via diode array, shown in Figure 1.55b.

Ac-T(4-**X**-Phe)PN-NH_2 **X** =	λ_{max}, nm	ε M^{-1} cm^{-1}
-SH, pH 4.0	249	13060 ± 960
-S^-, pH 8.5	277	16540 ± 1680
-S-S-glutathione disulfide, pH 4.0	241	10670 ± 1660
-SO_2^-, pH 4.0	225	11020 ± 170
	276	1430 ± 290
-SO_3^-, pH 2[a]	225[a]	n.d.[a]
	261[a]	
-S-NO[a]	251[a]	n.d.[a]

The derivatives of 4-thiophenylalanine representing different oxidation states of sulfur exhibited distinctive spectroscopic signatures observed by UV. A significant red shift in the absorption λ_{max} was observed on formation of the thiolate of 4-thiophenylalanine, from 249 nm to 277 nm. The sulfinic acid and sulfonic acid derivatives of 4-thiophenylalanine exhibited intense absorption at 225 nm, with a weaker absorption at 276 nm or at 261 nm, respectively. The S-glutathionylated

derivatives of 4-thiophenylalanine exhibited intense absorption at 241 nm, with similar intensity to the thiol species. The peptide containing 4-S-NO-thiophenylalanine exhibited a characteristic UV spectrum, with a λ_{max} at 251 nm. Combined, there was an apparent blue shift in the aromatic λ_{max} with increasing oxidation state, as the sulfur atom becomes less negatively charged.

With these distinctive shifts in the UV spectrum upon oxidation of the sulfur atom, 4-thiophenylalanine could potentially act as a direct spectroscopic probe for sulfur oxidation state. In contrast to current methods of detecting sulfur oxidation state,[10] 4-thiophenylalanine could potentially be used as a "real-time" direct probe of sulfur redox activity. 4-Thiophenylalanine, either incorporated into synthetic peptide sequences or expressed in proteins, could potentially be used for real-time imaging of the oxidative activity. Dramatic shifts were observed upon S-nitrosylation of the peptide containing 4-thiophenylalanine, with a unique λ_{max} at 251 nm. This unique spectral property of this derivative can potentially be used for "real-time" imaging of S-nitrosylation events, which are implicated in cardiac and neurodegenerative diseases.[48, 150, 151]

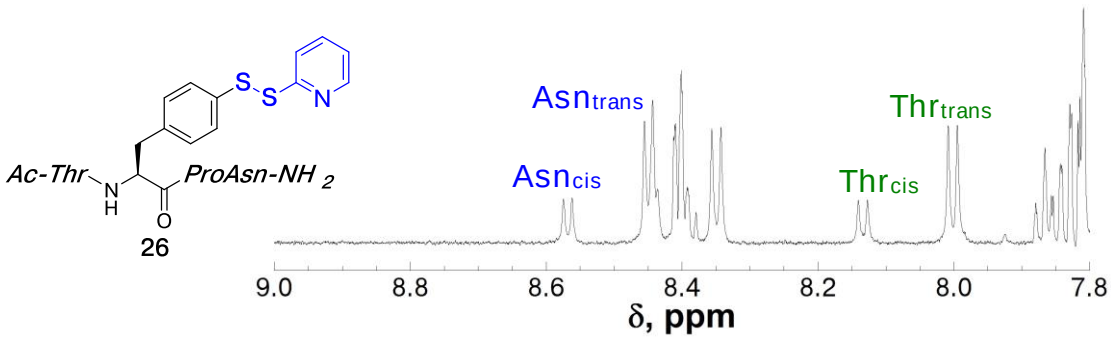


Figure 1.56 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K.

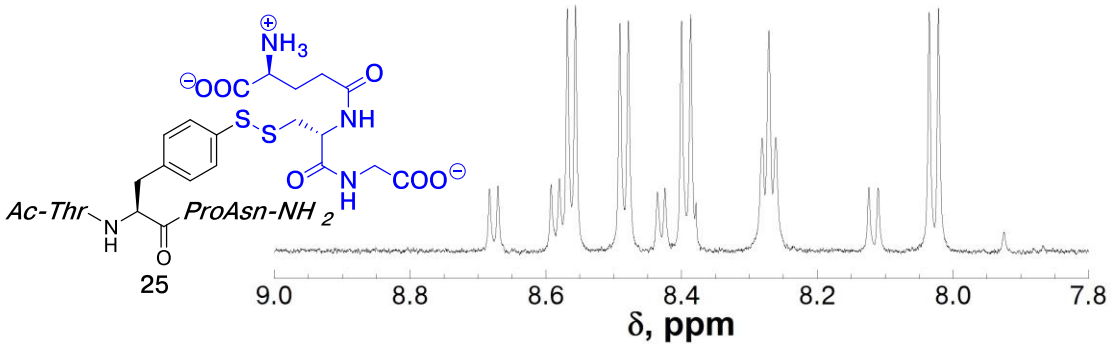


Figure 1.57 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-S-glutathione disulfide-thiophenylalanine
NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH$_2$ containing 4-S-S-glutathione disulfide-thiophenylalanine at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O at 298 K.

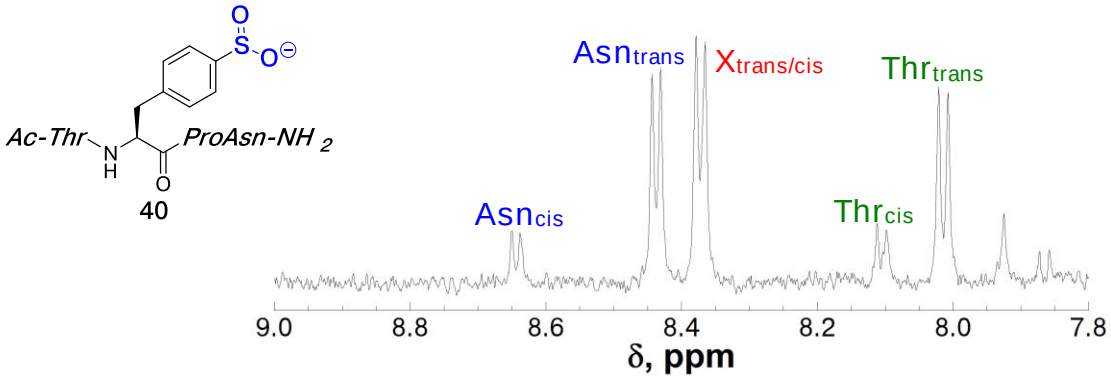


Figure 1.58 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-SO_2^--thiophenylalanine

NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH_2 containing 4-SO_2^--thiophenylalanine at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by TOCSY spectrum.

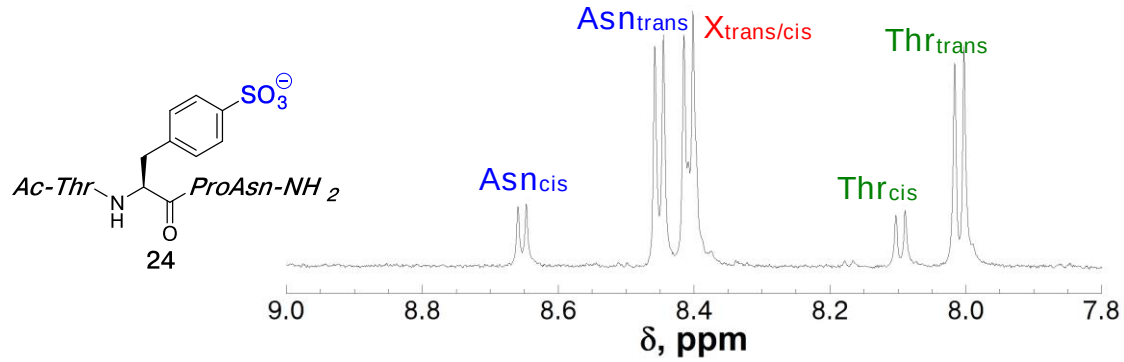


Figure 1.59 ^{1}H NMR spectrum of the amide region of the peptide Ac-ThrXProAsn-NH_2 containing 4-SO_3^--thiophenylalanine

NMR spectrum of the amide region for the peptide Ac-ThrXProAsn-NH_2 containing 4-SO_3^--thiophenylalanine at pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by TOCSY spectrum.

1.2.14 Aqueous azide-alkyne 1,3-cycloaddition of a model peptide containing of 4-S-propargyl-thiophenylalanine

Alkyne groups, like alkene functional groups, have been widely utilized as conjugation substrates in bioorthogonal reactions in peptides and proteins.[13, 51, 116] Although it is defined as a class of rapid and selective reactions with broad substrate scope, the term "'click' chemistry" is most commonly associated with the copper-catalyzed azide-alkyne cycloaddition that was independently described by Sharpless[118] and Meldal[152] (Figure 1.60).[116, 153, 154]Alkyne functional groups react rapidly with azides, regarded as 'click' reactions. Sharpless defined 'click' chemistry as a class of reactions that "must be modular, wide in scope, give very high yields, generates only inoffensive byproducts that can be removed by nonchromatographic methods, and be stereoselective."[155] A plethora of azide- and alkyne- functionalized reagents are commercially available. These reactions proceed in both organic and aqueous solvents, at mild temperatures, and with copper or ruthenium, produce 1,4-disubstituted or 1.5-disubstituted triazoles regioselectively.[51, 116] Azide-alkyne 'click' reactions have been widely utilized in organic synthesis, in fluorescent labelling proteins and peptides, in synthesis of glycoprotein conjugates, in modular drug design, in bioactivity assays, in high-throughput screening applications, and for generation of cellular imaging agents.[51, 116, 118, 156-158]

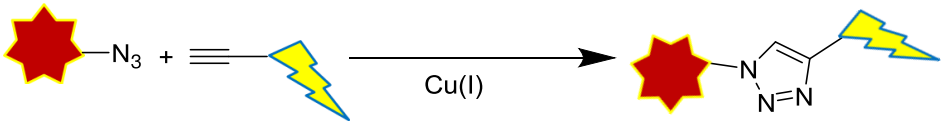


Figure 1.60 Generalized copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition Sharpless[154] and Meldal[152] independently demonstrated a copper-catalyzed azide-alkyne 1,3-cycloaddition reaction that yields a 1,4-disubstituted triazole. These reactions only require mild temperatures and are compatible with organic and aqueous solvents. Small molecules and biomolecules with azide or alkyne functional groups are widely commercially available.

A derivative of 4-thiophenylalanine was synthesized to contain a propargyl thioether functional group (Chapter 1.2.8) via alkylation reaction on solution. This amino acid contains a terminal alkyne as an aryl thioether, suggesting its application in azide-alkyne cycloadditions and other applications of alkynes (including Sonogoshira reactions with aryl iodides).[119, 139, 159] In addition, as an aryl thioether, 4-S-propargyl-thiophenylalanine could function as a locus for control of protein structure. Non-covalent interactions involving electron rich aromatic rings are well known and utilized in designed peptides and proteins.[18, 93, 160-163] The sulfur atom in thioethers, such as methionine, is also known to participate in non-covalent interactions.[7, 135, 164-166] Aryl thioethers have distinctive geometry that is often compared to oxygen-containing ethers, but there are key differences (Figure 1.61). The C–S–C bond angle of a an aryl thioether is ~100°, in contrast to the more sp^2-like bond angle for an aryl ether (~120°). Furthermore, aryl thioethers can be oxidized to sulfoxides and sulfones, which in crystal structures exhibits a structural preference about the aromatic C–C–S–C torsion angle (Figure 1.44), changing the attached group from being mostly co-planar to orthogonal to the aromatic ring.[139] Oxidation of the sulfur atom in aryl thioethers induces a change in the conformational preference in the torsion angle of

the ether linkage. Therefore, oxidation of 4-S-propargyl-thiophenylalanine inherently affects the presentation of the propargyl functional group in a manner that is not possible (or enforced) by propargyl tyrosine.

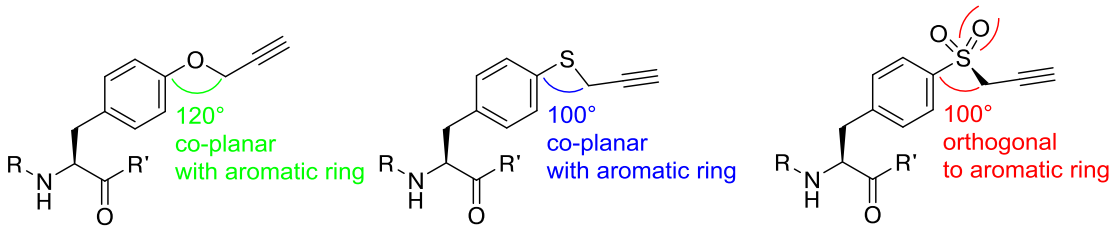


Figure 1.61 Conformational preferences about the aromatic C_{aro}–C_{aro}–X–Csp^3 torsion angle in crystal structures
Aryl ethers and aryl thioethers exhibit unique structural preferences. The bond angles for C_{aro}–X–Csp^3 (X = O, S) are distinct. In crystal structures in the Cambridge Structural Database, it was found that the ether linkage tends to be coplanar with the aromatic ring. In contrast, for oxidized thioethers the alkyl group to oriented perpendicular to the plane of the aromatic ring. See Figure 1.44 for additional details.

Given the potential role of 4-S-propargyl-thiophenylalanine derivatives as a means for conformational control of peptide structure, and as bioorthogonal "handles" in bioorthogonal reactions, we sought to examine utility of 4-thiophenylalanine derivatives in azide-alkyne 1,3-cycloaddition reactions. The model peptide containing 4-S-propargyl-thiophenylalanine was synthesized via solution phase alkylation reaction of propargyl bromide with the peptide Ac-T(4-SH-Phe)PN-NH_2, as described previously. The peptide containing 4-S-propargyl-thiophenylalanine was also oxidized in solution to generate the sulfone derivative. Both of these peptides containing 4-S-propargyl-thiophenylalanine were subjected to copper-catalyzed 1,3-cycloaddition reaction with 4-azidoaniline to determine if oxidation of the sulfur atom affected reactivity (Figures 1.62 and 1.63).

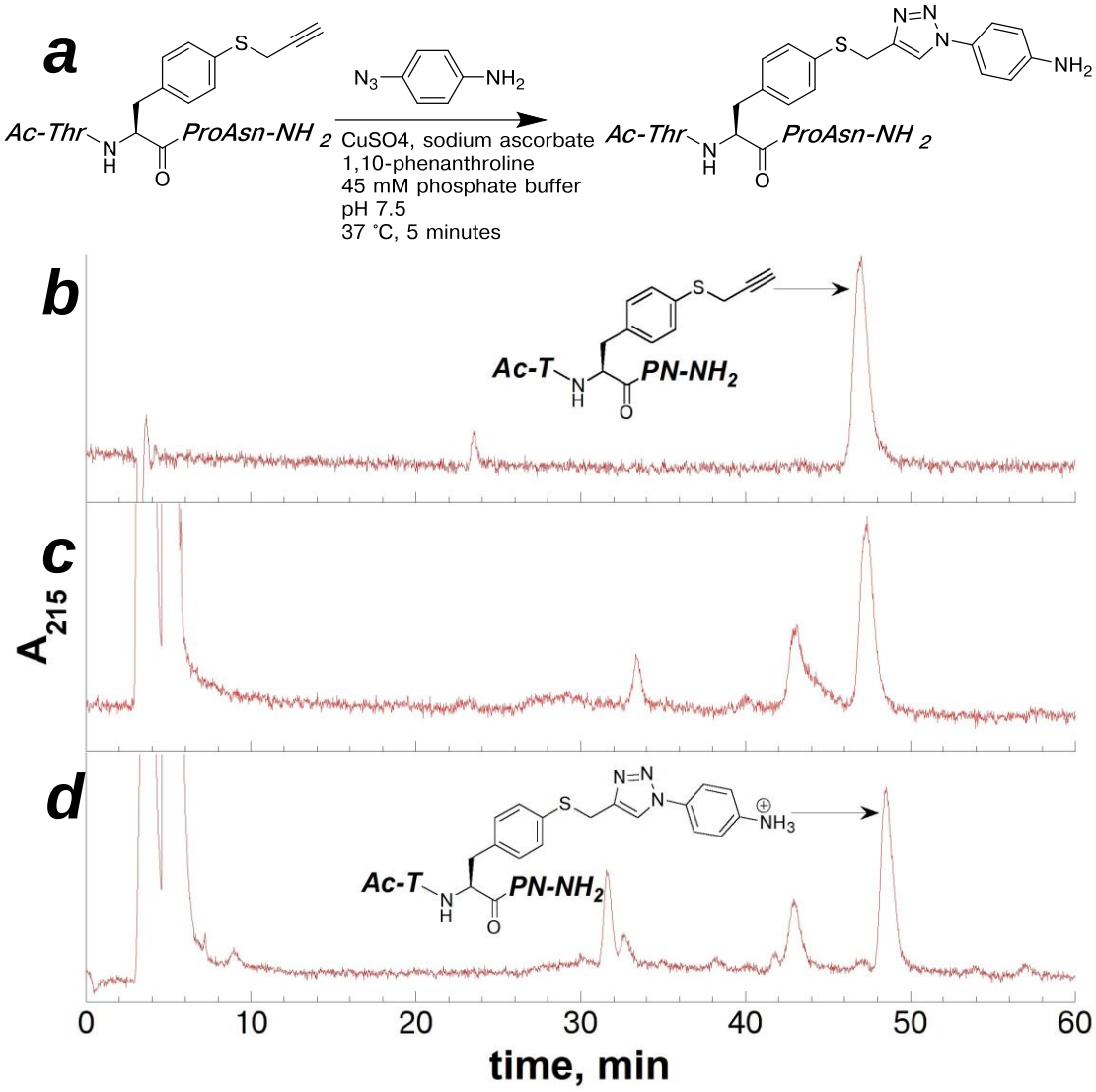


Figure 1.62 Copper-catalyzed 1,3-cycloaddition reaction on the peptide containing 4-S-propargyl-thiophenylalanine with 4-azidoaniline

(a) Scheme for the ‘click’ reaction on the peptide Ac-Thr(4-S-propargyl-Phe)ProAsn-NH_2 in solution with 4-azidoaniline; (b) HPLC chromatogram of the purified peptide Ac-T(4-S-propargyl-Phe)PN-NH_2; (c) HPLC chromatogram of the aqueous work-up of the peptide Ac-T(4-S-propargyl-Phe)PN-NH_2 and reaction reagents, prior to the addition of 4-azidoaniline; (d) HPLC chromatogram of the aqueous work-up of the crude reaction with the peptide Ac-T(4-S-propargyl-Phe)PN-NH_2 (70 μM final concentration) with 4-azidoaniline (100 mM final concentration) after allowing to incubate at 37 °C for 5 minutes in the presence of copper(II) sulfate (0.9 mM final concentration), sodium ascorbate (50 mM final concentration), and 1,10-phenanthroline (0.2 mM final concentration) in a solution of phosphate buffer (50 mM, pH 7.5) with 2% DMSO.
Crude reaction products were analyzed via HPLC using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

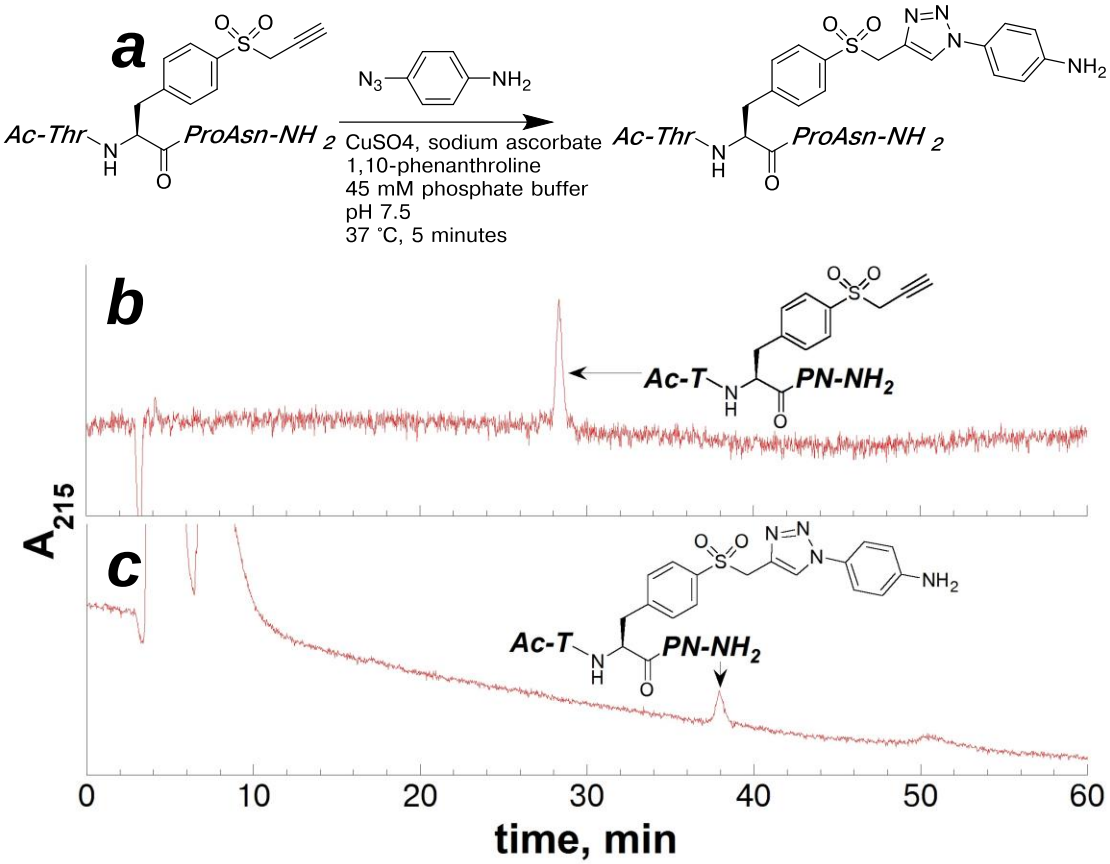


Figure 1.63 Copper-catalyzed 1,3-cycloaddition reaction on the peptide containing 4-SO_2-propargyl-thiophenylalanine with 4-azidoaniline
(a) Scheme for the 'click' reaction on the peptide Ac-Thr(4-SO_2-propargyl-Phe)ProAsn-NH_2 in solution with 4-azidoaniline; (b) HPLC chromatogram of the purified peptide Ac-T(4-SO_2-propargyl-Phe)PN-NH_2; (c) HPLC chromatogram of the aqueous work-up of the crude reaction with the peptide Ac-T(4-SO_2-propargyl-Phe)PN-NH_2 (70 μM final concentration) with 4-azidoaniline (100 mM final concentration) after allowing to incubate at 37 °C for 5 minutes in the presence of copper(II) sulfate (0.9 mM final concentration), sodium ascorbate (50 mM final concentration), and 1,10-phenanthroline (0.2 mM final concentration) in a solution of phosphate buffer (50 mM, pH 7.5) with 2% DMSO.
Crude reaction products were analyzed via HPLC using a linear gradient of 0-15% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The model peptides containing 4-S-propargyl-thiophenylalanine and the corresponding sulfone reacted rapidly under aqueous conditions with 4-azidoaniline to generate the triazole 'click' product. In the presence of copper(I), both 1,3-cycloaddition reactions were complete within 5 minutes at 37 °C. Combined with the observed geometric preferences of aryl thioethers, sulfoxides, and sulfones in crystal structures, 4-S-propargyl-thiophenylalanine and its corresponding oxidized derivatives provide a bioorthogonal "handle" with distinctive structural control. The structural preferences induced by oxidized thioethers of 4-S-propargyl-thiophenylalanine are not possible with propargyl tyrosine derivatives. Notably, propargyl tyrosine has been incorporated into expressed proteins via cell-free protein expression for utility in 1,3-cycloaddition bioorthogonal reactions.[167] The ready expression of proteins containing propargyl tyrosine suggests that 4-S-propargyl-thiophenylalanine could potentially be expressed in proteins using similar protocols, providing a means for incorporating this bifunctional amino acid into proteins.

1.3 Discussion

4-Thiophenylalanine exhibits distinctive spectroscopic, reactive, and structural properties from its natural oxygen analogue, tyrosine. Despite diverse potential applications, 4-thiophenylalanine has only been synthesized in 6 examples over more than 100 years. Among these examples, 4-thiophenylalanine was utilized for studying initial reaction steps in protein folding, for SAR studies in angiotensin II analogues to investigate the roles of aromatic amino acids on structure, in DNA topoisomerase for examining the nucleophilic role of tyrosine, in photo-initiated radical reactions, and for bacteriostatic therapeutics.[16, 17, 21, 23, 62, 168] The reported syntheses of 4-thiophenylalanine generated an amino acid that could not be readily incorporated into

peptides without using harsh or dangerous reaction conditions, and these synthetic challenges potentially inhibited the broader application of 4-thiophenylalanine. We sought to develop an alternative synthetic strategy, inspired by copper-mediated cross-coupling reaction approaches, to synthesize 4-thiophenylalanine within peptides.

We described a strategy in which commercially available Fmoc-4-iodophenylalanine was incorporated into peptides via solid-phase peptide synthesis, and the resin containing the peptide was subjected to copper-mediated cross-coupling conditions with thiolacetic acid to generate peptides contatining 4-S(acetyl)-phenylalanine.[169] Following cleavage and deprotection reactions, the peptides were subjected to reduction and thiolysis reaction in solution using DTT. In this manner, peptides containing 4-thiophenylalanine were synthesized in high conversion from peptides containing 4-iodophenylalanine in less than 24 hours using inexpensive, air-stable reagents, requiring only one purification of the resultant peptide. Earlier syntheses of 4-thiophenylalanine for incorporation into peptides required 6-9 synthetic steps in solution, generally incorporating a thiol protecting-group that would require harsh reaction conditions to remove from the peptide. In contrast, this solid-phase synthetic strategy was considerably less time-consuming and more practical in comparison to generating the protected amino acid for incorporation via solid-phase peptide synthesis. Using the solid-phase cross-coupling reaction, a series of peptides were synthesized containing 4-thiophenylalanine, including a proline-rich helix and a variant of the trp cage miniprotein. The cross-coupling reaction conditions were compatible with all canonical amino acid functional groups. Special consideration was required for peptides containing cysteine, where the trityl-protecting group was potentially unstable in the cross-coupling reaction conditions. In addition, this work

also represents the first reported copper-mediated cross-coupling reaction between an aryl halide and thiolacetic acid.

The trp cage miniprotein is a well-established, stabilized 20 amino acid protein, and the structure and stability of the miniprotein is determined by non-covalent interactions involving Trp6 and Tyr3 (Figure 1.27). The Y3(4-SH-Phe) variant of the trp cage miniprotein was synthesized using this copper-mediated cross-coupling strategy. Incorporating 4-thiophenylalanine in place of tyrosine in the trp cage miniprotein effectively replaces the hydroxyl group with a thiol. Therefore, the Y3(4-SH-Phe) variant of the trp cage miniprotein can be used to study the chalcogen effects on structural stability. In neutral form, the Y3(4-SH-Phe) trp cage miniprotin exhibited both structural and thermal stability similar to that of the native (tyrosine) trp cage miniprotein.[94-96, 103] These data demonstrate the compatibility of the cross-coupling reaction developed with complex peptide synthesis. The thermal stability differed between the thiol and thiolate forms of the Y3(4-SH-Phe) trp cage miniprotein by approximately 15 °C (Figure 1.28b), with the thiolate form of the protein being less stable. This reduced stability in the anionic form could potentially be due to a charge repulsion between the thiolate and the aromatic ring in the tryptophan residue. Alternatively, the reduced stability could be due to unfavorable repulsion between the electron-rich aromatic ring of 4-thiophenylalanine and the tryptophan aromatic ring, or due to disruption of a hydrophobic interaction between tyrosine and proline rings (Figure 1.27). Practical incorporation of 4-thiophenylalanine into peptides allowed for further study on the structural effects of the sulfur analogue of tyrosine. Furthermore, 4-thiophenylalanine could potentially be used to modulate the stability and structure of proteins.

Due to the challenges in prior syntheses 4-thiophenylalanine, the spectral, structural, and reactive properties of this amino acid were not previously characterized. With a facile approach to synthesize 4-thiophenylalanine, we sought to fully characterize the unique properties of this aryl thiolated amino acid. The pK_a of 4-thiophenylalanine was measured to be 6.4 in water. 4-Thiophenylalanine is more acidic than cysteine (pK_a = 8.5[109]), which allows for site-selective alkylation reactions on 4-thiophenylalanine in the presence of cysteine under acidic conditions. The UV spectrum of 4-thiophenylalanine indicated dramatic changes in the aromatic ring between neutral and thiolate forms of this amino acid, as shown by a red shift of the λ_{max} from acidic conditions (λ_{max} = 243 nm) to basic conditions (λ_{max} = 276 nm). These shifts in the UV spectra upon deprotonation of the thiol indicate significant shifts in the electronic distribution in the aromatic ring.

4-Thiophenylalanine contains an acidic, nucleophilic thiol, which is mostly in the anionic thiolate form under neutral pH conditions. 4-Thiophenylalaine was readily modified within peptides in solution using a series of alkylating reactions, and an approach was developed for conducting alkylation reactions on peptides on solid phase. In a competitive alkylation reaction, a peptide containing 4-thiophenylalanine was selectively alkylated in the presence of cysteine at pH 5.0. This difference in reactivity between indicates that 4-thiophenylalanine can potentially be used for site-selective modification in peptides and proteins containing cysteine. Several derivatives of 4-thiophenylalanine were synthesized via alkylation reactions, including the S-methyl, S-allyl, S-propargyl, S-2-nitrobenzyl, and S-*N*-ethyl succinimide derivatives of 4-thiophenylalanine. The S-2-nitrobenzyl functionality effectively “masks” the thiol group, which could be selectively removed via intense exposure to 365 nm light.[121]

The S-allyl and S-propargyl derivative of 4-thiophenylalanine were further utilized for bioorthogonal ligation reactions, including olefin cross-metathesis reactions and copper(I) mediated azide-alkyne 1,3-cycloaddition reaction. The variety of these alkylation reactions demonstrates the versatility of 4-thiophenylalanine as a functional "handle." With the enhanced acidity of 4-thiophenylalanine, or the ability to modify the thiol with bioorthogonal functional groups, this amino acid or related derivatives might be readily expressed in proteins via modified aminoacyl tRNA synthases[17, 63] and introduce a locus for site-selective modification in proteins.

Cysteine and methionine oxidation are crucial post-translational modifications in cell signaling and enzyme activation and inactivation.[5, 6, 10] With the unique spectral properties of 4-thiophenylalanine, we examined the potential spectroscopic signatures on different sulfur oxidation states of 4-thiophenylalanine. 4-Thiophenylalanine and alkylated variants were subjected to oxidation within peptides in solution to generate the sulfoxide, sulfone, asymmetric disulfides, sulfinic acid, sulfonic acid, and S-nitrosylated derivatives of 4-thiophenylalanine within peptides. Distinctive changes in the UV absorbance were observed upon oxidation of the sulfur atom, indicating 4-thiophenylalanine could potentially be used as a direct probe of sulfur oxidation state. Furthermore, the changes in the UV spectrum for the different sulfur oxidation states indicated unique changes in the electronic distribution in the aromatic ring, which can potentially modulate the strength non-covalent interactions with the aromatic ring of 4-thiophenylalanine.[7, 170]

The shifts in the UV spectrum and fluorescence emission spectrum in 4-thiophenylalanine and related derivatives indicated that the molecular orbitals in the aromatic ring are sensitive to electronic changes on the sulfur atom. Aromatic

interactions in proteins are important for biological recognition, protein folding and stability, and are widely utilized in drug design.[18, 93, 171] Modulation of aromatic electronics through substituent effects intrinsically effects the strength of aromatic interactions.[91, 172] 4-Thiophenylalanine and related derivatives provide a wide range of aromatic substituents that can dramatically shift the electronics of the aromatic ring, which can be utilized in designed proteins, peptides, or small molecule inhibitors.[7, 161, 173, 174] For example, if a tyrosine residue is known to participate in an aromatic interaction within a protein, replacing the tyrosine with 4-thiophenylalanine will affect this aromatic interaction, either strengthening or weakening it, or altering the geometric preference of the aromatic ring. Thioether derivatives of 4-thiophenylalanine can also be utilized for tuning non-covalent interactions in place of methionine. In cases where thioethers are known to influence reactive properties of aromatic rings, such as the Met56-flavin interaction in flavodoxin,[138] thioether and oxidized thioether derivatives of 4-thiophenylalanine can also be used to electronically modulate this reactivity.

While 4-thiophenylalanine was never found naturally in proteins, as Johnson & Brautlecht[16] originally hypothesized in 1912, but the sulfur analogue of tyrosine has provided insight and utility in proteins and peptides that is not possible with canonical cysteine or tyrosine. 4-Thiophenylalanine was previously utilized for applications in mechanistic enzymology, in SAR studies in peptides, in developing bacteriostatic small molecules, in protein folding studies.[17, 21, 22, 58, 62] With development of a more practical approach to synthesize 4-thiophenylalanine, we obtained the spectral, reactive, and structural properties of this unique amino acid and related derivatives. With more practical access to peptides containing 4-thiophenylalanine, this amino acid

can be broadly utilized in protein modification, in studying aromatic interactions, in examining electronic effects of post-translational modification, in expressed proteins, and for synthesis of hybrid materials. Furthermore, the diverse derivatives of 4-thiophenylalanine provided unique insights into the nature of aromatic interactions, which is discussed in greater detail in Chapters 2 and 3.

1.4 Experimental

1.4.1 Materials

Fmoc-L-amino acids were purchased from Novabiochem (San Diego, CA), Bachem (San Carlos, CA), or Chem-Impex (Wood Dale, IL). Rink amide MBHA resin, Fmoc-4-iodo-L-phenylalanine, and diisopropylethylamine (DIPEA) were purchased from Chem-Impex. Acetic anhydride (Ac_2O), trifluoroacetic acid (TFA), phenol, thioanisole, triisopropylsilane (TIS), sodium periodate ($NaIO_4$), copper(I) iodide, 1,10-phenanthroline, thiolacetic acid, methyl iodide, 2-aminoethanethiol, 2,2'-dithiodipyridine, lithium hydroxide, thiophenol, N-ethylmaleimide (NEM), nitrosonium tetrafluoroborate, and dithiothreitol (DTT) were purchased from Acros. Ethanedithiol (EDT) was purchased from Pfaltz & Bauer (Waterbury, CT). Methyltrioxorhenium (MTO) was purchased from Strem Chemicals (Newburyport, MA). Piperidine, Hoveyda-Grubbs catalyst (second generation), allyl iodide, and 2-nitrobenzyl bromide were purchased from Aldrich. Tri(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Hampton Research (Aliso Viejo, CA). O-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) was purchased from Senn Chemicals (San Diego, CA). Acetonitrile (MeCN), dimethylformamide (DMF), methylene chloride (CH_2Cl_2), methanol (MeOH), ether,

pyridine, *tert*-amyl alcohol, toluene, acetic acid, sodium chloride, sodium borohydride, magnesium sulfate, sodium hypochlorite, and hydrogen peroxide were purchased from Fisher. Deionized water was purified by a Millipore Synergy 185 water purification system with a Simpak2 cartridge. Solid-phase post-synthetic modification reactions were performed in capped disposable fritted columns (Image Molding), in disposable Eppendorf tubes (1.5 mL), or in glass vials (2 mL). All compounds were used as purchased with no additional purification.

1.4.2 Peptide Synthesis and Characterization

All peptides (0.1 or 0.25 mmol) were synthesized on a Rainin PS3 peptide synthesizer on Rink amide resin via standard Fmoc solid-phase peptide synthesis using HBTU as a coupling reagent. 60 minute coupling reactions were conducted with 4 equivalents of Fmoc- amino acid and HBTU. 3 equivalents of amino acid were used for Fmoc-4-iodo-L-phenylalanine. All peptides contained an acetyl group at the N-terminus and amide groups at the C-terminus.

The resin containing peptides were subjected to cleavage and deprotection reaction conditions for 2-4 hours under standard conditions (90% TFA/5% TIS/5% H_2O for Ac-TXPN, Ac-GPPXPPGY-NH_2, and KKHMCX, where X is 4-iodophenylalanine or 4-thiophenylalanine derivatives; 84% TFA/4% each of H_2O/phenol/thioanisole/ethanedithiol for trp cage peptides). TFA was removed by evaporation under nitrogen. Peptides were precipitated with cold ether, and the resultant mixture was centrifuged. The ether was removed and the precipitate containing the peptide was dried. The peptides were dissolved in water or buffer, and then filtered using a 0.45 μm syringe filter. The peptides were purified, reaction conversion was determined, and peptide purity were determined using reverse phase

HPLC on a Vydac C18 semi-preparative column (250 × 10 mm, 5-10 µm particle, 300 Å pore) or on a Varian Microsorb MV C18 analytical column (250 × 4.6 mm, 3-5 µm particle, 100 Å pore) using a linear gradient of buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA). Peptide purity was verified via reinjection on an analytical HPLC column. Peptides were characterized by ESI-MS (positive ion mode) on an LCQ Advantage (Finnigan) mass spectrometer. All purification conditions and observed masses of synthesized peptides are reported in Tables 1.10 and 1.11.

Concentrations of peptides were determined by UV, where 4-iodophenylalanine $\varepsilon_{280} = 280\ M^{-1}\ cm^{-1}$.[175] Concentrations of peptides containing 4-thiophenylalanine were determined via Ellman's test, or by UV using measured molar extinction coefficients ($\varepsilon_{280,\ pH\ 4.0} = 1440\ M^{-1}\ cm^{-1}$, $\varepsilon_{280,\ pH\ 5.0} = 2230\ M^{-1}\ cm^{-1}$, or $\varepsilon_{280,\ pH\ 8.5} = 15{,}600\ M^{-1}\ cm^{-1}$, *vide infra*).

Table 1.10. Purification and ESI-MS of peptides containing 4-iodophenylalanine and corresponding reaction products

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-T(4-I-Phe)PN-NH_2	60 minutes 0-45% buffer B	41.0	593.2	615.2 $(M+Na)^+$
Ac-T(4-SAc-Phe)PN-NH_2	60 minutes 0-45% buffer B	41.0	593.2	615.2 $(M+Na)^+$
Ac-T(4-SBz-Phe)PN-NH_2	60 minutes 0-45% buffer B	53.5	654.3	677.3 $(M+Na)^+$
Ac-T(4-SPh-Phe)PN-NH_2	60 minutes 0-45% buffer B	52.6	626.3	627.1 $(M+H)^+$
Ac-T(4-S-allyl-Phe)PN-NH_2	60 minutes 0-45% buffer B	44.4	590.3	591.0 $(M+H)^+$
Ac-T(4-SH-Phe)PN-NH_2	60 minutes 0-45% buffer B	30.1	550.2	573.1 $(M+H)^+$

Table 1.10 continued

Ac-GPP(4-I-Phe)PPGY-NH_2	80 minutes 0-55% buffer B	54.6	997.3	998.0 $(M+H)^+$
Ac-GPP(4-SH-Phe)PPGY-NH_2	80 minutes 0-55% buffer B	49.0	903.4	904.1 $(M+H)^+$
Ac-KKHMC(4-I-Phe)-NH_2	60 minutes 0-70% buffer B	36.3	959.3	480.8 $(M+2H)^{2+}$
Ac-KKHMC^{tBu}(4-I-Phe)-NH_2	60 minutes 0-75% buffer B	37.1	1015.4	508.9 $(M+2H)^{2+}$
Ac-KKHMC(4-SH-Phe)-NH_2	60 minutes 0-70% buffer B	31.9	865.4	866.4 $(M+H)^+$
Ac-KKHMC^{tBu}(4-SH-Phe)-NH_2	60 minutes 0-75% buffer B	39.4	921.4	461.9 $(M+2H)^{2+}$
Native trp cage miniprotein Ac-NLYIQWLKDGGPSSGRPPPS-NH_2	80 minutes 0-70% buffer B	51.5	2209.1	1106.2 $(M+2H)^{2+}$
Y3(4-I-Phe) trp cage miniprotein Ac-NL(4-I-Phe)IQWLKDGGPSSGRPPPS-NH_2	80 minutes 0-70% buffer B	67.1	2319.0	1160.9 $(M+2H)^{2+}$
Y3(4-SH-Phe) trp cage miniprotein Ac-NL(4-SH-Phe)IQWLKDGGPSSGRPPPS-NH_2	80 minutes 0-70% buffer B	60.7	2225.1	1113.9 $(M+2H)^{2+}$

Table 1.11. Purification and ESI-MS of modified peptides containing 4-thiophenylalanine

Peptide	**HPLC Conditions**	**Retention time (min)**	**Expected Mass**	**Observed Mass**
Ac-T(4-SMe-Phe)PN-NH_2	60 minutes 0-45% buffer B	34.7	564.2	587.2 $(M+Na)^+$
Ac-T(4-S-allyl-Phe)PN-NH_2	60 minutes 0-45% buffer B	43.0	590.3	613.2 $(M+Na)^+$
Ac-T(4-S(CH_2CH=$CHCH_2OH$)-Phe)PN-NH_2	60 minutes 0-45% buffer B	34.2	620.3	621.0 $(M+H)^+$
Ac-T(4-S-propargyl-Phe)PN-NH_2	60 minutes 0-55% buffer B	27.3	588.2	611.2 $(M+Na)^+$
Ac-T(4-S-propargyl-Phe)PN-NH_2 ‘click’ product with 4-azidoaniline	60 minutes 0-20% buffer B	48.5	722.3	745.4 $(M+Na)^+$
Ac-T(4-S(2-NO_2Bn)-Phe)PN-NH_2	60 minutes 0-45% buffer B	58.9	685.3	708.2 $(M+Na)^+$
Ac-T(4-S(NEM)-Phe)PN-NH_2				
Ac-T(4-S-S-(2-thiopyridyl) disulfide-Phe)PN-NH_2	60 minutes 0-45% buffer B	46.5	659.2	660.1 $(M+H)^+$
Ac-T(4-S-S-glutathione disulfide-Phe)PN-NH_2	60 minutes 0-50% buffer B	26.5	855.3	856.1 $(M+H)^+$

Table 1.11 continued

Ac-T(4-S(O)Me-Phe)PN-NH_2	60 minutes 0-20% buffer B	27.7	580.2	603.1 $(M+Na)^+$
Ac-T(4-SO_2Me-Phe)PN-NH_2	60 minutes 0-15% buffer B	36.3	596.2	619.1 $(M+Na)^+$
Ac-T(4-SO_2(propargyl)-Phe)PN-NH_2	30 minutes 0-10% buffer B	26.1	620.2	643.1 $(M+Na)^+$
Ac-T(4-SO_2(propargyl)-Phe)PN-NH_2 'click' product with 4-azidoaniline	60 minutes 0-15% buffer B	38.2	754.3	777.2 $(M+Na)^+$
Ac-T(4-SO_2(NEM)-Phe)PN-NH_2	10 minutes isocratic buffer A, then 30 minutes 0-40% buffer B	28.8	707.3	708 $(M+H)^+$
Ac-T(4-SO_2^--Phe)PN-NH_2	30 minutes 0-45% buffer B	9.4	581.2	583.2 $(M+2H)^+$
Ac-T(4-SO_3^--Phe)PN-NH_2	20 minutes isocratic buffer A, then 40 minutes 0-45% buffer B	9.2	598.2	599.1 $(M+H)^+$
Ac-T(4-SNO-Phe)PN-NH_2	15 minutes 0-45% buffer B	13.8	579.2	580 $(M+H)^+$

1.4.3 Copper-mediated cross-coupling reaction on peptides in solution phase

To the purified, lyophilized peptide containing 4-iodophenylalanine (1 μmol) was added *tert*-amyl alcohol (200 μL) and DIPEA (33 μL, 0.2 mmol). Copper(I) iodide (1.8 mg, 10 μmol), 1,10-phenanthroline (3.6 mg, 20 μmol), and thiolacetic acid were added (8.6 μL, 0.12 mmol) to the peptide in solution. The resultant mixture was sealed (either under air, or under nitrogen atmosphere, depending on the reaction conditions screened, Table 1.3), and then incubated on a heating block at 90 °C for 11 h. The resultant mixture was allowed to cool to room temperature. Water was added to the mixture (200 μL). The aqueous layer was washed with ether, then filtered on a 0.45 μm syringe filter to generate the peptide products. The resultant reaction products were characterized via direct injection on RP-HPLC. Additional peaks at t_R 46.9 min (Figure 1.64) and t_R 51.7 min (Figure 1.65) indicated undesired acetylation of the peptide during the solution phase reaction.

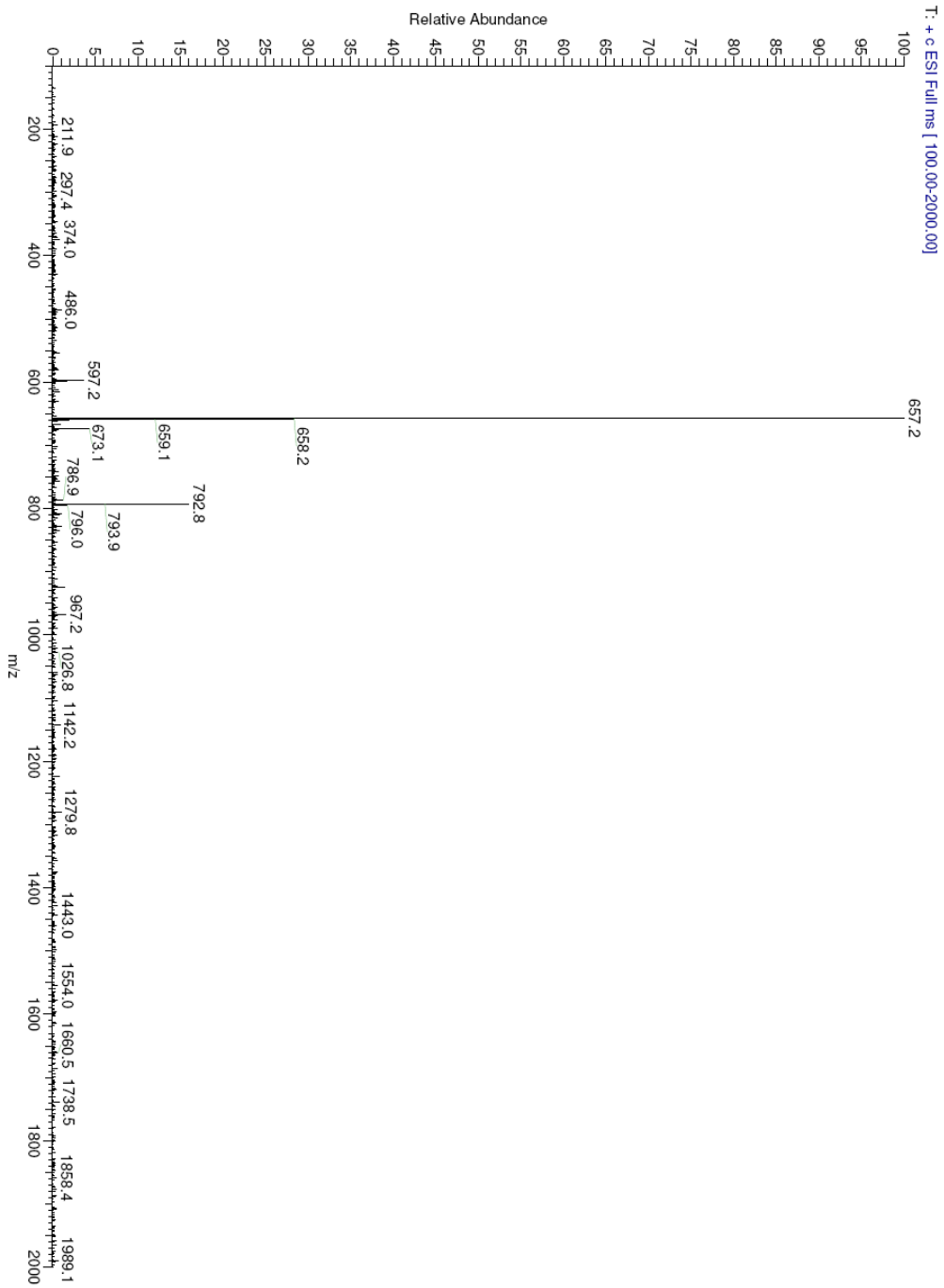


Figure 1.64 Mass spectrum of the product resulting from the cross-coupling reaction on the peptide containing 4-iodophenylalanine (49.6 min)
See Figure 1.12. The observed mass is consistent with the peptide Ac-T(4-thioacetyl-Phe)PN-NH_2 with an additional acetyl goup and a sodium ion (peptide + Ac + Na^+).

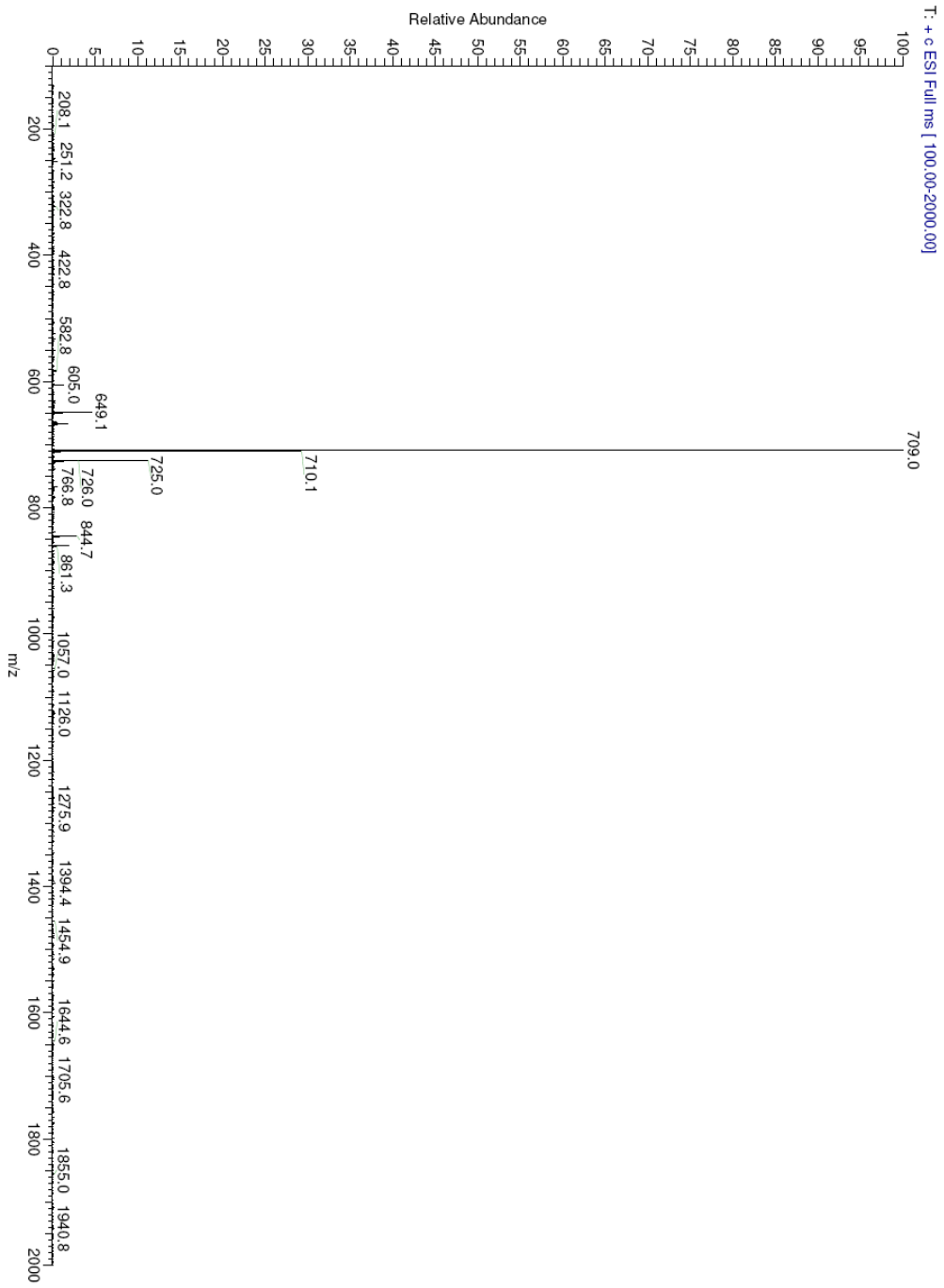


Figure 1.65 Mass spectrum of the product resulting from the cross-coupling reaction on the peptide containing 4-iodophenylalanine (51.7 min)
See Figure 1.12. The observed mass is consistent with the peptide Ac-T(4-I-Phe)PN-NH_2 with an additional acetyl group and a sodium ion (peptide + Ac + Na^+).

1.4.4 General copper-mediated cross-coupling reaction on peptides on solid phase

The resin with the peptide containing 4-iodophenylalanine (10-15 mg, 7-10 μmol) was placed in a glass vial, and toluene (200 μL) and DIPEA (33 μL, 0.2 mmol) were added. Copper(I) iodide (1.8 mg, 10 μmol) and 1,10-phenanthroline (3.6 mg, 20 μmol) were added to the solution containing the resin. Thiophenol (12.3 μL, 0.12 mmol), thiobenzoic acid (14.1 μL, 0.12 mmol), or allyl mercaptan (9.6 μL, 0.12 mmol) was added to the mixture. The vial was sealed and the resultant mixture was stirred in an oil bath set to 110 °C for 16-18 h. For reactions using thiobenzoic acid, the reaction temperature was set to 100 °C. The reaction mixture was allowed to cool and the reagents were removed via filtrations. The resin was washed with DMF (4 mL × 4), CH_2Cl_2 (4 mL × 2), and MeOH (4 mL × 2) and dried with ether. The resin was subjected to cleavage and deprotection reaction conditions using TFA. The TFA was removed via evaporation with nitrogen gas, and the peptide products were precipitated in ether. The resultant mixture was centrifuged, and the ether was removed. The precipitate was dried. The precipitate was dissolved in water or buffer (50 mM phosphate in water, pH 4.0) and characterized directly via HPLC.

1.4.5 General solution phase reduction of peptides containing 4-S(acetyl)-thiophenylalanine to generate peptides containing 4-thiophenylalanine

The resin containing the protected peptide with 4-S(acetyl)-thiophenylalanine was subjected to standard cleavage and deprotection reaction conditions with TFA. TFA was removed by evaporation, and the peptides were precipitated with ether. The resultant mixture was centrifuged, and the ether was removed. The precipitate contained a mixture of peptides with 4-S(acetyl)-thiophenyalalnine and the corresponding disulfides. To obtain the peptide containing 4-thiophenylalanine-

containing, the crude, precipitated peptide products were dissolved in 1 M phosphate buffer (pH 7.2, 50 μL). DTT was added (50 μL of a 100 mM solution in water) and the mixture was incubated at room temperature for 1 h to generate the peptides containing 4-thiophenylalanine.

1.4.6 General alkylation reactions on peptides in solution phase

The purified and lyophilized peptide containing 4-thiophenylalanine (0.3-0.8 μmol) was dissolved in 100 mM phosphate buffer (pH 8.0, 100 μL). The alkylating agent (methyl iodide, 1 μL, 16 μmol; allyl iodide, 1 μL, 11 μmol; a solution of 2-nitrobenzyl bromide, 5 μL, 11 mM solution in MeCN; or a solution of N-ethyl maleimide, 1.5 μL, 50 mM solution in MeCN) was added and the mixture was subjected to incubation at room temperature for 1 h to produce the alkylated peptide product. Excess 2-aminoethanethiol (approximately 5 mg) was then added to quench residual alkyl halide.

1.4.7 Competitive alkylation reactions on peptides in solution phase using N-ethyl maleimide

Stock solutions of peptides containing either 4-thiophenylalanine or cysteine were dissolved in water. Concentrations were determined via Ellman's test. The appropriate volumes of solutions of either the peptide containing 4-thiophenylalanine and/or the peptide containing cysteine (13 nmol) were diluted into a solution of TCEP (50 μM) in phosphate buffer (pH 6.5, 130 μL, 50 mM). A solution of N-ethylmaleimide (6.5 μL of a 1 mM solution in water) was added to the peptide. The resultant solution was subjected to incubation at room temperature for 10 minutes to generate the peptide products. The resultant solutions were characterized directly via

HPLC, or were quenched with 1 mL of buffer A (98% water, 2% MeCN, 0.06% TFA).

1.4.8 General alkylation reactions on peptides on solid phase

The resin containing the protected peptide with 4-S(acetyl)-thiophenylalanine (5-10 mg, 3-7 μmol) was placed in a capped disposable fritted column and THF (3 mL) was added. LiOH (7.2 mg, 300 μmol), DTT (23.1 mg, 150 μmol), ethanedithiol (25 μL, 300 μmol), and thiophenol (31 μL, 300 μmol) were added. The resultant mixture was agitated via rotary shaker at room temperature for 3 h. The solution was removed from resin, and then the resin was washed with MeOH (3 mL × 3). MeOH (3 mL), $NaBH_4$ (28.4 mg, 750 μmol), triethylamine (200 μL, 1.43 mmol), and allyl iodide (69 μL, 750 μmol) were added to the resin, and the resultant mixutre was stirred for 3 h at room temperature. The solution was removed from the resin, and then the resin was washed with MeOH (3 mL × 3) and dried with ether. The resin was subjected to cleavage and deprotection reaction conditions using TFA.

1.4.9 Bioconjugation reaction in solution phase on peptides containing 4-S-allyl-thiophenylalanine

The purified and lyophilized peptide containing 4-(S-allyl)-thiophenylalanine (0.3-0.8 μmol) was dissolved in phosphate buffer (50 mM, pH 8.0, 200 μL). For reaction solutions with $MgCl_2$, $MgCl_2 \cdot 6H_2O$ (10.5 mg, 52 μmol) was added to the peptide dissolved in phosphate buffer. Hoveyda-Grubbs II catalyst (2.6 mg, 4.2 μmol) was dissolved in *tert*-butanol (338 μL), and the resultant solution was heated and mixed by vortexing until fully dissolved. The Hoveyda-Grubbs catalyst solution (86 μL) was added to the peptide. The resultant solution was mixed via vortex, and then sparged with nitrogen gas. Allyl alcohol (3.6 μL, 53 μmol) was added, and the mixture

was then incubated for 2 h, either at room temperature or at 37 °C, to generate the peptide containing 4-S-1-butenol-thiophenylalanine, as an inseparable mixture of E and Z alkene isomers, as observed by NMR.

1.4.10 Oxidation reaction in solution phase to produce peptides containing 4-S(O)methyl-thiophenylalanine

To the purified peptide containing 4-S(methyl)-thiophenylalanine (0.3-0.8 μmol), a solution of MeCN:H_2O (3:5 v/v, 80 μL) was added. $NaIO_4$ was then added (2.4 μL of a 57 mM solution in water). The resultant mixture was stirred at room temperature for 24 h to generate the peptide containing 4-S(O)methyl-thiophenylalanine as an inseparable mixture of sulfoxide diastereomers. These diastereomers were inseparable by HPLC, but were observable via NMR.

1.4.11 General oxidation reaction to produce peptides containing sulfone derivatives of 4-thiophenylalanine in solution phase

To the purified peptide containing either 4-S(methyl)-thiophenylalanine (0.3-0.8 μmol), 4-S(N-ethylsuccinimide)-thiophenylalanine (0.3-0.8 μmol), or 4-S-propargyl-thiophenylalanine (0.3-0.8 μmol), a mixture of formic acid (90% in H_2O, 1 mL) was added. H_2O_2 was then added (30% in H_2O, 200 μL) and the mixture was stirred at room temperature for 24 h to produce the peptide containing the sulfone derivative of 4-thiophenylalanine.

1.4.12 Oxidation reaction in solution phase to produce peptides containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine disulfide

The purified peptide containing 4-thiophenylalanine (0.3-0.8 μmol) was dissolved in phosphate buffer (pH 8.0, 100 mM, 100 μL). A solution of 2,2′-dithiodipyridine (100 mM in MeCN, 20 μL) was added to the solution, and the

mixture was incubated at room temperature for 2 h to produce the peptide containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine.

1.4.13 Disulfide exchange reaction in solution phase to produce peptides containing 4-S-S-glutathione disulfide-thiophenylalanine disulfide

The purified peptide containing 4-S-S-(2-thiopyridyl) disulfide-thiophenylalanine (0.3-0.8 μmol) was dissolved in phosphate buffer (pH 8.0, 100 mM, 200 μL). A solution of glutathione (reduced, 100 mM in water, 20 μL) was added to the peptide. The mixture was incubated at room temperature for 2 h to generate the peptide containing 4-S-S-glutathione disulfide-thiophenylalanine disulfide.

1.4.14 Nonspecific oxidation reaction to produce multiple oxidized species of peptides containing 4-thiophenylalanine in solution phase

The purified peptide containing 4-thiophenylalanine (11 nmol) was dissolved in phosphate buffer (400 μL, pH 6.0, 50 mM) and the resultant solution was chilled on ice. A chilled solution of sodium hypochlorite in phosphate buffer (400 μL, 22 μM NaOCl, 50 mM phosphate, pH 6.0) was added to the peptide, and the solution was mixed via pipette for 5 seconds. The solution was then incubated on ice for 5 minutes to generate peptides containing a mixture of disulfides, the sulfonic acid, and the sulfinic acid derivatives of 4-thiophenylalanine.

1.4.15 Hydrolysis of a sulfone to selectively synthesize the peptide containing 4-SO_2^--thiophenylalanine

The purified peptide containing 4-SO_2(*N*-ethyl succinimide)-thiophenylalanine (0.3-0.8 μmol) was dissolved in a solution of 10 mM LiOH (100 μL). The resultant solution was incubated at room temperature for 1 h to generate the peptide containing the sulfonic acid derivative of 4-thiophenylalanine.

1.4.16 Oxidation reaction in solution phase to produce peptides containing 4-SO_3^--thiophenylalanine

Peptide containing 4-thiophenylalanine disulfide was obtained either from cross-coupled crude products (described previously), or by dissolving the peptide containing purified 4-thiophenylalanine (0.3-0.8 μmol) in phosphate buffer (pH 8.0, 100 mM, 100 μL) with iodine (0.5 mg) and incubating at room temperature for 1 h. To a solution of H_2O_2 (30% in water, 150 μL) was added a solution of methyltrioxorhenium (MTO, 100 μL of a 6 mM solution in MeCN). The mixture was vortexed and incubated at ambient temperature for 1 min. A solution of purified peptide Ac-T(4-S(disulfide)-Phe)PN-NH_2 (0.1-0.4 μmol) in MeCN (250 μL) was added to the MTO mixture. The resultant solution was then incubated at ambient temperature for 30 min to produce peptide containing 4-SO_3^--thiophenylalanine. The peroxide was quenched with excess $NaHSO_4$ (approximately 2 mg) and incubated at room temperature for 30 min. The reaction volume was reduced by evaporation of MeCN with nitrogen gas to reduce the volume to approximately 100 μL, and then 100 μL of either phosphate buffer (100 mM, pH 4.0) or water was added.

1.4.17 S-nitrosylation reaction in solution phase to produce peptides containing 4-SNO-thiophenylalanine

The purified peptide Ac-T(4-SH-Phe)PN-NH_2 (10 nmol) was dissolved in dry acetonitrile (8 μL) and the resultant solution was cooled to –78 °C in a bath of acetone and dry ice. Nitrosonium tetrafluoroborate was dissolved in dry acetonitrile (100 mM solution), and was cooled to –78 °C in a bath of acetone and dry ice. The solution of nitrosonium tetrafluoroborate in dry acetonitrile (2 μL, 100 mM) was added to the solution containing the peptide, and the resultant solution was mixed via pipette for 10 seconds. The solution was incubated at –78 °C in the bath of acetone and dry ice for 5

minutes to produce a mixture of the peptide containing the S-nitrosated derivative of the 4-thiophenylalanine and corresponding disulfides. The resultant reaction mixture was characterized via direct injection onto RP-HPLC.

1.4.18 General azide-alkyne 1,3-cycloaddition reaction in solution phase on peptides containing 4-S-propargyl-thiophenylalanine

The purified peptide containing either 4-S-propargyl-thiophenylalanine or 4-SO_2(propargyl)-thiophenylalanine (approximately 17 nmol; 70 μM final concentration) was dissolved in phosphate buffer (250 μL of 50 mM phosphate, pH 7.5). To the peptide-containing solution, copper(II) sulfate (25 μL of a 10 mM solution in water), sodium ascorbate (2.5 mg, 50 mM final concentration), and 1,10-phenanthroline (5 μL of a 10 mM solution in DMSO) were added. To this solution was subsequently added 4-azidoaniline·HCl (4.5 mg; 100 mM final concentration), and the resultant solution was allowed to incubate at 37 °C for 5 minutes to produce the conjugated peptide. The crude solution was diluted with deionized water (300 μL), and diethyl ether (1 mL) was added. After thorough mixing, the ether layer was removed via pipet, and CH_2Cl_2 (1 mL) was added. After thorough mixing, the aqueous layer was removed and filtered through a nylon filter (0.45 μm).

1.4.19 NMR Spectroscopy

Peptides were dissolved in buffer containing 5 mM phosphate (pH 4.0 or as indicated), 25 mM NaCl, 90% H_2O/10% D_2O, and 100 μM TSP. 100 μM TCEP was added for peptides containing thiophenylalanine to prevent disulfide formation during NMR experiments. Peptide concentrations were 10 μM-200 μM. NMR spectra were collected at 298 K on a Brüker AVN 600 MHz NMR spectrometer equipped with a triple resonance cryoprobe or a TXI probe. Spectra were internally referenced with

TSP. 1-D spectra were collected with a Brüker w5 watergate pulse sequence and a relaxation delay of 1.5-2 s. 2-D spectra were collected with a watergate TOCSY pulse sequence.

Well-resolved peaks in the NMR spectra were integrated after baseline normalization. $K_{trans/cis}$ was calculated based on the average integrated ratios of 2-3 pairs of peaks. In general, backbone amide peaks and/or Thr methyl peaks were used for integration.

1.4.20 UV-Vis Spectroscopy and Measurement of pK_a in Ac-T(4-SH-Phe)PN-NH_2

The purified peptide Ac-T(4-SH-Phe)PN-NH_2 was dissolved in buffer containing 50 mM phosphate and 100 μM DTT. Peptide concentrations were 45-65 μM. Absorbance spectra were collected on a Perkin-Elmer Lambda 25 UV-Vis spectrometer in a 1 cm cell. Absorbance spectra were collected from 310 nm to 220 nm with a slit width of 1 nm. After each UV measurment, the pH of each sample was verified with a pH electrode (Mettler Toledo). After measurements, the absence of disulfide formation was confirmed via HPLC. Data were the average of at least three independent trials. Error bars are shown and indicate standard error.

The pK_a of 4-thiophenylalanine was measured via the pH dependence of the molar extinction coefficient at the λ_{max} for the neutral and deprotonated forms (249 nm and 276 nm, respectively). The data were fit to Equation 1,[176, 177] using a non-linear least squares fitting algorithm (Kaleidagraph, version 4.1.1, Synergy Software), where ε = observed molar extinction coefficient, pH = pH of the sample, pK_a = acidity constant.

$$\varepsilon_{249} = \frac{\varepsilon_{249,thiolate} 10^{(pH - pKa)} + \varepsilon_{249,thiol}}{1 + 10^{(pH - pKa)}} \qquad (1)$$

1.4.21 Fluorescence Emission and Excitation Spectra for Ac-T(4-SH-Phe)PN-NH_2

Purified Ac-T(4-SH-Phe)PN-NH_2 was dissolved in buffer containing 50 mM phosphate (pH 4.0 or 8.5). Peptide concentrations were 100 μM with 125 μM DTT. Fluorescence spectra were collected in a 1 cm quartz cell (Starna) on a Photon Technology International fluorescence spectrometer model QM-3/2003 with a CW source and a Hamamatsu R928 PMT. All slit widths were 3 nm. Excitation scans were from 250 nm to 450 nm, with emission detected at 400 nm. Emission scans were collected from 300 nm to 500 nm with excitation at 295 nm. Data were the average of at least three independent trials. Data were background-corrected but were not smoothed. Error bars are shown and indicate standard error.

1.4.22 Circular Dichroism

CD experiments were conducted using a solution of peptides (20-65 μM) in 15 mM sodium phosphate buffer (pH 4.0, 7.0, or 8.5). TCEP (0.56 mM) was added to solutions of trp cage peptides containing 4-thiophenylalanine to prevent disulfide formation. After the CD experiments, the absence of disulfides was confirmed via HPLC. CD spectra were recorded in a 1 mm pathlength cell on a Jasco model J-810 spectropolarimeter. Thermal denaturation data were collected at 222 nm using a bandwidth of 4 nm. The temperature was increased in 1 ˚C intervals from 2 ˚C to 90 ˚C. Samples were allowed to equilibrate for 1 minute at each temperature prior to data collection. Data were collected with a 4 s response time. T_m data were smoothed using a 5-data-point smoothing window (KaleidaGraph 4.1.1); the smoothed data are shown in Figure 1.28.

T_m temperatures were determined using a slope for the 100% unfolded baseline ($d[\theta]_{222}/dT$) of –8.83 deg cm^2 $dmol^{-1}$/°C; and a slope for the 100% folded baseline ($d[\theta]_{222}/dT$) of +63.5 deg cm^2 $dmol^{-1}$/°C, based on prior work on the trp cage miniprotein.[96]

1.4.23 Cambridge Structural Database (CSD)

The Cambridge Structural Database (version 5.36 + 3 updates, released in May 2015) was searched for entries containing aryl ethers, aryl thioethers, aryl sulfoxides, or aryl sulfones (Ph–O–C^{sp3}, Ph–S–C^{sp3}, Ph–S(O)–C^{sp3}, or Ph–SO_2–C^{sp2}, respectively). The release of the database used in the current study contained over 710,000 molecules. Searches were set up and carried out using ConQuest (version 1.17 Cambridge Crystallographic Data Centre, 2014).

The search parameters for aryl ethers and aryl thioethers for each oxidation state of sulfur were defined as shown in Figure 1.66. The coordination number of sulfur was defined as indicated, and the coordination number of C (C1) was defined as 4 in all queries. Sulfoxides were defined using the two different structures indicated, and the results were combined. Only error-free, non-disordered structures where R < 0.10 were included, and powder pattern structures were excluded.

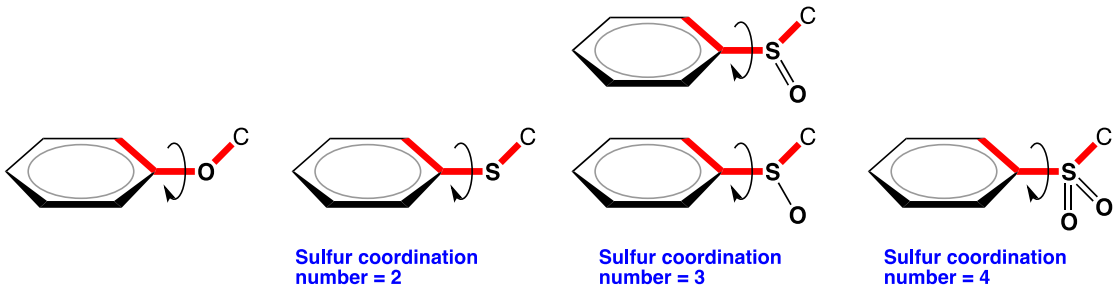


Figure 1.66 Defined structural parameters for the search query in the Cambridge Structural Database using Conquest
The torsion angle indicated in red was measured for all structures.

Torsion angles were measured for each structure obtained in the defined query (C3–C2–S–C1, as indicated in red in Figure 1.66, where C3 and C2 are aromatic carbons and C1 is the aliphatic carbon). The absolute value of the torsion angle was measured, and then a logical function was applied, where if the resultant torsion angle was greater than 90°, then 180° was subtracted from the torsion angle and the absolute value taken. Use of this logic function converts all torsion angles to a range of 0° to 90°, where the S–C1 or O–C1 bond is measured as between being between planar or perpendicular to the aromatic ring (Figure 1.66). In addition, bond lengths (S–C1 or O–C1) and bond angles (C2–S–C1 or C2–O–C1) were accumulated for each of these queries.

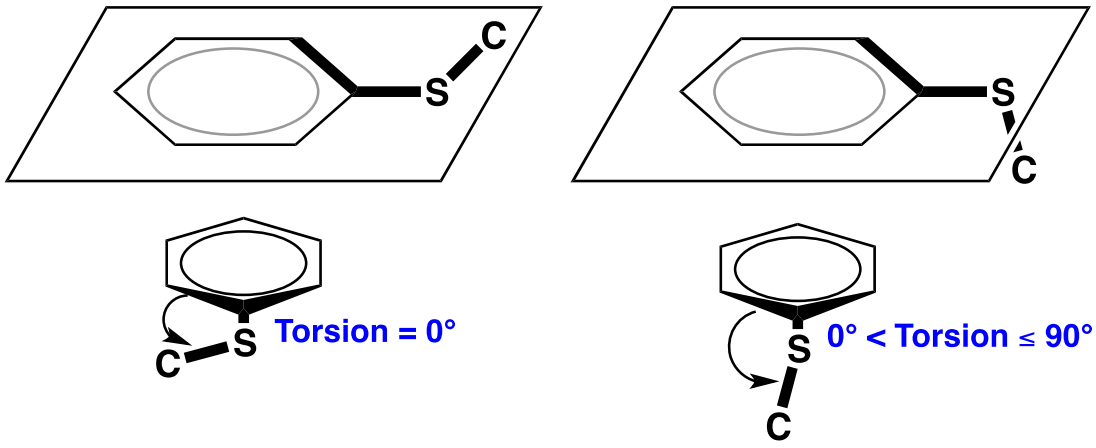


Figure 1.67 Defined torsion angle measurements from CSD
The absolute values of torsion angles obtained through ConQuest resulted in angles ranging from 180° to 0°. These torsion angles were converted using a logic function to range from 0°–90° to represent O–C or S–C bonds that were between the limiting cases of co-planar with the aromatic ring (left, torsion angle near 0°) or orthogonal to the aromatic ring (right, torsion angle near 90°).

Chapter 2

INSIGHTS INTO S–H/π AROMATIC INTERACTIONS: STUDIES ON BOC-4-THIOL-L-PHENYLALANINE-*TERT*-BUTYL ESTER VIA IR SPECTROSCOPY, X-RAY CRYSTALLOGRAPHY, AND *AB INITIO* CALCULATIONS

2.1 Introduction

As described by George Whitesides, "many biological molecules and aggregates derive much of their unique structure and function from noncovalent interactions: that is, from hydrophobic and ionic interactions and hydrogen bonds."[178] Whitesides was describing the nature of non-covalent interactions and the significant influence that numerous weak interactions can have on macromolecular structure. While protein structures are primarily driven to fold into active conformations through non-covalent interactions, including hydrogen bonding and formation of salt-bridges, there are more subtle and ubiquitous interactions that also play an important role.[19, 179]

Our understanding of non-covalent interactions, particularly those involving aromatic groups, has undergone a renaissance over the past few decades. The paradigm for protein folding and non-covalent interactions that dictate conformation of organic molecules has been generally described by hydrogen bonding (a largely electrostatic interaction) or the hydrophobic effect (dispersive interactions).[180] Recently, interactions such as cation/π,[162, 181-184] $n \rightarrow \pi^*$,[185-187] and halogen-bonding (a type of $n \rightarrow \sigma^*$ interaction)[188, 189] have been added to the repertoire of non-covalent interactions that influence protein and molecular structure and surpramolecular assembly. These and related non-covalent interactions are described either through electrostatic effects or through favorable orbital interactions, where filled HOMOs

donate electron density to unfilled LUMOs. Individually, these interactions can contribute approximately 0.25-5 kcal mol^{-1} or more to the stability of biological and organic structures.[181, 189, 190] With substantial loss of degrees of freedom upon protein folding, the entropic cost of forming the folded state must be compensated by the enthalpic stability provided by additive effects from multiple non-covalent interactions.

Although hydrogen bonds have been well characterized and described for more than 80 years, a re-evaluation of the nature of hydrogen bonds is required, in light of more recent identification of numerous non-covalent interactions that are driven primarily by orbital interactions (some do not involve any hydrogen atoms!). Hydrogen bonds have been described as an interplay of electrostatic and dispersive forces: classical hydrogen bonds have a greater electrostatic component, and non-conventional hydrogen bonds (often considered weaker) are driven by dispersive forces.[190, 191] However, these non-conventional interactions have been observed to have geometric preferences in donor-acceptor orientation, which is not completely described by a large dispersive component.[135, 189, 192] For example, halogens have been recognized to interact with electron-rich atoms or functional groups, which is not intuitive on first approximation.[189] The nature of the interaction is donation of electrons from the electron-rich functional group to the σ^* orbital of the halogen bond, which increases in size with larger halogen atoms (Figure 2.1).[189] In the context of conventional hydrogen bonds, the orbital component of stability may have been previously overlooked because the electrostatic interaction is predominant, and weaker hydrogen bonds were regarded as "non-directional" driven by dispersion. However, in hydrogen bond-type interactions that do not have significant electrostatic stability,

such as those involving S–H, C–H, or aromatic rings, the favorable orbital overlap interactions are significant and implicate concerted geometry and stability, which dispersive interactions do not.[193] With enhanced understanding of the fundamental driving energies of these more subtle, but directional, non-conventional interactions, these non-covalent interactions can be harnessed for rationally designed molecules, peptides, protein inhibitors, and supramolecular materials.

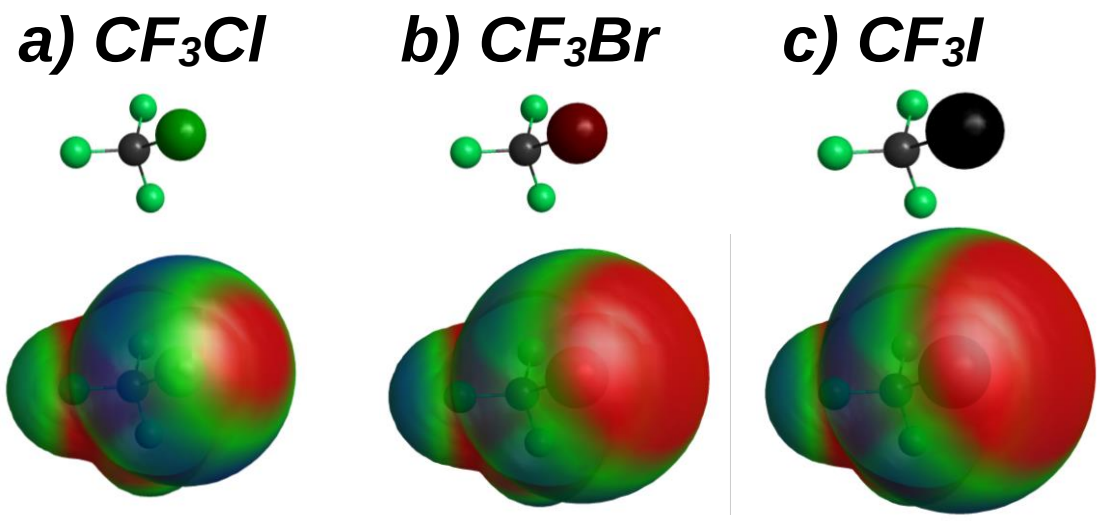


Figure 2.1 The halogen σ-hole: Non-covalent interactions via σ* interactions
The σ^* reveals a slightly electropositive surface potential on the end of the C–X bond, and the area of the "σ-hole" increases with larger halogens. The "σ-hole" allows the halogen atom to interact with electron-rich atoms or functional groups.[7, 194] With increased understanding of these non-covalent interactions, the definition of hydrogen bonding "donor-acceptor" interactions should be re-evaluated. Calculations for molecular orbitals were conducted using GAMESS (RHF 3-21G),[195] and visualized with MacMolPlt,[196] contoured at 0.001, colorized +0.01 (red) to –0.01 (blue).[196] (a) Calculated unoccupied orbital associated with the C–Cl bond in CF_3Cl ($\sigma^*_{C–Cl}$); (b) calculated unoccupied orbital associated with the C–Br bond in CF_3Br ($\sigma^*_{C–Br}$); (c) calculated unoccupied orbital associated with the C–I bond in CF_3I ($\sigma^*_{C–I}$).

2.1.1 The nature of hydrogen bonding and detection in biological molecules

The importance of the hydrogen bond in biology cannot be understated. Hydrogen bonds as they are known today were discussed in some early cases at the beginning of the 1900s,[197-201] but Linus Pauling's chapter on "hydrogen bonds" in his 1939 edition of *The Nature of the Chemical Bond*,[180, 202] is regarded as the juncture that brought hydrogen bonding into mainstream chemical knowledge.[203] In 1951, Pauling proposed the structures of an α-helix and a pleated β-sheet, based on backbone hydrogen bonding, and suggested that these repeating structures could be

integral in globular proteins such as hemoglobin and myoglobin.[204, 205] The formation of α-helices, β-sheets, and other secondary structural elements in proteins relies on intramolecular hydrogen bonding, and is known to be a primary driving force for protein folding.[204-206] In addition to main-chain hydrogen bonding, side chain groups in proteins also undergo extensive hydrogen bonding interactions, with other side chains, peptide backbone structures, and substrates.[206, 207] These local interactions play a major role in protein super secondary structure and tertiary structure.[206, 207] Beyond protein folding, hydrogen bonding is also implicated in substrate recognition, in folding of RNA aptamers, in stability of the DNA double helix, substrate and inhibitor binding to proteins, therapeutic efficacy of drug molecules, in crystal engineering, and in designed catalysts.

A "hydrogen bond," defined fundamentally, is a favorable interaction between a hydrogen-bearing donor (D) and an acceptor group (A). The hydrogen atom can be viewed as "shared" between the donor and acceptor groups, and lengthening of the D–H bond provides evidence of a hydrogen bond.[203] Idealized, conventional hydrogen bonding is described as a linear, monodentate formation D–H··A, but the geometry, distances, strength, and nature of the donor and acceptor groups can vary substantially (Figure 2.2).[203] The acceptor group (A) may be an electronegative atom, but, more generally, it can also be an electron-rich atom, bond, or system, such as a π-bond or a lone pair of electrons.[203] The strength of a hydrogen bond varies significantly, broadly between –0.5 and –45 kcal mol^{-1}, depending on the orientation and nature of the donor and acceptor groups, and on external competing interactions (gas-phase hydrogen bonds are much stronger than hydrogen bonds in solution).[190, 203] However, the vast majority of hydrogen bonds discussed here, those observed in solution in small

molecules and proteins involving main group elements, are less than –4 kcal mol^{-1}.[203] The nature of a hydrogen bond is largely attributed to an interplay between electrostatic interactions and dispersive forces. Stronger hydrogen bonds are suggested to have a greater electrostatic driving force with minor dispersive interactions. In contrast, weaker hydrogen bonds are attributed to decreased electrostatic contributions, and are predominantly driven by dispersion.[190] The distinguishing feature between hydrogen bonds and van der Waals interactions, as it has been described, is that hydrogen bonds are "directional" (where donors and acceptors align with specific geometry to maximize the interaction), while dispersive van der Waals interactions are "non-directional," and the geometry between the interacting groups is less defined.[190, 191]

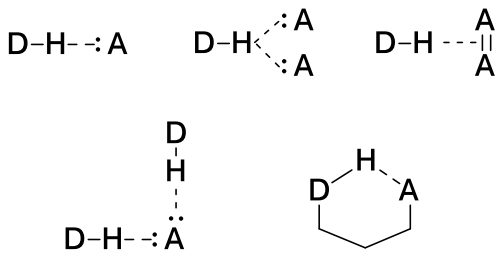


Figure 2.2 Examples of conventional hydrogen bond interactions
Idealized hydrogen bonds are linear with respect to the donor D–H··A acceptor groups, but the orientations and distances can vary depending on the nature of the donor and acceptor groups, and the competing interactions.[203]

There are three primary experimental means of determining if a hydrogen bond interaction is present within a molecule: x-ray crystallography or neutron diffraction, IR spectroscopy, and NMR spectroscopy. These experimental observations can be corroborated with *ab initio* calculations on the compound of interest, or compared to database analysis of published crystallographic data (namely the Protein Data Bank,

and the Cambridge Structural Database).[190, 203] Obtaining a solved crystal structure allows for direct measurement of bond lengths, interaction distances, and angles, and much of the discussion of hydrogen bonds and non-covalent interactions rely on observations in crystal structures. Indeed, Pauling referred to the hydrogen bonding interactions observed in crystal structures of amino acids in formulating his description of the structure of α-helices, and this allowed for accurate approximations of the dimensions of 3_{10}-helices[199, 204] However, a drawback associated with x-ray crystallographic data is that the donor D–H bond can often appear shorter in the solved crystal structure than it truly is, due to the fact that the electron density is displaced towards the donor atom rather than the hydrogen atom.[190, 203] This apparent displacement of hydrogen atom location is due to the fact that x-ray crystallography determines electron density, of which hydrogen atoms intrinsically have less than all other atoms. In order to exclude this potential error in hydrogen atom location, hydrogen bonds can also be identified by measuring only the heavy atoms, the distance D··A,[190, 203] but this measurement provides little information on the directionality and geometry of the hydrogen bond. For these reasons, neutron diffraction is currently the only means to determine accurate location of a hydrogen atom, and the direction of the D–H bond, within a molecular stucture.[190, 203]

Early observations of hydrogen bonds, prior to advances in x-ray crystallography, relied on IR spectroscopy for characterization.[200, 201] IR spectroscopy does not require a diffractable crystal, can be conducted on any phase of material (solid, liquid or gas), and is more definitive than x-ray crystallography in determining if a donor D–H bond participates in a hydrogen bond or not.[190] IR spectroscopy measures bond vibrations, and is sensitive enough to detect subtle changes in the

vibrations pertaining to the donor D–H bond.[190] If a donor D–H bond participates in a hydrogen bond, the D–H stretching frequency shifts to lower wavenumbers ("red shift") and generally becomes more intense when compared to the un-bound D–H stretching frequency.[208] Greater red shifts in the D–H stretching frequency can indicate a shorter H··A distance, suggesting a stronger interaction.[208] Frequency shifts in bonds pertaining to the acceptor group are also sometimes observed, such as in carbonyl hydrogen-bond acceptors.[208] In proteins, secondary structure can be approximated using amide bond vibrations, at 1650-1620 cm^{-1} and 1550-1520 cm^{-1}, where α-helices and β-sheets have distinct vibration modes.[209, 210] These vibrational signatures arise from the differences in hydrogen bonding patterns between these two different secondary structures.[209] However, these amide stretching vibrations can be misleading, given the complexity of this region in the IR spectrum for proteins, and NMR spectroscopy is generally more reliable for identifying hydrogen bonds and assigning secondary structure.[209]

^{1}H NMR spectroscopy can be more sensitive than IR methods for identifying weak hydrogen bonds, and the observed hydrogen bond can be assigned to a specific residue in proteins.[203] The chemical shift of a proton is directly related to the shielding effects induced by a proximal electronic field, such as a π-bond, halogen atom, or aromatic ring. ^{1}H NMR is especially useful for weaker hydrogen bonds, such as C–H or S–H bond donors, where the interactions are very weak and can be difficult to discern via IR methods.[203] In 1950, Huggins *et al.* observed changes in the chemical shift of the chloroform proton when mixed with a cosolvent, either acetone or triethylamine, and calculated the strength of this interaction by the observed chemical shift change (Δδ).[211, 212] Similar changes in the proton chemical shift were observed in

phenol derivatives with acetic acid.[211, 212] These observations in the ^{1}H NMR spectrum were consistent with expected shielding effects due to a hydrogen bond interaction, corroborated with known IR data at the time, and provided further general support in the theory of hydrogen bonding interactions.[203, 211, 212] Chemical shifts also exhibit different degrees of temperature dependence between solvent exposed protons and protons that are engaged in an hydrogen bonding interaction.[213] In addition to proton chemical shift changes and temperature dependence, deuterium exchange experiments can also be used to identify regions in proteins that participate in hydrogen bonding interactions.[203] Protons engaged in a hydrogen bond, or protons that are not solvent exposed, will exchange with deuterated solvent more slowly, and can be identified directly via ^{1}H NMR. Tamaoki *et al*. identified α-helical residues in endothelin-1 using this method.[214] Both deuterium exchange and temperature-dependence of proton chemical shifts were used to identify hydrogen bonded residues for structure determination of the cyclic peptide varv F, and these data corroborated with the x-ray crystallography data that was obtained from this same peptide.[215] Scalar coupling between ^{15}N and a carbonyl carbon (across a protein hydrogen bond N–H··O=C) has also been used to identify hydrogen bonds by NMR without relying on proton signals (which there are many in proteins).[216]

Discerning a "hydrogen bond" in a structure can be somewhat subjective. Use of the terms "weak" and "strong," or "distant" and "close" contacts, do not have clear and universal definitions. Some of the "weaker" hydrogen bonds are less than 4 kcal mol^{-1}, and have non-linear D–H··A geometry, but these "weak" donor-acceptor interactions can have significant influence on secondary structures in proteins. Depending on the energetic strength of the interaction, some may argue that the

weakest interactions are not hydrogen bonds at all, given the "lack of directionality," and are more similar to dispersive van der Waals interactions rather than a true donor-acceptor hydrogen bond.[190, 191, 203]

2.1.2 Aromatic π bonds as hydrogen bond acceptors

One of the ubiquitous hydrogen bond interactions in proteins and small molecules involves π bond acceptors, where the electron-rich π orbital in double or triple bonds acts as the acceptor for a donor D–H (Figure 2.3).[190] One of the first reports on hydrogen bonding interactions involving aromatic π bonds was reported by Yoshida and Osawa,[217] where phenolic O–H stretching frequencies were observed to shift in the presence of aromatic and heteroaromatic groups. Interest in aromatic π interactions have increased in recent decades, as improvements in x-ray crystallography allowed for more complete identification of aromatic interactions, both in small molecules and in proteins, and as the roles of aromatic rings in non-covalent interactions have become more apparent.[18, 203, 218] In addition to small molecule compounds, the aromatic amino acids tyrosine, phenylalanine, and tryptophan have been found to impact peptide and protein structure and substrate recognition via aromatic hydrogen bonding interactions.[18, 218]

Similar to hydrogen bonding interactions, cations can interact with the electron-rich faces of aromatic rings (Figure 2.3a). The cation/π interaction initially drew interest when it was discovered that K^+ ions engaged in tight interactions with benzene in the gas phase.[219] A charged ion was not expected to interact with a hydrophobic molecule such as benzene. The stability of this K^+/benzene complex was attributed to the quadrupole moment of the benzene molecule, which interacts favorably with a cation.[219] Kumpf & Dougherty,[220] in studying the selectivity of

potassium ion channels, performed calculations for complexes of benzene with various cations, and found that only K^+ interacted favorably as a 1:2 complex with benzene. Based on these calculations, a potential mechanism was proposed for how potassium channels are selective for specific ions, where the aromatic amino acids that line the potassium ion channel are selective for K^+ via cation/π interactions.[220] Although dispersive forces are known to play a role in these interactions,[219] the electrostatic component is a much more dominant driving force in cation/π interactions.[220, 221] With a greater electrostatic component to cation/π interactions, cations can be expected to align over the aromatic centroid where the greatest negative electrostatic potential is localized (Figure 2.3a).

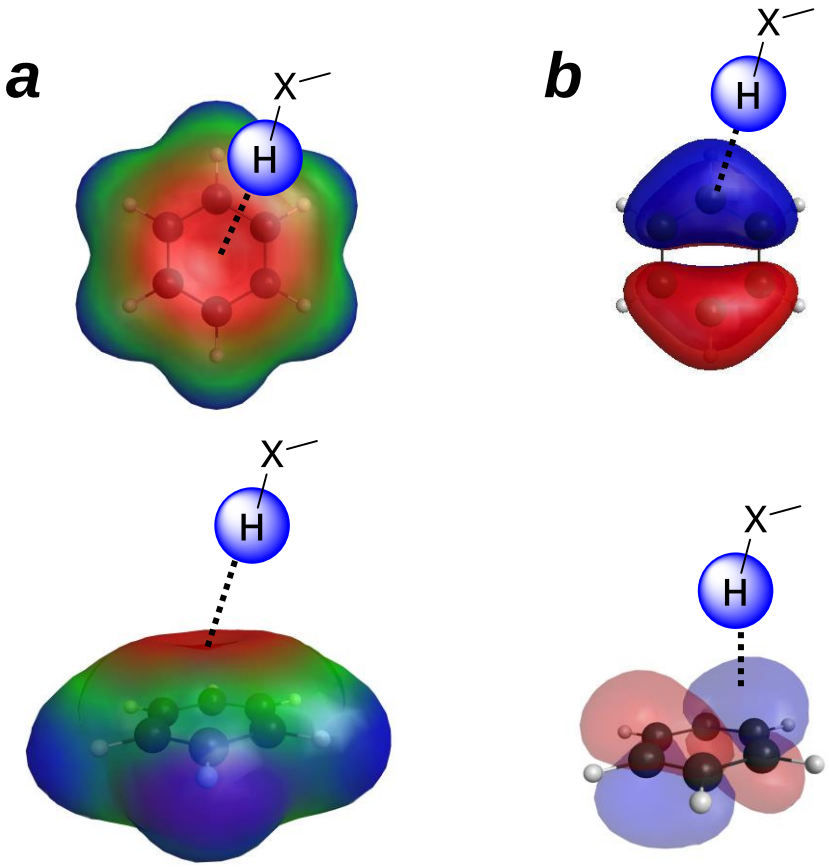


Figure 2.3 Hydrogen bond interactions with aromatic rings
Electron-rich groups can act as hydrogen bond acceptors, including aromatic rings. The electrostatic potential (a) shows an area of negative electrostatic potential (red, –0.025) where hydrogen atoms can interact. The C-H bonds around the edges of the aromatic ring have positive electrostatic potential (blue +0.025), which can act as donors in hydrogen bonds. Hydrogen bonds with aromatic rings are also described as an interaction between the hydrogen atoms and the highest occupied molecular orbital (b). Calculations were conducted using GAMESS (RHF 6311-G),[195] and visualized with MacMolPlt,[196] contoured at 0.002 in the electrostatic potential map, and 0.025 for the molecular orbitals.

As a π-acceptor, conjugated aromatic rings can accept any traditional hydrogen-bearing donor group, such as O–H, N–H, S–H, or C–H. The face of the aromatic ring acts as an acceptor, but the C–H bonds at the edge of the aromatic rings

are polarized and relatively electron-deficient, and can participate as hydrogen bond donors (Figure 2.3a).[218, 222] The dual-modality of aromatic rings allows for aromatic rings to interact with each other through edge-to-face or slipped stacking interactions (Figure 2.3a).[18, 161, 218] In 1990, Sanders & Hunter[223] described the nature of aromatic-aromatic interactions, with observations of slipped-stacking in porphyrins and other aromatic host-guest model systems, as a σ-π attraction, stabilized by both electrostatic and van der Waals components. In fact, structural motifs in proteins rely on these aromatic-aromatic interactions, including the "aromatic box" found in nicotinic acetocholine receptors (Figure 2.4b).[182, 224] An early analysis of crystal structures from the Protein Data Bank, Burley & Petsko (analyzing only 34 protein structures available at the time) observed, "the significance of the aromatic-aromatic interaction for protein structure stabilization is reflected in the strikingly high incidence of such groups: 80 percent of the aromatic pairs identified in proteins are involved in networks."[218] Furthermore, the importance of side chain-aromatic interactions in peptide structure stabilization was noted: "aromatic residues, and hence, aromatic pairs, are absent in regions where the polypeptide chain is disordered."[218]

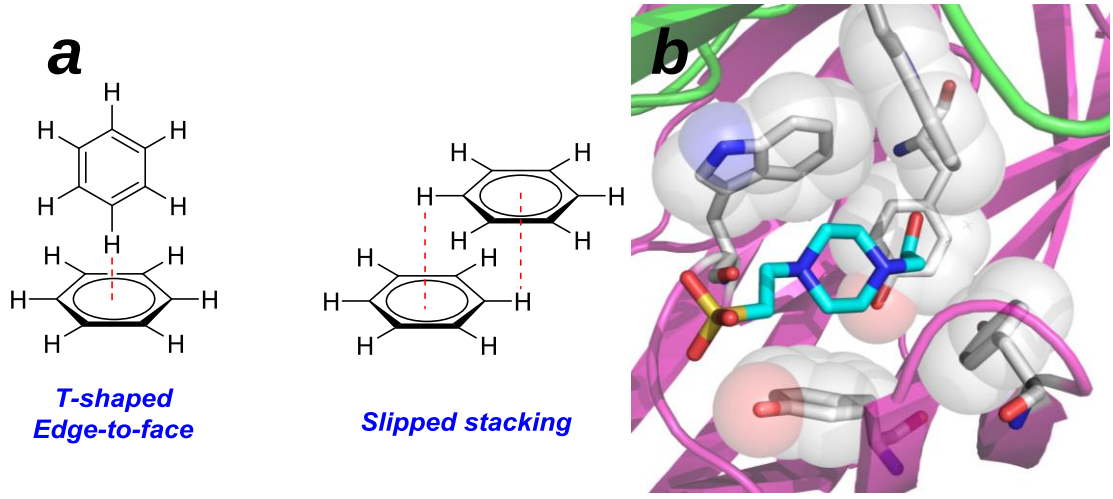


Figure 2.4 Interactions between aromatic rings[223]
(a) Geometries of aromatic-aromatic interactions; (b) extensive aromatic-aromatic interactions within the receptor binding site of the nicotinic acetocholine receptor protein (PDB ID: 1L9B).[224] In addition to acting as a π-acceptor on the face of the aromatic ring, the C–H bonds at the edge of the aromatic ring can act as hydrogen bond donors. Interactions between aromatic rings have been found extensively in proteins, including the "aromatic box" motif.[224]

2.1.3 Sulfur-aromatic interactions and their biological relevance

In the context of other main-group hydrogen bond donors and acceptors, non-covalent interactions based on sulfur are potentially the least understood. In crystallographic studies of X–H hydrogen bonds, both from the CSD and the PDB, the occurrence of sulfur is comparatively less than other main-group elements, limiting the analysis of trends in sulfur-interactions.[225] In early literature describing the hydrogen bond, the hydrogen bonding ability of sulfur was only tentatively considered,[226] but it has since been well established to play important roles in hydrogen bonding interactions.[7, 227, 228] Similar to alcohols, thiols can act both as hydrogen bond donors or acceptors, and both exhibit similar behavior in hydrogen bonding. However, sulfur has ready access to d-orbitals which oxygen does not, and it is a "softer" acid or base in comparison to oxygen. For example, in comparing ethers

and thioethers, the C–S σ* orbitals are larger than the C–O σ* orbitals, which affects the ability of the thioether group to participate as a conventional hydrogen bond acceptor.[7] The larger C–S σ* orbitals in this example allow for thioethers to participate in interactions that resemble "halogen-bonding," which is not accessible for oxygen analogues.[7] Furthermore, the bond angle C–O–H is approximately 109-120° (sp^3 versus sp^2), while the bond angle C–S–H is approximately 100°, which intrinsically affects the geometric preferences of hydrogen bonding involving alcohols and thiols, respectively.

Thiol hydrogen bonds have been identified via ^{1}H NMR spectroscopy, where self-association of alkyl thiols and aryl thiols was observed as a downfield chemical shift with increasing concentration.[226, 229] As with other X–H hydrogen bonds, thiol S–H hydrogen bonds are readily observed via IR spectroscopy as a red shift in the $\nu_{S–H}$ stretching frequency. The thiol S–H stretching frequency is a comparatively weak signal, but few other frequencies are found near 2590-2540 cm^{-1}, where most thiols are identified.[208] In substituted benzene thiols, the stretching frequency $\nu_{S–H}$ was observed to have a concentration- and temperature-dependent shift due to self-association via S–H···S or S–H/π interactions.[226, 230-232] Solvents were also observed to influence the $\nu_{S–H}$ stretching frequency, as thiols can participate in hydrogen bonding interactions with benzene, acetone, ethers, acetonitrile and even chloroform.[230, 233] Boxer and coworkers[233] examined N–H/π and S–H/π aromatic interactions through changes in the $\nu_{X–H}$ stretching frequency with respect to aromatic solvents. With increasingly electron-rich aromatic solvents, the $\nu_{S–H}$ stretching frequency in thiophenol was increasingly red shifted. The most substantial $\nu_{S–H}$ red shift, –40 cm^{-1} from carbon tetrachloride, was observed in hexamethylbenzene, the most electron-rich

aromatic solvent investigated.[233] S–H/π aromatic interactions have also been identified by IR in solid samples, where a $\Delta\nu_{\text{S–H stretch}}$ shift of –19 cm^{-1} was observed from a solution in chloroform to a microcrystalline sample in a KBr pellet. [234]

Numerous examples of sulfur-aromatic interactions have been discovered in proteins, encompassing methionine, cysteine, and cystine interactions with phenylalanine, tyrosine, and tryptophan.[7, 173, 222, 228] Sulfur-aromatic interactions were conventionally thought to be driven by hydrophobic effects (involving Met) or weak electrostatic effects (involving Cys), but have been found to be stronger and more prevalent in proteins than these weak intermolecular forces can completely rationalize.[7] Non-covalent interactions between methionine sulfur atoms and aromatic side-chains have been implicated in stabilizing α-helices in proteins, with estimates between 1-3 kcal mol^{-1}.[135, 136] Viguera & Serrano[235] found *i*,*i*+4 interactions between Cys or Met and Phe stabilized α-helices within model peptides in water, with up to 2.0 kcal mol^{-1} of stability for Phe-Cys interactions. A designed peptide containing a β-turn was stabilized based on a sulfur-aromatic interaction between Met and Trp.[236] Weaver *et al.*[228] found that 39% of reduced cysteine residues in proteins in the PDB interacted with a phenylalanine aromatic ring. An earlier, broader PDB study showed that nearly half of the sulfur-containing amino acids (Met, Cys, and cysteine) interacted with an aromatic ring,[173] suggestive of the importance of these interactions in protein folding. For example, a structurally and energetically significant Cys/Trp sulfur-aromatic interaction was found near the active site in the D2 dopamine receptor, which potentially plays a role in stabilizing the binding pocket.[174] In the conserved 60s loop in flavodoxins, mutations to Met56 has been found to influence the redox potentials of

the flavin cofactor, potentially by effecting a sulfur-aromatic interaction with the flavin (Figure 2.5).[138]

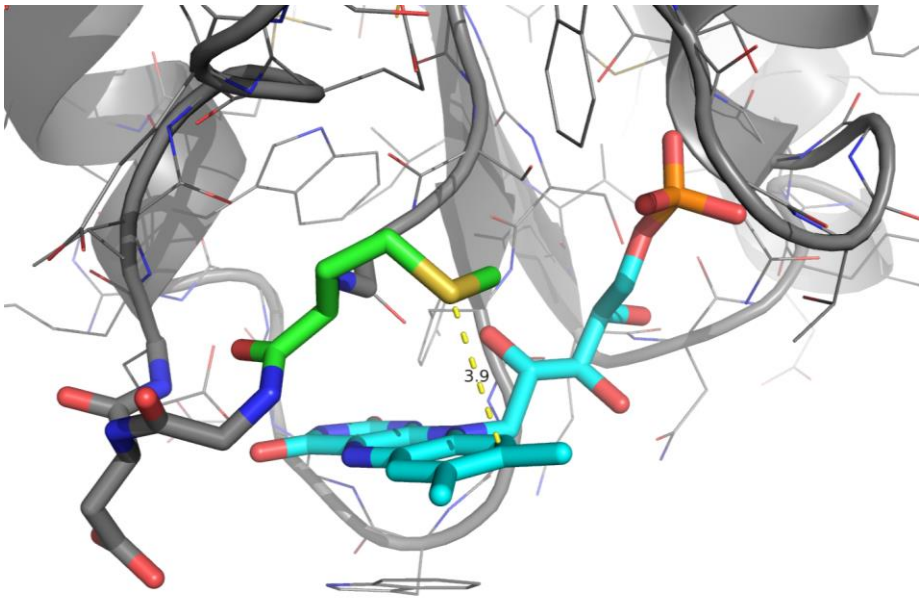


Figure 2.5 Sulfur-aromatic interaction in flavodoxin that influences the redox potential
Oxidized flavodoxin from *Clostridium beijerinckii* (PDB ID: 5NLL).[237] In flavodoxins, the 60s loop undergoes a conformational change upon reduction, where the amide bond between residues 57-58 switches from cis to trans. Residue 56 is typically Trp, Leu, or Met.[138] In flavodoxins containing Met56, the midpoint potentials are less negative than other flavodoxins, which may be due to the presence of a sulfur-aromatic interaction.[138]

S–H/π aromatic interactions have been generally described as weaker versions of O–H/π interactions that may be largely driven by van der Waals interactions, and yet S–H/π interactions are found in numerous examples of small molecules and proteins, and can have significant consequences on structure and stability.[7, 222, 227] In Biswal & Wategoankar's analysis of S–H binding to indoles and benzene, compared to O–H, N–H and C–H, it was found that S–H/π aromatic interactions were the

strongest in comparison to other X–H/π aromatic interactions, and yet these interactions were driven almost completely by dispersion.[164] In calculations performed by Zhou *et al.*[238], it was found that an $n \rightarrow \sigma^*$ interaction significantly stabilized S··O complexes, and it was suggested that $\pi_{Aro} \rightarrow \sigma^*$ contributions can potentially stabilize S··π or S–H/π aromatic interactions. Given that S–H/π aromatic interactions can have subtle but powerful influences on protein structure and function, we sought to identify the nature of these interactions using 4-thiophenylalanine, the sulfur analogue of tyrosine. IR, NMR, and *ab initio* studies of non-covalent interactions with 4-thiophenylalanine can complement the work on cysteine- and methionine-aromatic interactions, given that the sulfur atom has access to the electron-rich aromatic ring. Understanding the strength and geometry of S–H/π aromatic interactions involving aryl thiols can guide the usage of 4-thiophenylalanine and related derivatives in peptide and protein design, and provide fundamental insights into sulfur and thiol interactions. Given that methionine- and cysteine-aromatic interactions have been implicated in protein folding and function, we anticipate broad utility of 4-thiophenylalanine within the context of rationally designed proteins.

2.2 Results

2.2.1 Solution phase synthesis of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

A copper-mediated cross-coupling reaction was previously developed for the practical synthesis of peptides containing 4-thiophenylalanine using commercially available 4-iodophenylalanine.[169] Although this solid-phase synthesis is applicable to most peptides, the potential for side-chain deprotection of trityl groups and subsequent side reactions restricts the broader application of this approach. In order to avoid

potential issues with side-chain reactivity in resin-bound peptides, 4-thiophenylalanine can be synthesized as a protected amino acid for solid-phase peptide synthesis. Synthesis of 4-thiophenylalanine as a free amino acid in solution allows for incorporation into proteins via modified aminoacyl *t*-RNA synthases[17] or auxotrophic bacteria. In order to generate 4-thiophenylalanine in solution, a synthetic strategy was developed based on the copper-catalyzed cross-coupling reaction on Boc-4-iodophenylalanine (Figure 2.6). Protection of the carboxylic acid in Boc-4-iodo-L-phenylalanine was necessary prior the copper-mediated reaction with thiolacetic acid. Due to the potential for disulfide formation in the presence of strong bases, such as those used for hydrolysis of a methyl ester, the use of base-labile protecting groups was not practical for removal after formation of the aryl thiol. Therefore, the carboxylic acid was protected as a *tert*-butyl ester because it is acid-labile. Using this protecting-group strategy, thiolysis of the thioacetyl group under acidic conditions should prevent significant formation of disulfides, and the *tert*-butyl ester can be hydrolyzed under acidic conditions to avoid disulfide formation. Boc-4-iodo-L-phenylalanine was allowed to react with *tert*-butanol in the presence of dimethylaminopyridine (DMAP) to protect the carboxylic acid as a *tert*-butyl ester (Figure 2.7).

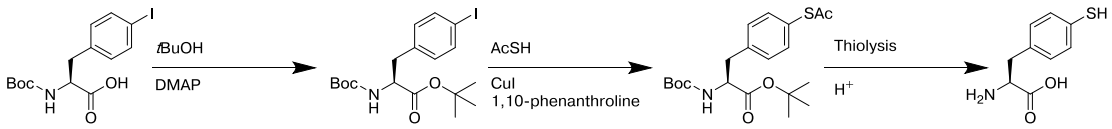


Figure 2.6 Synthetic strategy to obtain 4-thiophenylalanine
From commercially available Boc-4-iodophenylalanine, the carboxylic acid was protected as a *tert*-butyl ester. This protected amino acid can be subjected to copper-catalyzed cross-coupling conditions, similar to those described by Sawada.[88] After thiolysis and global deprotection under acidic conditions, 4-thiophenylalanine can be used directly in expression, or can be Fmoc-protected for incorporation into peptides via solid-phase peptide synthesis.

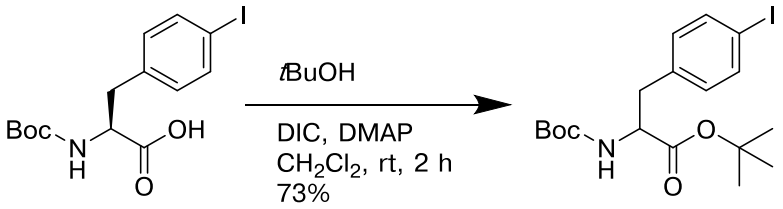


Figure 2.7 Initial synthesis of Boc-4-iodo-phenylalanine-*tert*-butyl ester
The carboxylic acid was protected as a *tert*-butyl ester via *tert*-butanol in the presence of DMAP. This reaction generated reasonable yield on the first attempt, but complete racemization of the product was observed by chiral HPLC.

Characterization of the Boc-4-iodophenylalanine-*tert*-butyl ester product via NMR and HRMS showed that the carboxylic acid was protected as a *tert*-butyl ester. However, chiral HPLC analysis of this product showed substantial racemization (23% ee, Figure 2.8). DMAP is known to cause racemization, particularly when used as a stoichiometric reagent. An alternative method for protecting the carboxylic acid as a *tert*-butyl ester was utilized (Figure 2.8)[239] with a smaller quantity of DMAP (0.3 eq versus >1 eq). Characterization of the product obtained by the scheme in Figure 2.9 showed retention of stereochemistry via chiral HPLC (Figure 2.10). The product yield of the alternative reaction scheme was also improved over the initial synthetic strategy (99% versus 73% product yield, respectively).

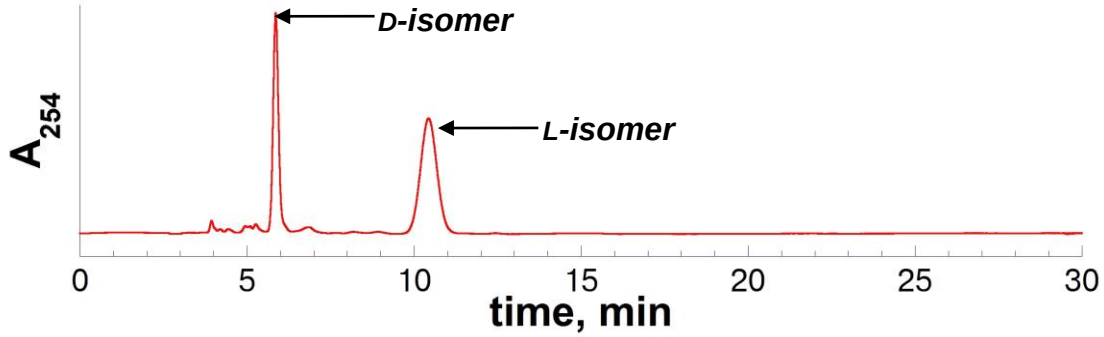


Figure 2.8 Chiral HPLC chromatogram of Boc-4-iodophenylalanine-*tert*-butyl ester product from the initial synthetic strategy

Chiral HPLC analysis of showed substantial racemization (23% ee) of the product Boc-4-iodophenylalanine-t*ert*-butyl ester obtained from the initial synthesis using stoichiometric DMAP. Analytical chiral HPLC chromatogram (UV detection at 254 nm) of purified Boc-4-iodophenylalanine-*tert*-butyl ester using a mixture of isocratic 20% isopropanol in hexanes over 30 minutes on a Daicel ChiralPak 1A column (250 × 4.6 mm, 5 µm particle, 1.0 mL/min).

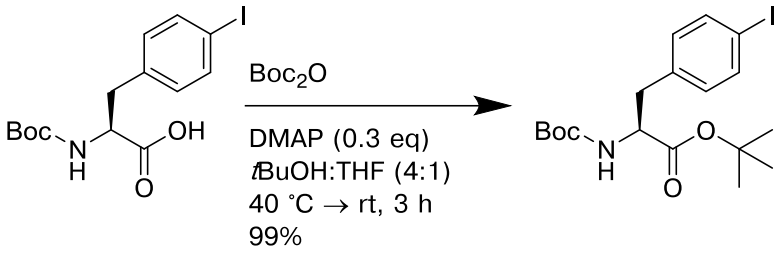


Figure 2.9 Optimized synthesis of Boc-4-iodo-L-phenylalanine-*tert*-butyl ester
An alternative strategy was used to protect the carboxylic acid as a *tert*-butyl ester, using a smaller quantity of DMAP (0.3 eq) and Boc_2O as an additional reagent. Reaction yields were also improved over the previous synthesis.

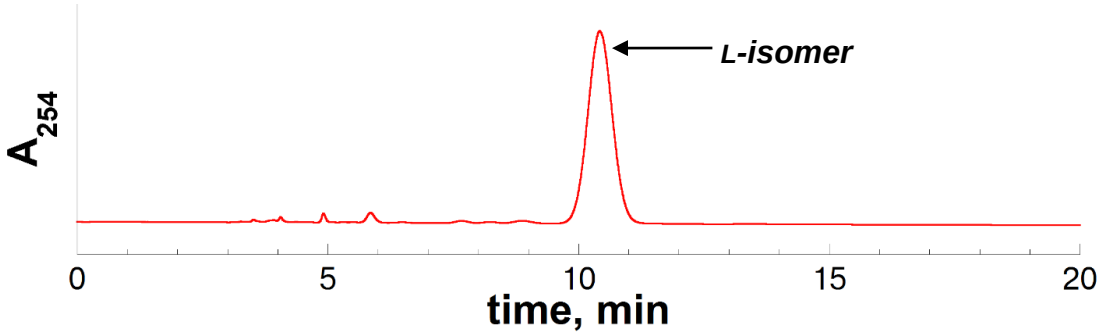


Figure 2.10 Chiral HPLC chromatogram of Boc-4-iodo-L-phenylalanine-*tert*-butyl ester using an alternative synthetic strategy
Using a smaller quantity of DMAP (0.3 eq) and Boc_2O as an additional reagent, chiral HPLC analysis showed that the Boc-4-iodo-L-phenylalanine-*tert*-butyl ester product retained enantiopurity (>98% ee). Analytical chiral HPLC chromatogram (UV detection at 254 nm) of purified Boc-4-iodo-L-phenylalanine-*tert*-butyl ester product using a mixture of isocratic 20% isopropanol in hexanes over 20 minutes on a Daicel ChiralPak 1A column (250 × 4.6 mm, 5 µm particle, 1.0 mL/min).

Boc-4-iodo-L-phenylalanine-*tert*-butyl ester was subjected to the copper-catalyzed cross-coupling reaction conditions with thiolacetic acid, similar to those described by Sawada[88] (Figure 2.11), which were used in the initial reaction screening for the synthesis of peptides containing 4-thiophenylalanine (Chapter 1). The products that resulted from the cross-coupling reaction were a mixture of the thioacetyl product and corresponding disulfides, as identified by ESI-MS, consistent with observations in

the cross-coupling reaction on resin-bound peptides and iodobenzene (Chapter 1.2.1).[169] To identify possible work-up conditions for thiolysis and reduction of the reaction products, 4-S(acetyl)-thiol-L-phenylalanine-*tert*-butyl ester and the corresponding disulfides were isolated and allowed to react with various reductants to produce Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. Initial attempts using sodium methanethiolate for thiolysis followed by reduction with zinc dust were unsuccessful.[240] However, reaction with methanethiolate followed by reduction using dithiothreitol was more successful (Figure 2.12), with the reduced thiol as the major reaction product observed by TLC. Characterization of the reaction product by LIFDI HRMS and ^{1}H NMR confirmed the presence of a free thiol, rather than the symmetric disulfide. The thiol proton was clearly visible in the ^{1}H NMR spectrum as a singlet at δ 3.41 ppm (Figure 2.13a). The aromatic region also differed for the thiol product in comparison to a sample of partially oxidized product (Figure 2.13b, mixture of disulfides and reduced thiol). Additional ^{13}C NMR and ^{1}H-^{13}C HMBC NMR characterization was also completed for the reduced form of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (shown in section 2.2.3, Figures 2.21 and 2.22). Detailed resonance assignments are included at the end of this chapter in section 2.4.4 (Table 2.19). The Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product crystallized during purification as colorless prisms from 20% ethyl acetate in hexanes, and characterization of these crystals will be further discussed in section 2.2.2.

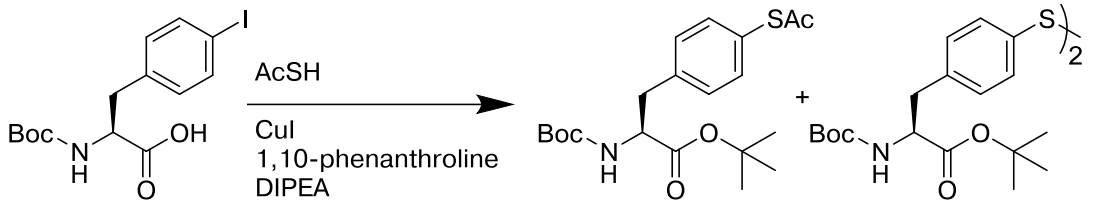


Figure 2.11 Initial copper-catalyzed cross-coupling reaction with Boc-4-iodo-L-phenylalanine-*tert*-butyl ester and thiolacetic acid

The copper-catalyzed cross-coupling conditions described by Sawada[88] were employed for this reaction, similar to the initial reaction conditions utilized with iodobenzene in Chapter 1.2.1.

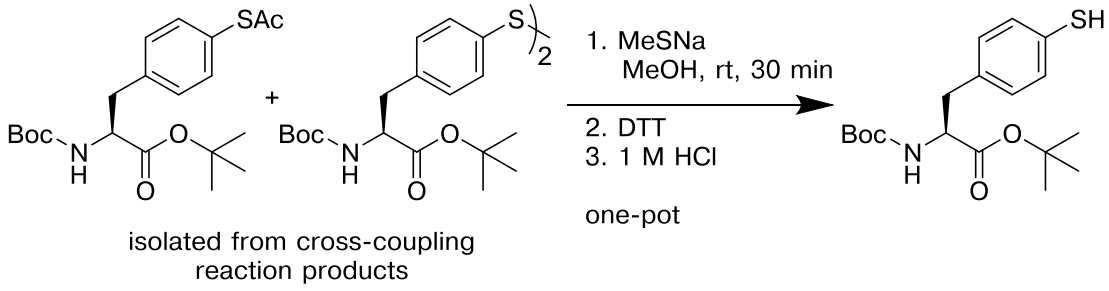


Figure 2.12 Thiolysis and reduction of isolated reaction products to produce Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

The products formed from the copper-catalyzed cross-coupling reaction were isolated from the crude reaction, and then combined as solution of Boc-4-S(acetyl)-thiol-L-phenylalanine-*tert*-butyl ester and corresponding disulfides. After reaction these products were allowed to react with sodium methanethiolate, followed by reduction with dithiothreitol, the Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product was acidified with 1 M HCl and extracted into ethyl acetates and purified by column chromatography. The purified product was verified by ^{1}H NMR and LIFDI-HRMS to be fully reduced.

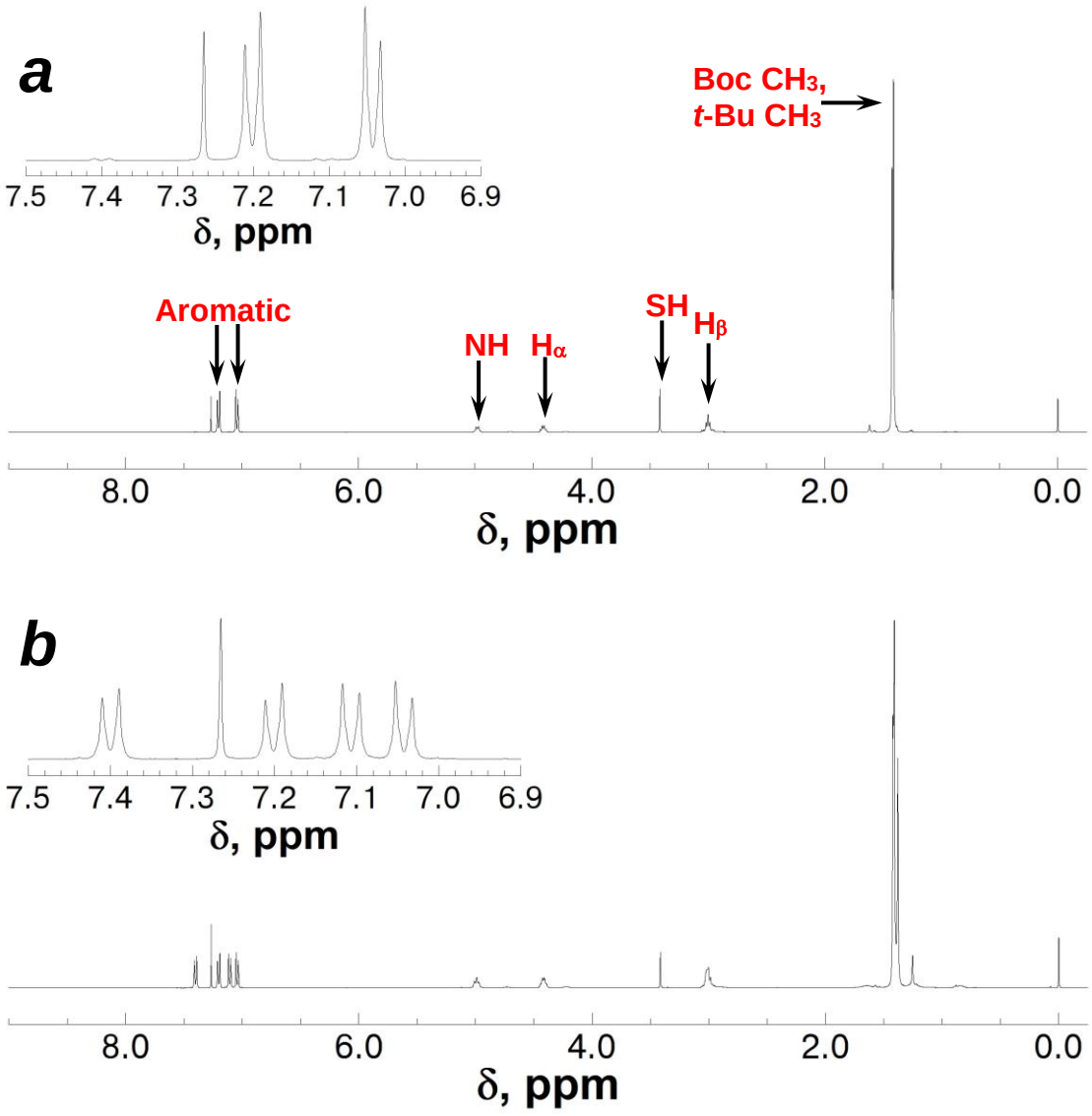


Figure 2.13 ^{1}H NMR Spectra of reduced and oxidized Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in $CDCl_3$

^{1}H NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as (a) the reduced thiol and (b) as a mixture of thiol and symmetric disulfide. The sample shown in (b) was obtained by allowing air oxidation of a sample of (a) over several days. The thiol proton was clearly visible at 3.41 ppm. The aromatic region also differs between the purified thiol and the mixture of thiol and disulfides (respective insets). Combined with mass spectrometry data, the product was confirmed as the reduced thiol species, as opposed to symmetric disulfides.

Having confirmed that Boc-4-thiol-L-phenylalanine-*tert*-butyl ester can be obtained from the isolated products from the cross-coupling reaction, the work-up conditions were optimized so that the 4-thiophenylalanine product could be purified directly from the crude reaction mixture. Boc-4-iodo-L-phenylalanine-*tert*-butyl ester was subjected to copper-catalyzed cross-coupling reaction conditions in either toluene or *tert*-amyl alcohol, according to the scheme in Figure 2.14; both of these reactions generated a mixture of Boc-4-S(acetyl)-thiol-L-phenylalanine-*tert*-butyl ester and the corresponding disulfides. The crude reactions were filtered over celite and the solvent was removed. The crude residue from each cross-coupling reaction was then subjected to sodium methanethiolate and DTT (reaction in toluene) or DTT alone (reaction in *tert*-amyl alcohol). By TLC, both of these reductive work-up conditions cleanly generated the reduced Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product. Based on these data, either toluene or *tert*-amyl alcohol can be used for the copper-catalyzed cross-coupling reaction, as was observed in the synthesis of peptides containing 4-thiophenylalanine on solid phase (Chapter 1.2.3). Furthermore, it was established that DTT was sufficient to reduce the crude mixture of reaction products to Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

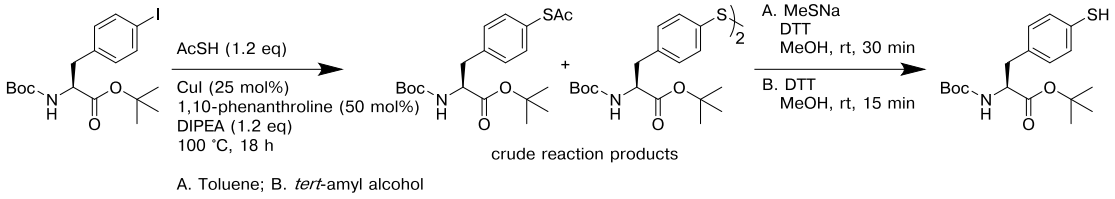


Figure 2.14 Initial screen of conditions for one-pot thiolysis and reduction of reaction products from copper-catalyzed cross-coupling reaction on Boc-4-iodo-L-phenylalanine-*tert*-butyl ester

Two different conditions were compared in the cross-coupling reaction, where either toluene or *tert*-amyl alcohol were used as a solvent. Reaction conversion was similar as observed by TLC. The reaction mixtures were cooled and filtered over celite and the solvent was removed. Reaction A (cross-coupling reaction in toluene) was allowed to react with sodium methanethiolate and DTT in methanol for 30 minutes at room temperature. Reaction B (cross-coupling reaction in *tert*-amyl alcohol) was allowed to react with only DTT in methanol of 15 minutes at room temperature. Both of these reaction schemes generated a single product, observed by TLC, which corresponded to Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

The optimized reaction conditions initially generated the thiol product in 68% yield, which decreased to 13-40% yield after multiple repurifications, although these product yields were inconsistent. The product isolated from this reaction sequence was a yellow oil rather than a colorless, crystalline material (see above). The ^{1}H NMR and mass spectrometry data from the yellow oil were consistent with data obtained from the colorless, crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product. It was proposed that the yellow color of the reaction product resulted from a charge transfer complex with iodine, where the iodine can potentially "leech" out of the reaction product over multiple repurifications over silica gel. The presence or absence of iodine from the reaction product can significantly affect the mass of the final product, depending on the extent of iodine removal, potentially resulting in very inconsistent isolated reaction yields. Various aryl thiols, sulfides, and disulfides have been previously examined as charge-transfer complexes with iodine, and the iodine charge-

transfer complex with thiophenol was reported to have a UV λ_{max} at 310 nm, consistent with the observed yellow color of the isolated Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.[241] Based on similar charge-transfer complexes with thiols,[241] it was proposed that the iodine associated with a lone pair of electrons on the sulfur atom, an $n \rightarrow \sigma^*$ interaction, also known as a "σ-hole" or halogen bonding interaction.[193, 241-243] In order to remove iodine from the proposed charge-transfer complex, which can potentially increase the likelihood of disulfide formation (via oxidation of the thiol with iodine) and obscure the accurate isolated yields from the cross-coupling reaction, the purified, yellow Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product was precipitated as a white, crystalline solid from hexanes. Multiple cycles of precipitation from hexanes were required to obtain the total isolated yield of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as a white solid (rather than as a yellow, oily charge-transfer complex) from the purified reaction product. With the additional precipitation step after column purification, the isolated product yields from the copper-catalyzed cross-coupling reaction were consistently 30%.

Reaction completion was verified by observed depletion of the Boc-4-iodo-L-phenylalanine-*tert*-butyl ester starting material, both by TLC and ^{1}H NMR of the crude reaction. The Boc- and *tert*-butyl ester methyl protons were observed in the ^{1}H NMR of the crude reaction, suggesting that the low isolated reaction yields were not due to potential Boc- or *tert*-butyl ester deprotection during the reaction conditions. Hartwig has previously reported copper(I):1,10-phenanthroline:thiophenolate complexes in reactions with aryl thiols and aryl iodides, using conditions similar to the optimized cross-coupling reaction.[244] Similar complexes were reported by Kaim, with observed aggregation of the complexes from solution.[245] During the copper-catalyzed

cross-coupling reaction to produce Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, large, black, needle-like crystals were observed to form on the sides of the reaction vial as the reaction progressed. These crystals were potentially insoluble complexes of the thiolate product with copper(I) and 1,10-phenanthroline, similar to those described by Hartwig.[244] A possible structure of this copper(I):1,10-phenanthroline:thiophenylalanine product complex is shown in Table 2.1, although multiple related species of this complex could have been present in the crude reaction products. In Hartwig's work, the isolated copper(I):thiophenolate crystals were allowed to react with an excess of aryl iodides to drive the reaction to formation of the desired aryl thioether products.[244] A similar strategy was utilized in an effort to improve the isolated product yield of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, where the crude cross-coupling reaction products were allowed to react with an excess of aryl thiolates or potential copper(I) ligands (Table 2.1). In this manner, the Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product can potentially be released from the copper(I) complex prior to the thiolysis and reduction step. Heating these reaction solutions, in order to mimic the metalation conditions during the cross-coupling reaction, was expected to break the copper-sulfur bonds or increase the rate of exchange with thiolates (or neocuprione) in solution. Work-up strategies on the crude reaction mixture using different thiolates or copper(I) ligands were screened for formation of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (Table 2.1). In addition to the work-up with excess thiols following the cross-coupling reaction, the amount of copper(I) catalyst was decreased to 10 mol%, which would reduce the overall formation of copper(I):1,10-phenanthroline:thiophenylalanine product complexes.

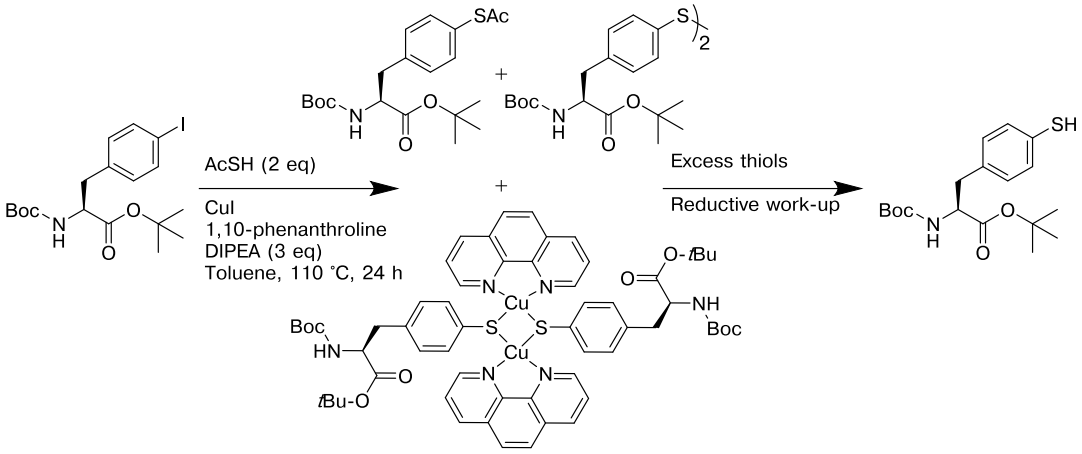
Table 2.1. Screening reaction conditions for the release of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product from the possible copper(I) complex

A crude reaction product mixture from the copper-catalyzed cross-coupling reaction was used to screen these work-up conditions, after verifying completion of the reaction by TLC. Following the reductive conditions described below, the products were isolated, either as Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as a charge-transfer complex with iodine, or as a mixture of the thiol and corresponding disulfides as charge-transfer complexes

Abbreviations: 4-mercaptophenylacetic acid (MPAA), dithiothreitol (DTT), thiolacetic acid (AcSH), diisopropylethylamine (DIPEA), copper(I) iodide (CuI), potassium acetate (KOAc), *tert*-butanol (*tert*-BuOH).

*"charge transfer product" refers to products isolated that were visually yellow and oily, indicating some unknown quantity of iodine still present in the sample.

n.d. indicates "not determined"



Entry	Catalyst Loading (2:1 1,10-phenanthroline:CuI)	Treatment on Crude reaction products	Charge transfer product* yield (both thiol and disulfide)	Charge transfer product* yield (thiol only)
1	25 mol% CuI	2 eq DTT, 2 eq Cs_2CO_3 30 min at 60 °C in THF	50%	n.d.
2	25 mol% CuI	2 eq DTT, 2 eq Cs_2CO_3 2 eq KOAc 3 h at 60 °C in THF	51%	n.d.
3	25 mol% CuI	2 eq DTT, 2 eq Cs_2CO_3 5 eq 4-hydroxythiophenol 3 h at 60 °C in THF	32%	n.d.

Table 2.1 continued

4	25 mol% CuI	2 eq DTT, 2 eq Cs_2CO_3 5 eq MPAA 8 h at 60 °C in THF	54%	n.d.
5	25 mol% CuI	2 eq DTT, 1 eq neocuprione 4 eq MPAA 8 h at 110 °C in THF	26%	n.d.
6	10 mol% CuI	2 eq DTT, 2 eq Cs_2CO_3 30 min at rt in THF	72%	33%
7	10 mol% CuI	1.5 eq DTT, 1.5 eq MPAA 1.5 eq Cs_2CO_3 2 h at 60 °C in THF	n.d.	34%
8	10 mol% CuI	1.5 eq DTT, 1.5 eq MPAA 1.5 eq Cs_2CO_3 2 h at 110 °C in THF	n.d.	31%
9	10 mol% CuI	1.5 eq DTT, 5 eq MPAA 2 eq Cs_2CO_3 3 h at 110 °C in *t*-BuOH	n.d.	38%

The isolated product yields were improved via reaction of the crude reaction products with excess aryl thiolates at elevated temperatures. Isolated reaction yields were improved by decreasing the amount of copper(I) catalyst (Table 2.1, entry 1 versus entry 6), consistent with the proposed formation of copper(I):1,10-phenanthroline:thiophenylalanine complexes. 4-Mercaptophenylacetic acid was selected as the aryl thiolate to compete with the Boc-4-thiol-L-phenylalanine-*tert*-butyl ester product for the copper(I) complex for two reasons: due to the relatively electron-rich aromatic ring, which promotes the binding ability of the thiolate with the copper(I) complex; and due to its hydrophilicity, where the 4-mercaptophenylacetic acid and its copper(I) complex will retain on a normal phase silica gel column and allow for more efficient purification. 4-Hydroxythiophenol coeluted with the product, which complicated purification without improving the reaction yield. Neocuprione was not found to be effective at competing with the product for the copper(I) complex.

The optimized reaction conditions are described in detail in the experimental section (Chapter 2.4.5).

The complications in the solution phase synthesis of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester demonstrated the utility and practicality of the solid phase approach to generate peptides containing 4-thiophenylalanine, described in Chapter 1. However, this solution-phase reaction sequence provides a novel synthetic route to obtain 4-thiophenylalanine as an amino acid, which can be deprotected for expression in proteins,[17] or protected for incorporation into peptides via solid-phase peptide synthesis. Based on this synthesis, Fmoc-4-S(propargyl)-thiophenylalanine was generated and incorporated into model peptides as a site for bioorthogonal conjugation.[139] In addition, the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester provided unique insight into the nature of S–H/π aromatic interactions. The crystal structure of this amino acid revealed a special geometry of non-covalent S–H/π aromatic interactions, which was further explored using solid state NMR, FT-IR, *ab initio* calculations, and comparison to other crystal structures in the CSD.

2.2.2 Crystallography of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

Boc-4-thiol-L-phenylalanine-*tert*-butyl ester crystallized during column purification using 20% ethyl acetate in hexanes. The product crystallized as a clear, colorless, orthorhombic crystal, with approximate dimensions of 0.19 mm × 0.20 mm × 0.33 mm. The x-ray crystal structure was solved to 0.77 Å resolution. The hydrogen atoms were located for the thiol and carbamate hydrogens based on observable electron density (Figure 2.15), which provided reasonable approximation of the thiol bond length and trajectory (Figure 2.16, Figure 2.17, and Figure 2.18).

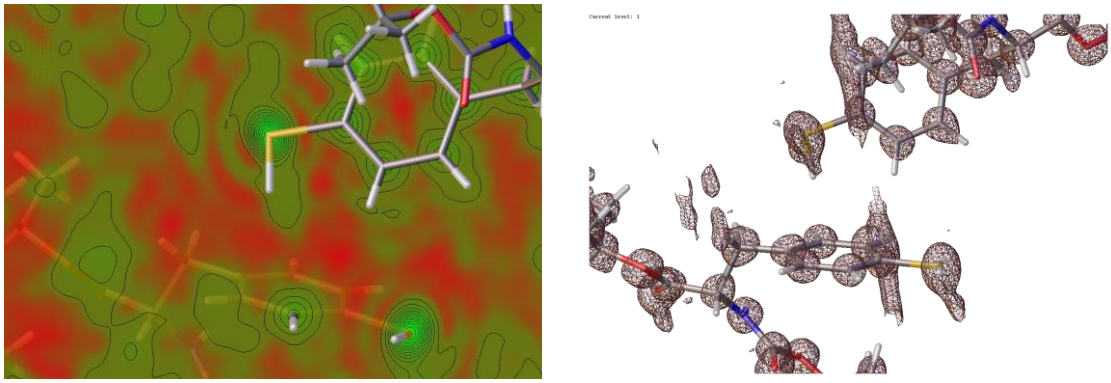


Figure 2.15 Electron density of crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester
The thiol hydrogen atom of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was located based on the observed electron density. Left: contours of 2Fo-Fc electron density map. The thiol S-H bond is within the plane of the contours. Green indicates electron density. Right: wireframe of 2Fo-Fc electron density map shown at 1 sigma.

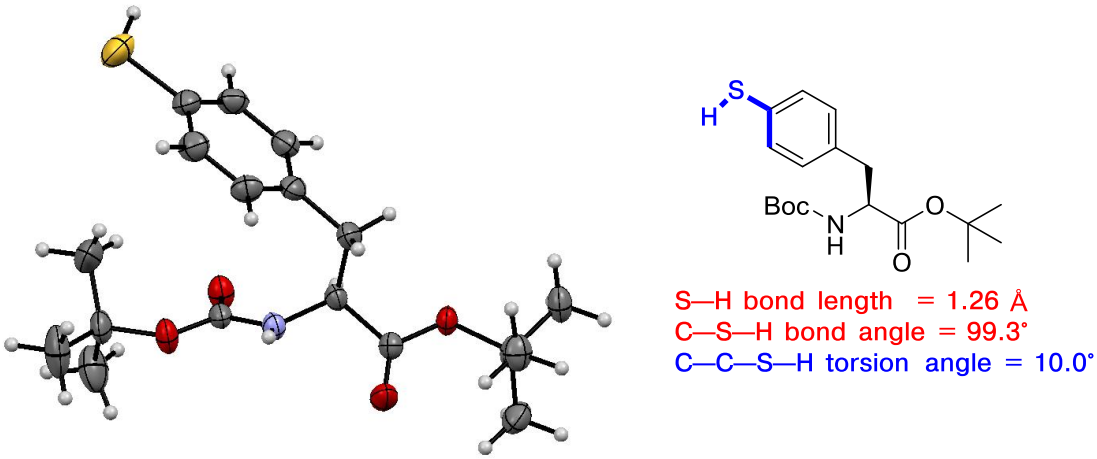


Figure 2.16 The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester
The thiol and carbamate hydrogen atoms of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were located based on the electron density map. Left: ORTEP diagram of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester with ellipsoids shown at 50% probability; right: relevant bond lengths and angles are shown in the schematic.

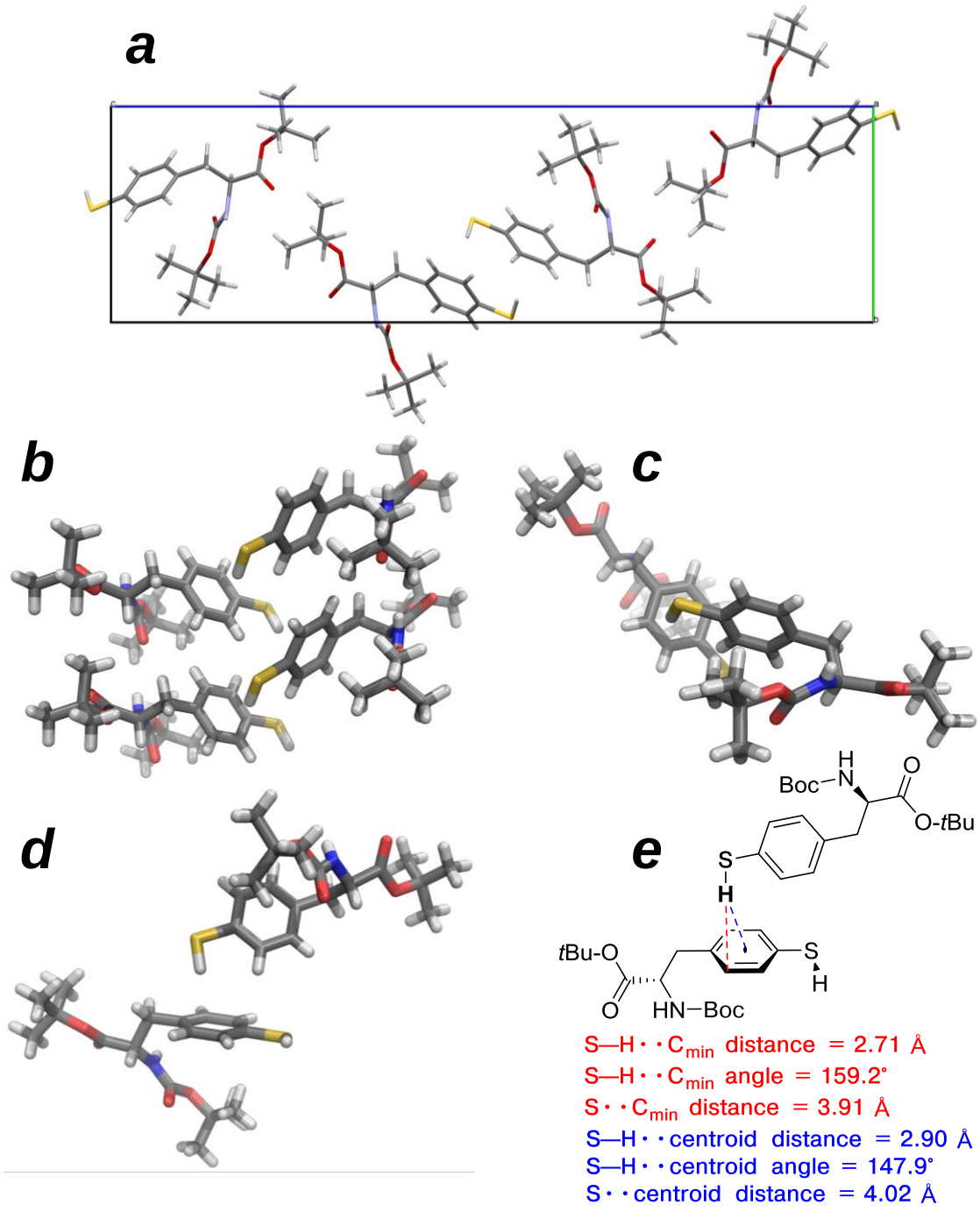


Figure 2.17 Geometry of an observed intermolecular S–H/π aromatic interaction in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

(a) The unit cell of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester revealed an intermolecular S–H/π aromatic interaction. The thiol S–H bond did not engage in a conventional hydrogen bond with either available carbonyl oxygen or the ester lone pairs of electrons; (b) the intermolecular S–H/π aromatic interaction stabilized the crystal packing in a "herring bone" pattern; (c, d) viewing down the thiol S–H bond shows alignment towards an aromatic bond or carbon, rather than the aromatic centroid; (e) schematic representation of the interacting Boc-4-thiol-L-phenylalanine-*tert*-butyl ester dimer and relevant intermolecular contact distances. C_{min} indicates the nearest aromatic carbon to the thiol hydrogen atom, and centroid indicates the centroid of the aromatic ring. The distance from the thiol hydrogen to the aromatic ring carbon is less than the ΣvdWr for carbon and hydrogen atoms (2.90 Å[203]). Both the thiol hydrogen and sulfur atoms are nearer to an aromatic ring carbon than the aromatic centroid.

The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester revealed an intermolecular S–H/π aromatic interaction with exceptionally close contacts between the thiol S–H and the interacting aromatic ring carbons. It might be expected that the thiol would participate in a conventional hydrogen bond with one of the carbonyl oxygen atoms or with the ester oxygen atom, each with two lone pairs that are typical hydrogen bond acceptors. Instead, the thiol S–H bond appeared to interact preferentially with the aromatic ring, and stacked within the unit cell as a "herring bone" pattern (Figure 2.17b). The distance from the located hydrogen atom and the centroid of the aromatic ring was 2.90 Å, at the sum of the van der Waals distance (ΣvdWr) for carbon and hydrogen.[203] However, the thiol hydrogen atom was found to be closer to one of the carbons in the aromatic ring, at 2.71 Å, which is less than the ΣvdWr for carbon and hydrogen. The alignment of the thiol bond was also more favorable with the aromatic carbon over the aromatic centroid, where the angle S–H–C_{min} (referring to the nearest aromatic carbon) was 159° and the angle S–H–centroid was 148°. Even with reasonable location of the hydrogen atom, x-ray crystallographic data does not provide an accurate location, due to the intrinsically low electron density

around a hydrogen atom. Location of the hydrogen atom based on the electron density map can bias the placement of the hydrogen nearer to the heavier atom, apparently shortening the bond length. The observed S–H bond length in the crystal structure was 1.26 Å, significantly shorter than the known bond length, 1.338 Å.[203] To correct for this heavy atom bias, the bond length was normalized to a standard length, placing the hydrogen at the appropriate location along the vector of the bond (Figure 2.18).[203] With the normalized S–H bond length, the contact between the thiol hydrogen atom and the nearest aromatic carbon is even closer, 2.63 Å.

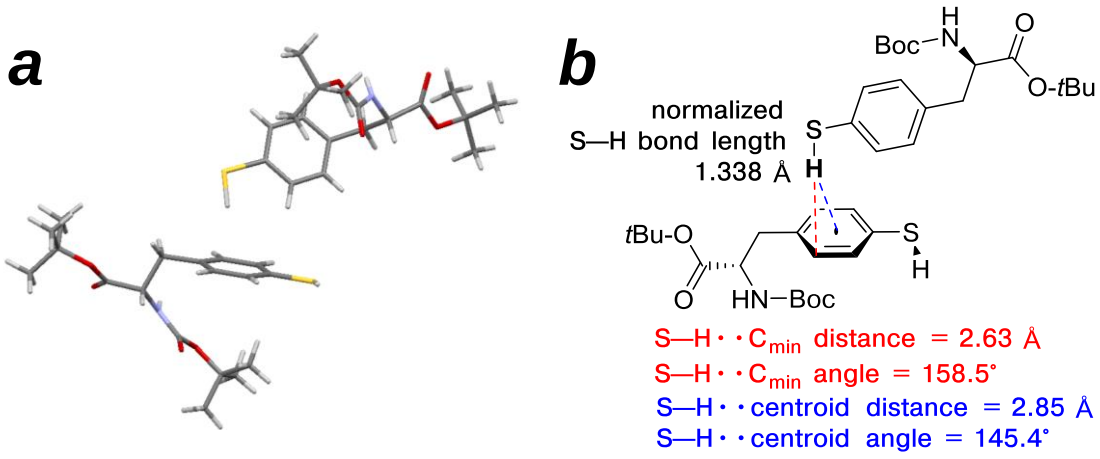


Figure 2.18 Geometry of the intermolecular S–H/π aromatic interaction in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester with a normalized thiol S–H bond length
(a) The thiol S–H bond was normalized to 1.338 Å[203] along the vector of the S–H bond to account for a heavy atom bias in the solved crystal structure; (b) schematic representation of the interacting Boc-4-thiol-L-phenylalanine-*tert*-butyl ester dimer and relevant intermolecular contact distances, with the adjusted position of the hydrogen atom. C_{min} indicates the nearest aromatic carbon to the thiol hydrogen atom, and centroid indicates the centroid of the aromatic ring.

In another effort to crystallize Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, an alternative packing structure was observed after a sample was crystallized from

methanol over three days at room temperature (Figure 2.19). This crystal structure did not show evidence of a S–H/π interaction, and instead the thiol S–H bond participated in an intermolecular hydrogen bond with the Boc- carbonyl oxygen atom, a conventional hydrogen bond.[190] However, this crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was found to be racemic, even though the mother liquor was measured to have 98% ee (verified via chiral HPLC analysis). It is plausible that these monoclinic, racemic crystals formed initially from a sample that contained trace D-amino acid, and the orthorhombic, enantiopure crystals formed more slowly, after the D-amino acid content has been removed from solution. Indeed, chiral analysis of the mother liquor from which the racemic crystals had formed was 98%ee, supporting this possibility.

The intermolecular distances and angles for the thiol S–H bond to the two different acceptors, either the phenyl ring as in Figure 2.17 or the carbonyl oxygen atom as in Figure 2.19, are compared in Table 2.2. Measured bond lengths, angles, and torsions for both crystal forms of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester are noted in Table 2.3, and compared to a known crystal structure of Boc-tyrosine methyl ester (CSD NIKZAK).[246]

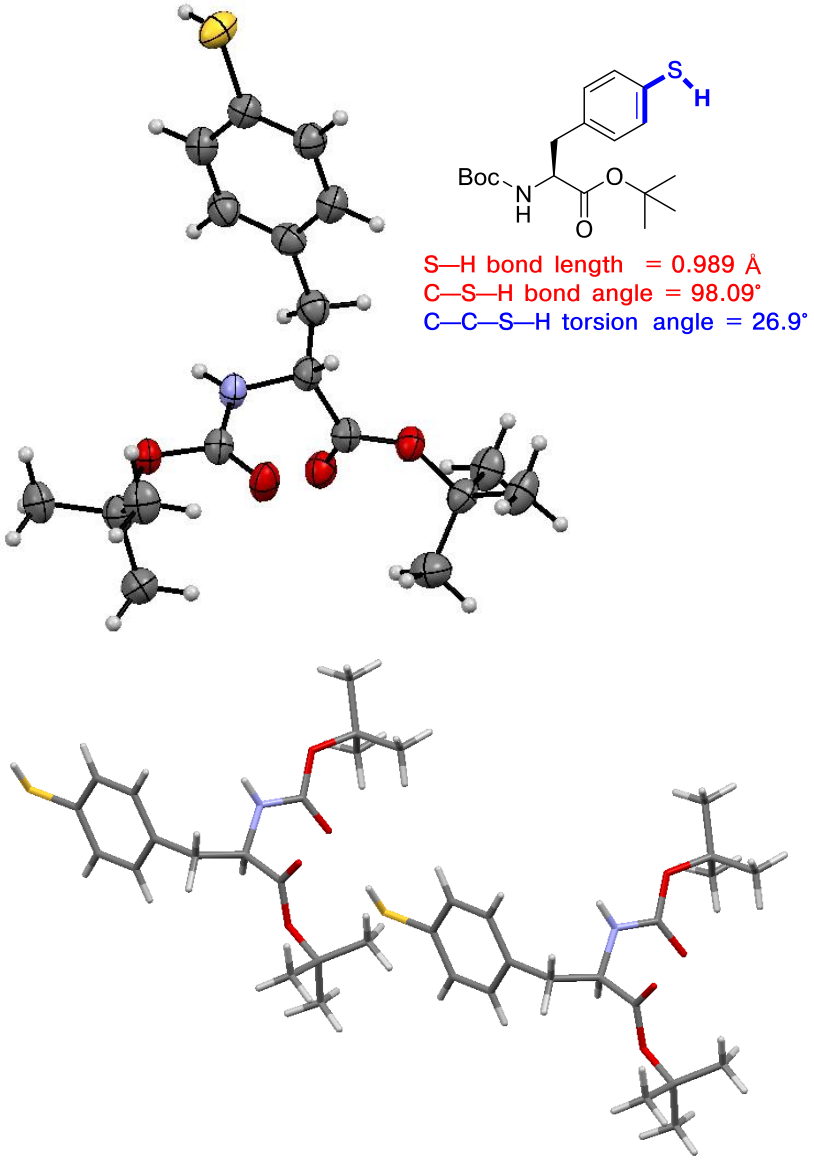


Figure 2.19 Crystal structure of Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester The sample used for crystallization was 98% ee, but these centrosymmetric, monoclinic crystals were racemic and had a different crystal packing than enantiopure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. (a) ORTEP diagram of Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester with ellipsoids shown at 50% probability; (b) relevant bond lengths and angles are shown in the schematic. The thiol S–H bond length is unusually short; (c) the crystal packing showed an intermolecular S–H/O interaction. This intermolecular interaction contrasts from the observations in the enantiopure, orthorhombic crystals shown in Figure 2.15-2.18.

Table 2.2. Comparison of intermolecular S–H interactions in different crystal forms of Boc-4-thiolphenylalanine-*tert*-butyl ester

Measured distances and angles were obtained from the crystal structures shown in Figure 2.17 & Figure 2.19. The comparison is between the acceptor of the S–H bond, either as the phenyl ring (Figure 2.17) or as the carbonyl oxygen atom (Figure 2.19). Cmin refers to the nearest aromatic carbon atom.

Parameter	Enantiopure S–H··π interaction Acceptor = phenyl ring	Racemic S–H··O interaction Acceptor = carbonyl O
S–H··A distance, Å	To centroid: 2.90 To C_{min}: 2.71	2.38
S–H··A angle, °	To centroid: 147.9 To C_{min}: 159.2	176.2
S··A distance, Å	To centroid: 4.02 To C_{min}: 3.91	3.37

Table 2.3. Bond lengths, angles, and torsions in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester

Measured parameters were obtained from the crystal structure shown in Figure 2.17 & Figure 2.19. The measurements for Boc-tyrosine methyl ester are previously published (CSD NIKZAK).[246] Hydrogen atoms are as reported in the CSD or as they were located in the x-ray crystal structure, and the S–H or O–H bond lengths are not normalized.

Parameter	Boc-4-thiol-L-phenylalanine-*tert*-butyl ester X = S	Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester X = S	Boc-tyrosine methyl ester X = O
C–X bond length, Å	1.77	1.78	1.37
X–H bond length, Å	1.26	0.99	0.82
C–X–H bond angle, °	99.3	98.1	109.5
C–C–X–H torsion angle, °	1.77	26.9	6.50
ϕ**, °**	-140.0	62.1	-71.3
ψ**, °**	173.2	-138.0	-7.7
χ_1**, °**	-50.5	68.4	-67.1

For the thiol that interacted with the carbonyl oxygen atom, the distance was shorter and alignment was more linear in comparison to the interaction observed with the thiol and phenyl ring. The increased electronegativity of the oxygen atom and greater electrostatic driving energy can account for this apparently stronger interaction. It is possible that additional interactions in the crystal packing of racemic Boc-4-thiolphenylalanine-*tert*-butyl ester stabilized the geometry and alignment of this S–H··O interaction, given that this interaction was only observed in racemic crystals.

In all three cases, the X–H bond (where X = S or O) was shorter than standard bond lengths,[203] which is not uncommon for x-ray crystallographic data. The C–S–H

bond angles for both forms of Boc-4-thiolphenylalanine-*tert*-butyl ester were typical of thiol bonds, which generally have tighter bond angles than their oxygen analogues.[60] This aspect about hydroxyls and thiols is sometimes overlooked in considering protein crystal structures or designing protein variants involving cysteine and serine. Both serine and cysteine can participate in conventional hydrogen bonds and in aromatic interactions in proteins, but have very different geometric preferences and different preferences for acceptor groups (cysteine is more likely to participate in aromatic interactions over serine).[247]

In the crystal structure of Boc-4-thiolphenylalanine-*tert*-butyl ester, intermolecular hydrogen bonds between the carbamate hydrogen and carbonyl oxygen atoms provide further stability in the crystal packing. Indeed, the enantiopure and racemic crystals have different overall crystal packing, and different intermolecular thiol interactions. The additional interactions observed in the crystal packing, which clearly influence the thiol hydrogen bond acceptors and geometry, can potentially obscure the fundamental nature of the thiol-aromatic interaction. In order to determine the energy and geometry of the S–H/π interaction, separate from additional interactions that stabilize the crystal packing in Boc-4-thiolphenylalanine-*tert*-butyl ester, *p*-thiocresol was crystallized and the structure was solved for comparison. *p*-Thiocresol can be viewed as a truncation of 4-thiophenylalanine, where the carbamate, amide, and *tert*-butyl groups are removed. *p*-Thiocresol can orient freely and the crystal packing should represent the most favorable interactions and geometry. It was discovered that commercially available *p*-thiocresol was sold as diffractable crystals (Sigma-Aldrich), and the crystal structure was obtained directly from commercially available material without further recrystallization (Figure 2.20).

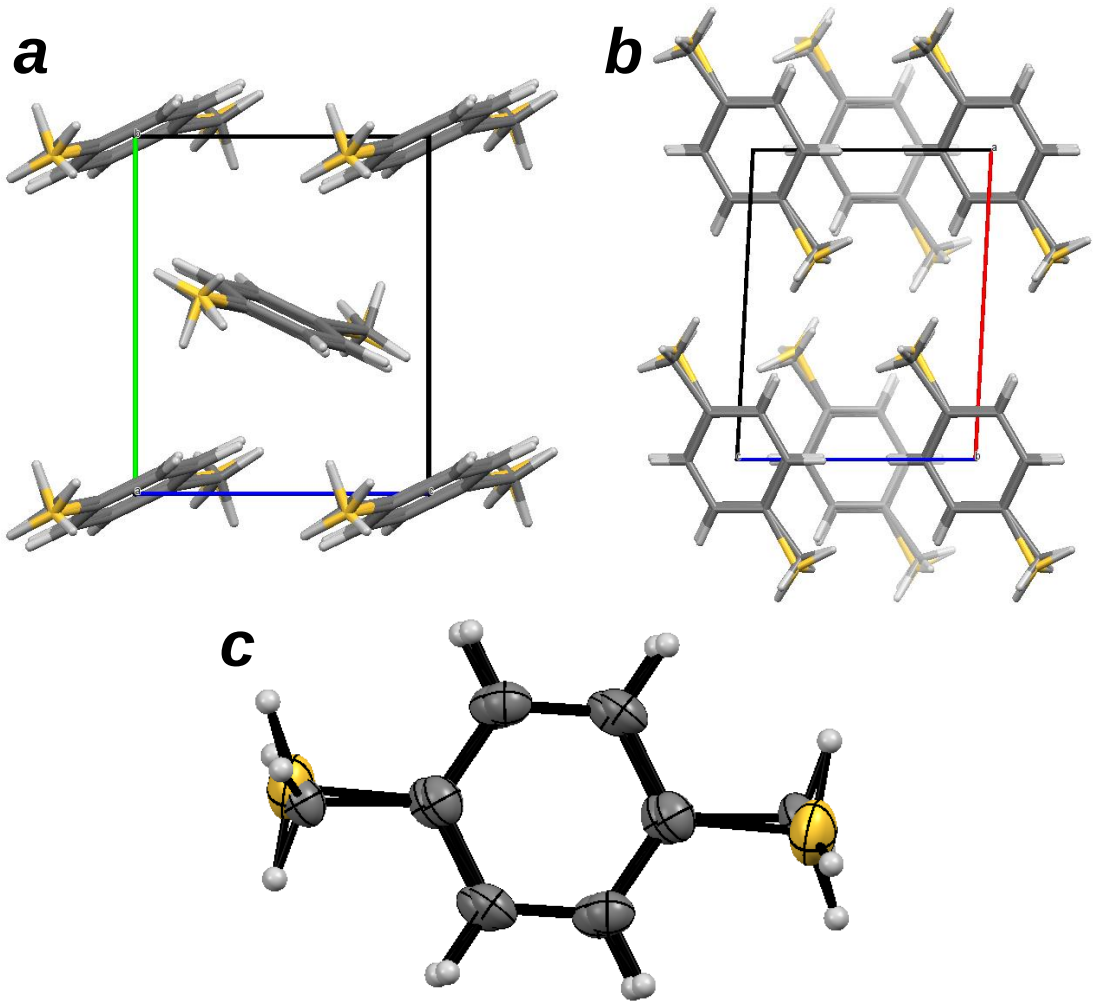


Figure 2.20 Crystal structure of *p*-thiocresol
Diffractable crystals of *p*-thiocresol were obtained directly from commercially available material (Sigma-Aldrich). (a, b) The unit cell for *p*-thiocresol is shown along two different axes. The staggered, columnar arrangement (right) was potentially the source of the symmetry observed in the electron density map, which rendered the methyl and thiol substituents indistinguishable; (c) the ORTEP diagram of *p*-thiocresol with ellipsoids shown at 50% probability shows the centrosymmetric crystal structure, where the thiol group and methyl group were indistinguishable by electron density.

Surprisingly, the space group of these *p*-thiocresol crystals was monoclinic, and the structure was centrosymmetric about the aromatic ring. It is possible that the columnar stacking of *p*-thiocresol in the crystal structure "blurred" the electron density

around the methyl and thiol groups, rendering these two substituents indistinguishable by electron density. Alternatively, disorder and similar sterics allow either the methyl or the thiol group to exist in either position. Unfortunately, the thiol hydrogen could not be uniquely located, given this centrosymmetric structure of *p*-thiocresol, and comparison of the S–H/π interaction with Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was not possible with *p*-thiocresol.

Having observed these short, intermolecular contacts within Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, this amino acid provided a unique opportunity for detailed insights into the nature of S–H/π aromatic interactions, in addition to the potential applications for this amino acid in protein modification. The geometry and strength of the S–H/π interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was further explored via NMR spectroscopy, FT-IR spectroscopy, and *ab initio* calculations, which are discussed in the following sections.

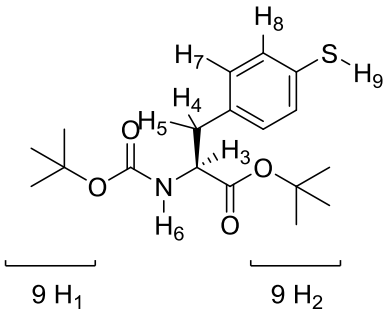
2.2.3 NMR Characterization of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol

In ^{1}H NMR spectroscopy, the chemical shift (δ) of a proton signal is dependent on the shielding effects from the local chemical environment, with respect to induced magnetic fields. The chemical shift can change depending on solvent interactions and shielding effects from the local environment. ^{1}H NMR spectra were obtained for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in solvents that could potentially interact with the thiol proton, thereby changing the shielding effects and changing the observed chemical shift. Chloroform-d has the potential to weakly interact with the thiol sulfur via a C–D interaction or with the thiol sulfur via a halogen bond; benzene-d_6 can potentially interact as an S–H/π or S/π aromatic interaction; methanol-d_3 can interact

as a traditional hydrogen bond acceptor S–H··O. These proton chemical shifts for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were compared to *p*-thiocresol in the same solvents, and the summarized resonance assignments are in Table 2.4 and Table 2.5, respectively.

Table 2.4. ^{1}H chemical shift assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as a function of solvent

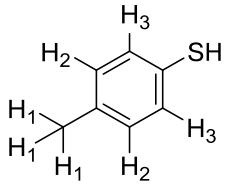
^{1}H chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HMBC data (Figure 2.22), and ^{1}H chemical shifts in other solvents were assigned by analogy. The $\Delta\delta$ with respect to the thiol resonance in $CDCl_3$ were calculated.



H#	Group	$CDCl_3$ δ, ppm	MeOH-d_3 δ, ppm	Benzene-d_6 δ, ppm
1	*tert*-butyl methyl	1.41	1.30	1.22
2	*tert*-butyl methyl	1.42	1.31	1.39
3	Hα	4.41	4.09	4.58
4	Hβ1	3.00	2.73	2.78
5	Hβ2	3.00	2.88	2.90
6	Carbamate	4.98	6.69	5.00
7	C_{Aro}, ortho	7.20	6.99	6.80
8	C_{Aro}, meta	7.04	7.10	6.88
9	Thiol	3.41	4.00	2.97
Δδ, ppm		0.00	+0.59	–0.44

Table 2.5. ^{1}H chemical shift assignments for *p*-thiocresol as a function of solvent

^{1}H chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC spectra (Figures 2.24, 2.25), and ^{1}H chemical shifts in other solvents were assigned by analogy. The Δδ with respect to the thiol resonance in $CDCl_3$ were calculated.



H#	Group	$CDCl_3$ δ, ppm	MeOH-d_3 δ, ppm	Benzene-d_6 δ, ppm
H1	Methyl	2.29	2.26	1.97
H2	C_{Aro}, ortho	7.18	7.15	6.97
H3	C_{Aro}, meta	7.04	7.03	6.73
SH	Thiol	3.37	4.03	3.03
Δδ, ppm		0.00	+0.66	–0.34

Based on prior IR studies on aryl thiols, chloroform is known to interact weakly with thiols, while benzene, acetone, and dioxane interact with thiols more strongly (observed as a greater red shift in the ν_{S-H} stretch, see Chapter 2.2.4).[230, 231] Compared to chloroform, the chemical shift of the thiol proton was shifted downfield by 0.59-0.66 ppm in the presence of a polar solvent (in methanol-d_3) for both *p*-thiocresol and Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. This downfield shift in the thiol proton chemical shift indicated a deshielding effect due to solvent interactions with methanol. IR studies indicated that benzene and substituted benzene derivatives interact strongly with aryl thiols, depending on the electron-donating ability of the

aromatic substituents.[233] In contrast, the thiol proton resonance was shifted upfield by 0.34-0.44 ppm in the presence of a benzene-d_6 for both *p*-thiocresol and Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. An upfield chemical shift change in the thiol proton in the presence of benzene-d_6 indicates a shielding effect from an aromatic ring. Aromatic solvents such as benzene are known to induce chemical shift changes upon interacting with specific protons, termed aromatic-solvent induced shift (ASIS).[248] In general, the chemical shifts of a given compound are shifted upfield in benzene-d_6 compared to $CDCl_3$. The ring current effects in aromatic rings causes a shielding effect on protons that are located above or below the plane of the aromatic ring, which shifts the proton resonance to a higher field upon interaction with the aromatic ring.[248] The differing chemical shifts of the thiol proton between benzene-d_6 and methanol-d_3 suggest that the thiol interactions with benzene are fundamentally different from the interactions with conventional hydrogen bonding solvents.

To gain further insight into the nature of the S–H/π interaction in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, solid-state NMR was used to determine the shielding effects on the aromatic ring and thiol proton in the crystal structure. Typical solid-state NMR produces broad, weak signals that are not useful for interpretation, due to anisotropy of the nuclei within a solid sample. Cross-polarization magic-angle spinning (CP-MAS) rotates the sample at high frequency (14 kHz) at 54.74° with respect to the magnetic field, which reduces the anisotropy and sharpens the resultant signals. Carbon nuclei have less anisotropy in solid samples than protons, and CP-MAS NMR is more often used for obtaining ^{13}C NMR spectra in solid samples. Comparisons of ^{13}C NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in solution ($CDCl_3$) and in crystalline form are shown in Figure 2.21. Resonance

assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester are shown in Table 2.6, based on correlations observed in solution ^{1}H-^{13}C HMBC NMR spectra (Figure 2.22). *p*-Thiocresol was also characterized using CP-MAS ^{13}C NMR, and the comparisons of ^{13}C NMR spectra for *p*-thiocresol in solution ($CDCl_3$) and in crystalline form are shown in Figure 2.23. Resonance assignments for *p*-thiocresol were based on correlations observed in solution ^{1}H-^{13}C HSQC NMR spectra and ^{1}H-^{13}C HMBC NMR spectra (Figure 2.24, Figure 2.25, and Table 2.7, respectively).

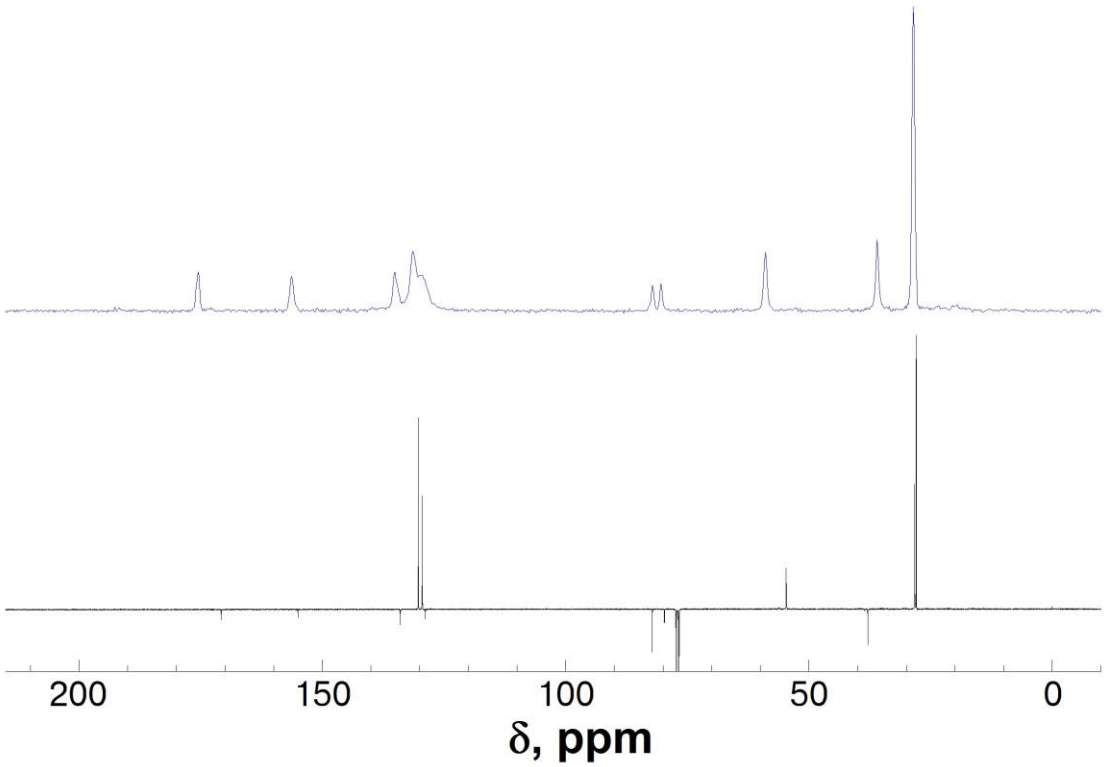


Figure 2.21 ^{13}C NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in crystals and in $CDCl_3$
^{13}C NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester via CP-MAS ^{13}C NMR of the crystalline sample (top, blue) and in $CDCl_3$ (bottom, black).

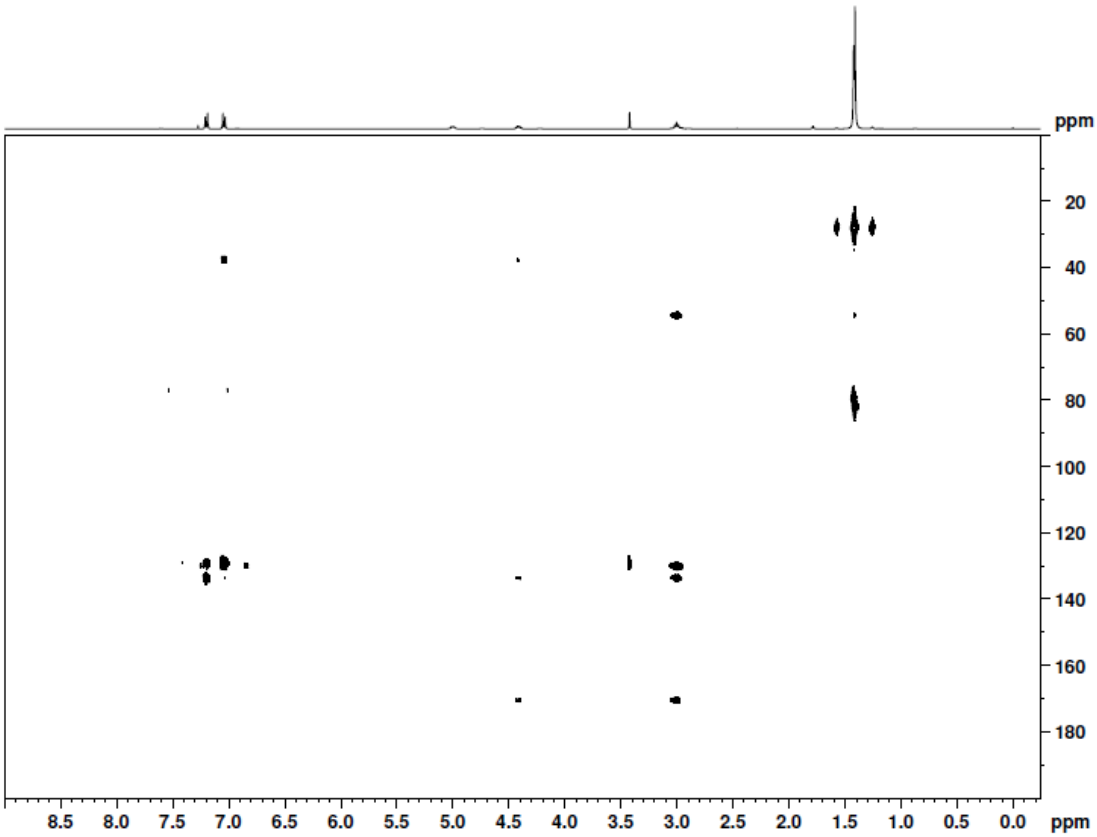
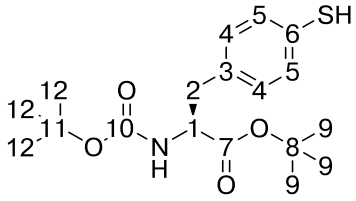


Figure 2.22 ^{1}H-^{13}C HMBC NMR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in $CDCl_3$

^{1}H-^{13}C HMBC NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as the fully reduced thiol in $CDCl_3$. The Hβ protons and thiol proton have clear correlations in the aromatic region, which allowed for assignment of each aromatic carbon and aromatic proton.

Table 2.6. ^{13}C Chemical shift assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

^{13}C chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HMBC data, and chemical shifts in solid state were assigned based on ^{13}C-^{1}H HETCOR experiments.



C#	Carbon	$CDCl_3$ δ, ppm	Solid State δ, ppm	Δδ, ppm
C7	Carbonyl, *tert*-butyl ester	170.8	175.4	4.7
C10	Carbonyl, Boc	155.0	156.3	1.3
C3	C_{Aro}, ipso	134.0	135.1	1.1
C4	C_{Aro}, ortho	130.3	131.4	1.2
C5	C_{Aro}, meta	129.5	131.4	1.9
C6	C_{Aro}, para	128.9	129.6	0.8
C8	C, *tert*-butyl ester	82.2	82.2	0.0
C11	C, Boc	79.7	80.4	0.7
C1	Cα	54.7	58.9	4.3
C2	Cβ	37.9	36.0	–1.9
C12	CH_3, Boc	28.3	28.6	0.3
C9	CH_3, *tert*-butyl ester	28.0	28.6	0.6

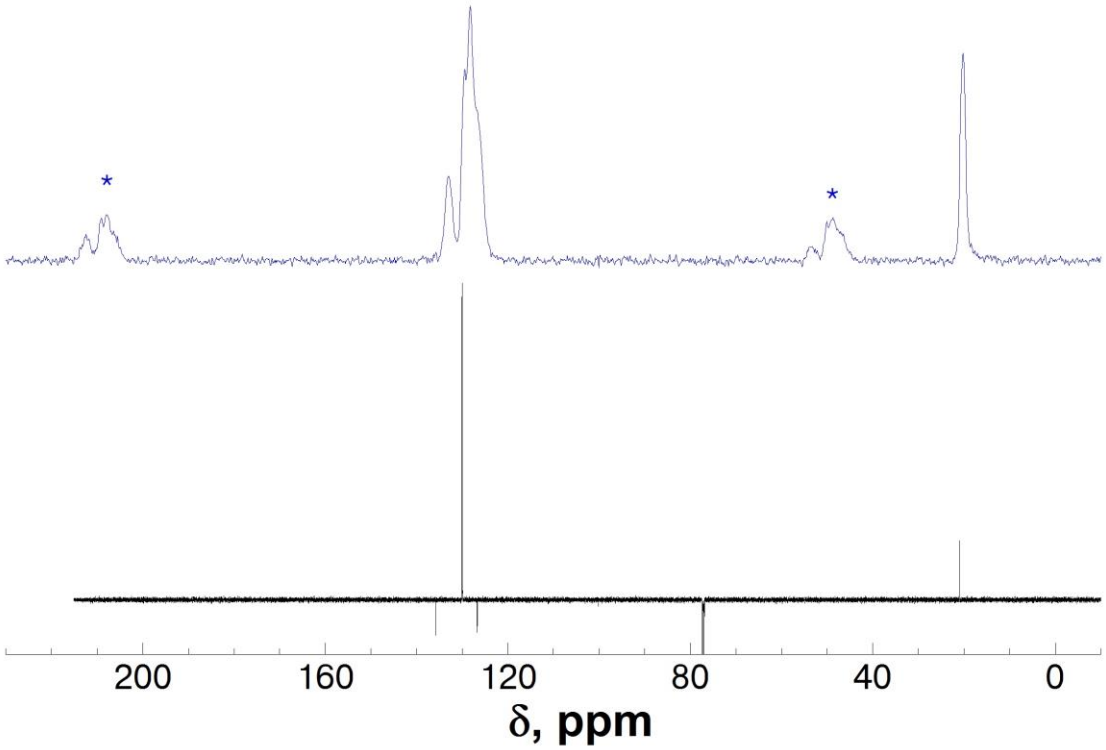


Figure 2.23 ^{13}C NMR spectra of *p*-thiocresol in $CDCl_3$ and in crystals
^{13}C NMR spectra of *p*-thiocresol via CP-MAS ^{13}C NMR of the crystalline sample (top, blue) and in $CDCl_3$ (bottom, black).
*denotes spinning side-bands of the aromatic signals.

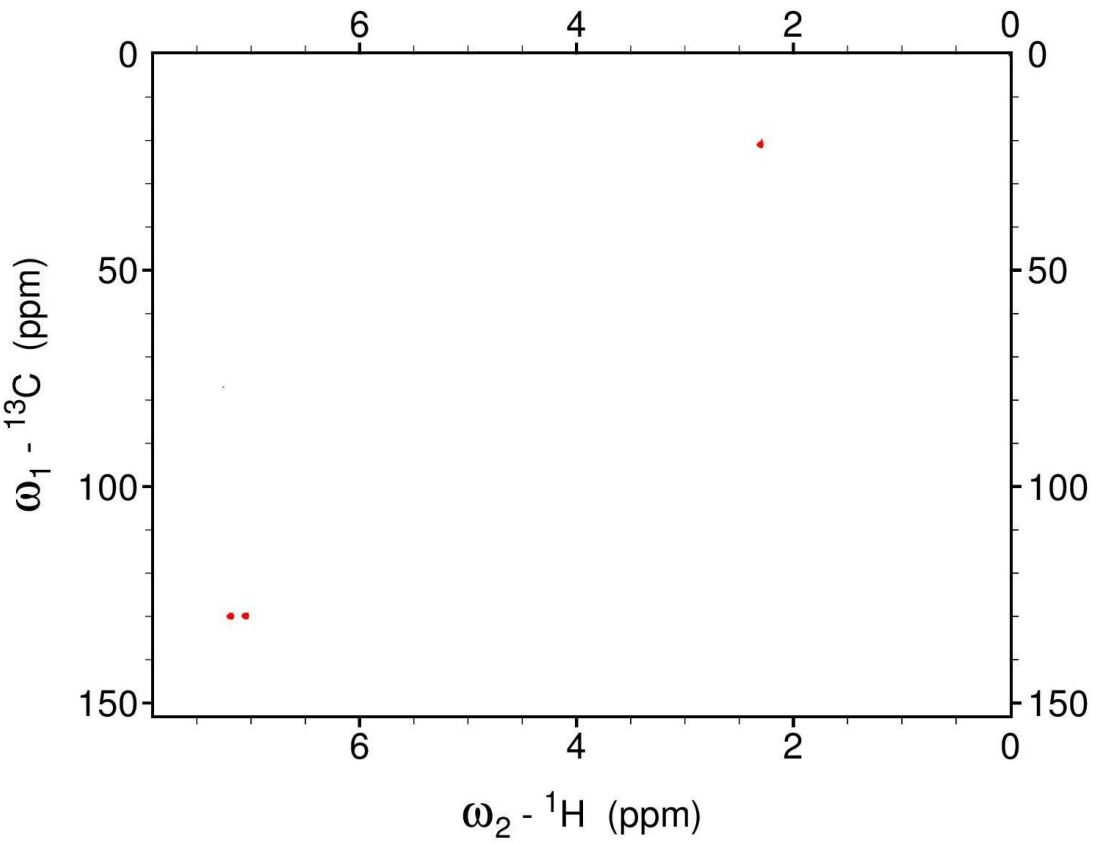


Figure 2.24 ^{1}H-^{13}C HSQC NMR spectrum of *p*-thiocresol in $CDCl_3$
^{1}H-^{13}C HMBC NMR spectrum of *p*-thiocresol in $CDCl_3$.

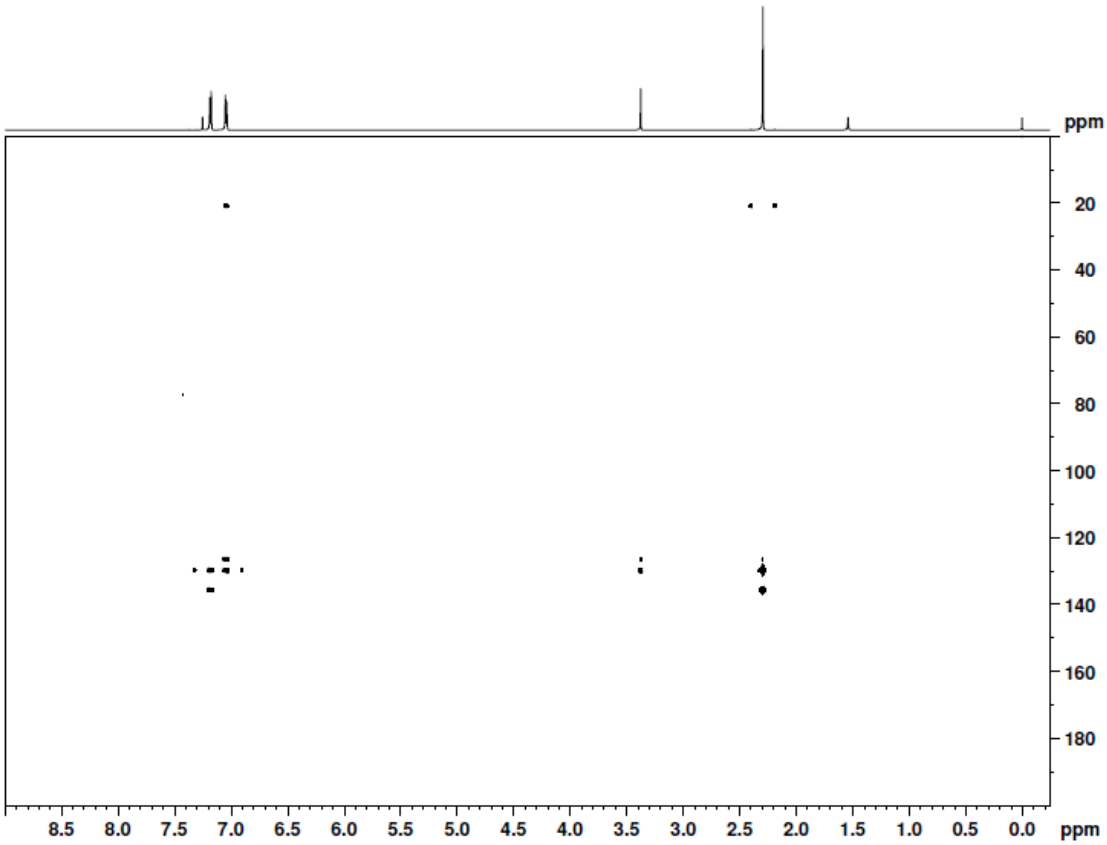
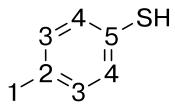


Figure 2.25 ^{1}H-^{13}C HMBC NMR spectrum of *p*-thiocresol in $CDCl_3$
^{1}H-^{13}C HMBC NMR spectrum of *p*-thiocresol in $CDCl_3$. The methyl protons and thiol proton have clear correlations in the aromatic region, which allowed for assignment of each aromatic carbon and aromatic proton.

Table 2.7. ^{13}C Chemical shift assignments for *p*-thiocresol in solution and solid samples

^{13}C chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC data. ^{13}C chemical shifts in crystalline material were assigned based on ^{13}C-^{1}H-HETCOR spectra.



C#	Carbon	$CDCl_3$ δ, ppm	Solid State δ, ppm	Δδ, ppm
C2	C_{Aro}, ipso	135.6	133.0	2.6
C3	C_{Aro}, ortho	129.9	129.5	0.4
C4	C_{Aro}, meta	129.8	128.3	1.5
C5	C_{Aro}, para	126.6	126.8	–0.2
C1	CH_3	20.9	20.2	0.7

One of the most significant differences in the ^{13}C NMR spectra in comparing the solution and crystalline forms of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was the chemical shift for the carbonyl carbon in the *tert*-butyl ester group ($\Delta\delta$ = +4.7 ppm). This chemical shift change was likely a result of the hydrogen-bonding interaction with the carbamate N–H, consistent with observations in the crystal structure. Changes in the ^{13}C chemical shifts were also evident in the aromatic carbons, to a lesser extent ($\Delta\delta$ 0.8-1.9 ppm), but the assignments in the crystalline sample for the aromatic carbons were not as reliable due to lower resolution, signal broadening, and spectral overlap. Interestingly, the ^{13}C chemical shifts for Cα and Cβ changed significantly ($\Delta\delta$ = 4.3 ppm and –1.9 ppm, respectively). The reason for this change is less clear, as the crystal structure does not provide evidence for any particular interactions that would affect the shielding for these carbon atoms. Potentially, the chemical shift changes are as a result of χ_1 conformational restriction within the crystal structure, in comparison to free rotation about χ_1 in solution.

Similar chemical shift changes were observed in the aromatic region of *p*-thiocresol that were observed in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. The C_{ispo} ^{13}C chemical shift was shifted downfield to a greater extent with *p*-thiocresol than for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester ($\Delta\delta$ 1.1 ppm vs 2.6 ppm, respectively). As noted, the spectral overlap and lower resolution of solid-state NMR spectra rendered the chemical shift assignments in the crystalline sample somewhat less reliable, and care must be taken in analysis of these signals.

As described previously, obtaining the ^{1}H NMR spectrum by solid-state NMR generally produces broad, poorly interpretable spectra. The use of Lee-Goldberg ^{13}C-^{1}H HETCOR solid-state NMR[249] can provide reasonable assignments in the ^{1}H NMR

via a 2D NMR experiment. This HETCOR experiment reveals proton signals indirectly through the carbon signals, which allows for improved resolution of the proton resonances in solid samples. The intensity of the correlations depends on the distance between the coupled atoms, rather than through-bond correlations. ^{13}C-^{1}H HETCOR experiments were conducted on solid-state samples of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester with mixing times of 300 μs or 1000 μs (Figure 2.26 and Figure 2.27, respectively). A HETCOR spectrum was also collected for *p*-thiocresol with a 300 μs mixing time (Figure 2.28). The chemical shift assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol for samples in $CDCl_3$ and in crystalline samples are given in Tables 2.8 and 2.9, respectively.

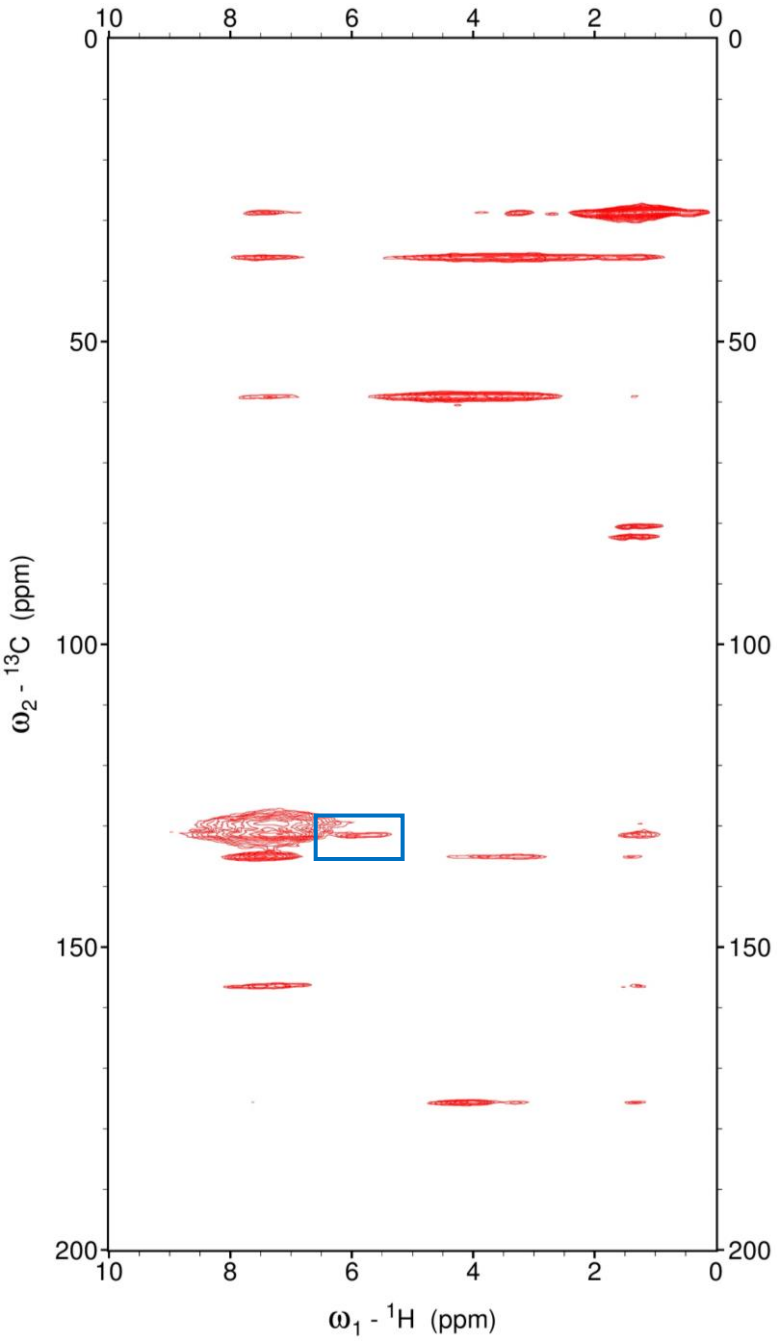


Figure 2.26 ^{13}C-^{1}H-HETCOR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in crystal form

^{13}C-^{1}H-HETCOR spectrum of the crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was obtained via Lee-Goldberg CP-MAS solid-state NMR with a spinning rate of 14 kHz. The mixing time was 300 μs.

The downfield thiol proton, correlating with an aromatic carbon, is boxed.

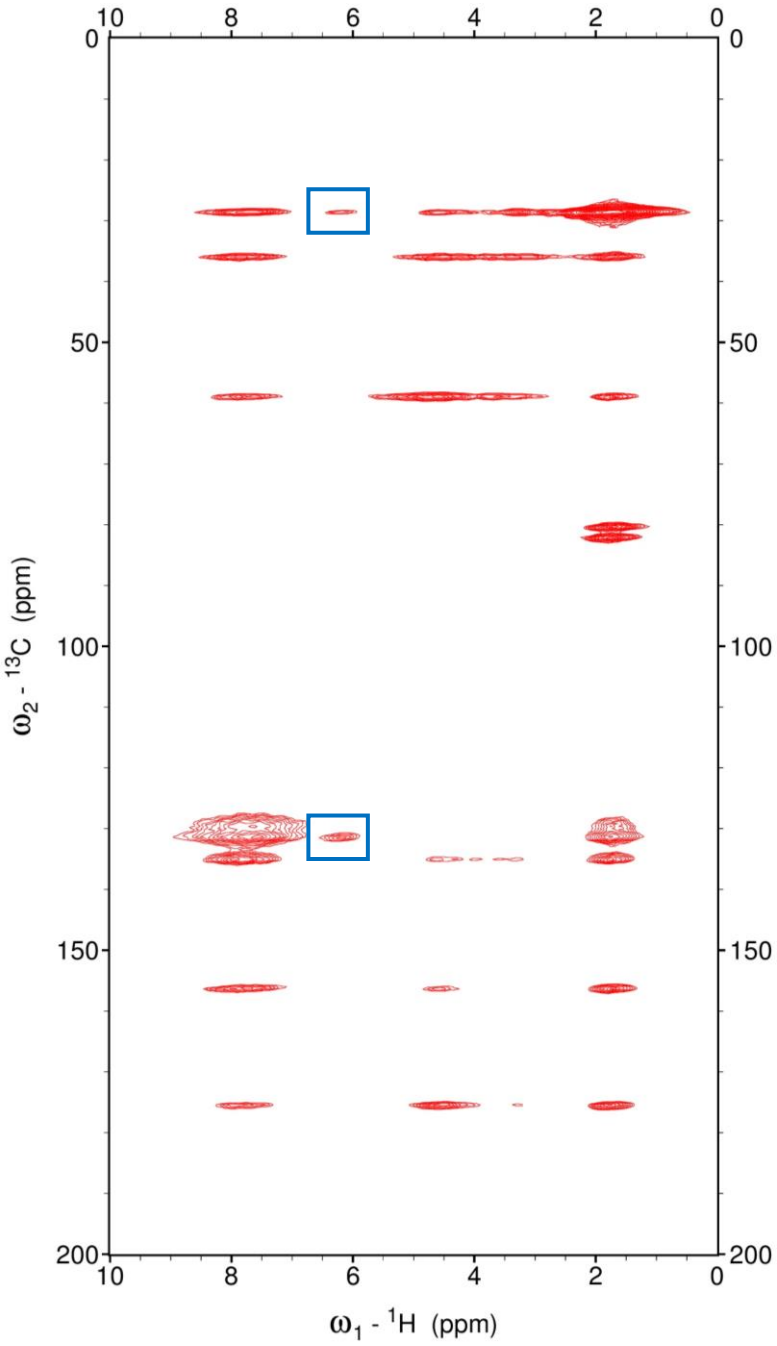


Figure 2.27 ^{13}C-^{1}H-HETCOR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in crystal form

^{13}C-^{1}H-HETCOR spectrum of the crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was obtained via Lee-Goldberg CP-MAS solid-state NMR with a spinning rate of 14 kHz. The mixing time was 1,000 μs. The downfield thiol proton, correlating with an aromatic carbon and to a *tert*-butyl methyl group, is boxed.

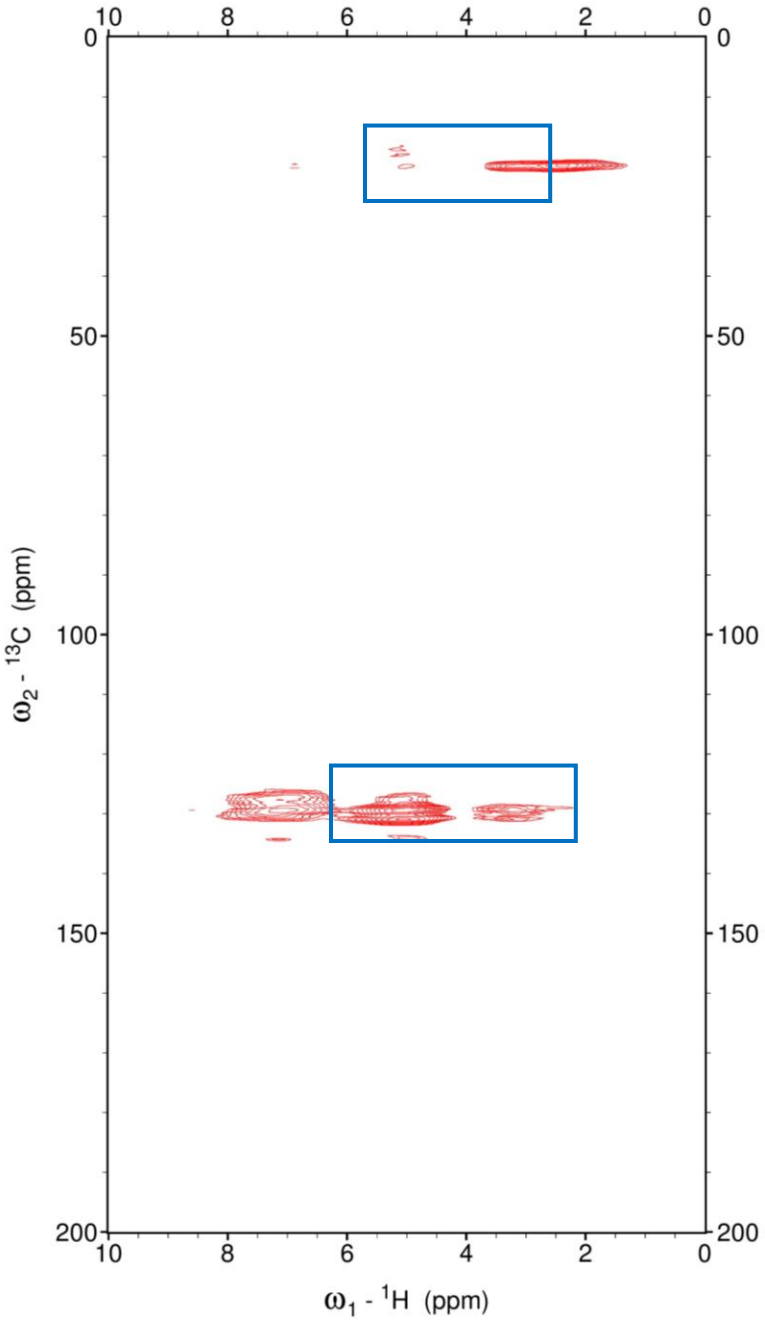
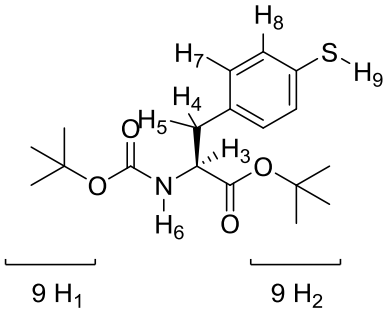


Figure 2.28 ^{13}C-^{1}H-HETCOR spectrum of *p*-thiocresol in crystal form
^{13}C-^{1}H-HETCOR spectrum of the crystalline *p*-thiocresol was obtained via Lee-Goldberg CP-MAS solid-state NMR with a spinning rate of 14 kHz. The mixing time was 300 μs. The two thiol protons, interacting with both the aromatic carbons and the methyl carbon, are boxed.

Table 2.8. ^{1}H chemical shift assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in solution and crystalline form

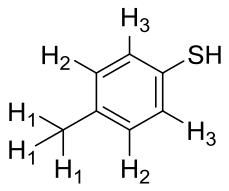
^{1}H chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HMBC data (Figure 2.22), and ^{1}H chemical shifts in the solid sample were taken directly from solid state ^{13}C-^{1}H-HETCOR data. The $\Delta\delta$ with respect to the resonances in $CDCl_3$ were calculated.



H#	Group	$CDCl_3$ δ, ppm	Solid-state δ, ppm	Δδ, ppm
1	*tert*-butyl methyl	1.4	1.7	0.3
2	*tert*-butyl methyl	1.4	1.7	0.3
3	Hα	4.4	4.6	0.2
4	Hβ1	3.0	3.3	0.3
6	Carbamate	5.0	7.4	2.4
7	C_{Aro}, ortho	7.0	7.6	0.6
8	C_{Aro}, meta	7.2	7.6	0.4
9	Thiol	3.4	6.2	2.8

Table 2.9. ^{1}H chemical shift assignments for *p*-thiocresol in ester in solution and crystalline form

^{1}H chemical shifts in $CDCl_3$ were assigned based on ^{1}H-^{13}C-HMBC data (Figure 2.25), and ^{1}H chemical shifts in the solid sample were taken directly from solid state ^{13}C-^{1}H-HETCOR data. The Δδ with respect to the resonances in $CDCl_3$ were calculated.



H#	Group	$CDCl_3$ δ, ppm	Solid-state δ, ppm	Δδ, ppm
H1	Methyl	2.3	2.4	0.1
H2	C_{Aro}, ortho	7.2	7.1	–0.1
H3	C_{Aro}, meta	7.0	7.1	0.1
SH	Thiol	3.4	5.0, 3.2	1.6,–0.2

Careful examination of both HETCOR spectra of the crystalline Boc-4-thiol-L-phenyalanine-*tert*-butyl ester showed one peak at 6.2 ppm that was assigned to the thiol proton, significantly shifted downfield in comparison to the sample in chloroform (Δδ = +2.8 ppm). The carbamate proton was also significantly shifted downfield in the crystalline sample (Δδ = +2.4 ppm compared to $CDCl_3$), consistent with the hydrogen bonding interaction observed in the crystal structure of Boc-4-thiol-L-phenyalanine-*tert*-butyl ester. Other protons were only marginally shifted in crystalline Boc-4-thiol-L-phenyalanine-*tert*-butyl ester based on the HETCOR spectrum.

Based on the HETCOR spectra, there appeared to be two different thiol hydrogen environments in crystalline *p*-thiocresol: one that interacts more strongly with the aromatic ring at 5.0 ppm (comparable to the resonance at 6.2 ppm Boc-4-

thiol-L-phenyalanine-*tert*-butyl ester), and one that does not interact as strongly with the aromatic ring at 3.2 ppm. The two apparent thiol resonances, representing the "interacting" and "non-interacting" forms of the thiol proton in *p*-thiocresol, are observed in the HETCOR spectrum ($\Delta\delta$ = 1.8 ppm).

2.2.4 FT-IR Spectroscopy of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol

Vibrational spectroscopy, including FT-IR spectroscopy, characterizes the bending and stretching of chemical bonds in molecules. The fundamental thiol S–H stretching frequency ($\nu_{S–H}$) is generally 2600-2550 cm^{-1}, depending the local environment and nature of thiol S–H bond interactions.[203, 232] Few other vibrational frequencies are observed near the thiol frequency range, and so the $\nu_{S–H}$, although typically weak, can be distinguished from most other functional groups. The $\nu_{S–H}$ is dependent on the concentration of the thiol compound in solution, as a result of self-association either by S–H/π interactions or by S–H$\cdot\cdot$S hydrogen bonding interactions at higher concentrations.[230, 232] The $\nu_{S–H}$ is also temperature dependent[230] and solvent dependent, where polar protic or aromatic solvents cause a red shift in the $\nu_{S–H}$ due to solvent interactions.[231, 233] As the $\nu_{S–H}$ decreases in frequency, the signal also typically sharpens and increases in intensity.[232, 233]

Saggu *et al.*[233] have previously examined the effect of aromatic solvents on the $\nu_{S–H}$ of thiophenol using 12 different benzene derivatives with a range of electron-donating substituents. In this work, the $\nu_{S–H}$ of thiophenol red shifted in a predictable manner with increasingly electron-rich aromatic rings. The greatest red shift in $\nu_{S–H}$ of thiophenol was in the presence of hexamethylbenzene, with a $\Delta\nu_{S–H}$ of –40 cm^{-1} relative to CCl_4.[233] In addition, the intensity of the $\nu_{S–H}$ signal increased with lower

frequencies. The work by Saggu *et al.*[233] demonstrates that the $\nu_{S–H}$ is strongly dependent on the aromatic electronic effects within the interacting aromatic ring. FT-IR spectroscopy can also be used to examine S–H/π aromatic interactions in solid samples. Rozenberg *et al.*[234] compared $\nu_{S–H}$ in microcrystalline and solution samples of N-(o-hydroxyphenyl)-3-sulfanylmethylpyrrolidin-2-one, and found a substantial red shift (–19 cm^{-1}) in the $\Delta\nu_{S–H}$ between solid and solution forms, consistent with an S–H/π aromatic interaction observed in the crystal structure. In a silylthiol compound with a crystallographically observed S–H/π interaction, Jabłońska *et al.*[250] observed two distinct $\nu_{S–H}$ bands in the solid-state FT-IR, separated by 31 cm^{-1}, which was attributed to "interacting" and "non-interacting" forms of the the thiol group.

Given the observations in the crystal structure and solid-state NMR spectra of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, it was expected that the $\nu_{S–H}$ should exhibit a red shift in comparison of the crystalline sample with a solution sample. Analysis of solution and solid samples by FT-IR requires little additional equipment, unlike solution and solid NMR analysis (Chapter 2.2.3). In addition, the magnitude of the $\nu_{S–H}$ red shift can provide a measure of interaction strength, both through comparison to calculated values and in comparison to numerous examples of thiol FT-IR studies in the literature.[230-234]

The FT-IR spectra were taken of crystalline samples of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as a KBr pellet, and compared to solution samples of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. A series of solvents was included in this study, in order to examine the effect of different hydrogen bonding partners on the $\nu_{S–H}$ red shift and signal intensity. CCl_4 and $CHCl_3$ have limited interactions with thiols, and served as the bases for comparison to other solvents and co-solvents. Thiol

interactions with carbonyl and esters were examined (acetone and ethyl acetate), as well as cyclic ethers (THF) and alcohols (methanol). Solvents were prepared as 10% and 25% cosolvents in CCl_4, or neat (for ethyl acetate and acetone). *p*-Thiocresol was also examined by FT-IT with the same solvents. All solution samples were taken with three independent trials and averaged after baseline correction using solutions of 200 mM Boc-4-thiol-L-phenylalanine-*tert*-butyl ester or *p*-thiocresol.

A summary of the IR data in different solvents, including thiol S–H ν_{max}, absorbance intensity, the signal full-width at half-maximum, and the $\Delta\nu_{S-H}$ red shift, is shown for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (Table 2.10) and *p*-thiocresol (Table 2.11). The absolute absorbance spectra for the thiol S–H stretching frequency in crystalline and in solution (25% cosolvents in CCl_4) forms are shown for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol in Figure 2.29. The spectra for crystalline samples were normalized with respect to the ν_{max} at 2925-2935 cm^{-1} to allow for direct comparison of the absolute intensities (solvent data all were collected at 200 mM sample concentration, but solid samples were based on arbitrary amounts of microcrystals). Absorbance spectra that were normalized with respect to the ν_{S-H} Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol are shown in Figure 2.30. Comparisons of the ν_{S-H} in each cosolvent system at 10% or 25% in CCl_4 (and neat, for acetone and ethyl acetate) for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol are shown in Figure 2.31 and Figure 2.32, respectively.

Table 2.10. IR data for the S–H stretching frequency in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

The ν_{max}, absorbance intensity, and full-width at half-maximum (FWHM) of the S–H stretching frequencies, and the $\Delta\nu_{max}$ relative to ν_{max} in $CHCl_3$ were obtained from the IR spectra shown in Figure 2.31. Spectra were collected with 1 cm^{-1} resolution, and were background-subtracted, baseline-corrected, and averaged over at least three independent trials (except for the crystalline data). Errors in absorbance were less than 10% unless indicated otherwise (shown in parentheses).
[a]Absolute intensity was normalized with respect to the ν_{max} at 2925-2935 cm^{-1} in 25% acetone/CCl_4. This approach was employed because the signal in the crystalline material was based on an arbitrary amount of solid, and can not be directly compared to solution samples (all at 200 mM). Normalization was conducted based on the C–H stretching frequency due to its similar signal intensity in all solution samples. This normalization allows direct comparison of the intensities of ν_{S-H} in solution and in solid state. The absolute absorbance for the sample is shown in brackets.
"n.d." was not determined due to low solubility
"br" indicates a broad signal either with a FWHM greater than 80 cm^{-1} or two apparently overlapping signals.

	ν_{max}, cm^{-1}	Intensity, Absorbance	FWHM, cm^{-1}	$\Delta\nu_{max}$, cm^{-1}
Crystalline	2538	0.2315[a] [0.0956]	18	–47
CCl_4	n.d	n.d.	n.d.	n.d
$CHCl_3$	2585	0.0037	21	0
Ethyl Acetate	2566	0.0195	33	–19
25% Ethyl Acetate in CCl_4	2567	0.0084	37	–18
10% Ethyl Acetate in CCl_4	2569	0.0033 (±0.0018)	52	–16
Acetone	2558	0.0200	55	–27
25% Acetone in CCl_4	2559	0.0088 (±0.0019)	49	–26
10% Acetone in CCl_4	2559	0.0052 (±0.0018)	57	–26
25% MeOH in CCl_4	2541	0.0045	br[c]	–44
	2497	0.0050		–88
10% MeOH in CCl_4	2554	0.0025	br[c]	–31
	2499	0.0026		–86
25% THF in CCl_4	2534	0.0057	75	–51
10% THF in CCl_4	2541	0.0034	br[c]	–44

Table 2.11. IR data for the S–H stretching frequency in *p*-thiocresol

The ν_{max}, absorbance intensity, and full-width at half-maximum (FWHM) of the S–H stretching frequencies, and the $\Delta\nu_{max}$ relative to ν_{max} in CCl_4 were obtained from the IR spectra shown in Figure 2.32. Spectra were collected with 1 cm^{-1} resolution, and were background-subtracted, baseline-corrected, and averaged over at least three independent trials (except for the crystalline data). Errors in absorbance were less than 10%.

[a]Absolute intensity was normalized with respect to the ν_{max} at 2925-2935 cm^{-1} in 25% acetone/CCl_4. This approach was employed because the signal in the crystalline material was based on an arbitrary amount of solid, and can not be directly compared to solution samples (all at 200 mM). Normalization was conducted based on the C–H stretching frequency due to its similar signal intensity in all solution samples. This normalization allows direct comparison of the intensities of ν_{S-H} in solution and in solid state. The absolute absorbance for the sample is shown in brackets.

"n.d." was not determined due to low solubility;

"br" indicates a broad signal either with a FWHM greater than 80 cm^{-1} or two apparently overlapping signals.

	ν_{max}, cm^{-1}	Intensity, Absorbance	FWHM, cm^{-1}	$\Delta\nu_{max}$, cm^{-1}
Crystalline	2563	0.0235[a] [0.0263]	18	–23
CCl_4	2586	0.0022	25	0
$CHCl_3$	2585	0.0051	21	–1
Ethyl Acetate	2567	0.0211	33	–19
25% Ethyl Acetate in CCl_4	2570	0.0093	35	–16
10% Ethyl Acetate in CCl_4	2571	0.0046	39	–15
Acetone	2558	0.0267	55	–28
25% Acetone in CCl_4	2561	0.0105	52	–25
10% Acetone in CCl_4	2565	0.0055	55	–21
25% MeOH in CCl_4	2521	0.0062	br[c]	–65
10% MeOH in CCl_4	2584	0.0027	br[c]	–2
	2548	0.0023		–38
	2518	0.0021		–21
25% THF in CCl_4	2530	0.0082	69	–56
10% THF in CCl_4	2531	0.0036	br[c]	–55

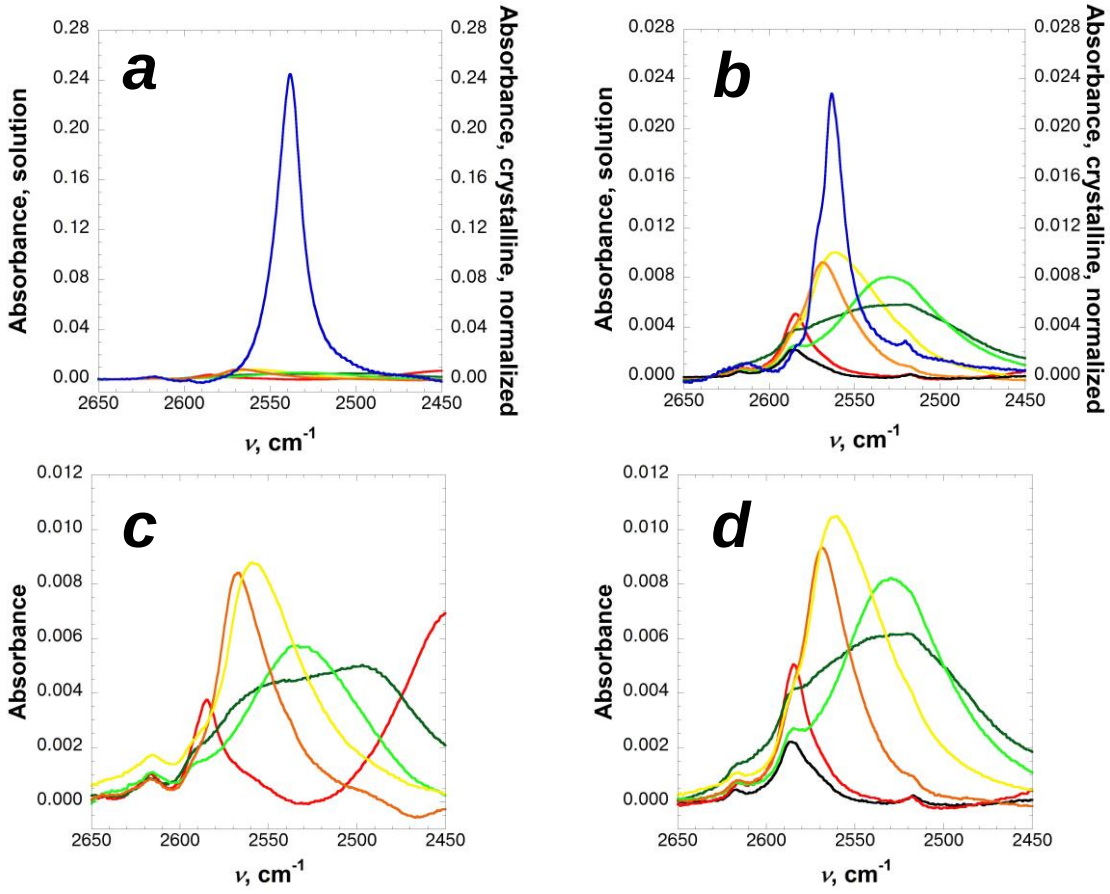


Figure 2.29 Comparison of the IR spectra of the S–H stretching frequency of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol.
In the crystalline data, the ν_{max} near 2928 cm^{-1} was normalized to the same absorbance as 25% acetone in CCl_4. The S–H stretching frequencies in solid and solution samples of (a, c) Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and (b, d) *p*-thiocresol. Charts c and d have the crystalline spectra omitted for clarity. Crystalline material (blue) was analyzed via pressed pellet method in anhydrous KBr. Solutions were prepared as 200 mM solutions of compounds in the specified solvent (CCl_4, black; $CHCl_3$, red; 25% ethyl acetate/CCl_4, orange; 25% acetone/CCl_4, yellow; 25% THF/CCl_4, light green; 25% methanol/CCl_4, dark green). Spectra were collected with 1 cm^{-1} resolution, and were background subtracted, baseline corrected, and averaged over at least three independent trials (except for the crystalline data). Additional data at different concentrations of cosolvents are shown in Figure 2.31, and are included in the appendix.

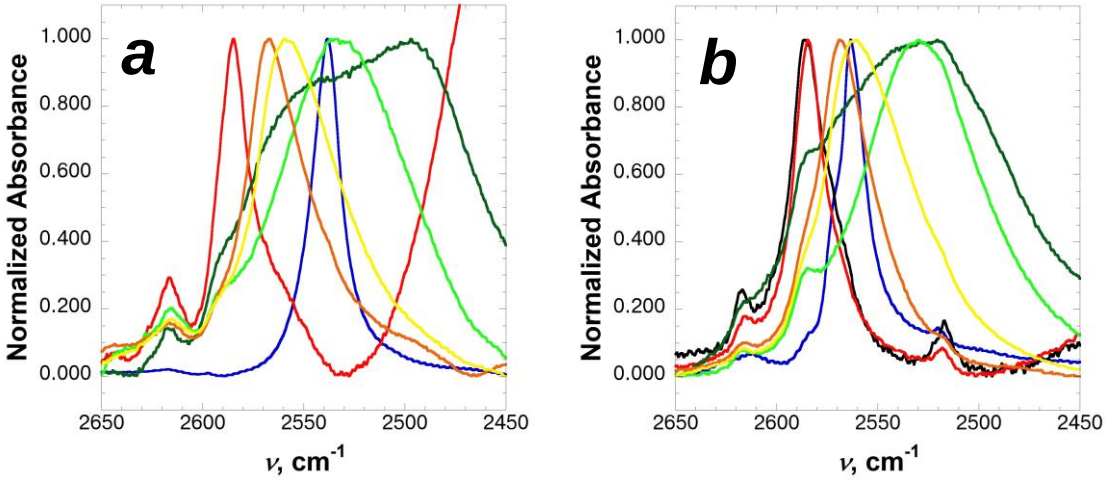


Figure 2.30 Normalized IR spectra of the S–H stretching frequency of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol.
The normalized absorbance of the S–H stretching frequency for (a) Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and (b) *p*-thiocresol. Crystalline material (blue, b only) was analyzed via pressed pellet method in anhydrous KBr. Solutions were prepared as 200 mM concentrations of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in the specified solvent ($CHCl_3$, red; 25% ethyl acetate/CCl_4, orange; 25% acetone/CCl_4, yellow; 25% THF/CCl_4, light green; 25% methanol/CCl_4, dark green). Spectra were collected with 1 cm^{-1} resolution, and were background subtracted, baseline corrected, and averaged over at least three independent trials (except for the crystalline data). Additional data at different concentrations of cosolvents are shown in Figure 2.31, and are included in the appendix.

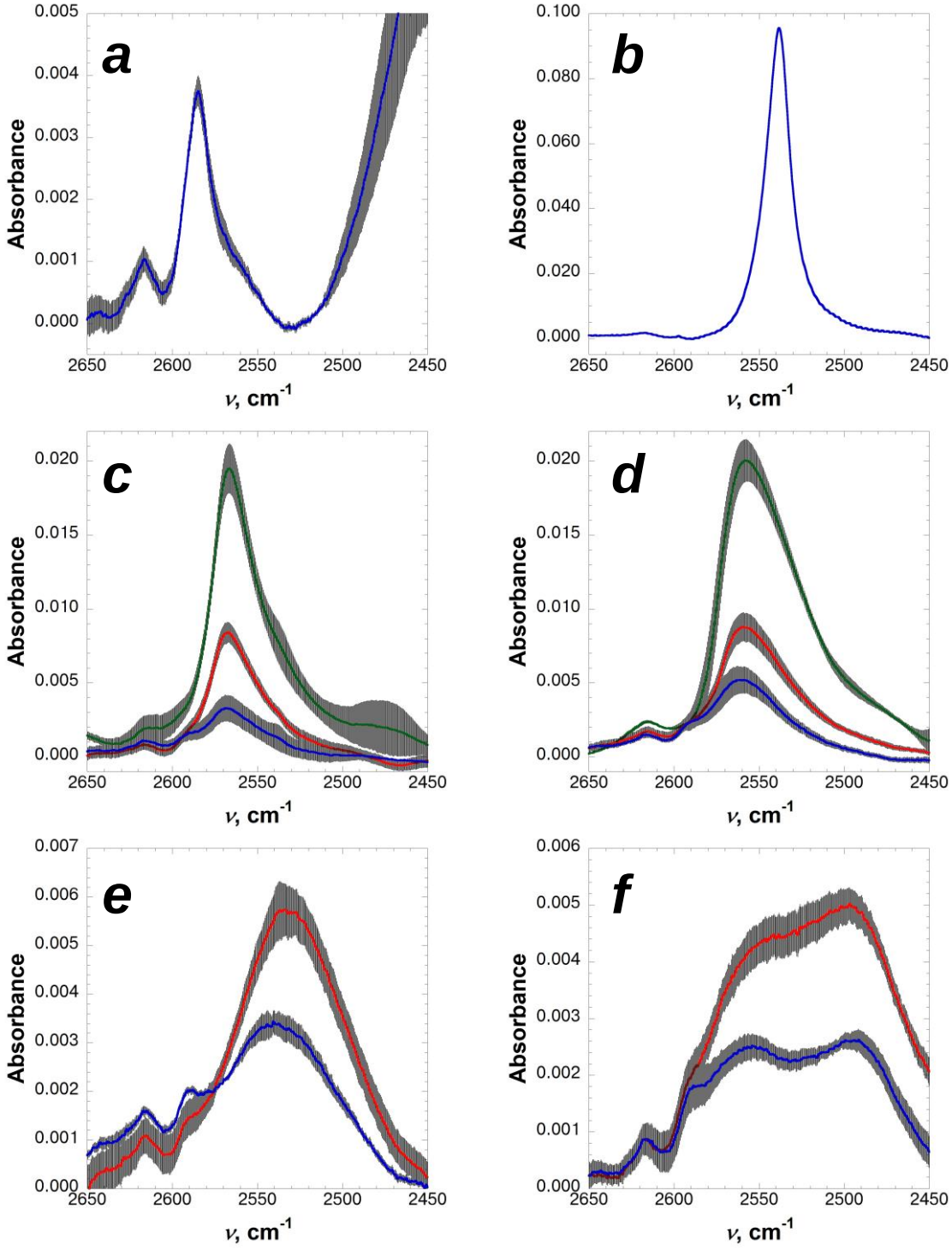
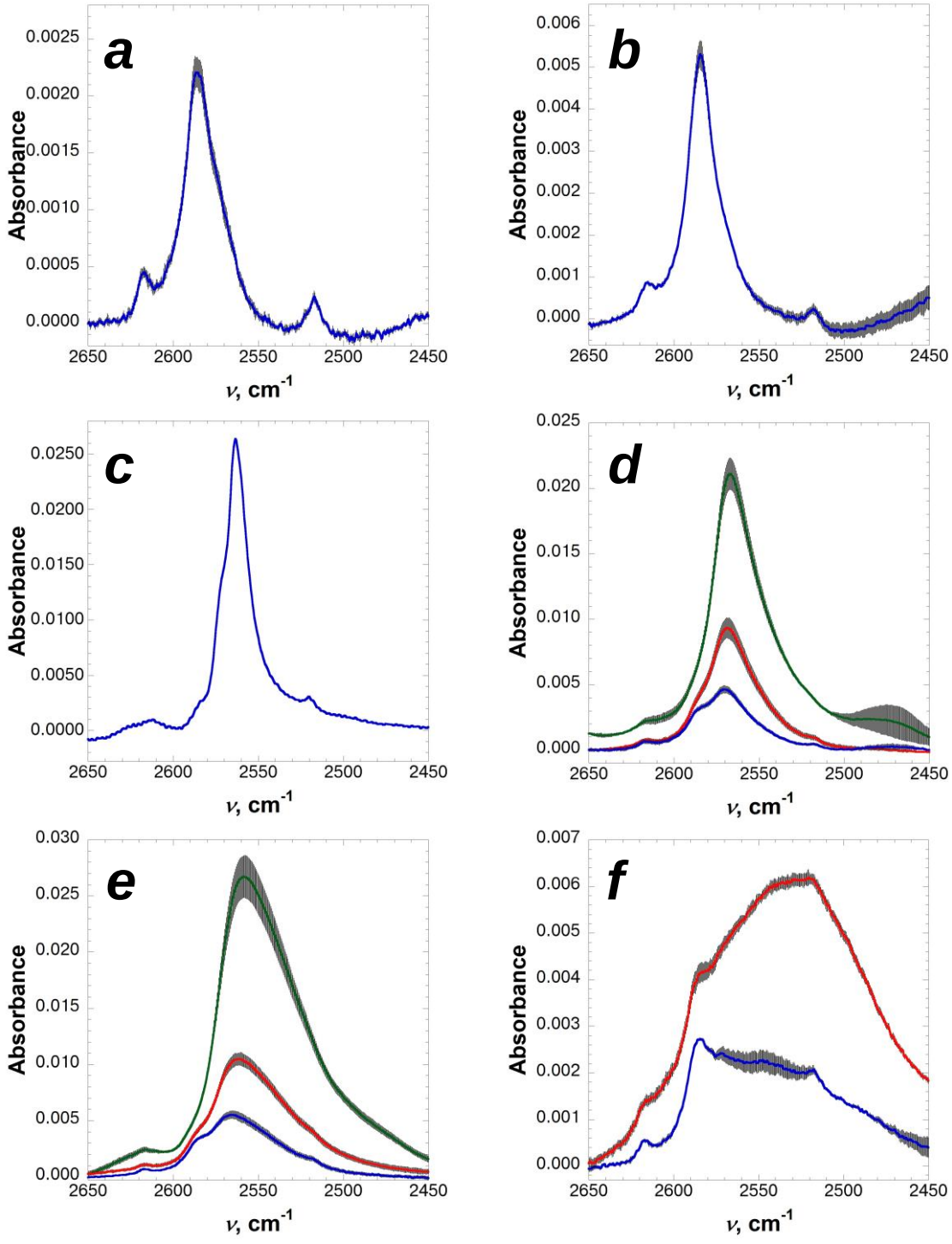


Figure 2.31 Dependence of ν_{S-H} signal intensity on cosolvent concentration for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

FT-IR spectra of the S–H stretching region of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in (a) $CHCl_3$; (b) crystalline form; (c) ethyl acetate in CCl_4 (green, 100% ethyl acetate; red, 25% ethyl acetate; blue, 10% ethyl acetate); (d) acetone in CCl_4 (green, 100% acetone; red, 25% acetone; blue, 10% acetone); (e) methanol in CCl_4 (red, 25% methanol; blue, 10% methanol); or (e) THF in CCl_4 (red, 25% THF; blue, 10% THF). Error bars indicate standard error.


a
b
c
d
e
f
Absorbance
ν, cm-1
2650
2600
2550
2500
2450

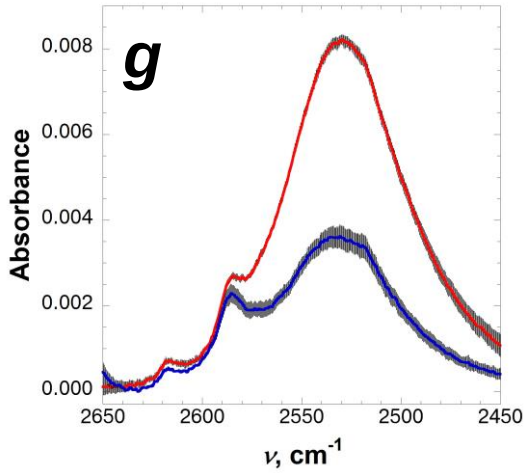


Figure 2.32 Dependence of $\nu_{S–H}$ signal intensity on cosolvent concentration for *p*-thiocresol

FT-IR spectra of the S–H stretching region of *p*-thiocresol (200 mM) in (a) CCl_4; (b) $CHCl_3$; (c) crystalline form; or (d) ethyl acetate in CCl_4 (green, 100% ethyl acetate; red, 25% ethyl acetate; blue, 10% ethyl acetate); (e) acetone in CCl_4 (green, 100% acetone; red, 25% acetone; blue, 10% acetone); (f) methanol in CCl_4 (red, 25% methanol; blue, 10% methanol); or (g) THF in CCl_4 (red, 25% THF; blue, 10% THF). Error bars indicate standard error.

The S–H stretching frequency is dependent upon hydrogen bonding or other non-covalent interactions, either with bulk solvent or as self-association, and is observed via FT-IR as a red shift in the $\nu_{S–H}$.[231] Due to low solubility of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in CCl_4, data were compared to the $\nu_{S–H}$ in $CHCl_3$. Using *p*-thiocresol, the $\nu_{S–H}$ in CCl_4 and $CHCl_3$ were found to be similar ($\Delta\nu_{S–H} = –1$ $cm^{–1}$). In both Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol, solvent dependent red shifts were observed in the $\nu_{S–H}$, as CCl_4 > $CHCl_3$ > ethyl acetate/CCl_4 > acetone/CCl_4 > THF/CCl_4 > methanol/CCl_4. For Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, the $\nu_{S–H}$ was red shifted with the conventional hydrogen bond acceptors, in ethyl acetate/CCl_4 ($\Delta\nu_{S–H} = –16$ to $–19$ $cm^{–1}$) and acetone/CCl_4 ($\Delta\nu_{S–H} = –26$ to $–27$ $cm^{–1}$). The $\nu_{S–H}$ signal intensity also increased for solutions of ethyl acetate or acetone

in CCl_4. The largest red shifts in solution for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were observed in methanol/CCl_4 ($\Delta\nu_{S-H}$ = –31 to –88 cm^{-1}), but these were observed as several, weak and broad signals. Similar broad and weak ν_{S-H} signals were observed for solutions of THF/CCl_4 ($\Delta\nu_{S-H}$ = –44 to –51 cm^{-1}). *p*-Thiocresol exhibited similar behavior to Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in all solutions, although the red shift in ν_{S-H} and signal intensities were generally less (Table 2.11). The ν_{S-H} absorbance intensity was also dependent on the amount of interacting cosolvent. The low intensity and width of the ν_{S-H} signals in the presence of methanol and THF suggest a different, disordered interaction with bulk solvent, while conventional hydrogen bond acceptors (carbonyl oxygen atoms) can interact more favorably with aryl thiols.

Strikingly, the $\Delta\nu_{S-H}$ for the crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was –47 cm^{-1}, greater than the observed red shift for *p*-thiocresol ($\Delta\nu_{S-H}$ = –23 cm^{-1}). The observed red shift in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was also greater than thiophenol in hexamethylbenzene ($\Delta\nu_{S-H}$ = –40 cm^{-1}),[233] or the red shift in the crystalline aryl thiol reported by Rozenberg ($\Delta\nu_{S-H}$ = –19 cm^{-1}).[234] The crystalline samples were based on an arbitrary amount of material in a pressed KBr pellet, and the signal intensities cannot be directly compared against the solution data. In order to make appropriate comparisons between the signal intensities in different solvents and the crystalline samples, the data from the crystalline samples were normalized with respect to the ν_{max} near 2928 cm^{-1} in 25% acetone/CCl_4, and these data are shown in Figure 2.30. After normalization, the ν_{S-H} signal intensity for the crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was observed to increase more than 15-fold over the absorbance in $CHCl_3$, while the ν_{S-H} absorbance in the

crystalline *p*-thiocresol increased to an intensity similar to that observed in acetone or ethyl acetate. The red shift in the ν_{S-H} for crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was greater than observed shifts in ethyl acetate or acetone, indicating that the thiol interaction with the aromatic ting in the solid state is both stronger and fundamentally distinct from interactions with carbonyl oxygen atoms in solution. The greater red shift and increase in absorbance intensity in the ν_{S-H} for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester over *p*-thiocresol suggests a stronger interaction and a more well-defined crystalline structure, consistent with observations in x-ray crystal structures (Chapter 2.2.2) and NMR spectra (Chapter 2.2.3).

In order to make comparisons for all solution and solid-state data, without normalization, the ratios of the ν_{S-H} against the ν_{max} near 1500 cm^{-1} were compared directly. A ring C–C "semi-circle" stretching vibration (ν_{C-C}) in *p*-disubstituted aromatic rings absorbs in the range 1520-1480 cm^{-1},[208] which should be consistent in intensity and ν_{max} for all solvent mixtures and for crystalline samples. The ν_{max} and absorbance intensity for the ring C–C stretching vibration, and the ratios of ν_{S-H}/ν_{C-C} for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol are shown in Table 2.12 and Table 2.13, respectively.

Table 2.12. Relative absorbance intensities for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

The intensity of the S–H stretching frequency at ν_{max} (2534-2586 cm^{-1}) was compared to the intensity of the aromatic C–C frequency at ν_{max} (1494-1497 cm^{-1}). The ν_{max} and absorbance intensities for $\nu_{S–H}$ and $\nu_{C–C}$ were obtained from the IR spectra shown in Figure 2.31 and data included in the appendix. Spectra were collected with 1 cm^{-1} resolution, and were background-subtracted, baseline-corrected, and averaged over at least three independent trials (except for the crystalline data).
"n.d." was not determined due to low solubility
"br" indicates a broad signal either with a FWHM greater than 80 cm^{-1} or two apparently overlapping signals.

	ν_{max}, cm^{-1}		Absorbance Intensity		S–H Intensity / C–C Intensity
	S–H	C–C	S–H	C–C	
Crystalline	2538	1496	0.0956	0.0886	1.08
CCl_4	n.d.	n.d.	n.d.	n.d.	n.d.
$CHCl_3$	2585	1495	0.0038	0.4859	0.01
Ethyl Acetate	2566	1497	0.0195	0.2908	0.07
25% Ethyl Acetate in CCl_4	2567	1495	0.0084	0.4079	0.02
10% Ethyl Acetate in CCl_4	2569	1494	0.0037	0.3775	0.01
Acetone	2558	1497	0.0200	0.2588	0.08
25% Acetone in CCl_4	2559	1495	0.0088	0.3813	0.02
10% Acetone in CCl_4	2559	1495	0.0052	0.4397	0.01
25% MeOH in CCl_4	2497	1496	0.0050	0.2845	0.02
25% THF in CCl_4	2534	1495	0.0057	0.3694	0.02
10% THF in CCl_4	2541	1494	0.0034	0.4044	0.01

Table 2.13. Relative absorbance intensities for *p*-thiocresol

The intensity of the S–H stretching frequency at ν_{max} (2530-2586 cm^{-1}) was compared to the intensity of the aromatic C–C frequency at ν_{max} (1493-1496 cm^{-1}). The ν_{max} and absorbance intensities for ν_{S-H} and ν_{C-C} were obtained from the IR spectra shown in Figure 2.32 and data included in the appendix. Spectra were collected with 1 cm^{-1} resolution, and were background-subtracted, baseline-corrected, and averaged over at least three independent trials (except for the crystalline data).
"n.d." was not determined due to low solubility
"br" indicates a broad signal either with a FWHM greater than 80 cm^{-1} or two apparently overlapping signals.

	ν_{max}, cm^{-1}		Absorbance Intensity		S–H Intensity / C–C Intensity
	S–H	C–C	S–H	C–C	
Crystalline	2563	1493	0.0263	0.1057	0.25
CCl_4	2586	1496	0.0022	0.1343	0.02
$CHCl_3$	2585	1495	0.0051	0.1826	0.03
Ethyl Acetate	2567	1496	0.0211	0.1368	0.15
25% Ethyl Acetate in CCl_4	2570	1496	0.0093	0.1540	0.06
10% Ethyl Acetate in CCl_4	2571	1496	0.0046	0.1448	0.03
Acetone	2558	1496	0.0267	0.1713	0.16
25% Acetone in CCl_4	2561	1496	0.0105	0.1465	0.07
10% Acetone in CCl_4	2565	1496	0.0055	0.1500	0.04
25% MeOH in CCl_4	2521	1496	0.0062	0.1402	0.04
25% THF in CCl_4	2530	1495	0.0082	0.1357	0.06
10% THF in CCl_4	2531	1496	0.0036	0.1349	0.03

The ν_{C-C} had a slight solvent dependence in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, which was not observed in *p*-thiocresol. In Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, the ν_{C-C} red shifted by up to –3 cm^{-1} in solution, and the crystalline ν_{C-C} blue shifted 20 cm^{-1} relative to $CHCl_3$. For *p*-thiocresol, the ν_{C-C} did not vary significantly in solution, and the crystalline ν_{C-C} red shifted –3 cm^{-1} relative to CCl_4. Shifting in the aromatic C–C "semi-circle" stretching vibration to higher

wavenumbers can be attributed to more stretching or contracting of the aromatic substituents, in this case the thiol C–S bond.[208] Kochi and coworkers[251] noted that large cations positioned over an aromatic ring carbon invoked a "bending" of the aromatic ring as the interacting carbon became more tetrahedral in geometry; the observed blue shift in the ν_{C-C} in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester is potentially suggestive of a distortion of the aromatic ring.

The absorbance intensity for the ν_{C-C} did not vary significantly in *p*-thiocresol, although there was some intensity variation in the ν_{C-C} for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, potentially due to signal mixing with aromatic C–H bonds. Both of the crystalline samples showed a higher ν_{S-H}/ν_{C-C} ratio than solution samples, verifying the large absorbance intensities in the ν_{S-H} for crystalline samples, and confirming the presence of a unique thiol-aromatic interaction.

Taken together, the observation of a lower frequency in the ν_{S-H} suggests a lowering in energy of the thiol bond, consistent with a non-covalent thiol interaction, whether it be hydrogen bonding or an aromatic interaction. Direct comparison of the ν_{S-H} between the crystalline *p*-thiocresol and Boc-4-thiol-L-phenylalanine-*tert*-butyl ester indicates a more pronounced interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, consistent with the observations from the crystal structures and NMR data. The red shift in ν_{S-H} was greater in crystalline samples than for solutions of acetone/CCl_4 of ethyl acetate/CCl_4, suggesting an interaction that is stronger or fundamentally distinct from a conventional thiol-carbonyl hydrogen bond (Figure 2.19), and that a stronger interaction plays a role in the crystal structure. The ν_{S-H} for crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (2538 cm^{-1}) compares to the

ν_{S-H} thiophenol in a solution of hexamethylbenzene/CCl_4 (2549 cm^{-1}),[233] consistent with an intermolecular S–H/π aromatic interaction.

2.2.5 *ab initio* calculations on a minimized structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

Having observed an S–H/π aromatic interaction with carbon-oriented geometry in the x-ray crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, and having observed significant shifts in the thiol proton via ^{1}H NMR and in the $\nu_{S-H,\ stretch}$ via FT-IR, we sought to gain further insight into the energetic contributions that drive this interaction via *ab initio* calculations. Thiol S–H/π aromatic interactions have been previously examined through DFT calculations and natural bond orbital (NBO) analysis, but these calculations have primarily focused on hydrogen sulfide, which can interact with aromatic rings in a bidentate manner, and inherently will interact with a different geometry than an thiol (R–S–H, R ≠ H).[164, 252] Prior theoretical studies have examined the different binding modes of sulfur (which can act as either a donor or acceptor),[253, 254] or have corroborated calculations with experimental IR data.[234, 250, 255] Biswal & Wategaonkar[164] calculated the interactions of H_2S with indole rings, and found that S–H/π interactions were preferred over the available N–H/S interactions. In addition, it found that S–H/π interactions were stronger than comparable X–H/π interactions, and attributed the increased strength to a greater dispersion component.[164] In fact, the term "dispersion-stabilized hydrogen bonds" was suggested to denote hydrogen bonds that are stabilized by dispersive forces as opposed to electrostatics, as the traditional term for a hydrogen bond implies.[164]

In this study on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, the underlying electrostatic and favorable orbital overlap contributions to stabilizing the

intermolecular S–H/π aromatic interaction can be determined via NBO analysis.[255, 256] Experimental S–H stretching frequencies in crystalline and solution samples of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester can be compared to calculated ν_{S-H} frequencies for given S–H bond lengths, which can provide an approximation for the S–H bond length in the x-ray crystal structure (see Chapter 2.2.2 regarding determination of thiol bond length in x-ray crystal structures.

With the aid of our collaborators, Dr. Sudipta Sinha, Dr. Shi Bai, and Dr. Sandeep Patel, the energies of the thiol S–H were calculated and compared with the IR stretching frequencies that were measured experimentally. In order to identify the key energies that contribute to stabilization of the S–H/π aromatic interaction, and to reduce computational time, the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was truncated for calculations, as shown in Figure 2.33. In the truncated dimer, there are two different thiol groups represented: one which actively interacts with the aromatic ring, and one that does not participate in any aromatic interaction. Comparing the calculated results of these two different thiol groups provides insights into the energetic components that drive S–H/π aromatic interactions. Furthermore, the charges for each atom in this dimer were calculated, as well as the natural bond orbitals (NBO) and associated energies.

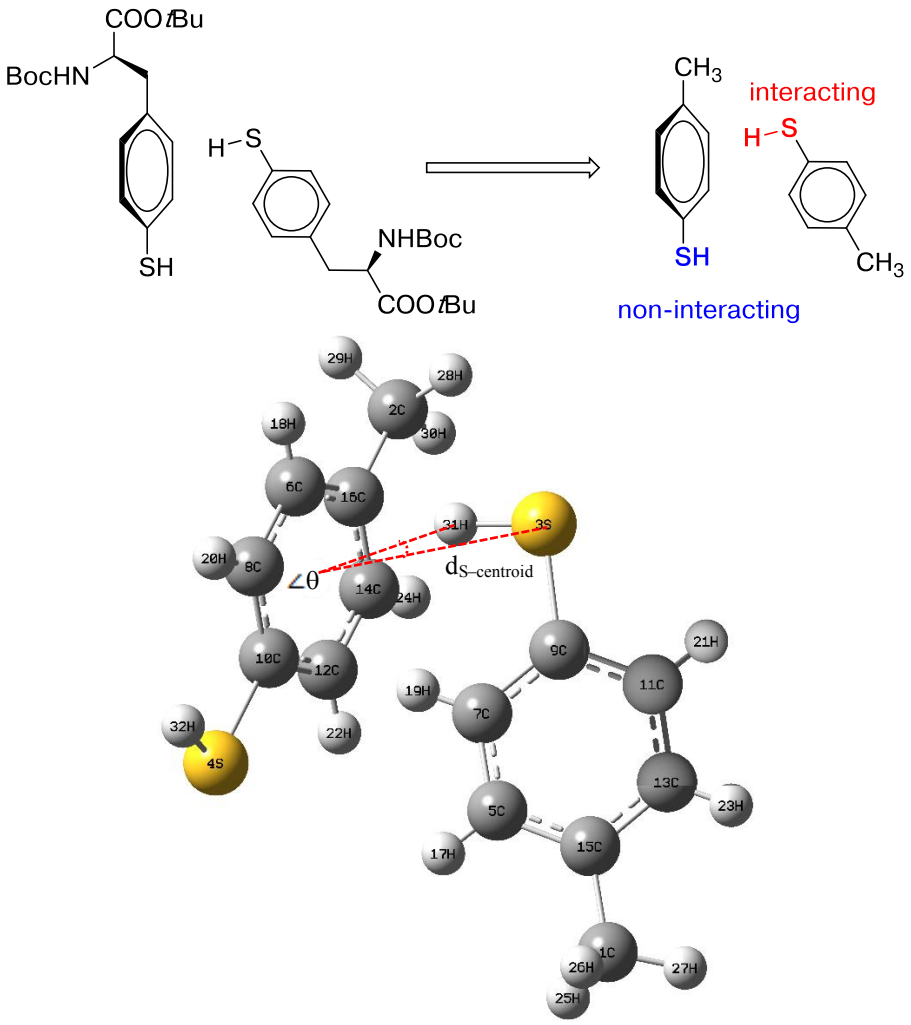


Figure 2.33 Truncation of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester for *ab initio* calculations
Based on the atomic coordinates from the orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (left), a *p*-thiocresol dimer was generated for calculations (right). The two thiol groups are either "interacting" or "non-interacting" with respect to the aromatic ring, as indicated.

Having observed a significant $\Delta\nu_{S-H}$ between crystalline and solution Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (described in previous section), we sought to calculate the associated S–H bond lengths for these observed frequencies. The x-ray crystal structure showed an S–H bond length of 1.26 Å, shorter than the standard bond

length for thiols (1.338 Å).[203] Since x-ray crystal structures are solved based on an electron density map, hydrogen atoms can be difficult to accurately locate due to low electron density. In cases where hydrogen atoms can be located, x-ray crystal structures generate models where polarized bonds appear shorter than their true nature, since the electron density of the bond is observed closer to the heavier atom.[203] Hydrogen atoms that participate in hydrogen bonds generally have longer bond lengths from the hydrogen-bearing bond, due to increased "sharing" of the proton between the two interacting heavy atoms.[203] The S–H stretching frequency depends on the S–H bond length, and so the ν_{S-H} can be calculated for a given bond length. These calculated ν_{S-H} can be compared to observed ν_{S-H} for solution and crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, obtained via Ft-IR spectroscopy (Chapter 2.2.4). With the aid of our collaborator, Dr. Shi Bai, the S–H bond lengths in the *p*-thiocresol dimer (truncated Boc-4-thiol-L-phenylalanine-*tert*-butyl ester) were varied and the predicted S–H stretching frequency was calculated (Table 2.14).

Table 2.14. Dependence of the S–H stretching frequency on the S–H bond length in an S–H/π aromatic interaction

The S–H bond length on the truncated *p*-thiocresol structure derived from the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (Figure 2.33) was varied, and the corresponding S–H stretching frequency and intensity was calculated for each bond length. In order to represent the "interacting" and "non-interacting" modes of the thiol group, the calculations were performed on the *p*-thiocresol dimer and on the further isolated monomer structure, respectively. The computational method and basis set for each calculation is shown.

Thiol structure	S–H bond length, Å	S–H stretching frequency, cm^{-1}	Intensity	method	basis set
monomer (non-interacting)	1.261	3319	0.95	B3LYP	cc-pvtz
	1.338	2722	0.21	B3LYP	cc-pvtz
	1.350	2641		B3LYP	cc-pvtz
	1.360	2575	1.5	B3LYP	cc-pvtz
	1.360	**2584**	**2.1**	**TPSSh**	**cc-pvtz**
	1.375	2477	2.1	B3LYP	cc-pvtz
	1.400	2311	2.7	B3LYP	cc-pvtz
dimer (interacting)	1.261	3320	19.1	B3LYP	cc-pvtz
	1.338	2714	29.5	B3LYP	cc-pvtz
	1.350	2634	31.8	B3LYP	cc-pvtz
	1.360	2568	33.8	B3LYP	cc-pvtz
	1.360	**2568**	**37.2**	**TPSSh**	**cc-pvtz**
	1.375	2470	37.2	B3LYP	cc-pvtz
	1.400	2305	43.6	B3LYP	cc-pvtz

The observed S–H stretching frequency in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (ν_{S-H} 2538 cm^{-1}) correlated to an S–H bond length of 1.36 Å. Surprisingly, the thiol stretching frequency for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester dissolved in chloroform (ν_{S-H} 2585 cm^{-1}) also correlated to an S–H bond length bond length of approximately 1.36 Å. Both of these calculated S–H bond

lengths are longer than the crystallographically observed S–H bond length, demonstrating the importance of normalizing polarized bond lengths to standard values.[203]

Longer S–H bonds correlated with lower S–H stretching frequencies, indicating that thiol hydrogen bonding (which can lengthen the S–H bond) is consistent with a red shift in the $\nu_{S–H}$. For a given S–H bond length, the calculated $\nu_{S–H}$ red shift was greater in the *p*-thiocresol dimer compared to the monomer. The calculated intensities of the $\nu_{S–H}$ were substantially greater in the dimer structure than the monomer, with a calculated 15-20-fold greater IR intensity in the interacting thiol over the non-interacting thiol. These calculations were consistent with observations in IR spectra, where the intensity of the $\nu_{S–H}$ increased more than 15-fold between solution (in $CHCl_3$) and crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (Table 2.10).

Additional calculations were performed on the *p*-thiocresol dimer in order to examine the energetic contributions that stabilized the observed S–H/π aromatic interaction, with the aid of our collaborators Dr. Sudipta Sinha and Dr. Sandeep Patel. To determine the role of electrostatics in the observed S–H/π interaction, the atomic charges were calculated for each atom in the *p*-thiocresol dimer at various levels of theory (B3LYP and MP2) and using various basis sets (6-311+G(2d,p) and aug-cc-vdz). The calculated atomic charges are listed in Table 2.15. To determine the extent of favorable molecular orbital overlap in the S–H/π interaction, the *p*-thiocresol dimer was also examined via NBO analysis. NBO analysis calculates the orbitals with the highest occupancy of electrons.[257] These analyses allow for identification of the geometries and energetics of all bonding and anti-bonding orbitals, and the interaction

energies between these orbitals.[257] NBO analysis has been utilized previously to identify orbital interactions involved in S–H/O interactions,[255] molecular orbital contributions to "blue-shifted hydrogen bonds,[258] orbital overlap contributions to ligand binding in acetocholinesterase via conventional and non-conventional hydrogen bonds,[259] and in characterizing $n \rightarrow \pi^*$ orbital interactions that stabilize peptide backbone conformations.[185-187] Indeed, a favorable orbital overlap interaction was observed in the *p*-thiocresol dimer between the S–H σ^* orbital and an aromatic π orbital (Figure 2.34). A summary of the stabilization energies due to electrostatics and orbital interactions for the *p*-thiocresol dimer is shown in Table 2.15.

Table 2.15. Calculated atomic charges on the *p*-thiocresol dimer

Atomic charges of the *p*-thiocresol dimer were calculated at B3LYP and MP2 levels of theories using 6-311+G(2d,p) and aug-cc-pvdz basis sets, obtained from the CHELPG method. The atom numbers correspond to those shown in Figure 2.33. The thiol groups are highlighted in yellow. The data in bold represents the thiol group that is interacting with the aromatic ring.

		B3LYP		MP2	
S.NO	Atom #	6-311+G (2d,p)	aug-cc-vdz	6-311+G (2d,p)	aug-cc-pvdz
1	C	-0.213	-0.191	-0.204	-0.228
2	C	-0.339	-0.310	-0.323	-0.356
3	**S**	**-0.344**	**-0.348**	**-0.345**	**-0.345**
4	S	-0.287	-0.289	-0.291	-0.290
5	C	-0.218	-0.207	-0.215	-0.222
6	C	-0.226	-0.218	-0.217	-0.228
7	C	-0.119	-0.106	-0.105	-0.120
8	C	-0.108	-0.098	-0.099	-0.110
9	C	0.094	0.093	0.079	0.087
10	C	0.072	0.076	0.065	0.068
11	C	-0.077	-0.065	-0.063	-0.075
12	C	-0.011	-0.004	-0.002	-0.010
13	C	-0.303	-0.286	-0.298	-0.309
14	C	-0.359	-0.341	-0.356	-0.367
15	C	0.254	0.242	0.246	0.257
16	C	0.327	0.315	0.315	0.329
17	H	0.121	0.115	0.121	0.126
18	H	0.139	0.133	0.138	0.143
19	H	0.095	0.087	0.087	0.095
20	H	0.083	0.077	0.080	0.084
21	H	0.151	0.141	0.147	0.153
22	H	0.131	0.122	0.126	0.132
23	H	0.141	0.129	0.138	0.145
24	H	0.154	0.143	0.153	0.160
25	H	0.051	0.044	0.048	0.055
26	H	0.054	0.048	0.053	0.059
27	H	0.059	0.054	0.058	0.064
28	H	0.106	0.098	0.102	0.111
29	H	0.071	0.062	0.066	0.076
30	H	0.090	0.083	0.088	0.096
31	**H**	**0.209**	**0.206**	**0.206**	**0.213**
32	H	0.202	0.196	0.203	0.209

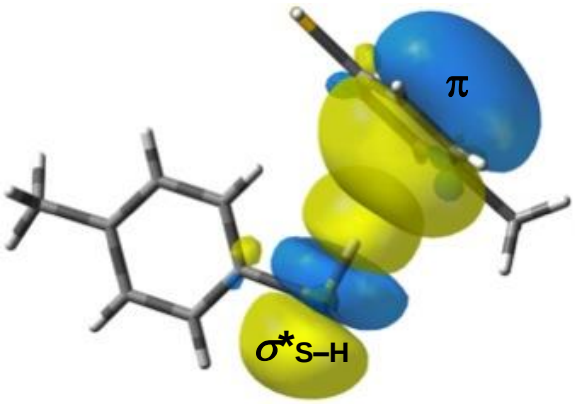


Figure 2.34 Favorable molecular orbital overlap that stabilizes the S–H/π interaction

NBO analysis of the *p*-thiocresol dimer revealed a favorable interaction between the S–H σ* orbital and an aromatic π orbital. This interaction between the orbitals stabilized the S–H/π interaction. The interaction energy between these molecular orbitals, calculated using various basis sets, is included in Table 2.16 (0.70-1.45 kcal mol^{-1} interaction energy).

Table 2.16. Summary of electronic polarization, electrostatic interaction, and NBO analysis on the *p*-thiocresol dimer

Calculations were performed on the *p*-thiocresol dimer at B3LYP and MP2 levels of theories using 6-311+G(2d,p) and aug-cc-pvdz basis sets. The "interacting" and "non-interacting" thiol groups are shown in red and blue, respectively.
*indicates the thiol group which is interacting with the aromatic ring.
Δq is the deviation of atomic charges from the *p*-thiocresol monomer and the calculated atomic charges from the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.
ΔΔq is the difference of atomic charges between the "interacting" and "non-interacting" thiol groups
E_{el} is the energy due to the electrostatic interaction
$E_{i\rightarrow j^*}$ is the interaction energy between donor NBO (σ_j^*) and acceptor NBO (σ_i)
$\langle \pi_{Aro} | \sigma^*_{S-H} \rangle$ is the energy between the orbitals π_{Aro} and σ^*_{S-H}
α is electron occupancy for the indicated orbitals

	S.NO	B3LYP 6-311+G (2d,p)	B3LYP aug-cc-pvdz	MP2 6-311+G (2d,p)	MP2 aug-cc-pvdz
Δq (e)	*S	0.0656	0.0715	0.0598	0.0640
	*H	-0.0211	-0.0263	-0.0094	-0.0254
	S	0.0083	0.0127	0.0053	0.0095
	H	-0.0140	-0.0165	0.0061	-0.0215
ΔΔq S (e)		-0.057	-0.059	-0.054	-0.055
ΔΔq H (e)		0.007	0.010	0.003	0.004
ΔE_{el}(kcal/mol)		-0.866	-0.745	-1.07	-0.72
$E^{(2)}_{i\rightarrow j^*}$ (kcal/mol)		0.70(0.75)	1.19 (1.26)	0.76 (0.80)	1.45 (1.52)
$\langle \pi_{Aro} \vert \sigma^*_{S-H} \rangle$		0.1048	0.1261	0.1070	0.1064
α	σ^*_{S-H}	0.0075	0.0079	0.00524	0.0056
	π_{Aro}	1.690	1.691	1.685	1.685

Using all basis sets, there were only marginal differences in charge between the "interacting" and "non-interacting" thiol groups. The sulfur atoms in the "interacting" thiol were slightly more negatively charge than the "non-interacting" sulfur atom, while the charges on the hydrogen atoms were nearly identical between interacting and non-interacting thiol groups (Table 2.15, highlighted). These perturbations in charge distribution suggests a minimal role for a change in polarization and an induced dipole in the thiol bond with formation of the S–H/π interaction. The energy of stabilization due to electrostatics was calculated to be 0.72-1.07 kcal mol^{-1} (E_{el}), depending on the basis set. Notably, these calculations are in the gas phase, which results in larger energetic magnitudes of electrostatic interactions.

NBO analysis on the *p*-thiocresol dimer revealed a significant and favorable in-phase orbital overlap interaction between the S–H σ^* orbital and an aromatic π orbital (Figure 2.34). The energy of stabilization was calculated using second order perturbation theory (see experimental section), and the energies of the interaction between the two orbitals based on these calculations are shown in Table 2.16 as $E^{(2)}_{i\rightarrow j^*}$. These calculations indicate a significant stabilization of the S–H/π interaction as a result of a $\pi\rightarrow\sigma^*$ orbital interaction, calculated to be 0.70-1.45 kcal mol^{-1}, depending on the basis set.

Combined, these calculations on the *p*-thiocresol dimer indicate that the S–H/π aromatic interaction is stabilized both by electrostatics and orbital overlap contributions. The NBO analysis suggests that the orbital overlap between the S–H σ^* and π orbitals is the primary stabilizing interaction, and predominates over electrostatic interactions, particularly when considered outside of the gas phase. The aromatic π orbitals are localized near carbon atoms and aromatic ring bonds, where

the greatest electrostatic potential is localized in the aromatic centroid. With an interaction that is driven by a favorable orbital overlap interaction, rather than driven by electrostatic effects, the interacting LUMO can be expected to orient towards the HOMO, in this case the S–H σ^* towards the degenerate aromatic π orbitals at the edge of the aromatic ring face. These calculations are consistent with the crystallographic observation that the S–H bond interacted preferentially with the aromatic ring carbons over the aromatic centroid.

2.2.6 Survey of the Cambridge Structural Database (CSD) for aromatic interactions

The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester showed an unexpected geometry of an S–H/π interaction, involving sub-van der Waals distances and near alignment of the S–H bond with an aromatic carbon. Most crystallographic and database analyses of X–H/π aromatic interactions determine geometry with respect to the aromatic centroid rather than individual aromatic carbons.[161, 190, 203, 225, 228] For an interaction that is driven primarily by electrostatic effects, the aromatic centroid is an appropriate reference for comparison. However, in hydrogen bonds that are driven by molecular orbital interactions, the aromatic centroid may *not* be the most appropriate reference for comparison, as the interaction is localized towards the π molecular orbitals (which are centered on individual carbon atoms, not the centroid). Some studies of crystal structures in the Cambridge Structural Database (CSD) have examined centroid-oriented versus π-oriented hydrogen bonds in C–H/π interactions,[260, 261] but none have considered these geometric differences with thiol S–H/π aromatic interactions. Potentially, carbon-oriented thiol S–H/π aromatic interactions are general in crystal structures, and can be distinguished from the

geometries in electrostatically driven interactions, such as cation/π aromatic interactions.

In order to compare the geometries of S–H/π and cation/π aromatic interactions (representing interactions primarily stabilized by molecular orbital overlap or electrostatic effects, respectively), the CSD was subjected to a search for crystal structures that contained 6-membered aromatic rings with an interacting thiol or group I cation (Li^+, Na^+, or K^+). Cation/π interactions have been well-established in both small molecules and proteins, and are driven primarily by electrostatic interactions.[162, 182, 183, 221, 262] Due to the significant electronic effects that stabilize cation/π interactions, the cations can be expected to localize near the aromatic centroid, where the region of negative electrostatic potential is greatest (Figure 2.35a). In contrast, our data on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester suggests that thiol-aromatic interactions are driven by favorable aromatic $\pi \rightarrow \sigma^*_{S-H}$ orbital interactions, and the thiol bond will be directed towards the aromatic carbons rather than the centroid (Figure 2.35b).

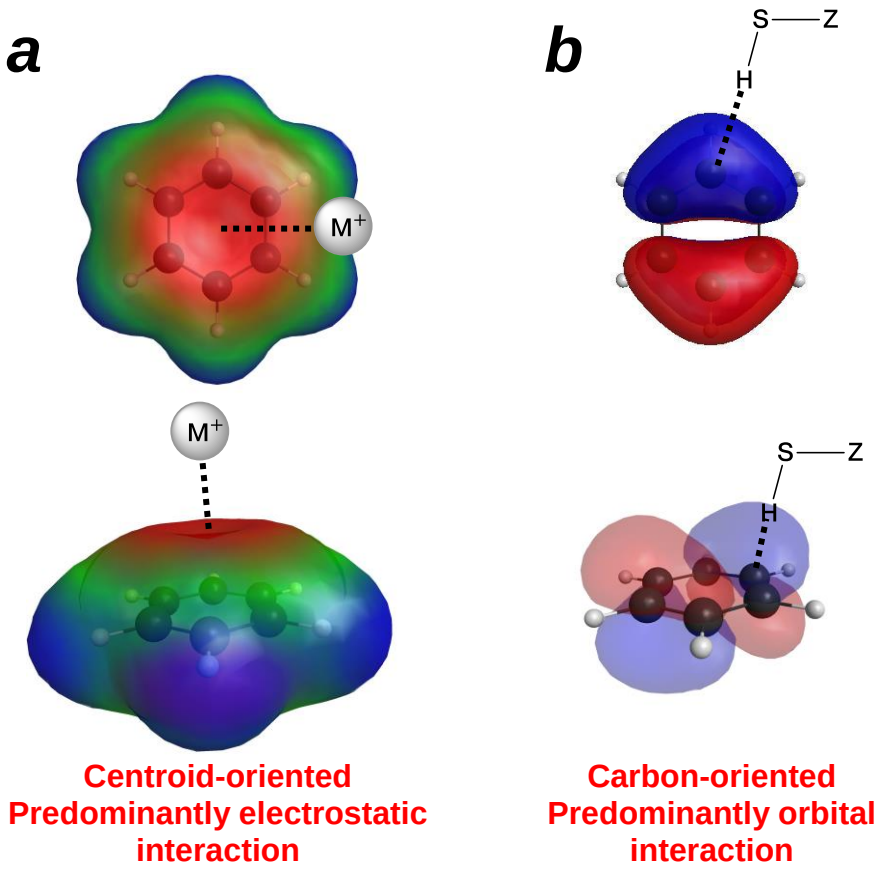


Figure 2.35 Distinct interaction geometries for cation/π and S-H/π aromatic interactions
Interactions with aromatic rings that are driven by electrostatic effects or molecular orbital overlap can potentially have different modes of binding and geometric preferences. Calculations of electrostatic potential and molecular orbitals were conducted using GAMESS (6-311G),[195] and visualized with MacMolPlt.[196] (a) Cation/π interactions are driven by electrostatic effects, with a preference for the aromatic centroid. The surface is shown at a contour of 0.002, with negative and positive electrostatic potentials (–0.025 to +0.025) shown in red and blue, respectively; (b) the highest occupied molecular orbital of benzene, contoured at 0.02. Based on IR and x-ray crystallographic data on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, that the geometry of S–H/π interactions has directionality towards the aromatic ring carbons, given the favorable $\pi \rightarrow \sigma^*_{S-H}$ molecular orbital interaction.

We obtained crystal structures deposited in the CSD containing X/π aromatic interactions, where X is a group I cation (Li^+, Na^+, or K^+, generally M^+) or a thiol

group (Z–S–H, Z = non-hydrogen). Only cation/π or S-H/π interactions with 6-membered aromatic rings were examined and compared. The crystal structures obtained from the CSD were based on the distance from the M^+ or thiol sulfur atom to the centroid of the aromatic ring ($d_{M+–centroid}$ or $d_{S–centroid}$ respectively), which was generously defined as <4.80 Å. Hydrogen atoms cannot be reliably located from x-ray crystallographic data, due to the low electron density around hydrogen atoms, and so only heavy atoms were used for the initial search parameters.[203]

For the crystal structures obtained in this initial search of the CSD, distance parameters were measured between the aromatic ring and the interacting atom (M^+ or thiol H or S; generally, X) in order to compare geometries of cation/π and thiol/π aromatic interactions. The distances for comparison included: the distance from X to the nearest aromatic carbon (defined as $d_{X–Cmin}$); the distance from X to the plane of the aromatic ring ($d_{X–plane}$); and the distance from the aromatic centroid to the projection of X onto the plane of the aromatic ring (radius, r). The radius was calculated from $d_{X–plane}$ and $d_{X–centroid}$ via the Pythagorean theorem. For the thiol groups, measurements were also obtained for the angle S–H–centroid, the Z–S–H bond angle, and the S–H bond length. Graphical representations of the measured parameters are shown in Figure 2.36.

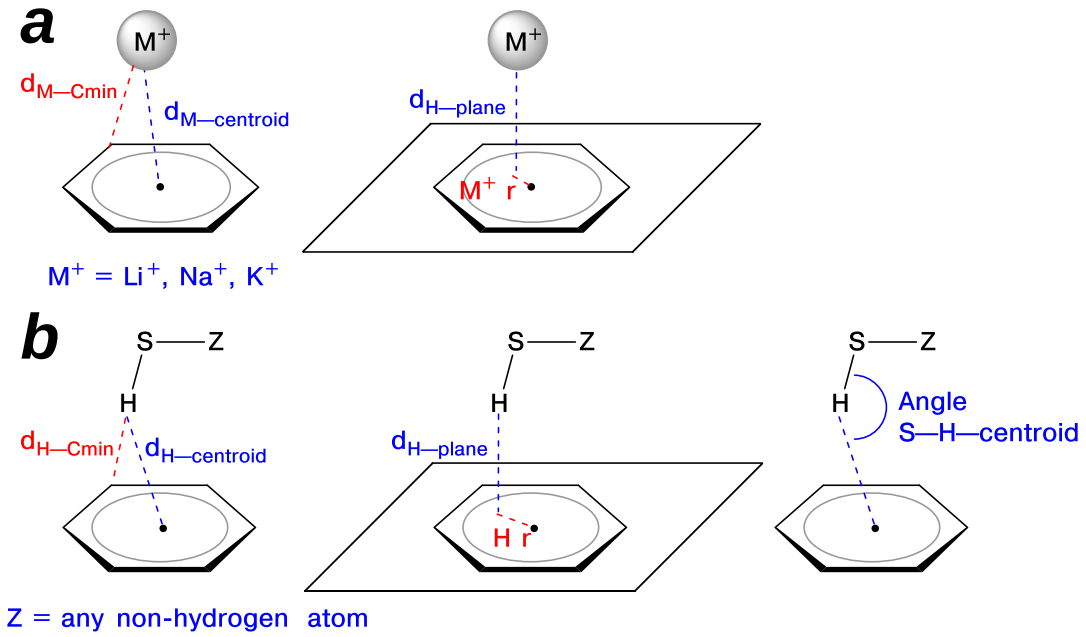


Figure 2.36 Comparison of geometries for cation/π and S–H/π interactions with aromatic rings

From the crystal structures obtained from the CSD, distances were measured in order to compare cation/π and S–H/π aromatic interactions. Generally, the distances between the interacting atom (X) and the aromatic ring were measured. (a) Measurements obtained for cation/π aromatic interactions involving Li^+, Na^+, or K^+ cations (generally, M^+); (b) Measurements obtained for S–H/π interactions; only hydrogen atom measurements are shown in this figure, but sulfur atoms were measured in the same manner. Thiol groups could be bound to Z, which was any non-hydrogen atom.

X = Li^+, Na^+, or K^+; thiol sulfur or hydrogen atom

$d_{X–Cmin}$: the distance from X to the nearest aromatic carbon

$d_{X–centroid}$: the distance from X to the centroid of the aromatic ring

$d_{X–plane}$: the normal distance from X to the plane of the aromatic ring

r, radius: the distance from the aromatic centroid to the projection of X onto the plane of the aromatic ring, calculated as $\sqrt{(d_{X–centroid})^2–(d_{X–plane})^2}$

The crystal structures of cation/π and S–H/π interactions from the CSD were parsed in order to compare interaction geometries for only thiol groups or cations that were closely interacting with the aromatic ring. The region for interaction was defined

as a "cylinder" in reference to the aromatic ring, where only structures where atom X was located within the defined cylinder were retained for further analysis (Figure 2.37). The dimensions of the interaction region were specific to each interacting atom X. For M^+ and thiol hydrogens, the base of the cylinder (restriction on r) was set to 2× the radius of an aromatic ring ($r \leq 3.00$ Å). For thiol sulfur atoms, an additional 1.20 Å allowance was added to account for the S–H bond length ($Sr \leq 4.20$ Å). The height of the cylinder (restriction on $d_{X-plane}$) was defined as 1.25 × the sum of the van der Waals radii (ΣvdWr) between atom X and carbon (Figure 2.37). Historically, an interaction distance that was less than the ΣvdWr for the two interacting atoms was a criterion for identifying hydrogen bonds, but it has since been recognized that interactions can occur at longer distances than the ΣvdWr.[190] The cylinder restriction identified the closer contact distances in S–H/π and cation/π interactions, with interactions above the plane of the aromatic ring, either near ring carbons or the centroid, and allowed for close interactions outside of the plane of the ring (including "edge-on" interactions). The restrictions on the distances r and d_{plane} for each interacting atom X are shown in Figure 2.37.

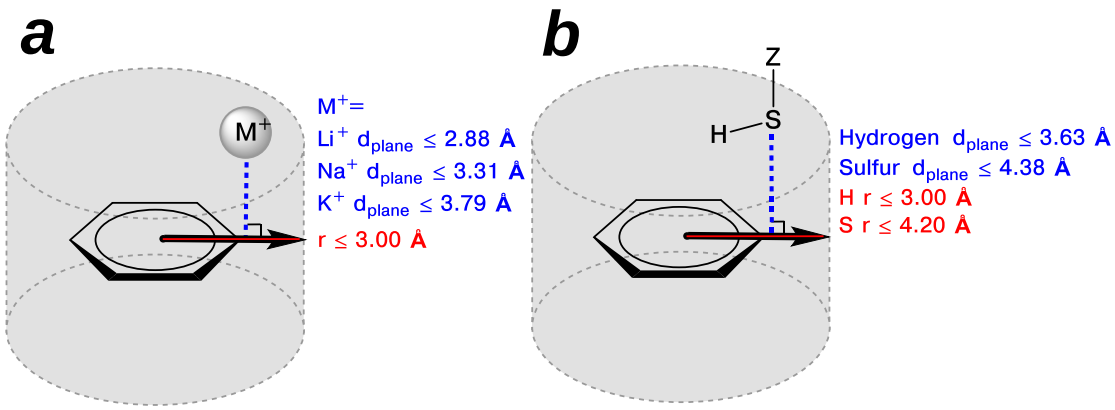


Figure 2.37 Identifying crystal structures obtained from the CSD with closely interacting cation/π or S–H/π aromatic interactions

A region of interaction was defined as a cylinder based on the plane of the aromatic ring, in order to identify and compare only structures with close cation/π or S–H/π aromatic interactions. (a) Defined cylinder for cation/π interactions, where r ≤ 3.00 Å, and d_{plane} ≤ 1.25 × the ΣvdWr for carbon and the cation; (b) defined cylinder for S–H/π interactions, where H r ≤ 3.00 Å, S r ≤ 4.20 Å, and d_{plane} ≤ 1.25 × the ΣvdWr for carbon and the atom S or H.

ΣvdWr for carbon and Li^+, 2.30 Å[203, 263]
ΣvdWr for carbon and Na^+, 2.65 Å[203, 263]
ΣvdWr for carbon and K^+, 3.03 Å[203, 263]
ΣvdWr for carbon and hydrogen, 2.90 Å[203]
ΣvdWr for carbon and S, 3.50 Å[203]

For the structural details that are provided by x-ray crystal structures, there are intrinsic limitations in structural information based on electron density. Hydrogen atoms cannot be accurately located via x-ray crystallography, and in many cases the hydrogen atom can only be modelled or estimated.[203] Measurements in the crystal structure based on heavy atom location, including sulfur, are much more reliable than measurements with hydrogen atoms. Without knowing how the hydrogen atoms were located or resolved in the x-ray crystal structures from the CSD, the cylinder restriction was applied to crystal structures containing thiols, separately, in two ways:

either using the hydrogen atom as a reference, or the sulfur atom for reference, and these data were termed "sulfur reference" and "hydrogen reference," respectively (Table 2.17). In general, the thiol hydrogen atom was used as a reference for the cylinder restriction for thiol structures, because it provides an approximation for the distance and directionality of the thiol S–H σ* orbital in S–H/π interactions.

Due to the nature of x-ray crystallography with data collection of electron density maps, X–H bond lengths often appear shorter than their true nature. With little electron density around hydrogen atoms, the electrons in σ bonds and heavy atoms are more resolved than hydrogen atoms, which biases the electron density map towards the heavier atoms in polarized bonds, and the bond appears shorter.[203] To compensate for this bias, the bond lengths in the crystal structure can be normalized to standard bond lengths by moving the hydrogen atom to the appropriate position along the vector of the X–H bond. For all structures obtained from the CSD with identified thiol-aromatic interactions, two data sets of structures and measurements were analyzed separately: one with the reported coordinates for hydrogen atoms, and one with the hydrogen coordinates adjusted to normalize the S–H bond length to 1.338 Å.[203] Normalization of the S–H bonds and correction of the hydrogen atom coordinates were conducted using ConQuest. The crystal structures with normalized S–H bond lengths (termed "normalized") were also limited to structures within the defined interaction region, the cylinder restriction on the radius and d_{plane} for hydrogen and sulfur atoms.

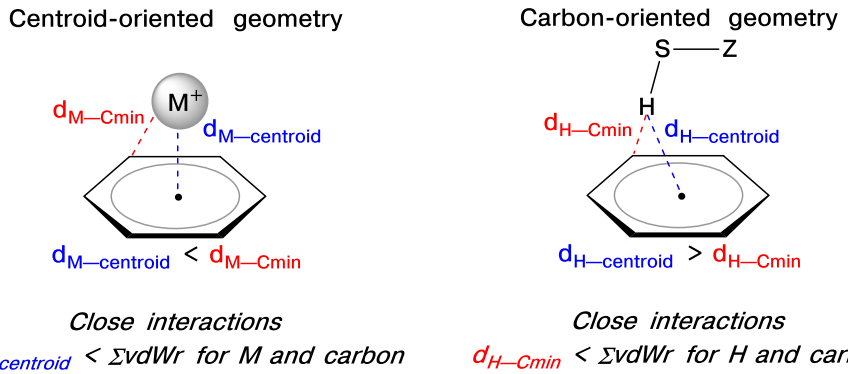
After identifying structures with close cation/π and S–H/π interactions using the "cylinder" restriction, the remaining crystal structures were manually examined and annotated for specific criteria that could potentially alter the geometry of the

cation/π or S–H/π interactions. Analysis for cation/π or S–H/π interactions were limited to included only 6-membered, isolated, aromatic rings. For example, structures where the aromatic ring was part of a polycyclic, conjugated system, such as naphthalene, were excluded from the study. However, structures where the aromatic ring was part of a complex macrocycle, such as in a cyclophane, were included for further analysis. Structures with carbanions within or adjacent to the interacting aromatic rings were also excluded from this analysis, since the addition of a negative charge within the aromatic ring could potentially bias the strength or geometry of the interaction. In thiol structures, only thiol S–H bonds directed towards the ring were included in the analysis (defined as $(d_{S-plane} - d_{H-plane}) \geq -0.10$ Å), so that only S–H/π interactions were compared to cation/π interactions. Thiol S–H bonds directed away from the aromatic ring could potentially result from an S···π interaction, which was not part of the present study. Some thiol crystal structures from the CSD included tricoordinate, hypervalent sulfur atoms, or thiol S–H bond lengths and angles that were unreasonable (including thiol bond lengths 0.86 Å > S–H or S–H > 1.42 Å; thiol bond angles 80° > X–S–H or X–S–H > 120°), and these structures were also excluded from our analysis. Cation structures were also examined for unreasonable geometries. An interaction with a crown ether could potentially modulate the coordination sphere around the cation, or sterically affect an ideal cation/π interaction. Cations that interacted with both crown ethers and aromatic rings were annotated, but included in the present study. The complete list of all structures with annotations are included in the appendix. The number of resultant structures after each level of parsing are shown in Table 2.17.

After the cylinder restriction and manual parsing of structures and entries, the measured distances $d_{X–centroid}$ and $d_{X–Cmin}$ were compared to identify geometric differences between cation/π and S–H/π aromatic interactions. The distances $d_{X–Cmin}$ or $d_{X–centroid}$ were compared to the ΣvdWr for carbon and X, as this comparison identifies the frequency and localization of shorter contacts between aromatic rings and the interacting atom X. Statistics were also gathered for where $d_{X–Cmin} < d_{X–centroid,}$ to determine the frequency where X interacts preferentially with ring carbons over the aromatic centroid. These data are included in Table 2.17, and are grouped by each interacting atom X and each level of parsing from the initial data sets from the CSD. Some x-ray crystal structures had multiple, non-redundant interactions within the same structure, and these are reported as "entries," rather than "structures" in Table 2.17.

Table 2.17. Comparison of $d_{X–centroid}$ and $d_{X–Cmin}$ for cation/π and S–H/π interactions in crystal structures from the CSD

Comparisons were made to identify whether cation/π and S–H/π aromatic interactions have centroid- or carbon-oriented geometries. Quantities of occurrences where the $d_{X–Cmin}$ and $d_{X–centroid}$ distances were less than the ΣvdWr for carbon and X, or where $d_{X–Cmin} < d_{X–centroid}$. The percentage of the represented category is shown in parentheses. For all thiol structures, the $d_{H–Cmin}$ and $d_{H–centroid}$ distances were used for comparison (not $d_{S–Cmin}$ and $d_{S–centroid}$). Quantities and statistics are shown after each level of parsing, where "cylinder restriction" indicates all structures where atom X was within a defined region of interaction (see description above, Figure 2.37), and where "manual parsing" indicates only structures that did not have unusual or potentially altered geometry due to external coordinating groups (see above description and experimental section).
"H reference" and "S reference" refers to thiol structures where the hydrogen atom or sulfur atom was used for cylinder restrictions.
"normalized" refers to thiol structures where the S–H bond was normalized to 1.338 Å.



	X/π aromatic interaction	Cylinder Restriction		Manual Parsing	
		Structures	**Entries**	**Structures**	**Entries**
$d_{X–Cmin} < d_{X–centroid}$	Li^+	28 (76%)	72 (83%)	17 (89%)	40 (93%)
	Na^+	48 (77%)	79 (75%)	36 (75%)	61 (77%)
	K^+	124 (57%)	227 (56%)	110 (57%)	201 (56%)
	Thiol H, H reference	99 (90%)	164 (90%)	68 (87%)	111 (87%)
	Thiol H, S reference	152 (93%)	378 (95%)	92 (90%)	186 (92%)
	Thiol H, normalized, H reference	98 (88%)	158 (89%)	66 (85%)	106 (87%)
	Thiol H, normalized, S reference	155 (92%)	386 (94%)	81 (91%)	199 (95%)

Table 2.17 continued

$d_{X-Cmin} < \Sigma vdWr$	Li^+	14 (38%)	25 (29%)	7 (37%)	17 (40%)
	Na^+	5 (8%)	7 (7%)	4 (8%)	5 (6%)
	K^+	21 (10%)	35 (9%)	19 (10%)	32 (9%)
	Thiol H, H reference	58 (53%)	81 (45%)	45 (58%)	62 (49%)
	Thiol H, S reference	64 (39%)	89 (22%)	47 (46%)	65 (32%)
	Thiol H, normalized, H reference	66 (59%)	98 (55%)	47 (60%)	71 (58%)
	Thiol H, normalized, S reference	72 (43%)	106 (26%)	46 (52%)	62 (30%)
$d_{X-centroid} < \Sigma vdWr$	Li^+	8 (22%)	14 (16%)	2 (11%)	3 (7%)
	Na^+	7 (11%)	11 (10%)	5 (10%)	7 (9%)
	K^+	43 (20%)	92 (22%)	38 (20%)	84 (23%)
	Thiol H, H reference	18 (16%)	28 (15%)	18 (23%)	27 (21%)
	Thiol H, S reference	18 (11%)	28 (7%)	18 (18%)	27 (13%)
	Thiol H, normalized, H reference	24 (22%)	35 (20%)	24 (29%)	35 (20%)
	Thiol H, normalized, S reference	24 (14%)	35 (9%)	18 (20%)	23 (11%)

In order to determine whether the manual parsing parameters introduced unexpected biases in the data, the frequencies of occurrences between the entries included after the cylinder restriction and after the manual parsing were compared. For all interacting atoms investigated, M^+ and thiol S and thiol H, there was generally good agreement in the statistics before and after the manual parsing restrictions. Given reasonable agreement in all categories of aromatic interactions, hereafter only the structures included after manual parsing are examined and analyzed.

In more than 87% of the structures and entries, the thiol hydrogen atoms were nearer to the aromatic ring carbons than the aromatic centroid, regardless of how the data was parsed or normalized. In cation/π entries, most cations were also nearer to the aromatic ring carbons than to the aromatic centroid, with Li^+ exhibiting the highest

frequency for centroid-oriented geometry (93%), followed by Na^+ (77%) and K^+ (56%). It was expected that all of the cations would exhibit centroid-oriented geometry where $d_{X–Cmin} > d_{X–centroid}$, suggesting strong electrostatic effects, and so observing these relatively high frequencies of carbon-oriented geometry along a periodic trend was intriguing. Based on these comparisons, Na^+/π and K^+/π aromatic interactions showed higher frequency of centroid-oriented geometry over $Li^+/$ and S–H/π aromatic interactions, which have higher frequencies of carbon-oriented geometry.

The ΣvdWr for two interacting atoms was a threshold for identifying hydrogen bonds, where interaction distances that were less than the ΣvdWr of the respective atoms (sub-ΣvdWr) were considered to be hydrogen bonding.[203] While interactions are known to occur at distances greater than the ΣvdWr, this threshold can provide a cut-off for identifying exceptionally close interaction distances, and providing insight into the nature of cation/π and S–H/π interactions as they become stronger.

In entries representing the three different cation/π interactions, there was a lower frequency of sub-ΣvdWr contact distances with aromatic rings than thiol hydrogen atoms, generally less than 10%, with two major exceptions. In Li^+/π interactions, 40% of entries exhibited sub-ΣvdWr distances with carbon-oriented contacts; in K^+/π interactions, 23% of entries exhibited sub-ΣvdWr distances with centroid-oriented contacts. These data suggest that while most cation/π interactions are not exceptionally close contacts, smaller cations exhibit close-contacts with aromatic ring carbons and large cations exhibit close-contacts with the aromatic centroid. This trend suggests different binding modes for different cations. A purely electrostatic interaction is insufficient to describe these trends with different interaction orientations for different cations.

S–H/π aromatic interactions had higher frequencies of sub-ΣvdWr aromatic contacts that were carbon-oriented over centroid-oriented. Where thiol hydrogen atoms were identified within the defined region of interaction with the aromatic ring, 49% of entries were carbon-oriented sub-ΣvdWr distances, while only 21% of entries were centroid-oriented sub-ΣvdWr distances. With normalizing S–H bond lengths, the frequency of close contacts between thiol hydrogens and aromatic carbons increased to 58% of entries (while the frequency of centroid-oriented sub-ΣvdWr distances remained 20%). Overall, entries with S–H/π interactions had higher frequencies of sub-ΣvdWr contact distances with aromatic rings than cation/π interactions, with a significantly higher percentage of entries with carbon-oriented geometry. The differing interaction geometry and higher frequency of close contacts with S–H/π interactions over cation/π interactions suggests fundamental differences in the nature of these associations, which cannot be attributed to differences in electrostatic effects. If an electrostatic interaction is so much stronger than a dispersive interaction, then how can thiol hydrogen atoms to interact with an aromatic ring at such close contact distances with such a defined orientation?

In order to obtain further insights into the distinctive geometries of these aromatic interactions, the frequencies of the measured distances from crystal structures were compared via histograms. Based on the statistics in Table 2.17, it was expected that the distance $d_{X–centroid}$ for cation/π aromatic interactions would be normally distributed near the ΣvdWr. For the distance $d_{X–Cmin}$, it was hypothesized that thiol hydrogen interactions would be normally distributed at distances that were sub-ΣvdWr, while cation/π interactions would be normally distributed beyond the ΣvdWr. The distributions of the radius distance can also reveal the orientation of the

interacting atom X with respect to the aromatic centroid, where shorter radius distances indicate closer alignment with the centroid. Carbon-oriented geometries should be distributed near 1.40 Å, indicating the interacting atom is oriented directly over the aromatic carbon atoms and bonds; centroid-oriented geometries should be distributed near 0 Å. Distributions of the distances $d_{X–centroid}$, $d_{X–Cmin}$, and radius for cation/π and S–H/π aromatic interactions are shown in Figures 2.38-2.40. The statistics for the distances $d_{X–centroid}$, $d_{X–Cmin}$, and radius for cation/π and S–H/π aromatic interactions are included in Table 2.18.

The thiol data shown in the following histogram charts were restricted to the interacting region with reference to the hydrogen atom (the cylinder restriction was applied to the hydrogen atom, not the sulfur atom), and the S–H bond lengths were as published in the CSD (not normalized). Additional comparisons with S–H/π aromatic interactions, using the data sets with normalized S–H bond lengths or using the data sets where the region of interaction was defined by the sulfur atom, are shown later in this chapter and in the experimental section.

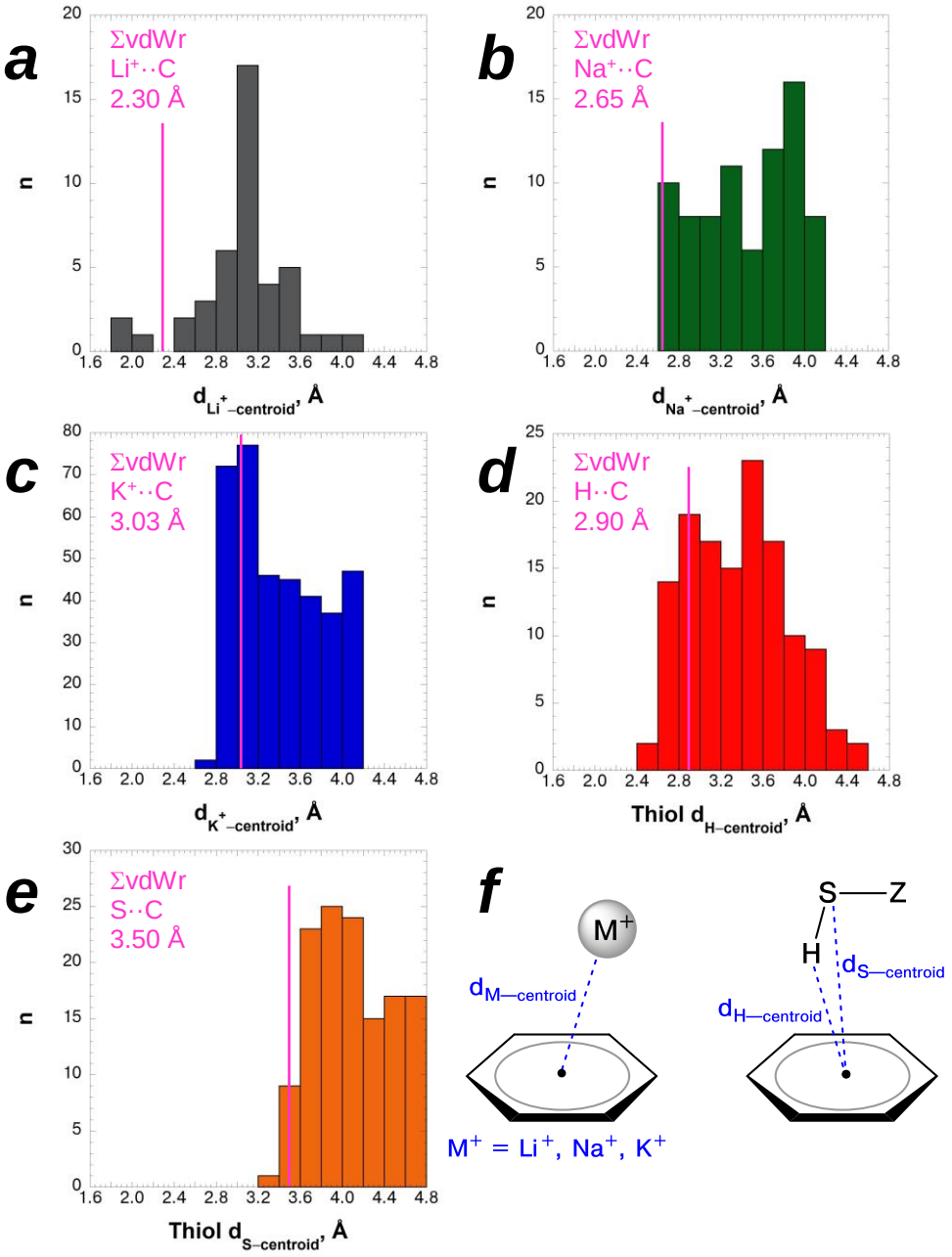


Figure 2.38 Distributions of the distance from the aromatic centroid to atom X for cation/π and S–H/π interactions ($d_{X–centroid}$, X = Li^+, Na^+, K^+, S, or H)

Histogram charts for the frequency of given distances from the aromatic centroid to the interacting atom X (X = Li^+, Na^+, K^+, thiol S, thiol H) for cation/π and S–H/π aromatic interactions; (a) frequency of observed distances for $d_{Li^+\text{-centroid}}$; (b) frequency of observed distances for $d_{Na^+\text{-centroid}}$; (c) frequency of observed distances for $d_{K^+\text{-centroid}}$; (d) frequency of observed distances for $d_{H\text{-centroid}}$, where the hydrogen atom was used as a reference for the cylinder restriction and the S–H bond length were not normalized; (e) frequency of observed distances for $d_{S\text{-centroid}}$, where the hydrogen atom was used as a reference for the cylinder restriction; (f) scheme showing these parameters for cation/π and S–H/π interactions.
The sum of the van der Waals radii (ΣvdWr) for respective atoms are indicated in each chart.
The data set includes only entries within the region of interaction (cylinder restriction) and were manual parsed (see p 96-97 for parsing criteria).

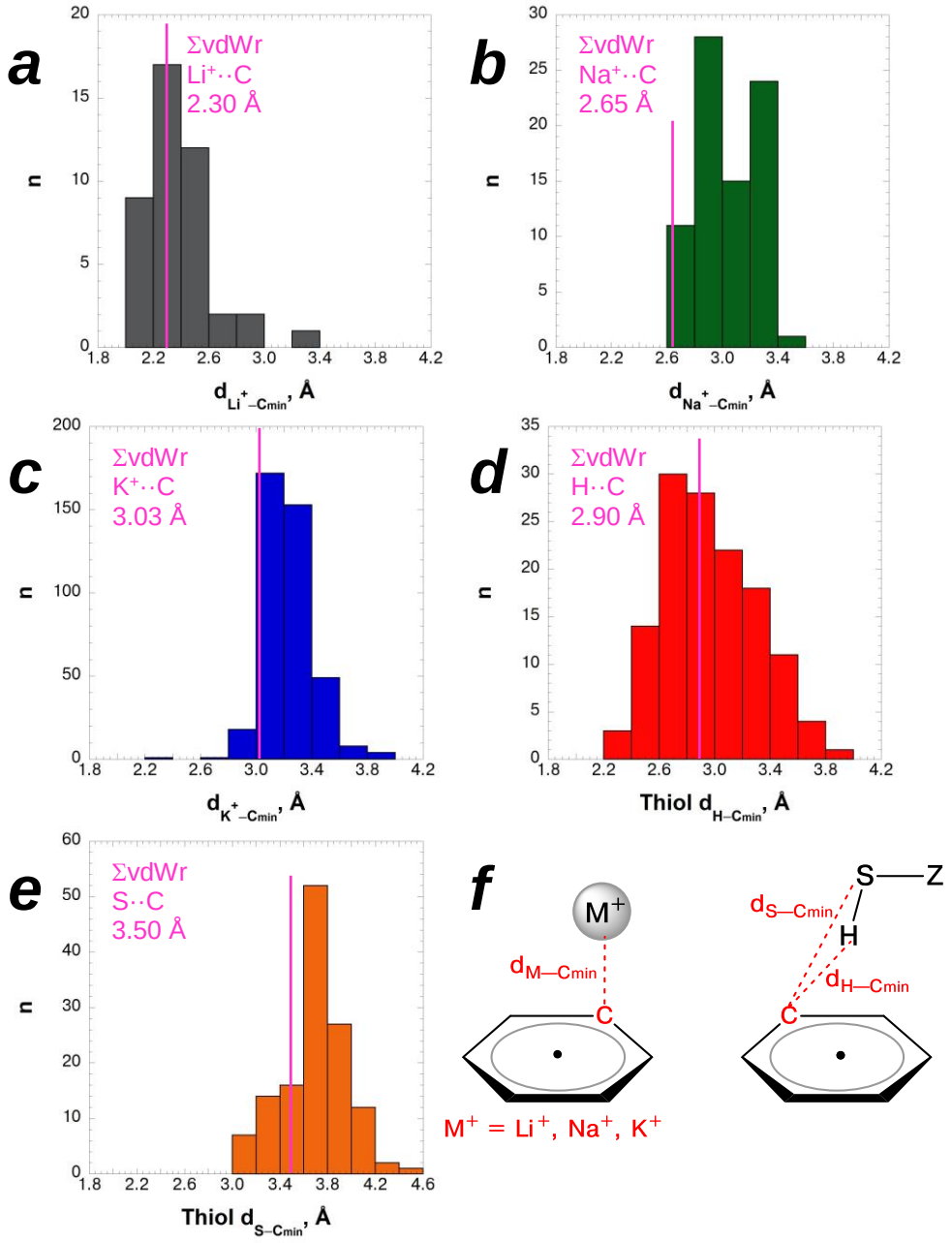


Figure 2.39 Distributions of the distance from the interacting atom X to the nearest aromatic carbon for cation/π and S–H/π interactions ($d_{X–Cmin}$, X = Li^+, Na^+, K^+, S, or H)

Histogram charts for the frequency of given distances from the interacting atom X (X = Li^+, Na^+, K^+, thiol S, thiol H) to the nearest aromatic carbon for cation/π and S–H/π aromatic interactions; (a) frequency of observed distances for $d_{Li^+\text{-}Cmin}$; (b) frequency of observed distances for $d_{Na^+\text{-}Cmin}$; (c) frequency of observed distances for $d_{K^+\text{-}Cmin}$; (d) frequency of observed distances for $d_{H\text{-}Cmin}$, where the hydrogen atom was used as a reference for the cylinder restriction and the S–H bond length were not normalized; (e) frequency of observed distances for $d_{S\text{-}Cmin}$, where the hydrogen atom was used as a reference for the cylinder restriction; (f) scheme showing these parameters for cation/π and S–H/π interactions.
The sum of the van der Waals radii (ΣvdWr) for respective atoms are indicated in each chart.
The data set includes only entries within the region of interaction (cylinder restriction) and were manual parsed (see p 96-97 for parsing criteria).

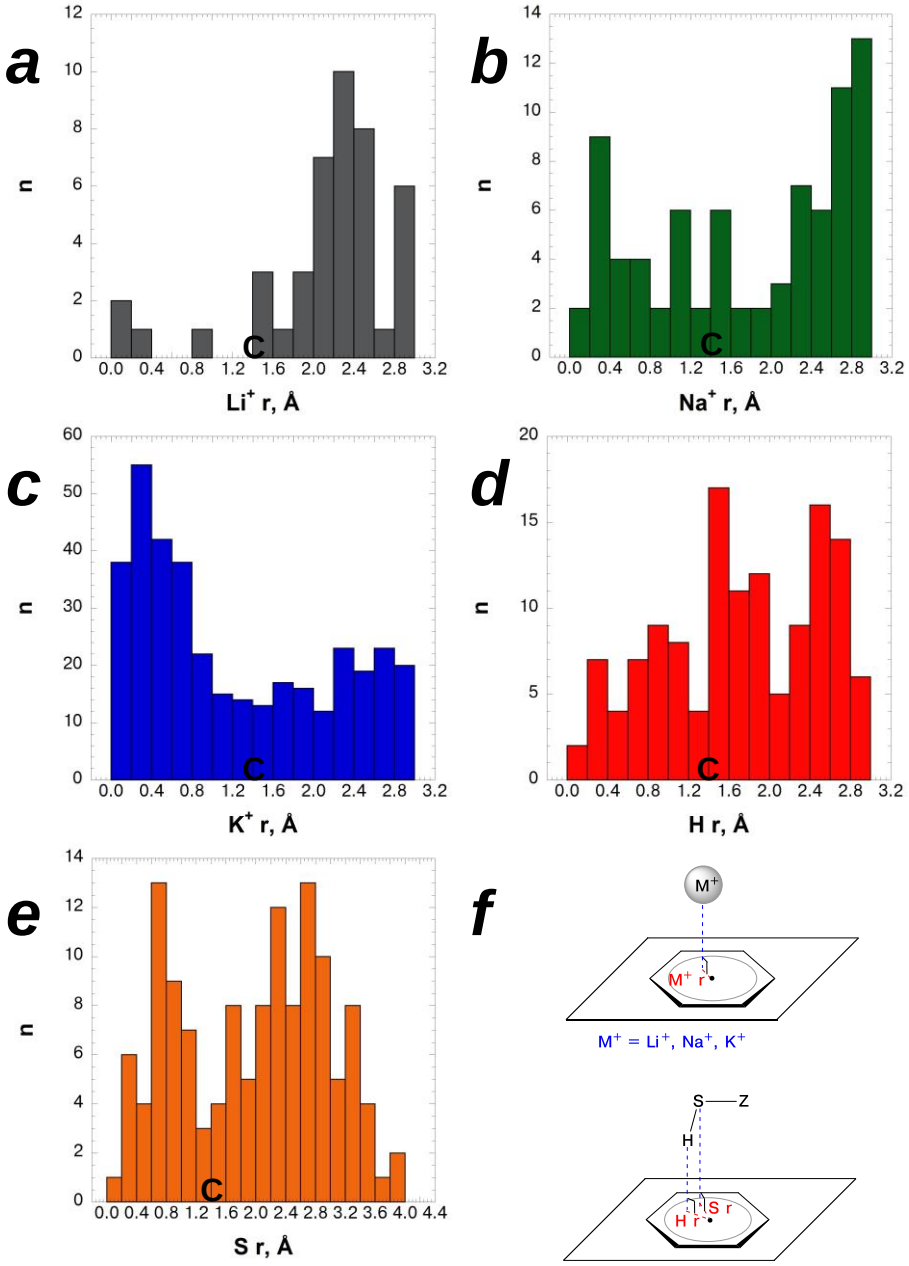


Figure 2.40 Distributions of the distance from the aromatic centroid to the projection of the interacting atom X within the plane of the aromatic ring (radius) for cation/π and S–H/π interactions ($d_{X–Cmin}$, X = Li^+, Na^+, K^+, S, or H)

Histogram charts for the frequency of given distances for the radius of the interacting atom X (X = Li^+, Na^+, K^+, thiol S, thiol H) for cation/π and S–H/π aromatic interactions; the radius is defined as the distance from the aromatic centroid to the projection of the atom X onto the plane of the aromatic ring; (a) frequency of observed distances for Li^+ r; (b) frequency of observed distances for Na^+ r; (c) frequency of observed distances for K^+ r; (d) frequency of observed distances for thiol hydrogen r, where the hydrogen atom was used as a reference for the cylinder restriction and the S–H bond length were not normalized; (e) frequency of observed distances for thiol sulfur r, where the hydrogen atom was used as a reference for the cylinder restriction; (f) scheme showing these parameters for cation/π and S–H/π interactions.

The location of the ring carbon atom (r = 1.40 Å) is indicated as C in each chart. The data set includes only entries within the region of interaction (cylinder restriction) and were manual parsed (see p 96-97 for parsing criteria).

Table 2.18. Statistics for geometry of cation/π and S–H/π interactions in crystal structures from the CSD

Statistics for measured distances in cation/π and S–H/π aromatic interactions, including the distance from X to the aromatic centroid ($d_{X–centroid}$), X to the nearest aromatic carbon ($d_{X–Cmin}$), and the aromatic centroid to X projected to the plane of the aromatic ring (X r), where X = M^+, thiol S, or thiol H.
The sum of the van der Waals radii (ΣvdWr) for respective atoms are indicated in each chart.
The data set includes only entries within the region of interaction (cylinder restriction) and were manual parsed (see p 96-97 for parsing criteria). The statistics for the thiol hydrogen atom where the thiol S-H bond length is indicated at the bottom of the chart. N/A indicates no value

X/π aromatic interaction		Mean	Standard Deviation	Median	Mode	Minimum	Maximum
Li^+ ΣvdWr = 2.30 Å	$d_{Li^+–centroid}$, Å	3.050	0.425	3.043	3.022	1.947	4.139
	$d_{Li^+–Cmin}$, Å	2.395	0.234	2.338	2.253	2.125	3.250
	Li^+ r, Å	2.115	0.675	2.255	N/A	0.109	3.000
Na^+ ΣvdWr = 2.65 Å	$d_{Na^+–centroid}$, Å	3.440	0.468	3.538	3.944	2.469	4.172
	$d_{Na^+–Cmin}$, Å	3.002	0.217	2.967	2.969	2.618	3.400
	Na^+ r, Å	1.849	0.944	2.233	N/A	0.000	2.995
K^+ ΣvdWr = 3.03 Å	$d_{K^+–centroid}$, Å	3.407	0.408	3.321	2.853	2.781	4.192
	$d_{K^+–Cmin}$, Å	3.226	0.172	3.205	3.134	2.250	3.843
	K^+ r, Å	1.221	0.934	0.938	0.000	0.000	2.976
Thiol H ΣvdWr = 2.90 Å	$d_{H–centroid}$, Å	3.376	0.488	3.500	3.417	2.537	4.499
	$d_{H–Cmin}$, Å	3.004	0.366	2.964	2.472	2.200	3.912
	H r, Å	1.863	0.782	1.887	1.884	0.174	2.987
Thiol S ΣvdWr = 3.50 Å	$d_{S–centroid}$, Å	4.101	0.371	4.060	3.725	3.324	4.798
	$d_{S–Cmin}$, Å	3.692	0.266	3.693	3.737	3.087	4.407
	S r, Å	1.976	1.000	2.133	N/A	0.082	3.828
Normalized Thiol H ΣvdWr = 2.90 Å	$d_{H–centroid}$, Å	3.279	0.494	3.295	3.469	2.412	4.478
	$d_{H–Cmin}$, Å	2.874	0.342	2.831	3.130	2.167	3.830
	H r, Å	1.722	0.795	1.749	N/A	0.000	2.999

In all cation/π and S–H/π aromatic interactions, the average distances from the aromatic centroid to the interacting atom were further than the ΣvdWr. The distributions of $d_{X–centroid}$ distances were all generally distributed beyond the ΣvdWr as well. The average distances from Li^+ or thiol hydrogen atoms to the nearest aromatic ring carbon were near the ΣvdWr for X and carbon. In contrast, average Na^+ and K^+ distances to the nearest carbon were much greater than the ΣvdWr for X and carbon (13% or 6% longer than ΣvdWr, respectively). These trends are consistent in comparing the histogram charts between M^+ and thiol H distances $d_{X–Cmin}$ (Figure 2.39), where Li^+ and thiol hydrogen atoms are reasonably well distributed about the ΣvdWr, while nearly all Na^+ and K^+ distances are greater than the ΣvdWr. Consistent with the data shown in Table 2.17, Li^+ cations and thiol hydrogen atoms are more likely to interact closely with aromatic ring carbons than larger cations.

The radius r of the interacting atom X, X r, identifies the localization of the atom X relative to the centroid of the ring, regardless of the contact distance (see scheme, Figure 2.40f). For interactions that are driven by electrostatic effects, which were expected to localize near to the aromatic centroid, the radius would be shorter (indicating the atom is localized closer to the aromatic centroid). For interactions with carbon-oriented geometry, the radius would be expected to be near 1.4 Å, indicating the interacting atom is located nearer to the radius of a benzene ring. For cation/π aromatic interactions, the X r increased with smaller ionic radii of the cation. This periodic trend identified that K^+ cations were distributed closer to the aromatic centroid (average K^+ r = 1.22 Å), while Li^+ cations were generally nearer to the *outside* of the aromatic ring (average Li^+ r = 2.11 Å). 32% of entries with K^+/π aromatic interactions had radius distances less than 0.5 Å, indicating that K^+ cations

had a significant likelihood of localizing directly over the centroid of the aromatic ring. The thiol hydrogen and sulfur atom distributions over the aromatic ring (X r) were more dispersed (Figure 2.40 d and e, respectively). Thiol sulfur atoms had two localizations, either inside or outside of the radius of the benzene ring, and the hydrogen atoms did not have any particular localization over the plane of the aromatic ring. The mean and mode of the distance H r were 1.87 Å and 1.88 Å, respectively, which localized hydrogen atoms slightly outside of the radius of the aromatic ring, using the hydrogen atom coordinates that were reported in the CSD.

Although the distance from a donor atom to the aromatic centroid is the general measure for strength of an aromatic interaction,[190] these data suggest that the distance to the nearest aromatic ring carbon can be a more relevant parameter in some cases. When an interaction between an atom and an aromatic ring is weakly electrostatic, orbital interactions may compete or dominate, which can potentially draw the vector of the donor group to the edges of the aromatic ring face, and away from the aromatic centroid. With the standard model of centroid-oriented aromatic interactions, these carbon-oriented interactions may appear weaker and less directional than their true nature.

While the histogram charts of X r shows localization of the interacting atoms over the plane of the aromatic ring, these charts do not distinguish between the closely interacting atoms and the longer-range interactions. To gain further insight into the fundamental differences in geometry between closely interacting cation/π and S–H/π interactions, the coordinates of each interacting atom, M^+, thiol H, and thiol S, were compared relative to the centroid *and* to the plane of the aromatic ring (Figure 2.41). These comparisons identify where the interacting atoms were localized over the plane

of the ring, with consideration of the stronger interactions with closer contact distances to the aromatic ring. By correlating the distance $d_{X-plane}$ with respect to the distance X r, the X-axis represents the plane of the aromatic ring, and the origin represents the aromatic centroid (Figure 2.41f). The correlations between $d_{X-plane}$ and X r were compared to the threshold of "close" contacts, the ΣvdWr for each respective atom X and carbon. The ΣvdWr was projected as a circle where the origin was at the radius of benzene (r = 1.4 Å, d_{plane} = 0.0 Å). The correlations for each cation/π and S–H/π aromatic interaction are shown in Figure 2.41. For comparison, the coordinates associated with the S–H/π aromatic interaction observed in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester are indicated in the charts.

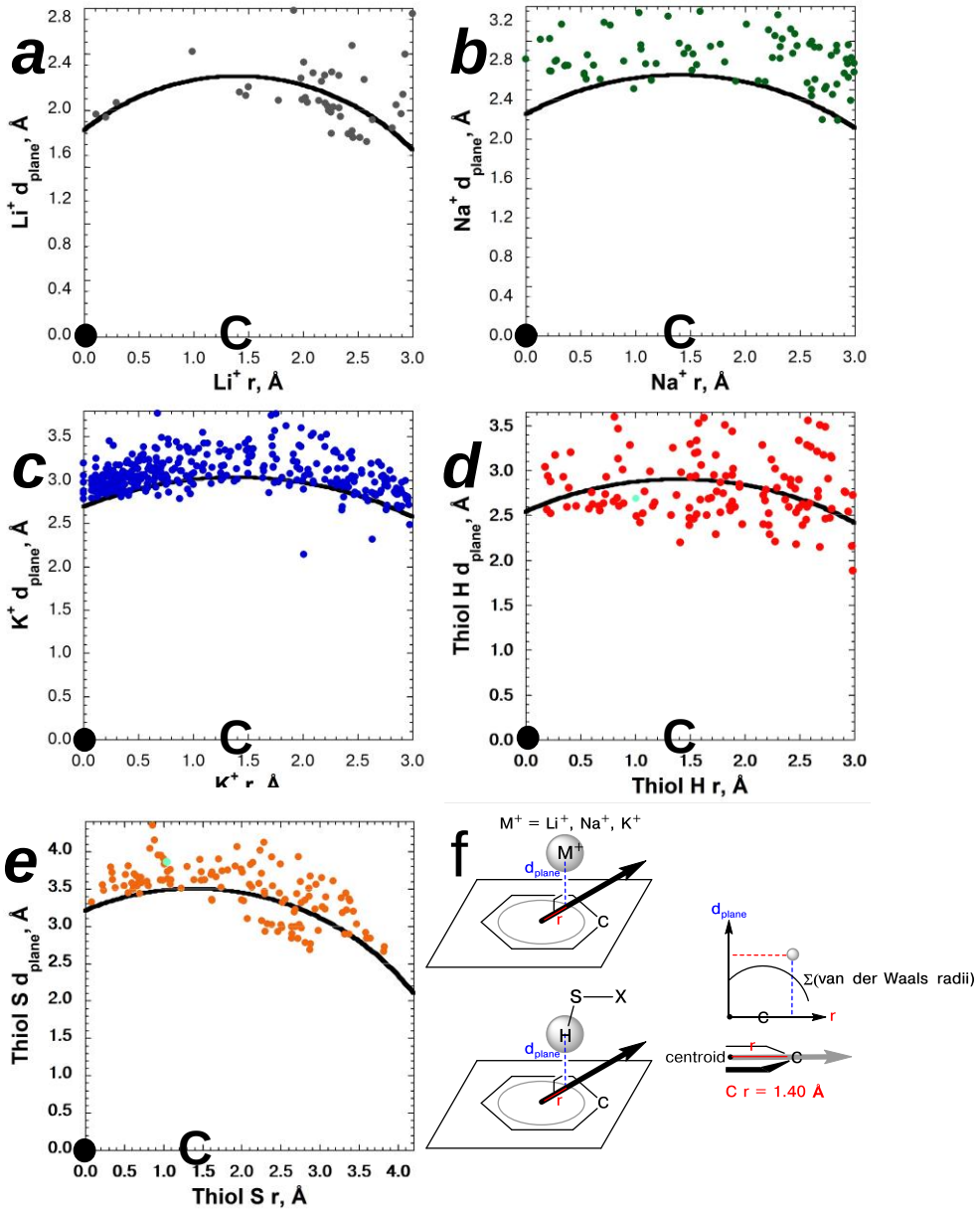


Figure 2.41 Localization of interacting atoms with respect to the aromatic ring for cation/π and S-H/π aromatic interactions

The atomic coordinates of the interacting atom, X, are indicated with reference to the plane of the aromatic ring for cation/π and S–H/π aromatic interactions. The distance $d_{X-plane}$ is correlated against the distance X r (radius), where X = Li^+, Na^+, K^+, thiol H, or thiol S. The X-axis represents the plane of the aromatic ring, and the origin represents the aromatic centroid. The location of aromatic centroid (●) and ring carbons (C) are indicated within the plots, where **C** is located at 1.40 Å from the origin. The black line indicates the ΣvdWr for carbon and the respective interacting atoms; points below the black line are less than the ΣvdWr. (a) atomic coordinates of Li^+ with respect to the aromatic ring; (b) atomic coordinates of Na^+ with respect to the aromatic ring; (c) atomic coordinates of K^+ with respect to the aromatic ring; (d) atomic coordinates of the thiol hydrogen atom with respect to the aromatic ring, (e) atomic coordinates of the thiol sulfur atom with respect to the aromatic ring, (f) scheme showing these parameters for cation/π and S–H/π interactions.
The data set for S–H/π aromatic interactions used in these projections were restricted on the hydrogen atom ("cylinder restriction," see experimental section and Figure 2.37 for details) and manually parsed for geometric biases and anomalies; the S–H bond lengths were not normalized.
The cyan data points that are indicated in the charts with thiol hydrogen and thiol sulfur coordinates indicate the coordinates for the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.
ΣvdWr for carbon and Li^+, 2.30 Å[203, 263]
ΣvdWr for carbon and Na^+, 2.65 Å[203, 263]
ΣvdWr for carbon and K^+, 3.03 Å[203, 263]
ΣvdWr for carbon and hydrogen, 2.90 Å[203]
ΣvdWr for carbon and S, 3.50 Å[203]

Consistent with observations from the histogram charts and statistics on measured interaction distances, cation/π and S–H/π aromatic interactions exhibited distinctive geometric orientations with respect to the aromatic ring. Na^+ and K^+ cations did not interact as closely with the aromatic ring as thiols, with less than 10% of entries below the threshold of the ΣvdWr for carbon and M^+. Na^+/π interactions did not appear to localize over the centroid or the aromatic ring carbon, and were generally dispersed over the plane of the aromatic ring (Figure 2.41b). A significant fraction of K^+ cations localized over the aromatic centroid (32%, where K r < 0.5 Å), but only 4 of these entries with centroid-oriented geometry were at distances less than

the ΣvdWr. Intriguingly, the Li^+ cation/π aromatic interactions that were well outside of the radius of the aromatic ring were also some of the closest contacts (Figure 2.41a). However, the small sample size of the data set for Li^+/π aromatic interactions must be considered with these observed trends (43 entries).

In contrast to the larger cation/π aromatic interactions of K^+ or Na^+, nearly half of the thiol hydrogen atoms interacted with the aromatic ring carbons at distances that were less than the ΣvdWr for carbon and hydrogen (Figure 2.41d). Of the 20 shortest contact distances in S–H/π aromatic interactions ($d_{H–Cmin}$), 6 entries were localized above the aromatic ring carbon (1.15 Å < H r < 1.65 Å). Four of these selected entries of S–H/π aromatic interactions are shown as examples in Figure 2.42. All of the thiol sulfur atoms with sub-ΣvdWr contact distances with the aromatic ring were localized above the aromatic carbons and outside of the aromatic ring, potentially allowing for the thiol hydrogen and the aromatic ring to interact. The data set of S–H/π aromatic interactions used in this comparison was based on hydrogen atom coordinates reported in the CSD, and was manually parsed to exclude any thiol structures where the thiol hydrogen was directed away from the aromatic ring (an indication of S··π aromatic interactions). Based on these combined data, the S–H/π aromatic interactions have a greater tendency for carbon-oriented geometry with sub-ΣvdWr contact distances, while large cation/π aromatic interactions were centroid-oriented at distances at or greater than the ΣvdWr for carbon and the respective cations.

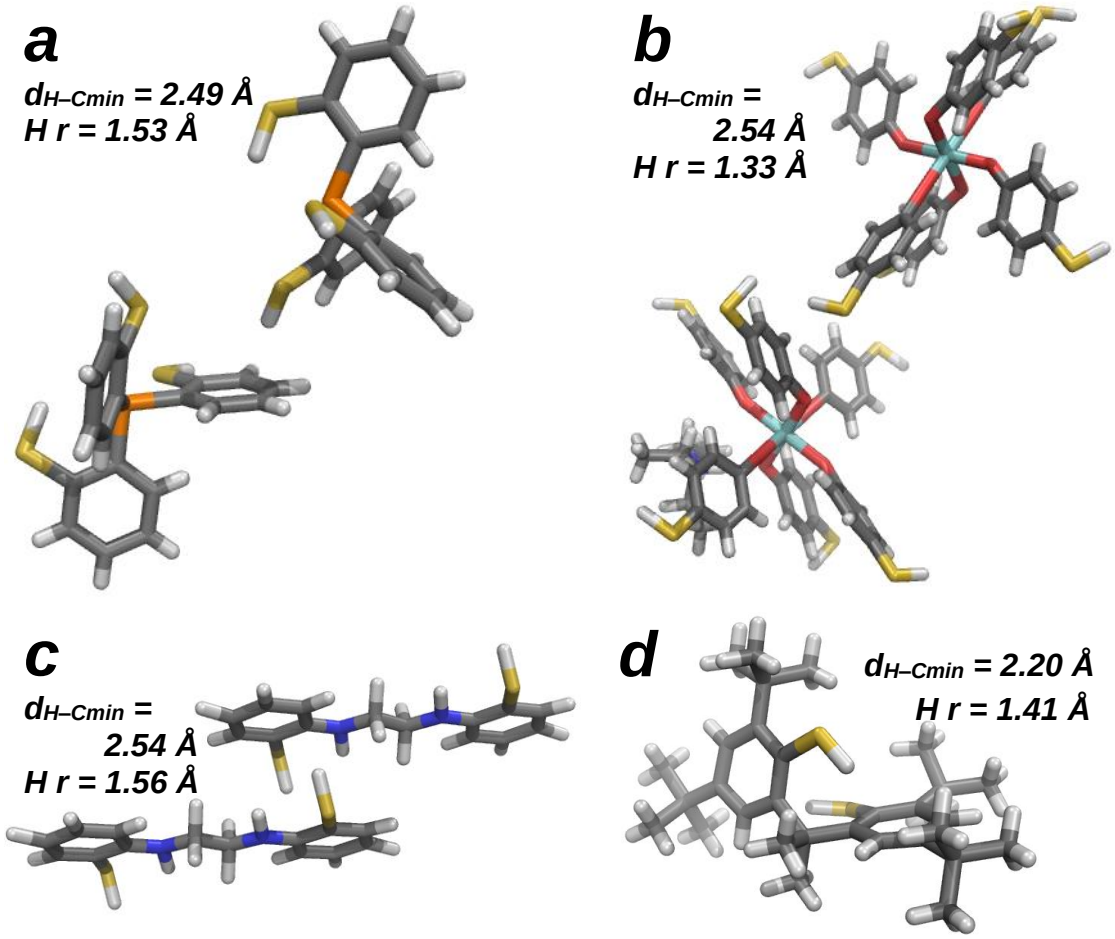


Figure 2.42 Close contacts in S–H/π aromatic interactions in crystal structures from the CSD
Six of the 20 shortest S–H/π aromatic contacts in crystal structures from the CSD involve hydrogen atoms positioned over the aromatic ring carbon. Four of these crystal structures from the CSD with close S–H/π contacts are shown: (a) COLDEN; (b) GAGTUF; (c) YULZUA; (d) WANCIX.
The ΣvdWr for carbon and hydrogen is 2.90 Å.[203]

Based on the observed favorable molecular orbital overlap interaction in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (Figure 2.34), the close contact distances and carbon-oriented geometries in crystal structures with S–H/π aromatic interactions from the CSD are also potentially due to an aromatic $\pi \rightarrow \sigma^*$

thiol orbital overlap interaction. Determining the direction of the thiol S–H bond provides an approximation for where the S–H σ^* orbital is localized, and the plausibility of an $\pi \rightarrow \sigma^*$ orbital overlap. If the close contacts in S–H/π aromatic interactions exhibit S–H bond directionality towards the aromatic ring carbon, then the aromatic $\pi \rightarrow \sigma^*$ thiol interaction is potentially general in S–H/π aromatic interactions. In order to examine this possibility, the vector of the S–H thiol bond was correlated with respect to the aromatic ring (Figure 2.43), based on the measurements S r, H r, $d_{S-plane}$ and $d_{H-plane}$. The S–H thiol bond for each entry from the CSD is represented as a vector, where the arrow points along the trajectory of the thiol bond towards the hydrogen atom location, as reported in the CSD.

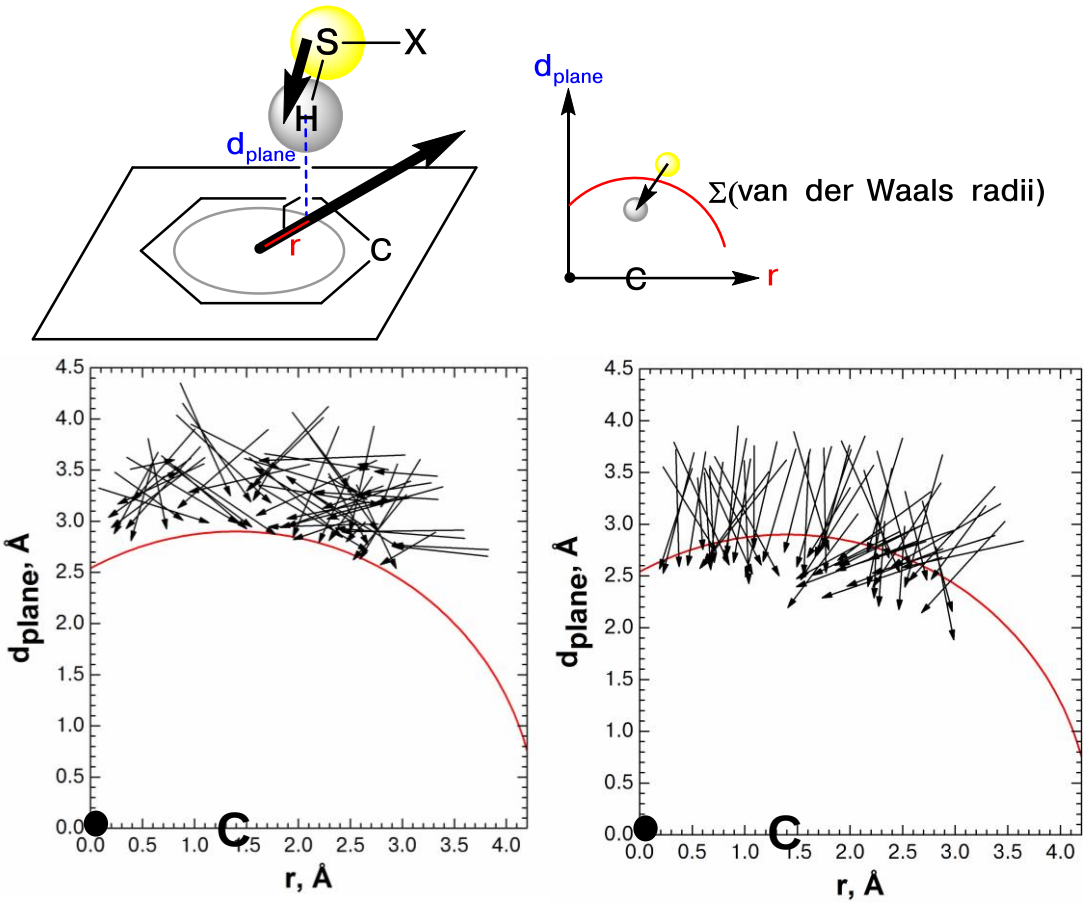


Figure 2.43 Direction of the thiol S–H bonds relative to the plane of the aromatic ring

The vector of the thiol S–H bond can approximate the location of the σ^* orbital in the context of the aromatic ring. The thiol S–H bond is represented as a vector, where the arrow points to the hydrogen atom location. The origin (●) and edge (C) of the aromatic ring are indicated within the plots, where C is located at 1.40 Å from the origin. The red line indicates the sum of the van der Waals radii for carbon and hydrogen (2.90 Å), with the origin at (0, 1.4).

Thiol hydrogen atoms are as reported in the CSD, using the hydrogen atom for reference in the cylinder restriction, and the data set was manually parsed (see experimental section and p. 96-97 for details).

Left: Only thiol S–H bonds where d_{H-Cmin} was greater than 2.90 Å (ΣvdWr for carbon and hydrogen)

Right: Only thiol S–H bonds where d_{H-Cmin} was less than 2.90 Å (ΣvdWr for carbon and hydrogen)

The geometric differences between the short and long range contacts in S–H/π aromatic interactions were quite clear from the vector diagrams: no discernible trend was observed for the long range contacts, and many of the close contacts were directed towards the aromatic ring carbon. The trend of carbon-directed, sub-ΣvdWr distances for S–H/π aromatic interactions in crystal structures from the CSD strongly suggests that the molecular orbital overlap basis for interaction is general.

As noted earlier in this section and others, the hydrogen atom location in x-ray crystal structures can be unreliable, due to the low electron density around hydrogen atoms and biasing for shorter X–H bond lengths towards the heavier atom.[203] To correct for the bond length bias, it is generally advised that short X–H bond lengths should be normalized to a standard bond length, 1.338 Å for S–H bond.[203] Histogram charts and comparisons between the atomic coordinates relative to the aromatic ring were also generated based on the data set containing S–H/π aromatic interactions with normalized S–H bond lengths. These comparisons are shown in Figure 2.44 and 2.45.

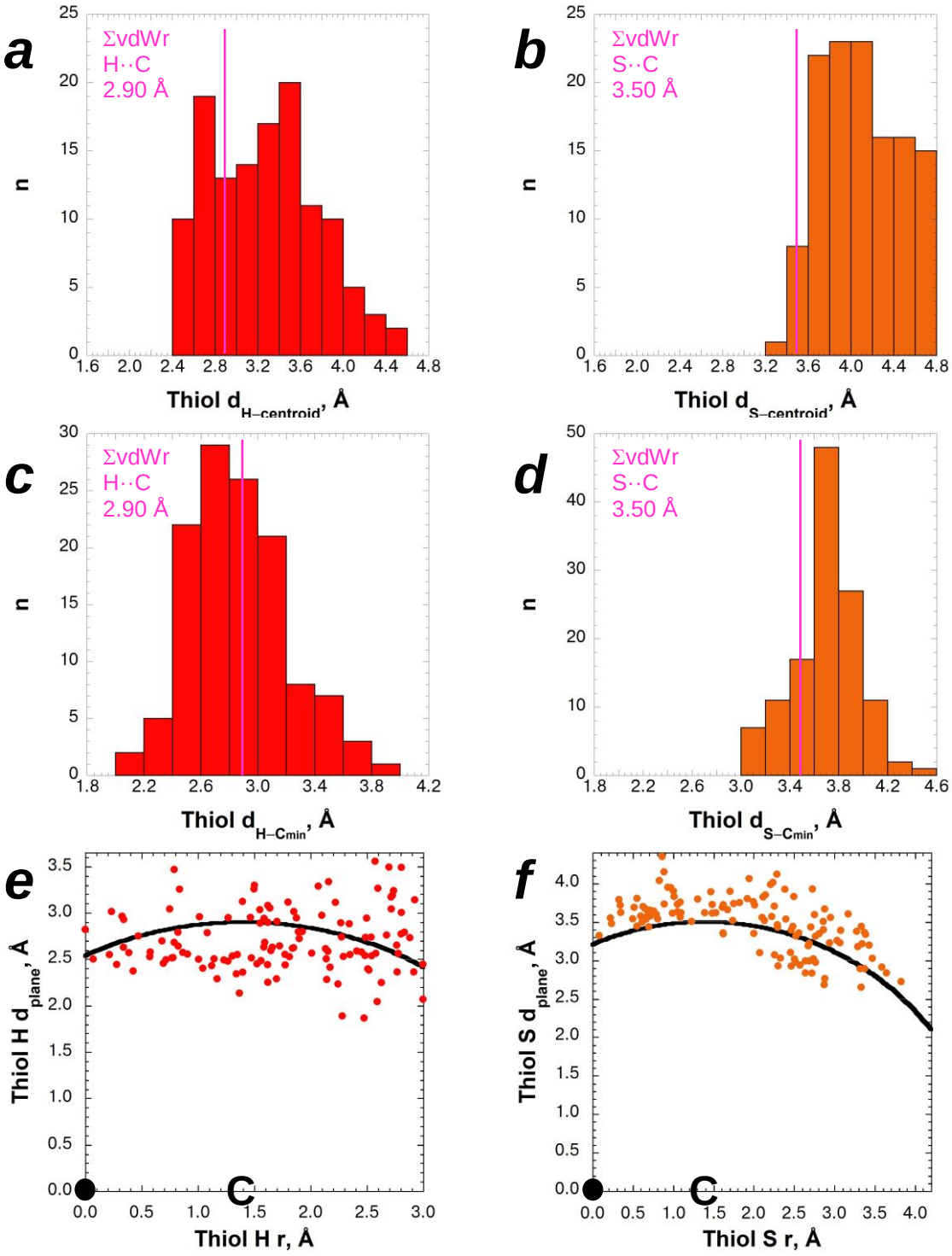


Figure 2.44 S–H/π interactions in the CSD with normalized S–H bond lengths: Histogram charts and correlation of atomic coordinates with respect to the

aromatic ring

The thiol hydrogen atom coordinates were adjusted to that the S–H bond length was normalized to 1.338 Å along the vector of the S–H bond (normalization conducted using ConQuest).[203] The hydrogen atom was used for reference in the cylinder restriction ($d_{H-plane} \leq 3.63$ Å, H r ≤ 3.00 Å). (a, b) Histogram charts for the frequency of distances from the aromatic centroid to the interacting atom (thiol H or thiol S, respectively); (c, d) Histogram charts for the frequency of distances from the interacting atom (thiol H or thiol S, respectively) to the nearest aromatic carbon atom; (e, f) the atomic coordinates of the interacting atom (thiol H or thiol S, respectively) are indicated with reference to the plane of the aromatic ring. The distance $d_{X-plane}$ is correlated against the distance X r (radius). The X-axis represents the plane of the aromatic ring, and the origin represents the aromatic centroid. The location of aromatic centroid (●) and ring carbons (C) are indicated within the plots, where **C** is located at 1.40 Å from the origin. The black line indicates the ΣvdWr for carbon and the respective interacting atoms; points below the black line are less than the ΣvdWr. The data set for S–H/π aromatic interactions used in these projections were restricted on the hydrogen atom ("cylinder restriction," see experimental section and Figure 2.37 for details) and manually parsed for geometric biases and anomalies; the S–H bond lengths were normalized to 1.338 Å.

ΣvdWr for carbon and hydrogen, 2.90 Å[203]

ΣvdWr for carbon and S, 3.50 Å[203]

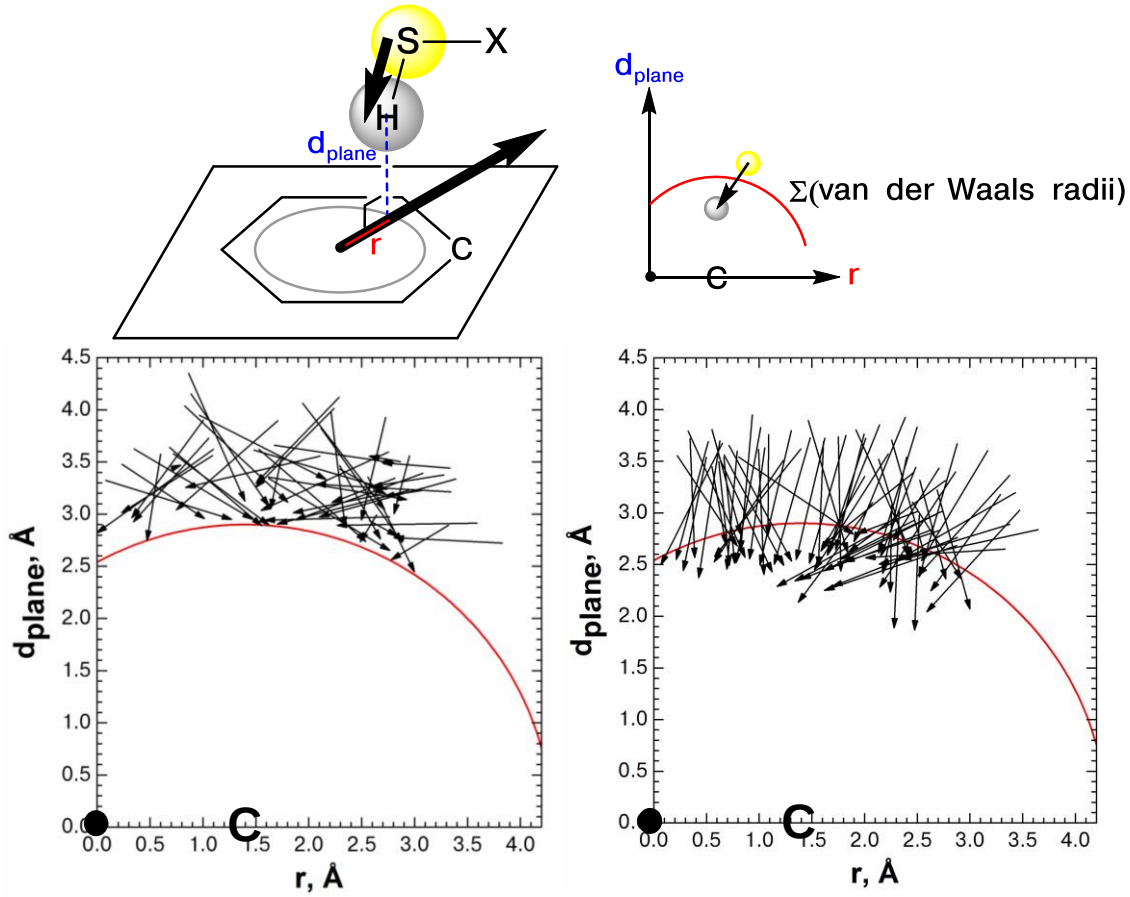


Figure 2.45 Direction of the normalized thiol S–H bonds relative to the plane of the aromatic ring

S–H bond can approximate the location of the σ^* orbital in the context of the aromatic ring. The thiol S–H bond is represented as a vector, where the arrow points to the hydrogen atom location. The origin ( ) and edge (**C**) of the aromatic ring are indicated within the plots, where C is located at 1.40 Å from the origin. The red line indicates the sum of the van der Waals radii for carbon and hydrogen (2.90 Å), with the origin at (0.0, 1.4).

The coordinates of the thiol hydrogen atoms were adjusted so that the S–H bond length was 1.338 Å, and the data set was limited to the region of interaction using the hydrogen atom for reference in the cylinder restriction, and the data set was manually parsed (see section 2.4.10 for details).

Left: Only thiol S–H bonds where d_{H-Cmin} was greater than 2.90 Å (ΣvdWr for carbon and hydrogen)

Right: Only thiol S–H bonds where d_{H-Cmin} was less than 2.90 Å (ΣvdWr for carbon and hydrogen)

The trends in distributions and atomic coordinates relative to the plane of the aromatic ring were reasonably consistent between thiol structures with reported and normalized S–H bond lengths. For the data set where the S–H bond length was normalized, there were substantially more close contacts between hydrogen atoms and the aromatic ring. The mean distance $d_{H–Cmin}$ was 2.87 Å, less than the ΣvdWr for carbon and hydrogen (Table 2.18), and the histogram plot of this distance $d_{H–Cmin}$ showed a high distribution of distances near or below the ΣvdWr. In the diagrams representing the S–H bond as a vector over the plane of the aromatic ring, significantly more of the close contacts are carbon-oriented, including many examples of the longer range contacts at the ΣvdWr threshold also exhibit carbon-directed geometry (Figure 2.45).

Taking into consideration the intrinsic bias of electron density maps towards heavy atoms, and normalization of the S–H bond length to standard values, there is a significant trend of carbon-oriented geometry in S–H/π aromatic interactions in crystal structures from the CSD. The carbon-oriented geometries are consistent with a general aromatic $\pi \rightarrow \sigma^*$ thiol molecular orbital overlap interaction that stabilizes the S–H/π interaction, consistent with our observations in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. The carbon-oriented geometries observed in S–H/π aromatic interactions are distinct from the centroid-oriented geometry for large cation/π aromatic interactions in crystal structures, potentially due to a greater electrostatic effect that stabilizes cation-aromatic interactions (which overwhelms any orbital interactions). Li^+/π aromatic interactions also exhibited unique geometric preferences in comparison to large cation/π interactions, with many examples of sub-ΣvdWr distances that were carbon-oriented. Although Li^+ cations are much smaller

than Na^+ and K^+, which decreases the favorable dispersive interactions, it is remarkable that these small cations were so frequently localized below the ΣvdWr, and outside of the radius of the aromatic ring. The periodic trend from carbon-oriented to centroid-oriented cation/π aromatic interactions with increasing ionic radius suggests that additional factors may play a role in the nature and geometric preferences of these interactions, potentially attributed to the "hardness" or "softness" of the interacting cation (described further in Chapter 2.3.5).

2.3 Discussion

The nature of X–H/π interactions has been commonly described as an interplay of electrostatic and dispersive forces, which largely depend on the donor and acceptor atom electronegativity.[190] However, dispersion does not account for the strength and directionality that is observed in non-conventional hydrogen bonds, such as halogen bonds or S–H/π interactions. S–H/π interactions have been likened to conventional hydrogen bonds, and thus compared to O–H/π or N–H/π interactions where the electrostatic driving force is much greater. However, if the nature of S–H/π interactions is fundamentally different than O–H/π or N–H/π interactions, then a different perspective may be necessary in considering these aromatic π interactions. This present study of the intermolecular interactions in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester provided insights into the underlying energy and fundamental driving forces of S–H/π aromatic interactions, and how these interactions provide unique opportunities in design applications.

2.3.1 Crystallographic observation of an intermolecular S–H/π aromatic interaction

The x-ray crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was solved to 0.77 Å resolution. Electron density around the thiol and carbamate nitrogen allowed for location of these hydrogen atoms (Figure 2.15). The bond thiol S–H bond length was calculated to be 1.26 Å, shorter than the standard S–H bond length, 1.338 Å.[203] The C–S–H bond angle was measured to be 99°, typical for thiols, with the thiol bond nearly co-planar with the aromatic ring (10° out of plane). Intermolecular hydrogen bonds were observed in the crystal structure, including the Boc- carbonyl oxygen with both the carbamate hydrogen and with an aromatic hydrogen (O··H distances 2.66 Å and 2.75 Å, respectively). Interestingly, the thiol group in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester did not participate in a conventional hydrogen bond interaction with either available carbonyl oxygen atom or the ester oxygen atom.

The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester revealed a close intermolecular contact between the thiol proton and an aromatic ring. Alignment of the thiol S–H bond towards the aromatic ring is considered "ideal" for an S–H/π interaction, with the assumption that the interaction is driven by electrostatic effects. In the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, the interaction of the thiol and the aromatic centroid was 2.90 Å, at the sum of the van der Waals radii (ΣvdWr) for carbon and hydrogen atoms. The alignment of the thiol S–H bond with the aromatic centroid was 148°. However, we observed a closer interaction with an individual ring carbon, at a distance of 2.71 Å, well below the ΣvdWr for carbon and hydrogen. The alignment of the thiol S–H bond with the ring carbon was also more linear at 159°. While linearity of an interaction is not a direct indicator of the strength of an interaction, some have suggested that linearity between a donor bond

and acceptor group is ideal geometry for hydrogen bonds.[190, 225] In addition to the thiol-aromatic interaction, a close intermolecular interaction was also observed between one of the aromatic C–H bonds and an individual aromatic carbon atom (Figure 2.46a), at a distance of 2.84 Å. The combined S–H/π and C–H/π interactions in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester allowed for an intermolecular, bidentate aromatic interaction (Figure 2.46), which stabilized the overall crystal packing within the unit cell. For both of the observed S–H and C–H/π aromatic interactions, there was a greater preference for interaction with the edge of the aromatic ring face over the centroid, suggesting a stronger interaction with the ring carbons.

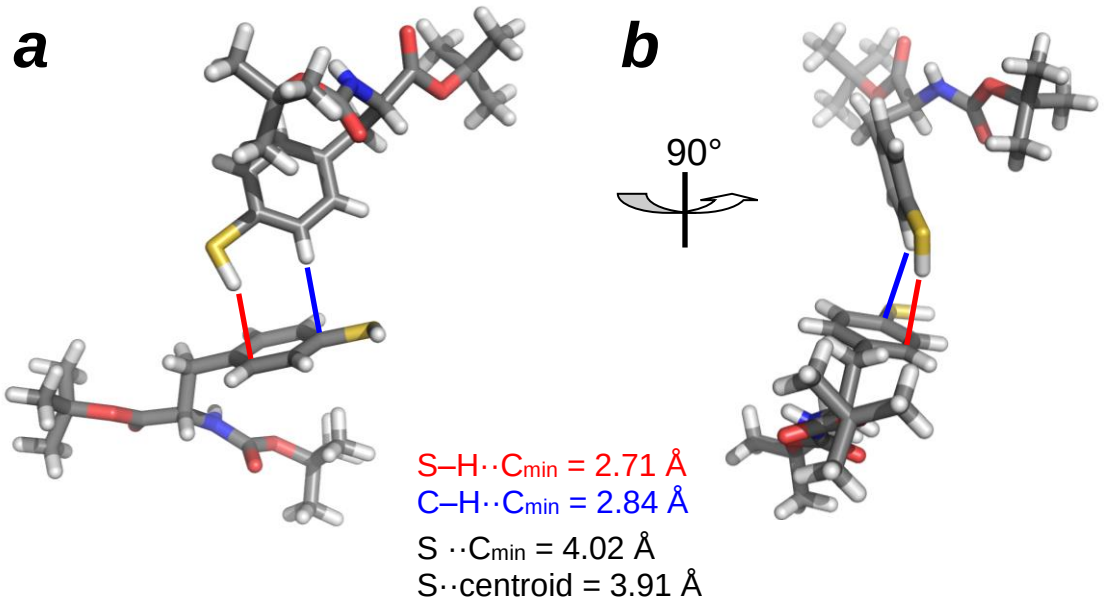


Figure 2.46 Intermolecular aromatic interactions in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester
The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester revealed intermolecular X–H/π contacts between aromatic rings, with distances that were less than the ΣvdWr for carbon and hydrogen. The alignment of the S–H and C–H bonds are directed towards individual aromatic ring carbon atoms, rather than the centroid of the aromatic ring. Perspectives between (a) and (b) are rotated about the y axis 90°. Due to intrinsic errors in hydrogen atom location from electron density maps, distance measurements between heavy atoms can be compared. Consistent with the observations with the hydrogen atom, the sulfur atom is more closely associated with the edge of the aromatic ring face over the aromatic centroid.

As with any crystal structure obtained from x-ray diffraction data, there is intrinsic uncertainty about location of hydrogen atoms, which have low electron density. The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester had some electron density around the sulfur atom that allowed for approximate location and directionality of the thiol S–H bond (as well as the carbamate amide hydrogen). However, the electron density is biased towards the heavier atom, and the calculated X–H bond lengths from x-ray diffraction generally appear shorter, particularly with

weaker hydrogen bonds.[203] To account for this bias in x-ray crystallographic data, it is recommended that the calculated are normalized to a standard bond length, where the hydrogen atom is shifted along the vector of the bond to the appropriate location.[203] With correction of the thiol S–H bond length to 1.338 Å,[203] the interaction between the thiol hydrogen atom and aromatic ring carbon is even closer, at 2.63 Å, while the distance of interaction with the aromatic centroid is nearly the same, at 2.85 Å (Figure 2.18). Alternatively, comparisons between interactions with only heavy atoms can be examined, which excludes any error in directionality of the thiol S–H bond. The distance between the thiol sulfur atom and the aromatic centroid was measured to be 4.02 Å, while the distance to the nearest aromatic carbon was shorter, 3.91 Å (Figure 2.46). Combined, measurements for both calculated and normalized crystal structures showed a closer association between the thiol and the edge of the aromatic ring face rather than the centroid.

The synthesis of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was also suggestive of unique interactions involving the thiol group. Boc-4-thiol-L-phenylalanine-*tert*-butyl ester formed disulfides in solution over a few days when allowed to stand in air, but as a crystalline solid, could resist air oxidation for a few weeks. A close intermolecular contact in the solid, crystalline 4-thiophenylalanine involving the thiol proton could potentially stabilize the thiol against reaction with oxygen in the air.

Desiraju & Steiner point out that hydrogen bond interactions with aromatic rings are more ambiguous in terms of the "acceptor."[190] It is generally followed that measurements and comparisons with X–H/π aromatic interactions are between the hydrogen atom and the aromatic centroid.[190, 225, 228] Measurement and comparison of

distances between a hydrogen atom and an aromatic centroid are potentially accurate for interactions that are primarily driven by electrostatic interactions. Indeed Desiraju & Steiner[190] described aromatic centroid-directed hydrogen bonds as "good hydrogen bond geometry," while an almost linear O–H··C_{aro} interaction, solved by neutron diffraction, was described as "more irritating geometry." However, we observed close S–H/π aromatic interactions within the crystal structure that were not centroid-oriented, with preferred geometry towards the edge of that aromatic ring face, at distances well below the ΣvdWr for carbon and hydrogen. Potentially, alternative measurements and comparisons may be required for interactions that are not driven by electrostatic interactions, since the preferred geometry can be quite different.

In order to examine the fundamental nature of S–H/π aromatic interactions, and to discern the underlying stabilizing features that dictate preferred geometries, Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was further characterized as a model compound via NMR, FT-IR, and *ab initio* calculations. Characterization of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester through these methods can be used to determine the strength and fundamental nature of the S–H/π aromatic interaction. Preferred geometries of thiol-aromatic interactions were compared in crystal structure deposited in the CSD, and were compared to cation-aromatic interactions, where the electrostatic effects are much greater.

2.3.2 Solution and solid-state NMR analysis of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester: A significant down-field chemical shift in the thiol proton in solid-state

In the ^{1}H NMR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, the thiol proton was observed as a sharp singlet at 3.41 ppm in $CDCl_3$. Proton chemical shifts are sensitive to solvent, particularly for hydrogens that can participate in

hydrogen bonds. In methanol-d_3, a downfield chemical shift was observed in the thiol proton ($\Delta\delta$ = +0.59-0.66 ppm) for both Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol, indicative of a deshielding effect from the bulk solvent (as a result of hydrogen bonding with methanol). Benzene and aromatic solvents are known to interact with aryl thiol protons, based on prior IR and ^{1}H NMR studies. In benzene-d_6, the thiol proton shifted upfield for both Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol ($\Delta\delta$ = –0.34-0.44 ppm). An upfield chemical shift change is consistent with a shielding effect as a result of aromatic ring current effects,[166, 248] which can be distinctive from a conventional hydrogen bond. Having established the two different forms of interaction with the thiol proton via NMR, the nature of the interaction in the crystal structure was examined via solid state NMR.

In the crystal structure of *p*-thiocresol, the thiol group was indistinguishable from the methyl group by electron density, and the thiol interaction was obscured. In the solid-state ^{13}C-^{1}H-HETCOR spectrum, two different chemical environments were observed for the thiol proton in *p*-thiocresol, at 3.2 ppm and 5.0 ppm. These two different environments can potentially result from two different modes of interaction, both "bound" (5.0 ppm) and "un-bound" (3.2 ppm) forms of the thiol proton. In the ^{13}C-^{1}H-HETCOR spectrum of deuterated *p*-thiocresol, these peaks were not present, further validating the assignment that both of these resonances are due to the thiol proton.

Examination of the ^{13}C-^{1}H-HETCOR spectrum of crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester showed the thiol proton at 6.2 ppm. The thiol proton was significantly shifted downfield in comparison to the sample in $CDCl_3$ ($\Delta\delta$ = +2.8 ppm), consistent with a relatively strong hydrogen bonding interaction. Downfield

chemical shifts can also result from changes in the acidity of the thiol proton due to changes in the electronics of the aromatic ring.[166, 264] Downfield chemical shifts also result from deshielding effects from proximal aromatic rings, which are strongest when the proton is within the plane of the aromatic ring. The carbamate proton was also significantly shifted downfield ($\Delta\delta$ = +2.4 ppm), consistent with the observed hydrogen bond in the crystal structure. Comparing all of the chemical shift changes in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester between crystalline and solution samples, the most significant shifts involved the thiol, carbamate, and carbonyl carbons, which are consistent with the observed hydrogen-bonding interactions in the crystal structure (Figure 2.47).

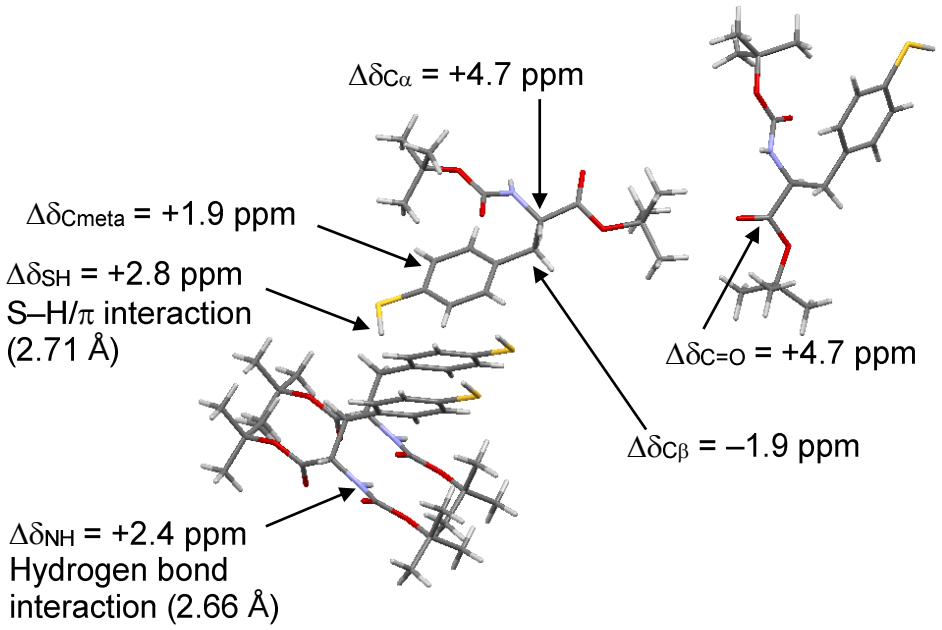


Figure 2.47 Chemical shift differences in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester between solid-state and solution NMR
Chemical shift changes greater than 1.5 ppm between solid and solution NMR in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. Distances and interactions are indicated where identified.

Strikingly, the chemical shifts of the thiol protons were shifted downfield in the crystalline form compared to solution samples, in both *p*-thiocresol and Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. The solution data in benzene-d_6 suggested that the thiol proton would shift slightly upfield (shielding effect) in the presence of aromatic interacting groups, but instead the interaction in the solid state deshielded the thiol proton. Aromatic groups can be deshielding if the interacting proton is in the plane of the aromatic ring, outside of the ring, directed towards the edge of the ring face, an edge-on interaction. Alternatively, the thiol proton could potentially have a different chemical shift due to changes in the energy levels of its own aromatic ring, resulting from participation in the intermolecular interaction.[264] While the downfield

chemical shift change in the thiol proton within solid-state samples was significant and consistent with an hydrogen bonding interaction, potentially due to an aromatic interaction, identification of the fundamental nature of the thiol-aromatic interaction would require further studies via vibrational spectroscopy and *ab initio* calculations.

2.3.3 Vibrational spectroscopy of the thiol S–H bond in solution and solid Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

Hydrogen bonding interactions have been extensively studied via FT-IR, where the X–H bond stretching frequency is sensitive to the nature and strength of non-covalent interactions, either as self-association or with bulk solvent.[208, 232-234] The extensive prior literature on thiol interactions by IR allows for ready comparison to published data on thiophenol, *p*-thiocresol, and solid samples containing thiols.[230-234] The S–H stretching frequency (ν_{S-H}) occurs where there are few other vibrational frequencies (2590-2550 cm^{-1}), and can be readily distinguished from other vibrations. The ν_{S-H} can red shift substantially in the presence of interacting solvents, and the absorbance intensity generally increases with greater $-\Delta\nu_{S-H}$.[208]

In order to examine the nature of the thiol interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, a series of solution samples were prepared using CCl_4 with cosolvents containing different hydrogen bonding partners, including carbonyl groups (acetone and ethyl acetate), an alcohol (methanol), or an ether (THF). Several solutions were prepared containing different amounts of cosolvent. These IR spectra for solution samples were compared against the IR spectrum of the crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in a pressed pellet of KBr. IR spectra for *p*-thiocresol in solution and as crystalline samples were also obtained for comparison.

A direct correlation of the absorbance intensity of ν_{S-H} was observed with respect to the amount of "interacting" cosolvent in CCl_4 for both Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol. The ν_{S-H} red shifted significantly with all cosolvents, although the increase in ν_{S-H} absorbance was only observed with ethyl acetate and acetone. The red shift in ν_{S-H} for solutions of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was –18 cm^{-1} in 25% ethyl acetate and –26 cm^{-1} in 25% acetone. *p*-Thiocresol exhibited similar $\Delta\nu_{S-H}$ in these solvents containing carbonyl groups (–16 cm^{-1} in 25% ethyl acetate; –25 cm^{-1} in 25% acetone). The increased $-\Delta\nu_{S-H}$ for solutions of acetone/CCl_4 compared to solutions of ethyl acetate/CCl_4 is potentially due to the increased dipole moment of the carbonyl in acetone, and a stronger S–H/O interaction. For both compounds, solutions in methanol/CCl_4 and THF/CCl_4 showed the lowest ν_{S-H}, but the absorbance intensities were considerably lower and the signals were much broader, suggesting that the thiol interaction was not well defined. Interestingly, the aromatic C-C frequency at 1500 cm^{-1} blue shifted substantially ($\Delta\nu_{C-C}$ = 20 cm^{-1}) in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, indicating an effect on "semi-circle" stretching in the aromatic ring. Not only is the thiol stretching frequency influenced by interaction with an aromatic ring, but the aromatic ring is also influenced by interaction with the thiol group. Interestingly, the IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in $CHCl_3$, which is known to interact favorably with aromatic rings via C–H/π interactions, also exhibited an additional blue-shifted band for this frequency.

Crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester exhibited a substantial red shift in ν_{S-H} (–47 cm^{-1} compared to $CHCl_3$) with a 15-fold increase in absorbance intensity. Crystalline *p*-thiocresol also showed a significant red shift in ν_{S-H} in

comparison to CCl_4 ($\Delta\nu_{S-H} = -22$ cm^{-1}). The increased red shift in ν_{S-H} for crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester over *p*-thiocresol suggests a stronger and more well-defined interaction, consistent with observations from x-ray crystal structures and solid-state NMR data. The greater red shift in ν_{S-H} for crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester over solutions with acetone or ethyl acetate suggests a fundamentally distinct, and potentially stronger interaction.

In the work by Saggu *et al.*,[233] the greatest red shift in ν_{S-H} for thiophenol was in the presence of hexamethylbenzene, with a $\Delta\nu_{S-H}$ of –40 cm^{-1} relative to CCl_4. Neat samples of thiophenol are reported to have a S–H stretch ν_{max} at 2569 cm^{-1} (similar to crystalline *p*-thiocresol, $\nu_{S-H} = 2563$ cm^{-1}), with a $\Delta\nu_{S-H}$ of –22 cm^{-1} compared to dilute solutions in CCl_4.[232] Between solution and solid forms of a compound containing an aryl thiolate, Rozenberg *et al.*[234] observed a red shift in ν_{S-H} of –19 cm^{-1}, consistent with formation of an S–H/π aromatic interaction in the crystal structure (Figure 2.48a). Jabłońska *et al.*[250] observed two ν_{S-H} in a crystalline silylthiol compound, ascribed to "interacting" and "non-interacting" forms of the thiol, separated by 31 cm^{-1} (Figure 2.48a). We observed a greater red shift in ν_{S-H} (–47 cm^{-1}) between crystalline and solution Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, consistent with a strong S–H/π aromatic interaction observed in the crystal structure (Figure 2.48c).

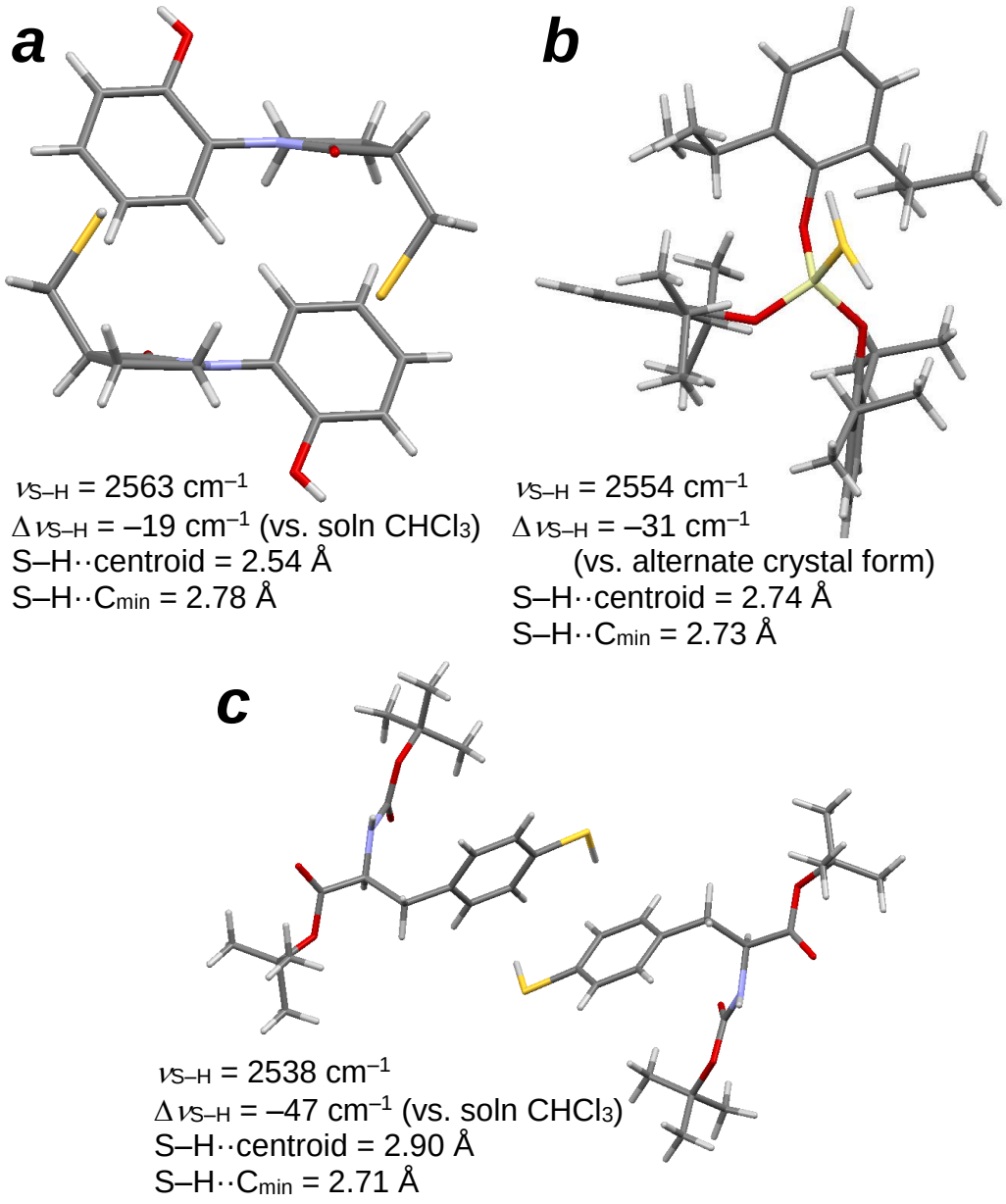


Figure 2.48 S–H/π aromatic interactions observed via crystallography and FT-IR

Thiol S–H/π aromatic interactions in crystal structures with corresponding solid-state thiol S–H stretching frequency. (a) N-(*o*-hydroxyphenyl)-3-sulfanylmethylpyrrolidin-2-one, viewed down the thiol S–H bond;[234] (b) Tris(2,6-diisopropylphenoxy)silanethiol, with two different thiol interaction modes shown in the crystal structure;[250] (c) Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (this work).

The experimentally observed differences in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester between solid and solution forms can be compared to calculated ν_{S-H} using *ab initio* calculations. The ν_{S-H} relates to a thiol S–H bond length (influenced by strong non-covalent interactions), which can be calculated and compared to the crystallographic observations. In addition, atomic charges and molecular orbitals can be calculated from the crystal structure, and an energy component analysis can be conducted on the observed S–H/π aromatic interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

2.3.4 *ab initio* calculations on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester: an orbital overlap interaction stabilizes the intermolecular S–H/π aromatic interaction

Both ^{1}H NMR and FT-IR studies on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester suggested that the crystal form was stabilized by an intermolecular interaction involving the thiol S–H bond, which differed from conventional hydrogen bond interactions. A significant downfield chemical shift was observed between solution ($CHCl_3$) and solid-state ^{13}C-^{1}H-HETCOR spectrum ($\Delta\delta$ 2.6 ppm), and a significant ν_{S-H} red shift ($\Delta\nu_{S-H}$ = –47 cm^{-1}) was observed via solid-state IR. For both NMR and FT-IR experiments, Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in the crystalline form exhibited greater shifts than any of the solvent effects examined, indicating a uniquely strong interaction. The S–H bond observed in the x-ray crystal structure was not oriented towards the aromatic centroid, but towards an individual aromatic ring carbon, suggesting that the S–H/π aromatic interaction was stabilized by an orbital overlap interaction between the thiol S–H σ* and the aromatic π orbitals.

ab initio calculations have been used to characterize X–H/π aromatic interactions through calculation of atomic charges, geometry and energies of

molecular orbitals, and the overall energy of stabilizing interactions.[164, 228, 265] Computational studies on thiol interactions have examined the energies and geometry S-H/O=C interactions,[255] S/π aromatic interactions,[238] or H_2S/π aromatic interactions.[164, 254, 266, 267] While thiol interactions are often overlooked and considered weak and driven by dispersion, calculations have shown that the interaction between hydrogen sulfide and benzene is 2.85 kcal mol^{-1}, similar in energy to the ammonia-benzene interaction.[254, 267] Notably, in calculations on hydrogen sulfide with indole or benzene rings, both hydrogen atoms are directed towards the aromatic ring, so that hydrogen sulfide interacts in a bidentate manner (increasing the overall stability of the complex).[164, 252, 254] However, thiol S–H/π aromatic interactions have not been well characterized via *ab initio* calculations, and the unique geometry and strength observed in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester provided an ideal model system for further computational study.

The crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was truncated to a *p*-thiocresol dimer, in order to save computational time, and the atomic charges and molecular orbitals were calculated using multiple basis sets with the aid of our collaborators, Dr. Sudipta Sinha and Dr. Sandeep Patel. Comparisons were made between the two thiol groups in the *p*-thiocresol dimer, where one thiol interacted with the aromatic ring, and the other did not (Figure 2.49). Natural bond orbital (NBO) analysis was conducted on the *p*-thiocresol dimer to identify the extent of orbital overlap between the σ^* and the aromatic π orbitals (Figure 2.49). Indeed, a favorable overlap interaction was observed, with an interaction energy of 0.70-1.45 kcal mol^{-1} (depending on the basis set used for calculation). The energy of interaction in vacuum due to electrostatics was calculated to be 0.72-1.07 kcal mol^{-1}, less than the interaction

due to aromatic π→S–H σ^* interaction. These calculations indicate the observed intermolecular S–H/π aromatic interaction in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was primarily stabilized by a favorable orbital overlap interaction, which is consistent with observed alignment of the S–H bond towards an aromatic ring carbon atom (where the aromatic orbitals are localized).

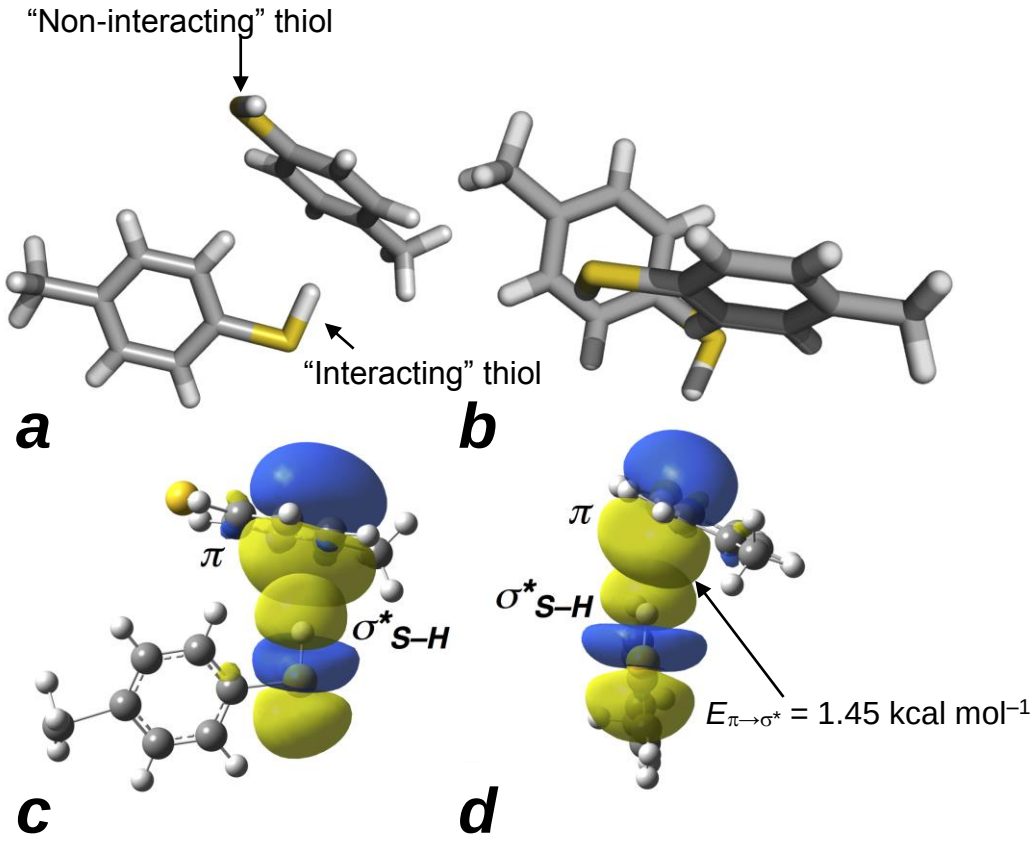


Figure 2.49 A molecular orbital overlap interaction that stabilizes an intermolecular S–H/π aromatic interaction in crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester
Different views of the aromatic π→σ* thiol interaction in the *p*-thiocresol dimer, a truncation of the x-ray crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. (a,b) Different views of the *p*-thiocresol dimer used in calculations; (c,d) different views of the favorable orbital overlap interaction that stabilized the intermolecular interaction, calculated to have an interaction energy of 0.70-1.45 kcal mol^{-1} using various basis sets.

Desiraju & Steiner state that hydrogen bonds are stronger and more idealized as an interaction D–H··A becomes more linear (where angle DHA approaches 180°).[191] They argue that with decreasing linearity of the hydrogen bond there is a decrease in the electrostatic component of the interaction, and the “hydrogen bond”

becomes a non-directional van der Waals interaction with increasing deviation from linearity.[190, 191] The strength of the intermolecular S–H/π interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was quite significant, with an observed $\Delta\nu_{S-H} = -47$ cm^{-1} between the solution and crystalline forms. In the crystal structure, the thiol S–H bond aligned more directly with an aromatic carbon, rather than the aromatic centroid. Desiraju & Steiner described carbon-oriented geometry of X–H/π interactions:[190]

> *"In the structure of the trifluoroethanol **50**, pairs of molecules are linked by mutual aromatic hydrogen bonds with a relatively well-centred geometry…This can be considered as a good hydrogen bond geometry. A more irritating geometry is found in the low-temperature neutron crystal structure of alkynol **51**, where a hydroxyl group points almost linearly at an individual carbon atom of the acceptor group…"*

In these examples, the structure with "irritating" carbon-oriented geometry was confirmed via IR to be a rather strong interaction, with a $\Delta\nu_{OH\ stretch}$ of –61 cm^{-1}. The calculations performed on these two structures, centered and off-centered geometries, showed that the centered geometry was energetically more favorable, and the off-centered geometry was labelled as curiosity. The *ab initio* calculations on the thiol-aromatic interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester supported the notion of a carbon-oriented interaction, as the interaction that stabilized the *p*-thiocresol dimer was found to be predominantly molecular orbital interactions over electrostatic effects.

In a PDB analysis of serine- and cysteine-aromatic interactions, the preferred orientations of phenylalanine aromatic rings were different between serine and cysteine,[247] indicating different means of energetic stability. That is, serine O–H/π interactions are potentially driven by greater electrostatic effects, while cysteine S–H/π interactions are potentially driven by molecular orbital interactions, which

manifests as distinctive geometric preferences between serine or cysteine and phenylalanine. The underlying energetic contributions that stabilize donor-acceptor interactions can have significant consequences in the geometry of interaction, which can be implicated in understanding protein folding and protein-protein interactions, in designing small molecule drugs, and in designing mutant enzymes with altered or enhanced functions. For example, designed small molecule protein inhibitors that interact through X–H/π aromatic interactions will have slightly different interaction geometries, depending on the relative contributions of electrostatic or molecular orbital effects through the X–H bond. These subtle differences in interaction geometry can affect the overall inhibitor binding energy and the efficacy of the inhibitor. In addition, the extent of molecular orbital overlap in aromatic π→S–H σ^* interactions can potentially affect the electron distribution of the aromatic ring, which can have consequences in chromophores or aromatic rings involved in electron transfer. Therefore, the roles that S–H, O–H, or N–H bonds play in modulating enzyme activity through X–H/π aromatic interactions can depend on the favorable molecular orbital overlap interactions.

2.3.5 Comparison of interaction geometries in S–H/π and cation/π aromatic interactions in crystal structures from the CSD

The x-ray crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester exhibited carbon-oriented geometry that was stabilized by an aromatic π→σ^* thiol orbital interaction. Our analysis of this crystal structure, and associated characterization in solution and in the solid state by FT-IR and NMR, is one example of a thiol-aromatic interaction with carbon-oriented geometry. In order to determine if our observations were general to all thiol-aromatic interactions, we searched for

crystal structures in the CSD containing S–H/π aromatic interactions and analyzed the distances and geometries of the interactions. In prior analyses of X–H/π aromatic interactions in crystal structures from the CSD and PDB,[225, 228, 234] a close contact distance between the centroid of the aromatic ring and the interacting atom served as a basis for the presence, and strength, of an interaction. However, based on our observations of the S–H/π interaction in Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, measurements and distances based on a centroid-oriented geometry may not be the most reasonable basis for determining interaction strength, particularly when electrostatic effects are not predominant. Nishio has previously examined the geometry of centroid-oriented versus carbon-oriented C–H/π aromatic interactions in crystal structures,[260, 261] but structural and geometric analysis in S–H/π aromatic interactions has never been established.

Based on the model we developed to describe the observed geometry in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, molecular orbital interactions predominate over electrostatic effects in S–H/π interactions, which manifests as a carbon-oriented geometry (where alignment for the orbital interaction is most favorable). A higher frequency of close-range contacts that are carbon-directed in S–H/π aromatic interactions in crystal structures from the CSD can support this hypothesis. In contrast, in interactions that are driven by electrostatic effects, such as the interaction between a cation and an aromatic ring, our model predicts that centroid-oriented geometry should be preferred because there is greater electronegative potential in the center of the aromatic ring rather than on the edges of the ring face (Figure 2.35). In our analysis of crystal structures from the CSD, the distances and measurements of S–H/π aromatic interactions (driven by molecular

orbital overlap) were compared to cation/π interactions (driven by electrostatic effects) to determine if the geometries differed between these two types of non-covalent interactions.

The structural analysis of crystal structures from the CSD demonstrated that S–H/π and Li^+/π interactions exhibited carbon-oriented geometry, particularly in crystal structures with the closest contact distances. 49% of entries containing S–H/π aromatic interactions from the CSD had hydrogen-carbon distances that were less than the sum of the van der Waals radii (sub-ΣvdWr), and this percentage is greater when the S–H bond length is corrected to a standard value[203] (normalized, 58%). Li^+/π interactions were localized outside of the aromatic ring, suggesting competitive interactions with other functional groups in the crystal structures. In contrast, K^+/π interactions exhibited centroid-oriented geometry, and did not interact as closely with the aromatic ring as S–H/π interactions. While K^+ has a significantly larger radius than Li^+ or hydrogen, significantly fewer structures exhibited sub-ΣvdWr distances between K^+ and the aromatic ring, regardless of centroid- or carbon-oriented geometry (Figure 2.50). Although it was initially expected that most cations would be localized near the aromatic centroid, the trend for centroid-oriented geometry was periodic, $K^+ > Na^+ > Li^+$, with a significant percentage of Li^+/π interactions that were carbon-directed (40% entries with $d_{Li+–Cmin} < \Sigma vdWr$ for lithium and carbon).

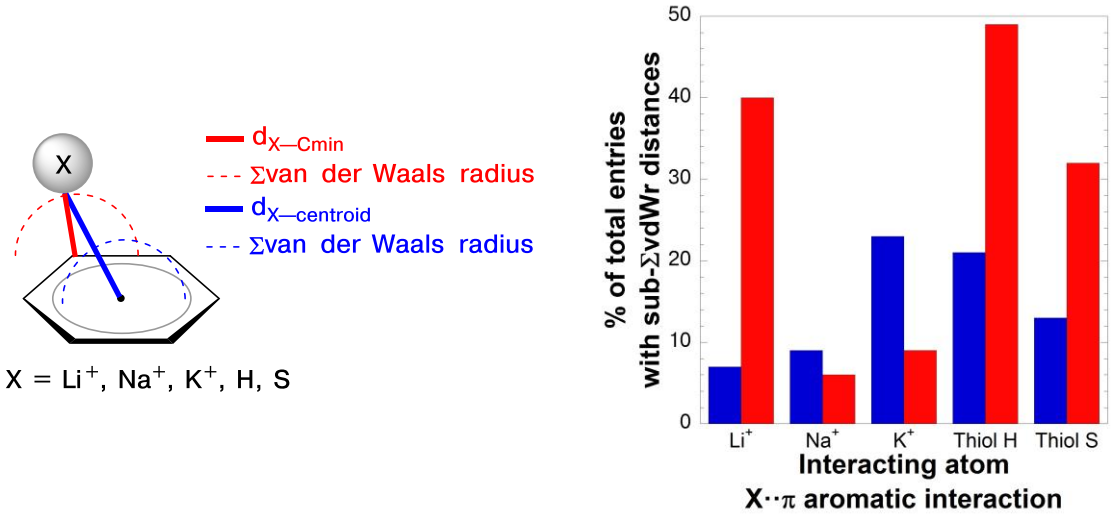


Figure 2.50 Percent of distances centroid-directed and carbon-directed geometries that are sub-ΣvdWr in cation/π and S–H/π interactions
Frequency of entries in the CSD where the distance $d_{X\text{-centroid}}$ (blue) or $d_{X\text{-Cmin}}$ (red) are sub-ΣvdWr (less than the sum of the van der Waals radii for carbon and the respective atom, X = Li^+, Na^+, K^+, thiol hydrogen, or thiol sulfur). The data sets of CSD entries shown are after restricting entries to a region of interaction (using reported hydrogen atom coordinates as a reference for thiol structures) and after manual parsing.

·Kochi and coworkers[251] have previously examined the geometric preferences for cation/arene interactions using x-ray crystallography. In this work,[251] alkali metals, Na^+, K^+, Rb^+ and Cs^+, were crystallized with hexakis(methoxymethyl)benzene, and the distance and localization of the cations were measured relative to the aromatic ring face. In these cation complexes, the C–O–C ethers coordinated the alkali metal over the plane of the aromatic ring. All cations studied oriented towards the aromatic ring, and the larger cations oriented with less favorable geometry to the ether lone pairs with greater geometric preference towards the aromatic ring. K^+ cations were oriented over the aromatic centroid, while Rb^+ and Cs^+ cations interfaced with the aromatic ring over individual carbon atoms. The difference in binding geometry was described as "π-modality" (centroid-oriented) or "σ-modality" (carbon oriented).

The observation of different geometric preferences for these alkali metal cation/π interactions can potentially be attributed to Pearson's hard-soft acid-base theory (HSAB).[268, 269] Atomic "hardness" is a trend of electron affinity and ionization potential, and relates to the energies of valence electrons and vacant orbitals, which describes trends in non-covalent interactions, ligand coordination to metals, and regioselectivity in nucleophilic substitution reactions.[270, 271] "Hard" atoms have higher energy outer vacant orbitals (LUMO orbital) with lower energy valence electrons (HOMO orbital), and are more likely to interact with other "hard" atoms.[269] "Soft" atoms have lower energy vacant orbitals with higher energy valence electrons, and are more likely to interact with other "soft" atoms. The "hardness" of an atom or functional group is calculated by (ionization potential – electron affinity)/2, or more generally $(E_{HOMO} - E_{LUMO})/2$.[269, 272] As a general trend, hard-hard or soft-soft associations are more stable than soft-hard or hard-soft interactions.[269, 272] Softer atoms have greater flexibility in electron redistribution and orbital interactions may be more important for interaction, while electrostatic contributions are more important in reactivity and interactions for harder atoms.[269, 270, 272] Alkene or aromatic π orbitals have higher energy occupied molecular orbitals than alkanes, and C=C bonds behave as "soft" functional groups.[269] The chemical "softness" of aromatic rings, alkenes, nitriles, carbonyls, and alkynes is part of the explanation for why these functional groups are excellent for coordination with "softer" transition metals, such as Au^{+} or Pd^{+2}.[270] "Hard" acids, such as Fe^{+3} or Mg^{2+}, coordinate preferentially with "hard" bases, such as hydroxides or ethers.[270] Chemical "hardness" can also describe or predict the reactivity of functional groups in organic reaction mechanisms, such as stability of disulfides (soft-soft S–S bond) over sulfenyl esters (soft-hard S–O bond),

or the regioselectivity of nucleophiles with propiolactone (hard nucleophiles attack the carbonyl carbon, soft nucleophiles attack an alkyl carbon).[271]

The chemical hardness of the alkali acids examined in this work progresses as $Li^+ > Na^+ > K^+$, which correlates with the observed geometric preferences of cation/π interactions in crystal structures from the CSD. Aromatic rings are "soft" bases with higher energy occupied molecular orbitals, and will interact more favorably with "soft" acids according to HSAB theory. Li^+ is a harder cation compared to Na^+ and K^+, and it may engage preferentially with competing "hard" functional groups (i.e. ether oxygen atoms). In the crystal structures analyzed with Li^+/π aromatic interactions, 15 of the 19 structures (79%) involved a competing $Li^+\cdots O$ ether interaction. K^+/π interactions exhibited the highest percentage of centroid-oriented geometry (32% K^+ located within a 0.5 Å radius of the aromatic centroid), while Na^+/π interactions were dispersed over the plane of the ring. Both Na^+ and K^+ can interact more favorably with an aromatic ring, since they are softer acids than Li^+, but the centroid-oriented geometry suggests that these cation/interactions are primarily driven by electrostatic effects.

Kochi and coworkers' observation of carbon-oriented aromatic interactions with softer alkali metals, Rb^+ and Cs^+,[251] can also be rationalized with HSAB theory (Figure 2.51). The relatively "harder" cations will interact with aromatic rings with a greater electrostatic basis and "softer" cations interact with greater orbital effects. Na^+ and K^+ are comparatively harder acids than Rb^+ and Cs^+, and interact with the aromatic centroid because electrostatic effects are greater. The softer cations, Rb^+ and Cs^+, interact with π molecular orbitals (manifesting as carbon-oriented geometry) because molecular orbital interactions predominate over electrostatic effects with

softer acids. Kochi and coworkers' also observed that with increasing cationic radius, favorable interaction geometry with the ether oxygen atoms decreased,[251] consistent with HSAB theory (ether oxygen atoms are "harder" bases than π orbitals [270]). Indeed, while cation/π interactions involving alkali metals have been attributed to largely electrostatic interactions,[183] our data indicates an orbital component that significantly effects the interaction geometry of these electrostatic interactions.

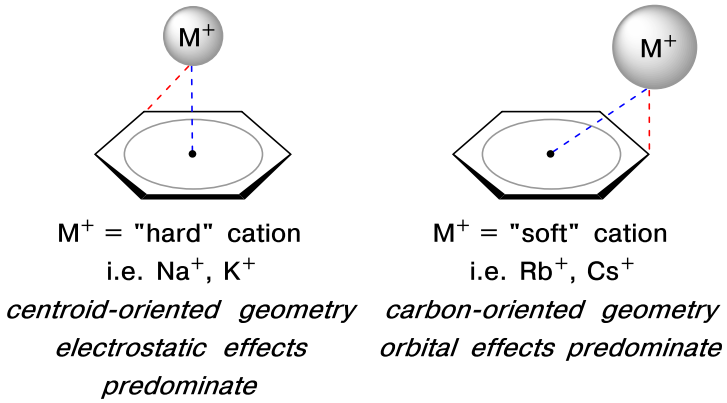


Figure 2.51 Cation-aromatic interaction geometry depends on chemical "hardness"
Our CSD analysis indicated centroid-oriented geometry with K^+/π interactions, dispersed orientation with Na^+/π interactions, and Li^+/π interactions that were completely outside of the aromatic ring (Figure 2.40). Li^+ cations are harder than Na^+ and K^+ cations, and potentially interact more favorably with harder bases, such as oxygen atoms. In contrast, softer cations can interact with aromatic rings (soft bases), but with geometry that depends on the relative "softness" of the cation. Kochi noted in crystal structures carbon-oriented geometries with Rb^+ and Cs^+/π interactions, and centroid-oriented geometry in K^+/π interactions, with indeterminate directionality with Na^+/π interactions.[251]

As in cation/π aromatic interactions, our model suggests that X–H/π aromatic interactions may also have differing geometric preferences for aromatic interactions depending on the atom X. In the analysis of crystal structures from the CSD, we

observed a significant percentage of carbon-directed, close contact distances with S–H/π aromatic interactions, in contrast to the traditional descriptions of "ideal" centroid-oriented X–H/π interactions.[190, 225, 228] Carbon-oriented geometries of X–H/π interactions have been suggested in a few studies of the CSD and PDB,[225, 247] but described as less "ideal" compared to centroid-oriented geometry.[190, 273] Malone *et al.*[225] characterized geometric preferences in X–H/π aromatic interactions, including X = O, N, and C, and noted that configurations with carbon-oriented geometries were among the most common. In crystal structures from the PDB, serine/π and cysteine/π interactions exhibit different interaction geometries with aromatic rings,[247] potentially due to the differing directionality of O–H and S–H interactions. Nishio explored carbon-directed versus centroid-directed C–H/π aromatic interactions, and found that C–H bonds were directed towards the aromatic centroid rather than aromatic carbons, with many examples of sub-ΣvdWr interaction distances.[260, 261] Potentially, if the analysis completed by Nishio was expanded to other X–H/π interactions, the donor atom influence on carbon-oriented versus centroid-oriented geometry would become more apparent. For example, the differing geometric orientations between C–H and S–H/π aromatic interactions could be due to the fact that C–H and S–H bonds would have different energies and volumes of the $\sigma^*_{X–H}$ orbital, thus influencing the strength and directionality of these non-covalent interactions.

A close contact distance between an atom and the aromatic centroid is the general standard for identifying the existence and strength of an aromatic interaction,[190, 203] but the focus on centroid-oriented interactions has potentially biased analysis against alternative, more favorable geometries (i.e. carbon-oriented interactions). In the CSD analysis by Malone *et al.*[225] on X–H/π aromatic interactions

in crystal structures, the search parameters excluded any structure where the X–H–*centroid* angle was less than 150°, so that only structures with idealized alignment of the X–H bond and the centroid where characterized.[225] Steiner states that linear, centroid-oriented X–H/π interactions are "ideal,"[190] but the X–H–centroid angle restriction potentially excluded many relevant structures with alternative orientations from Malone's study.[225] For example, the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester would have been excluded from Malone's study[225] (angle S–H–centroid = 147.9°), and the geometry of this interaction with close contact distances would have been completely disregarded.

In the study by Malone *et al.*[225], there were "too few" structures of S–H/π aromatic interactions "to be statistically significant," but a carbon-oriented geometry was noted in the few examples that were analyzed. Specifically, Malone[225] commented "S–H interactions all involve ring carbons rather than the centre, which may show a link between the nature of the donor heteroatom and the preference for the formation of a C–H···S interactions at the edge of the ring, or suggest that the d orbitals of sulfur are in some way involved in the interaction." Our analysis of the CSD, with 127 structures with S–H/π aromatic interactions, were consistent with Malone's[225] tentative suggestion: we observed a higher percentage of carbon-oriented geometry over centroid-oriented geometry, particularly with the short-range contact distances between the thiol hydrogen and the aromatic ring (49% $d_{H–Cmin} < 2.90$ Å, 21% $d_{H–centroid} < 2.90$ Å). Our observation of a periodic trend in cation/π aromatic interactions, predicted by HSAB theory, is consistent with orbital interactions influencing even strongly electrostatic interactions.

2.3.6 A new class of donor-acceptor interaction

As noted in the previous section, the definition of a conventional hydrogen bond is described primarily as an electrostatic effect between a donor and acceptor atom or group, with minor contributions from other favorable interactions.[203] The strength of a hydrogen bond is measured by the close proximity of the interacting groups, where closer distances between donors and acceptors represent stronger interactions.[203] The proximity of interacting groups can be measured directly from a crystal structure, or as a shift in the stretching frequency of the X–H bond (via IR), or as a change in the chemical shift in the donor or acceptor (via NMR).[203] To accurately measure interactions, it is crucial to identify the exact nature of the donor and acceptor groups; where the electron density is located and how it is donated to an acceptor. Identifying the donor and acceptor, and the energetic contributions that stabilize the interaction, provides insight into the preferred geometry for how the donor and acceptor groups will spatially interact. The subtle preferences in geometry between interacting groups can have significant consequences in orientation of protein-protein interfaces, small molecule interactions with a protein binding pocket, or therapeutic efficacy of designed drug molecules.

The nature of the donor and acceptor groups in the context of aromatic interactions, particularly aromatic interactions involving sulfur, remains an open question.[190, 191] Specifically, the nature of electron donation from aromatic rings has not been clearly defined, and this ambiguity prevents accurate measurements of donor-acceptor interactions, and determination of interaction strength. As stated by Desiraju & Steiner on aromatic hydrogen bonds[190]:

"A theoretical consensus model for the X–H···Ph hydrogen bond is yet to emerge. In particular, the bond geometry that represents the global energy minimum is still to be established. Earlier methods favoured the ideal ring centroid contact, whereas more recent computations prefer contacts to the centres of individual C–C bonds."

Interactions involving sulfur atoms, either as S–H/π or S··π, are also incompletely understood in terms of the nature of the donor or acceptor groups. As recently described by Biswal, "the H-bond acceptor strength of Sulfur, directionality and the electrostatic/dispersive nature of SCHB [sulfur-centred hydrogen bonds] are still inconclusive and debatable."[254] S–H/π aromatic interactions have been likened to weaker versions of O–H/π interactions, due to decreased electrostatic effects,[190] but recent evidence has demonstrated that sulfur-interactions can be significantly stabilizing. Biswal & Wategaonkar[164] calculated H_2S/indole interactions, and found that the strength of an S–H/π aromatic interaction was comparable or greater than similar C–H, N–H, or O–H/π interactions. The calculated red shift in the S–H stretching frequency was comparable to the red shift for the O–H bond, indicating similar interaction strength in spite of decreased electrostatic effects.[164] For S–H/π aromatic interactions, the fundamental donor-acceptor interaction has never been completely described, and so the energy and geometric orientation of these interactions have not been well characterized.

Identifying the nature of the donor and acceptor groups in S–H/π aromatic interactions can elucidate the nature of the interaction energy and determine the geometric preferences for these interactions. Our analysis of crystal structures from the CSD demonstrates that a significant percentage of S–H/π aromatic interactions have carbon-oriented geometry with contact distances below the sum of the van der

Waals distance (Figure 2.43), indicating that these interactions have directionality that can not be explained by dispersion. Our *ab initio* calculations on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester indicate a substantial stabilizing energy due to a favorable orbital overlap, indicating that the donor-acceptor interaction is aromatic $\pi \rightarrow \sigma^*_{S-H}$. With the understanding that S–H/π aromatic interactions are driven by favorable orbital overlap, and not electrostatic effects or dispersion, it would be expected that the vector of the thiol S–H bond would orient towards the aromatic ring carbons (where the π orbitals are located) rather than the centroid, consistent with our findings from the CSD.

This model of aromatic interactions driven by a substantial orbital overlap energy can be extended to other X–H/π aromatic interactions, and can describe different preferred orientations. For example, an O–H bond is more polarized than an S–H bond, and the electrostatic effects predominate in O–H interactions. In contrast, the σ^* orbital of an S–H bond is spatially different compared to O–H bonds (Figure 2.52), and favorable orbital overlap is more important for S–H interactions. Based on this model, the O–H bonds can be expected to orient towards an aromatic centroid, or with an electronegative atom (i.e. a "hard" base). Due to the greater contribution of orbital alignment in S–H interactions, the S–H bonds can be expected to orient towards aromatic carbon bonds, or with a "soft" base. Notably, thiols can potentially engage in non-covalent interactions through a "σ-hole" on *either* end of the S–H bond (Figure 2.52).

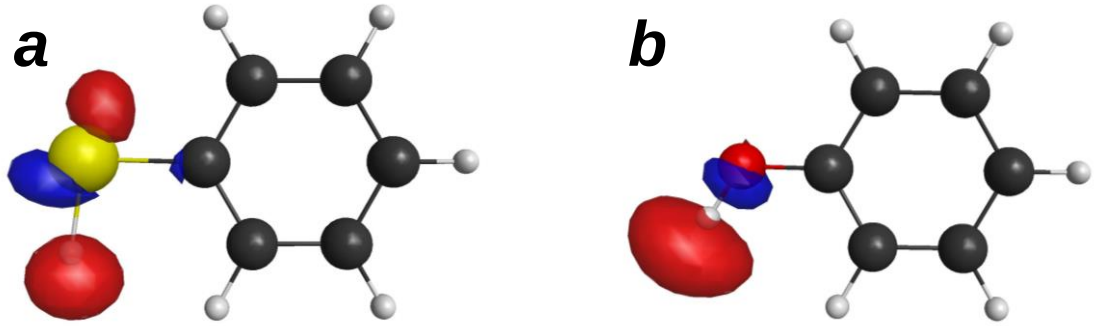


Figure 2.52 Comparison of $\sigma^*_{X–H}$ molecular orbitals in thiophenol and phenol
Thiol and alcohol bonds have different molecular orbitals and electrostatic effects, which can have significant consequences for ideal geometry with interacting atoms or groups. Calculations for molecular orbitals were conducted using GAMESS (RHF 6-31G),[195] and visualized with MacMolPlt (contoured at 0.1).[196] (a) Calculated unoccupied orbital associated with the S–H bond in thiophenol ($\sigma^*_{S–H}$). Note the two $\sigma^*_{S–H}$ orbitals that can participate in interactions (S–H/π or S··π); (b) calculated unoccupied orbital associated with the O–H bond in phenol ($\sigma^*_{O–H}$).

Described in section 2.1.3, thiol-aromatic interactions have been implicated in stabilizing α-helices and β-sheets, functional activity of dopamine receptors, in modulating redox potentials in flavodoxin, and in the structure and activity of therapeutic molecules.[7, 138, 174] Understanding the fundamental energy and geometric preferences of thiol-aromatic interactions can guide our design applications and elucidate the mechanisms of protein folding and function. The molecular orbital basis for S–H interactions also provides unique opportunities with thiol interactions that are not available with alcohol groups, since O–H groups can be expected to interact through greater electrostatic effects. The more significant contribution of orbital overlap effects in S–H/π interactions suggests that thiols can be utilized in modulating fluorescent properties or electron-transfer ability of aromatic rings, in a manner that may not be possible with O–H/π interactions. For example, the flavin cofactor in flavodoxin engages in a close S··π interaction with Met56 (Figure 2.5),[138] and

mutating this methionine to a selenomethionine can potentially lower the redox potential of the flavin by influencing the electron distribution.[274]

Based on this study of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, and the observation of an S–H/π aromatic interaction via x-ray crystallography, IR, NMR, and *ab initio* calculations, our results indicate that S–H/π interactions should not be regarded as weak or "non-directional," but understood that they are driven by underlying fundamental forces that differ from the traditional definition of hydrogen bonds. Hydrogen bonds are conventionally described by an interplay of electrostatic and dispersive forces, where weaker hydrogen bonds are attributed to decreased electrostatic effects, and described as "non-directional."[190] Our data strongly suggest that S–H/π aromatic interactions are driven by a favorable aromatic $\pi \rightarrow \sigma^*$ thiol orbital overlap with the aromatic π molecular orbitals with an interaction energy that is comparable to other X–H/π interactions. *ab initio* calculations on Boc-4-thiol-L-phenylalanine-*tert*-butyl ester showed up to 1.5 kcal mol^{-1} in stabilization due to the observed orbital overlap; the observed ^1H NMR chemical shift changes and observed IR ν_{S-H} frequency changes indicate significant hydrogen bonding effects upon formation of the intermolecular S–H/π interaction; the crystallographically observed geometry of carbon-oriented and close contact distances were consistent with thiol-aromatic interactions in the CSD, suggesting the generality of this interaction geometry. Furthermore, our analysis of cation/π aromatic interactions suggests that orbital effects can influence interaction geometry in strongly electrostatic cation/π aromatic interactions. Potentially, the definition of hydrogen bonds and donor-acceptor interactions should be re-evaluated as an interplay of orbital effects and electrostatic contributions.

2.4 Experimental

2.4.1 Materials

Boc-4-iodo-L-phenylalanine and di-*tert*-butyl dicarbonate (Boc_2O) was purchased from Chem-Impex (Wood Dale, IL). 4-Dimethylaminopyridine (DMAP), and *p*-thiocresol were purchased from Sigma. N,N′-Diisopropylcarbodiimide (DIC), *tert*-butanol, Copper(I) iodide, and 1,10-phenanthroline, were purchased from Acros. Dithiothreitol (DTT) was purchased from Fisher. 4-Mercaptophenylacetic acid (MPAA) was purchased from Alfa Aesar. All solvents were purchased from Fisher. Thin layer chromatography was conducted using Silicycle glass-backed plates (silica gel, 250 μm, 60 Å, F254). Flash chromatography was performed using 230-400 mesh (32-63 μm, 60 Å) silica gel from Silicycle.

2.4.2 Synthesis of Boc-4-iodo-D,L-phenylalanine-*tert*-butyl ester

Boc-4-iodo-L-phenylalanine (500 mg, 1.28 mmol) and DMAP (187 mg, 1.53 mmol) were placed in a glass vial (20 mL). CH_2Cl_2 (10 mL), DIC (200 μL, 1.53 mmol), and *tert*-butanol (150 μL, 1.53 mmol) were added to the solution, and the solution was stirred at room temperature for 2 hours. After the reaction was complete, the mixture was added to water (50 mL), and the crude product was extracted with CH_2Cl_2 (50 mL × 3). The crude product was concentrated under reduced pressure, and then purified via column chromatography (0-1% methanol in CH_2Cl_2 (v/v)) to yield Boc-4-iodo-D,L-phenylalanine-*tert*-butyl ester (417 mg, 0.93 mmol) as a colorless oil in 73% yield. Enantiopurity was determined via chiral HPLC (23% ee). The ^{1}H NMR and ^{13}C NMR spectra of the resultant product corresponded with literature values.[275]

2.4.3 Synthesis of Boc-4-iodo-L-phenylalanine-*tert*-butyl ester

Boc-4-iodo-L-phenylalanine (300 mg, 0.77 mmol) was dissolved in tetrahydrofuran (307 µL). Boc_2O (434 mg, 1.99 mmol) and *tert*-butanol (1.23 mL) were added, and the mixture was warmed to 40 °C to allow reagents to completely dissolve. DMAP (28 mg, 0.23 mmol) was added, and the mixture was stirred at room temperature for 3 hours, or until the disappearance of starting material was observed via TLC. Dilute HCl (1 M, 20 mL) was added, and the product was extracted with CH_2Cl_2 (20 mL × 3). The combined organic layers were dried over sodium sulfate and filtered. The solvent was removed under reduced pressure. The crude product was purified via column chromatography (0-1% methanol in CH_2Cl_2 (v/v)) to yield Boc-4-iodo-L-phenylalanine-*tert*-butyl ester (340 mg, 0.76 mmol) as a colorless oil in 99% yield. Retention of enantiopurity was verified via chiral HPLC (>98% ee). The ^{1}H NMR and ^{13}C NMR spectra of the resultant product corresponded with literature values.[275]

2.4.4 Synthesis of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester by reducing isolated cross-coupling reaction products

A mixture of purified Boc-4-S(acetyl)-thiol-L-phenylalanine and corresponding disulfides (100 mg combined, approx. 0.286 mmol) was dissolved in methanol (2.86 mL). Sodium methanethiolate (24 mg, 0.343 mmol) was added, and the solution was stirred at room temperature for 30 minutes under an atmosphere of nitrogen. DTT was added and stirred for 10 minutes. Dilute HCl (1 M, 30 mL) was added, and the product was extracted with ethyl acetate (20 mL × 3). The combined organic layers were dried over sodium sulfate and filtered. The solvent was removed under reduced pressure. The crude product was purified via column chromatography (0-30% ethyl acetate in hexanes (v/v)) to yield Boc-4-thiol-L-phenylalanine-*tert*-butyl ester. The product was

confirmed to be the reduced thiol via mass spectrum and ^{1}H NMR (see Figure 2.13).

HRMS (LIFDI-TOF) m/z: calcd for $C_{36}H_{52}N_2O_8S_2$ 353.1661, found 353.1673.

Detailed resonance assignments are shown in Table 2.19.

Table 2.19. NMR Resonance Assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

A purified sample of crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was dissolved in $CDCl_3$ containing tetramethylsilane as an internal reference. ^{1}H NMR resonances were assigned with the aid of COSY spectra. ^{13}C NMR resonances were assigned with the aid of HMBC spectra. Refer to Chapter 2.2.3 for additional details.

Atom #	Assignment	δ, ppm	Multiplicity	*J* coupling, Hz
SH	Thiol	3.41	s	
NH	Carbamate	4.98	d	8.0
H5	H_{Aro}, meta	7.20	d	8.1
H4	H_{Aro}, ortho	7.04	d	8.1
H1	Hα	4.41	ddd	7.7, 6.1, 6.1
H2A, H2B	Hβ	3.00	ddd	13.8, 6.2, 6.0
H12	Methyl, Boc	1.42	s	
H9	Methyl, *tert*-butyl ester	1.41	s	

Atom #	Assignment	δ, ppm
C7	Carbonyl, *tert*-butyl ester	170.8
C10	Carbonyl, Boc	155.0
C3	C_{Aro}, ipso	134.0
C5	C_{Aro}, meta	130.3
C4	C_{Aro}, ortho	129.5
C6	C_{Aro}, para	128.9
C8	C, *tert*-butyl ester	82.2
C11	C, Boc	79.7
C1	Cα	54.7
C2	Cβ	37.9
C12	CH_3, Boc	28.3
C9	CH_3, *tert*-butyl ester	28.0

2.4.5 Optimized copper-mediated cross-coupling reaction on Boc-4-iodo-L-phenylalanine-*tert*-butyl ester

Boc-4-iodo-L-phenylalanine-*tert*-butyl ester (400 mg, 0.89 mmol), copper(I) iodide (17 mg, 89 μmol), and 1,10-phenanthroline (32 mg, 180 μmol) were placed in an oven-dried glass vial with a stir bar. Toluene (1.8 mL) and DIPEA (470 μL, 2.7 mmol) were added, and the mixture was stirred at room temperature under nitrogen for 5 minutes. Thioacetic acid (128 μL, 1.8 mmol) was added to the solution at room temperature, the vial was sealed, and the mixture was heated to 110 °C in an oil bath and stirred for 24 hours. During this time, the reaction darkened to a red-brown color as iodine was formed. The solution was cooled to room temperature, and the solvent was removed under reduced pressure. The crude residue was redissolved in *tert*-butanol (1.8 mL), and 4-mercaptophenylacetic acid (MPAA, 150 mg, 0.89 mmol) was added. The reaction mixture was stirred at 110 °C for 6-9 hours. The solution was then cooled to room temperature, and DTT (150 mg, 0.97 mmol) and cesium carbonate (320 mg, 0.98 mmol) were added. This mixture was stirred at room temperature for 1 hour. The crude mixture was diluted with ethyl acetate (10 mL) and added to dilute HCl (1 M, 20 mL). The crude product was extracted with ethyl acetate (25 mL × 3) and the combined organic layers were dried over Na_2SO_4. The solvent was removed under reduced pressure, and the crude product was redissolved in CH_2Cl_2. The product was purified via column chromatography (0-10% ethyl acetate in hexanes (v/v)) to yield Boc-4-thiol-L-phenylalanine-*tert*-butyl ester as a separable mixture of free thiol and disulfides.

The disulfide products were recovered, and the solvent removed under reduced pressure. The resultant yellow oil was redissolved in THF (1.8 mL), and DTT (150 mg, 0.97 mmol) and Cs_2CO_3 (320 mg, 0.98 mmol) were added. The solution was

stirred for 1 hour at room temperature. The crude mixture was diluted with ethyl acetate (10 mL) and added to dilute HCl (1 M, 20 mL). The crude product was extracted with ethyl acetate (25 mL × 3) and the combined organic layers were dried over Na_2SO_4. The solvent was removed under reduced pressure, and the crude product was redissolved in CH_2Cl_2. This additional Boc-4-thiol-L-phenylalanine-*tert*-butyl ester generated from recovered disulfide products was purified via column chromatography (0-10% ethyl acetate in hexanes (v/v)), and combined with the product isolated from the initial purification.

The free thiol product was prone to form a charge transfer complex with residual iodine from the reaction, resulting in a yellow oil that rapidly formed disulfides. In order to disrupt the charge transfer complex and remove the iodine, the combined Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was precipitated from hexanes, or recrystallized from 25% ethyl acetate in hexanes (v/v), and the white, crystalline solids were filtered. The product after reduction of recovered disulfides and removal of iodine was obtained in 30% yield (95 mg, 0.27 mmol).

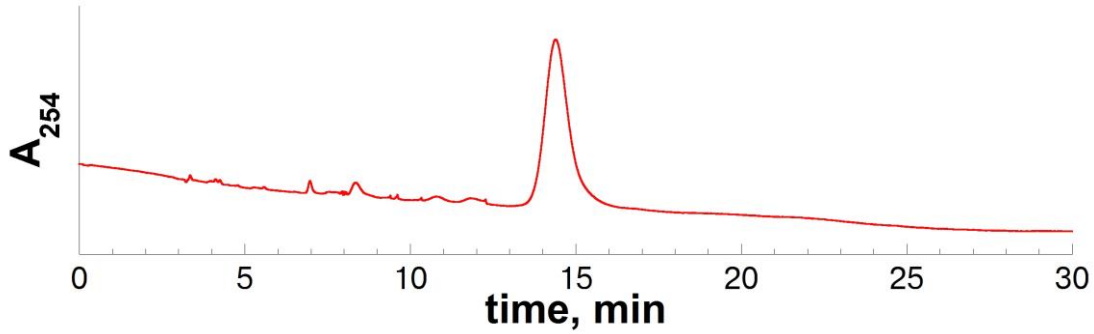


Figure 2.53 Chiral HPLC chromatogram of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

ChiralPak 1A column, 250 × 4.6 mm, 5 μm particle, 1.0 mL/min flow rate, isocratic 10% isopropanol/hexanes). The peaks t 8.4 min and 14.4 min are D- and L-isomers of Boc-4-thiolphenylalanine-*tert*-butyl ester, respectively (94% ee).

2.4.6 X-ray crystallography

Orthorhombic, enantiopure crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were obtained via slow evaporation at room temperature over one week of a solution of approximately 20 mg of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in 10 mL 25% ethyl acetate in hexanes. Diffractable crystals of *p*-thiocresol were used directly from commercially available material (Sigma) without recrystallization.

Crystallographic data was obtained with the assistance of Dr. Glenn P. A. Yap. The crystals were selected and mounted on plastic mesh using oil flash-cooled to the data collection temperature. Data were collected on a Bruker-AXS APEX II Duo CCD diffractometer with graphite-monochromated Mo–Kα radiation (λ = 0.71073 Å) for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester and *p*-thiocresol. Unit cell parameters were obtained from 60 data frames, 0.3° ω, from three different sections of the Ewald sphere, and refined with the entire diffraction data set. The systematic absences in the data and the unit cell parameters were uniquely consistent to the reported space groups. The data sets were treated with absorption corrections based on redundant multiscan data.[276] The structures were solved using direct methods and refined with

full-matrix, least-squares procedures on F^2.[277] The absolute structure parameter for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester refined to nil indicating the true hand of the data had been determined. The compound molecule of *p*-thiocresol resides on an inversion center and the full molecule was refined at half-occupancy with symmetry related atoms treated with equal atomic displacement parameters and with a minimal DAMP value for convergence. All non-hydrogen atoms were refined with anisotropic displacement parameters. All hydrogen atoms were treated as idealized contributions. Scattering factors and anomalous dispersion coefficients are contained in the SHELXTL program library.[277] Molecular graphics were generated using Mercury CSD.[278] The crystallographic information files have been deposited as CCDC 1416010 to 1416011.

2.4.7 NMR Spectroscopy

NMR spectra of solution samples were collected on a Brüker 400 MHz NMR spectrometer equipped with a cryogenic QNP probe or a Brüker 600 MHz NMR spectrometer equipped with a 5-mm Brüker SMART probe. Samples were dissolved in deuterated solvents purchased either from Acros or Cambridge Isotope Laboratories, Inc. (Tewksbury, MA).

All solid-state NMR data were obtained with the assistance of Dr. Shi Bai. ^{13}C CP/MAS experiments were conducted using a 4.0 mm HX or a 3.2 mm HCN solid-state magic-angle-spinning (MAS) probe on a Bruker Avance III spectrometer operating at a proton Larmor frequency of 500.13 MHz. All spectra were obtained at 298 ± 2 K and CP/MAS experiments were carried out with a sample spinning rate of 10,000 ± 2 Hz with a 4 mm rotor or 14,000 ± 2 Hz with a 3.2 mm rotor. TPPM decoupling during CP/MAS data acquisition was provided by a 104.2-kHz proton

decoupling field. For CP/MAS spectra, the contact time was 3 ms. Solid-state ^{13}C-^{1}H HETCOR spectra were acquired with a 3.2 mm HCN MAS probe at a spinning rate of 14,000 Hz. The mixing time was set to either 300 or 1000 μsec. During the t_1 evolution period, proton homonuclear decoupling with the FSLG scheme was applied to enhance proton resolution.[249]

All ^{13}C chemical shifts were referenced externally via the resonance of adamantane at an isotropic chemical shift relative to tetramethylsilane (TMS) of 38.55 ppm.

Solid samples of *p*-thiocresol were used as purchased, consistent with the crystals used in X-ray diffraction and in obtaining the crystalline FT-IR spectrum. Solid samples of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were obtained by precipitation from hexanes, and the solids were filtered and dried. The solids were redissolved in 20% ethyl acetate in hexanes and the solution was warmed until all of the material had dissolved, and the sample was allowed to recrystallize. The sample appeared to contain a mixture of solids and crystals. ^{13}C chemical shift assignments for *p*-thiocresol were obtained from HSQC experiments in $CDCl_3$, and were compared by analogy to the chemical shifts observed in the solid-state NMR spectra. ^{13}C chemical shift assignments for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester were obtained directly from solid-state ^{13}C-^{1}H HETCOR spectra.

2.4.8 FT-IR Spectroscopy

All spectra (64 scans) were recorded on a Nicolet Magna-IR 750 FT-IR spectrometer equipped with a liquid nitrogen-cooled MCT/A detector at a spectral resolution of 1 cm^{-1}. Solution samples were filled in a gas-tight IR cell (100 μL volume) equipped with quartz windows, and contained 200 mM of Boc-4-thiol-L-

phenylalanine-*tert*-butyl ester or *p*-thiocresol. *p*-Thiocresol was used as purchased without further purification (Sigma-Aldrich). Solvent spectra were subtracted from the sample spectra, and the baseline was manually corrected using the OMNIC FT-IR Software (Nicolet). Crystalline samples were measured using the pressed pellet method with anhydrous KBr.

2.4.9 *ab initio* calculations

All of the calculations were performed using the Gaussian 09 program package.[279] To reduce calculation time, the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester was truncated to remove all atoms so that only the dimeric and monomeric *p*-thiocresol molecules remained (Figure 2.33). Dr. Shi Bai performed the calculations on the *p*-thiocresol dimer for the correlation between ν_{max} and S–H bond length, and Dr. Sudipta Sinha performed all calculations for atomic charges and natural bond orbital (NBO) analysis. GAUSSVIEW was utilized for graphics preparation for this study.[279]

The structure was optimized at B3LYP level of theory with 6-311+G(2d,p) basis set followed by the frequency calculations to ensure that all of the structures were the true minima. The atomic charges were computed using CHELPG (CHarges from Electrostatic Potentials using a Grid-based method) at both B3LYP and MP2 level of theories for both crystal structure and optimized structures.[280] The NBO analysis was performed using the NBO program available in the Gaussian 09 software at B3LYP and MP2 levels of theory using various basis sets for the experimental crystal structure of the molecule.[256] The magnitude of the interaction energy, $E^{(2)}_{i \rightarrow j^*}$, between a donor NBO, σ^*_j, and an acceptor NBO, σ_i, can be estimated using the second order perturbation theory:

$$E_{i\rightarrow j^*}^{(2)} = q_i \frac{|\langle \sigma_j^* | \hat{F} | \sigma_i \rangle|^2}{\varepsilon_j^* - \varepsilon_i},$$

where *i* and *j* denote the donor and acceptor orbitals, q_i is the occupancy of bonding orbital, ε_i and ε_j^* correspond to their respective orbital energies, and $\hat{F}$ is the Fock operator respectively.

2.4.10 Cambridge Structural Database (CSD) and Search Parameters

A search was conducted on the Cambridge Structural Database (version 5.36, released in November, 2014) for entries containing non-bonded contacts between thiols (motif Z–S–H, Z = any non-hydrogen atom) or cations (Li^+, Na^+, or K^+) and a 6-membered aromatic ring containing only carbon atoms. The release of the database used in the current study contained over 710,000 molecules. Searches were conducted using ConQuest (version 1.17, Cambridge Crystallographic Data Centre, 2014).

The initial search parameters placed restrictions on the distance between the aromatic-centroid and the interacting atom, either the thiol sulfur atom for thiol-aromatic interactions or Li^+, Na^+, or K^+ (generally, M^+) for cation-aromatic interactions (Figure 2.54). Inter- and intramolecular contacts (2 bonds or more) were included within the search parameters. Only error-free, non-disordered structures where $R < 0.10$ were included in the search, and powder pattern structures were excluded. For cation-aromatic interactions, the range was defined as $d_{M^+\text{–centroid}} \leq 4.2$ Å. For thiol-aromatic interactions, the range was defined as $d_{S\text{–centroid}} \leq 4.8$ Å. The interacting 6-membered aromatic ring contained only carbon atoms, and heteroaromatic rings were excluded from this study. For thiol-aromatic interactions, two lists were generated with different coordinates for hydrogen atom location: one with coordinates as reported in the CSD, and one where the S–H bond was normalized

to 1.338 Å (the coordinates of the hydrogen atom were adjusted along the vector of the S–H bond).[203] Data and statistics obtained from the thiol structures with fixed S–H bond lengths are indicated as "normalized" entries.

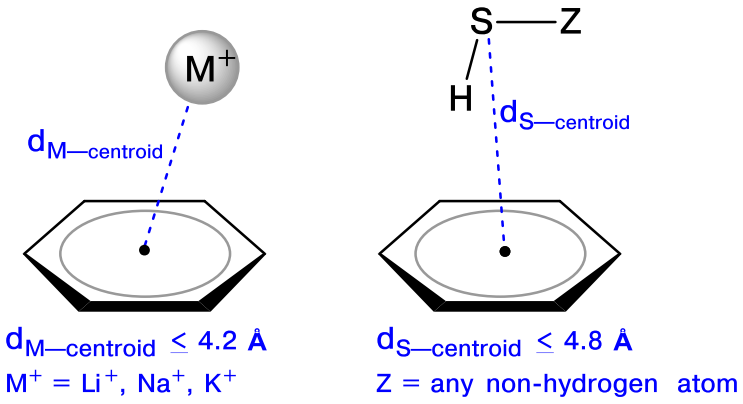


Figure 2.54 Initial search parameters for cation/π and S–H/π interactions in the Cambridge Structural Database using Conquest.
The initial search parameters on the CSD for thiol-aromatic and cation-aromatic interactions. $d_{X–centroid}$ is defined as the distance from the interacting atom to the aromatic centroid, where M^+ or thiol atom S. Aromatic rings were 6-membered, and heteroaromatic groups were excluded. The search parameters for cation-π and S–H/π aromatic interactions are indicated.

For all entries obtained from the initial search parameters, distances and bond angles were calculated. The measurements on the crystal structures included distances from the interacting atom (M^+, thiol S, or thiol H, generally referred to as "X") to the aromatic ring. The measurements included: $d_{X–centroid}$, distance to the calculated centroid of the aromatic ring; $d_{X–C_{min}}$, the shortest distance from X to any aromatic carbon atom; and d_{plane}, the normal distance from X to the plane of the aromatic ring (Figure 2.55). The radius (r) was defined as the distance from the aromatic centroid to the projection of X in the plane of the aromatic ring, and was calculated from

measured distances as $\sqrt{d_{X-centroid}^2 - d_{plane}^2}$. For the thiol entries, measurements were also obtained for the bond length S–H (range 0.86 Å-1.42 Å), for the Z–S–H bond angle (range 80°-120°), and for the angle S–H–centroid. Thiol structures that had S–H bond lengths or Z–S–H bond angles outside of these criteria were excluded.

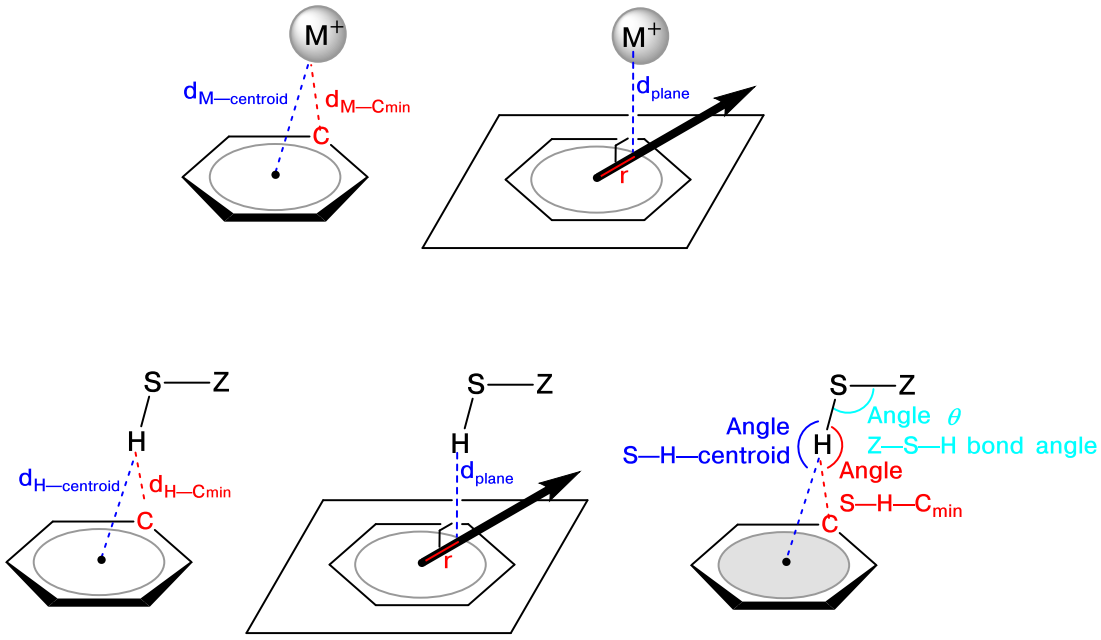


Figure 2.55 Defined distances and measurements accumulated from crystal structures in the CSD within defined parameters
$d_{X–centroid}$ refers to the distance between the atom X (where X = M^+, S or H) and the centroid of the aromatic ring; $d_{X–C_{min}}$ refers to the distance between the atom X (where X = M^+, S or H) and the nearest carbon atom of the aromatic ring; d_{plane} refers to the distance between the atom X (where X = M^+, S or H) and the plane of the aromatic ring; r refers to the projected distance from atom M^+, S, or H within the plane of the aromatic ring to the centroid, which was calculated from measured distances as $\sqrt{d_{X–centroid}^2 - X\,d_{plane}^2}$; angle θ refers to the bond angle Z–S–H, where Z is any non-hydrogen atom; angle S–H–centroid refers to the intermolecular angle S–H···centroid, where the centroid is the center of the aromatic ring; angle S–H–C_{min} refers to the intermolecular angle S–H···C_{min}, where C_{min} is the nearest aromatic carbon atom.

In order to identify and study entries with close cation- or thiol-aromatic interactions, a geometric "cylinder" was defined with respect to the aromatic ring (Figure 2.37). Only entries with interacting atom coordinates within the defined cylinder were examined further. The base of the cylinder was defined by r ≤ 3.00 Å for M^+ and thiol hydrogens. The size of the base was designed to be slightly larger than twice the radius of benzene. For thiol sulfur atoms, the base of the cylinder was

defined as $r \leq 4.20$ Å, which was twice the radius of benzene (3.00 Å) and an additional 1.20 Å to account for the S–H bond. The height of the cylinder was defined for each X as $d_{plane} \leq (1.25 \times \Sigma vdWr$ for carbon and X). The ionic radii for Li^+, Na^+, and K^+ were defined as 0.60 Å, 0.95 Å, and 1.33 Å, respectively;[263] the van der Waals radii for C, H, and S were defined as 1.70 Å, 1.20 Å, and 1.80 Å respectively.[203, 263] Using these definitions, the ΣvdWr for C··Li^+, C··Na^+, C··K^+, C··H, and C··S were defined as 2.30 Å, 2.65 Å, 3.03 Å, 2.90 Å, and 3.50 Å respectively.[203, 263] The cylinder restrictions and defined distances used for this study are shown in Figure 2.37.

After applying the cylinder restriction on the atomic coordinates of each M^+, hydrogen, or sulfur atom, the individual structures were manually examined and annotated for features that could potentially effect the geometry of the cation- or thiol-aromatic interaction. Aromatic rings that were part of a complex, conjugated, polycyclic system were excluded (i.e. naphthalene), while aromatic rings involved in larger marcocycles were included (i.e. cyclophanes). Aromatic rings or groups with carbanions were excluded, since this could cause an asymmetric distribution of electron density, and effect geometry of interaction. For thiol structures, only entries were the thiol bond was directed towards the aromatic ring were included (where $d_{S–plane}$ was greater than $d_{H–plane}$), as to only study S–H/π interactions (not S··π interactions). Thiol structures with sulfur atoms that were tri-coordinate or charged were excluded from this study, as this can significantly alter the strength and geometry of the thiol-aromatic interaction. Redundant measurements within a structure or structures with questionable features or geometry were excluded. Structures where M^+ was also interacting with a crown ether were included from further study. The entries and structures obtained from the CSD, following the cylinder restrictions, are shown

in Tables 2.20-2.26. The entries where manually annotated with the following criteria: (A) polycyclic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (C) where thiol groups were "pointed away" from the plane of the aromatic ring (as defined by $(d_{S-plane} - d_{H-plane}) \geq -0.10$ Å); (D) the thiol sulfur atom is apparently tri-coordinate or exceeding standard valency; (E) redundant structures or entries; (F) unusual structures; or (G) where M^+ was interacting with a crown ether. The number of resultant structures after each level of parsing are shown in Table 2.20.

The statistics, charts, and correlations in Chapter 2.2.6 focused on the thiol hydrogen interactions with aromatic rings in crystal structures from the CSD. As described above, the hydrogen atoms are generally not well resolved in x-ray crystal structures due intrinsically low electron density. While measuring contact distances between heavy atoms is more reliable and accurate from x-ray crystal structures, this heavy atom measurement can not distinguish between S–H/π and S··π aromatic interactions. For the thiol structural and geometric comparisons, the initial data set was parsed based on hydrogen atoms within a defined region of interaction (a cylinder based on the plane of the aromatic ring), for both modelled hydrogen atoms and "normalized" hydrogen atoms (where the hydrogen atom was re-positioned along the vector of the S–H bond so that the S–H bond length was set to a standard value, 1.338 Å).[203] In addition, the thiol data set was parsed based on the sulfur atoms within a defined region of interaction, since heavy atom location is more reliable that hydrogen atom location (these data sets were termed "sulfur restricted" for the cylinder restriction on the sulfur atom). The charts, correlations, and statistics for S–H/π aromatic interactions based on the sulfur restricted data sets (containing, separately,

CSD reported hydrogen atom locations and “normalized” hydrogen atom locations) were generally consistent with the hydrogen restricted data sets, and are shown in Figures 2.56 and 2.57. All of these data sets were manually parsed as described above.

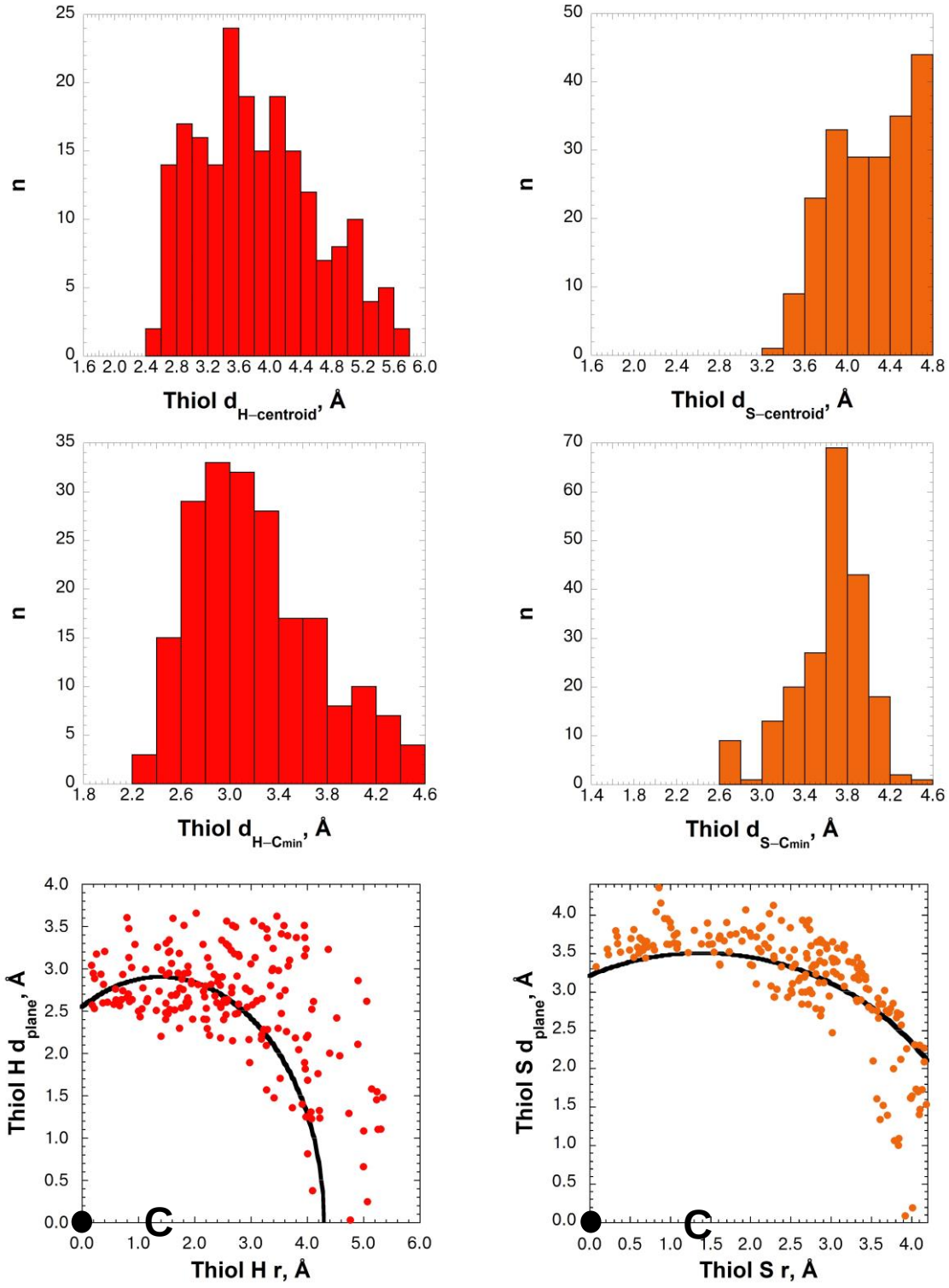


Figure 2.56 S–H/π interactions in the CSD: reported S–H bond lengths using

the sulfur atom as a reference for cylinder restriction

Plots of $d_{X-centroid}$, d_{X-Cmin}, and atomic coordinates of X with respect to the plane of the aromatic ring, where X = thiol H or thiol S. Thiol hydrogen atoms are as reported in the CSD, using the sulfur atom for reference in the cylinder restriction (Thiol S $d_{plane} \leq$ 4.38 Å, Thiol S $r \leq 4.20$ Å), and hydrogens were not restricted. The black line indicates the sum of the van der Waals radii for carbon and the respective atoms or ions, where entries below the line are less than the sum of the van der Waals radii (Σ van der Waals radii for C and: Thiol H = 2.90 Å; Thiol S = 3.50 Å).

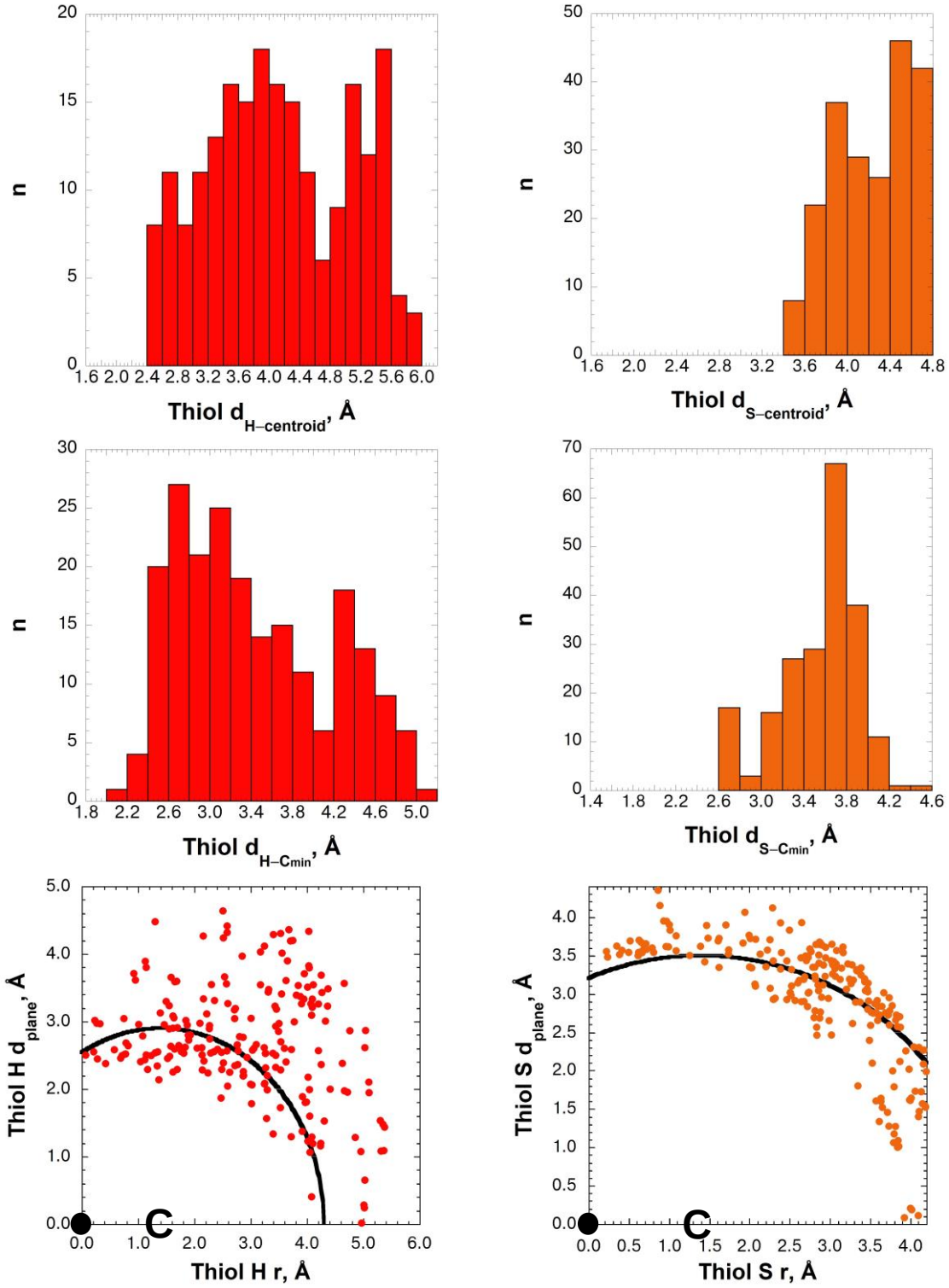


Figure 2.57 S–H/π interactions in the CSD: normalized S–H bond lengths using the sulfur atom as a reference for cylinder restriction.

Plots of $d_{X–centroid}$, $d_{X–Cmin}$, and atomic coordinates of X with respect to the plane of the aromatic ring, where X = thiol H or thiol S. Thiol hydrogen atoms were normalized to 1.338 Å along the vector of the S-H bond,[203] using the sulfur atom for reference in the cylinder restriction (Thiol S $d_{plane} \leq 4.38$ Å, Thiol S $r \leq 4.20$ Å), and hydrogens were not restricted. The black line indicates the sum of the van der Waals radii for carbon and the respective atoms or ions, where entries below the line are less than the sum of the van der Waals radii (Σ van der Waals radii for C and: Thiol H = 2.90 Å; Thiol S = 3.50 Å).

Table 2.20. Cation- and thiol-aromatic interactions obtained from the CSD database study

The numbers of entries and structures identified after each level of restriction are reported. For thiol structures and entries, data was accumulated for the reported S-H bond and for normalizing all S–H bonds to be 1.338 Å (noted as "normalized" in the table). These entries were restricted based on a geometric "cylinder" that was defined based on the aromatic ring, and entries were retained for further study where the atom of interest (Li^+, Na^+, K^+, thiol H, or thiol S) was within the defined "cylinder." These entries were then manually examined and annotated with specific criteria (described above), and the "manual parsing" excludes entries fitting criteria A, B, C, D, E, and F (criteria A1 and G were included).

	Initial Search		**Cylinder Restriction**		**Manual Parsing**	
	Structures	**Entries**	**Structures**	**Entries**	**Structures**	**Entries**
Li^+	82	198	37	87	19	43
Na^+	99	153	62	105	48	79
K^+	225	432	217	404	193	362
X-SH H reference	186	490	110	182	78	127
X-SH S reference	186	490	163	398	102	202
X-SH, normalized, H reference	186	490	111	178	79	122
X-SH, normalized, S reference	186	490	169	409	89	210

Chapter 3

ELECTRONIC CONTROL OF PROLINE CIS-TRANS ISOMERISM VIA A C–H/π AROMATIC INTERACTION: INSIGHTS INTO THE NATURE OF C–H/π INTERACTIONS

3.1 Introduction

Protein folding and functional conformations of enzymes are dictated by non-covalent interactions among the amino acid side chains and peptide backbone. Nature has evolved a selection of amino acids that can tailor these non-covalent interactions in proteins and their substrates, and each amino acid has specific advantages or roles to play in dictating tertiary structure. Hydrophobic amino acids such as leucine and phenylalanine aid in collapse of the hydrophobic core of proteins, where self-association of hydrophobic groups and exclusion of water overcomes some of the greatest entropic costs between unfolded and folded conformations of a given protein.[207] However, hydrophobic effects are non-directional, and may not enforce the defined conformations that are necessary for substrate recognition and enzyme active sites. Local, enthalpically driven conformations sometimes preclude global protein folding, where secondary structures must form before the hydrophobic collapse of the protein.[94, 179, 281] Defined protein structures are entropically costly with unfavorable loss of degrees of freedom, and enthalpically favorable interactions are necessary for fine-tuning protein structure. There are many non-covalent, enthalpically favorable interactions that Nature employs, including salt bridges, hydrogen bonding, and many other weak associations.[160, 281] While these stabilizing interactions may only be worth

0-4 kcal mol^{-1} individually, but combined, can easily overcome the energetic cost of protein folding.

There are many steps along the trajectory of protein folding, a major rate-determining step involves cis-trans isomerism of proline in some proteins.[282-284] Proline plays a unique role in protein structure, as its cyclic nature invokes conformational restraint on its allowed dihedral angles. The conformational restriction in proline decreases the energy difference between trans and cis amide bond conformations, and the sterically unfavorable cis-amide bond conformation is much more accessible for proline.[285, 286] Proline cis-trans isomerism is important for defining structures of loops and turns (cis-prolines are specifically involved in type VI β-turn formation), or for coordinating a metal ion, or for switching between active and inactive forms of an enzyme.[282, 287]

In surveys of protein crystal structures, aromatic amino acids are over-represented immediately adjacent to cis-prolines, suggesting that aromatic residues may play a role in stabilizing cis-proline.[285, 286] Numerous model peptide studies are consistent with the surveys of protein crystal structures.[90, 288, 289] In one study, Ac-TXPN-NH_2 model peptides were studied via NMR for cis-populations of peptides containing various aromatic amino acids as position X.[91] In this work, it was found that peptides containing tyrosine and tryptophan had higher cis-proline populations in comparison to peptides containing 4-nitrophenylalanine or 4-trifluoromethylphenylalanine. In this manner, cis-trans isomerism of proline could be tuned or controlled via electronic substituent effects.[91]

A C–H/π aromatic interaction resembles a conventional hydrogen bond, but with decreased polarization about the C–H bond, is comparatively weaker due to

decreased electrostatic driving energy.[290] The energetic components that drive C–H/π interactions have not been well described, but they have been attributed to largely dispersive interactions, sometimes called van der Waals interactions or charge-transfer interactions.[106] However, C–H/π aromatic interactions have specific directionality that cannot be explained by dispersive van der Waals interactions,[191] and additional energetic contributions must be stabilizing these weak noncovalent interactions.[106] As stated by Desiraju & Steiner: "it is important to note that the grey area between hydrogen bonds and van der Waals interactions occurs in the region of the C–H···π interaction."[190] We hypothesize that C–H/π interactions stabilize aromatic-cis-proline motifs in peptides and proteins, where the interaction between the prolyl Hα and the electron-rich aromatic ring compensates for the entropic cost of forming the β-turn motif.

To examine this possibility, we aim to expand on prior work[91] on aromatic-cis-proline motifs in model peptides to determine the scope of electronic influence on proline cis-trans isomerism. With this expanded scope of aromatic amino acids, we will identify key, stabilized structures that will allow for in depth analysis of the aromatic-cis-proline interaction. We will characterize these key structures using thermodynamic analysis, linear free-energy relationships, and x-ray crystallography in order to gain insight into the fundamental nature of prolyl C–H/π aromatic interactions. In light of the S–H/π aromatic interaction work (described in detail in Chapter 2), we proposed that favorable orbital overlap interactions are a stabilizing feature in C–H/π aromatic interactions, suggesting a general mechanism for weak, non-covalent interactions, where electrostatic contributions are significantly decreased. With this general model for X–H/π aromatic interactions, we can use these

design principles to design stabilized β-turn motifs for molecular recognition, or to synthesize novel biomaterials based on non-covalent supramolecular assembly.[178, 291]

3.1.1 Cis-trans isomerism of proline in proteins

In proteins and peptides, the amide bonds that link amino acids are either in trans (ω = 180°) or cis (ω = 0°) conformations, due to the partial double-bond nature of the amide bond.[292] The trans amide bond conformation predominates for all amino acids, as steric clash between the Cα and Cβ in adjoining amino acids renders the cis amide bond highly unfavorable (Figure 3.1a).[293] Cis amide bonds are infrequent in proteins, where less than 0.05% of PDB crystal structures have a non-proline cis amide bond, but 5-6% of prolines are in the cis conformation, depending on the parameters for a given study.[285, 286, 294] Based on these experimental statistics from protein crystal structures, the energetic difference between trans and cis conformations can be estimated: ~2 kcal mol^{-1} for Xaa-cis-Pro and ~5 kcal mol^{-1} for Xaa-cis-(non-Pro).[293] Proline is the only natural amino acid with a cyclic peptide backbone, which invokes conformational restriction in the pyrrolidyl ring. The trans and cis conformations are more similar for proline than for non-proline residues, both sterically and energetically (Figure 3.1b). Proline is noteworthy in that its equilibrium between cis and trans conformations is observable via NMR at room temperature in short peptides.[285] However, cis-trans isomerism of proline is a rate-determining step in protein folding, such as in RNase A or the G3P coat protein in filamentous bacteriophages.[295] Designated proline isomerase enzymes catalyze transitions between cis and trans conformers of proline, via stabilization of the transition state.[283, 287]

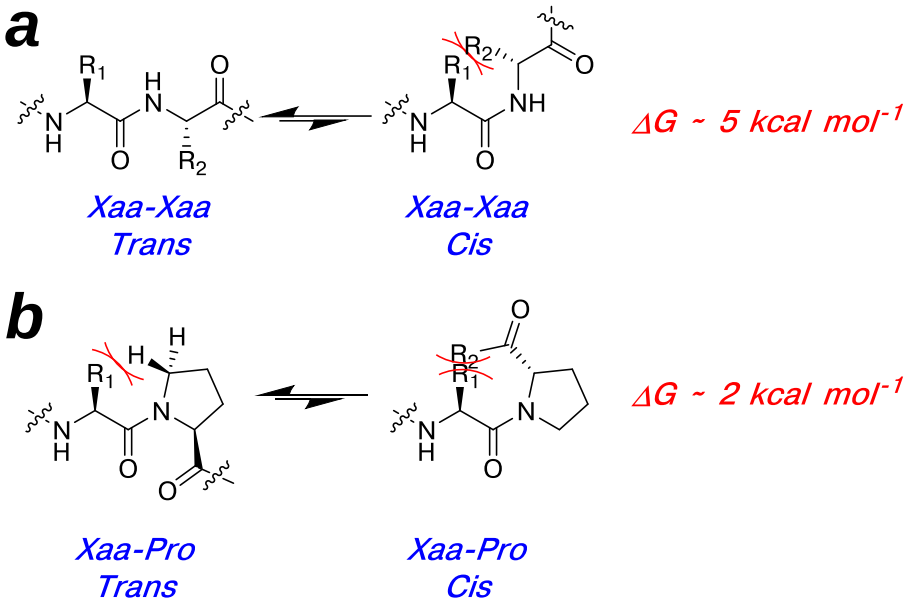


Figure 3.1 Cis-trans isomerism of amide bonds
(a) Cis-trans isomerism between two non-proline residues; (b) cis-trans isomerism between a non-proline and proline amide bond. The trans conformation (left) predominates in all amino acids, due to unfavorable steric clash between Cα and Cβ in the cis conformation (right). The cyclic nature of proline renders the trans conformation less favorable in comparison to other amino acids, due to steric clash between the Cβ and Cδ (b, left). With proline, the steric hindrance between cis and trans conformations are more similar than for other amino acids, rendering the cis conformation more energetically accessible.
Values for ΔG were calculated from statistics in PDB crystal structures[293] for frequency of amino acid, although model peptide studies show slightly different calculated energies.[296]

What then, fundamentally, causes a proline residue to be in the cis conformation? Is there a sequence dependence for cis-prolines, and what is the functional relevance for this conformation? Pal & Chakrabarti surveyed the Protein Databank (PDB) for cis-prolines in high resolution protein crystal structures, and found that cis-prolines were most likely to be preceded by either proline or an aromatic amino acid (Figure 3.2).[285] These trends are generally consistent with other

analyses of the PDB, depending on the sample size and search parameters.[286, 293, 297] The residues that were least likely to precede a cis-proline were β-branched amino acids.[285] The higher occurrence of aromatic amino acids preceding cis-proline was attributed to a prolyl C–H/π aromatic interaction.[285] In addition, non-prolyl cis amide bonds (<0.05% occurrence in protein crystal structures[293, 298]) have a higher likelihood of an aromatic residue preceding these cis-amino acids. These trends suggest that the same interactions that stabilize cis-prolines may also stabilize non-prolyl cis amide bonds (Figure 3.3).[285]

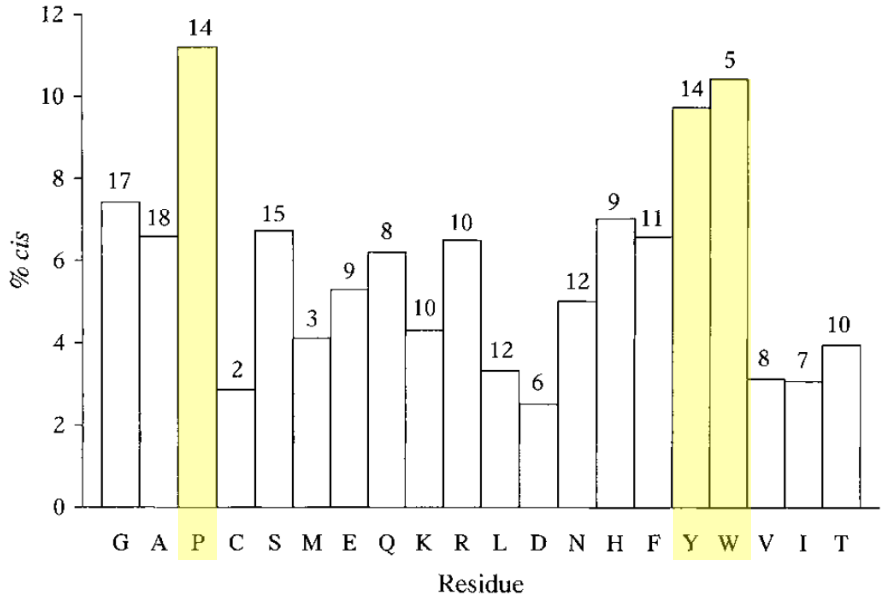


Figure 3.2 Distribution of X_{aa}-cis-Pro amide bonds[285]
Pal & Chakrabarti's analysis of the PDB for cis-prolines in protein crystal structures (≤ 2.0 Å resolution).[285] The frequency of amino acids preceding a cis-proline are shown, with the number of instances indicated for each residue. The three amino acids with highest frequency for preceding a cis-proline (Pro, Tyr, Trp) are highlighted. (Reprinted from Figure 2 of Pal, D.; Chakrabarti, P. Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations, *J. Mol. Biol*, **1999**, *294*, 271-288, with permission from Elsevier.)

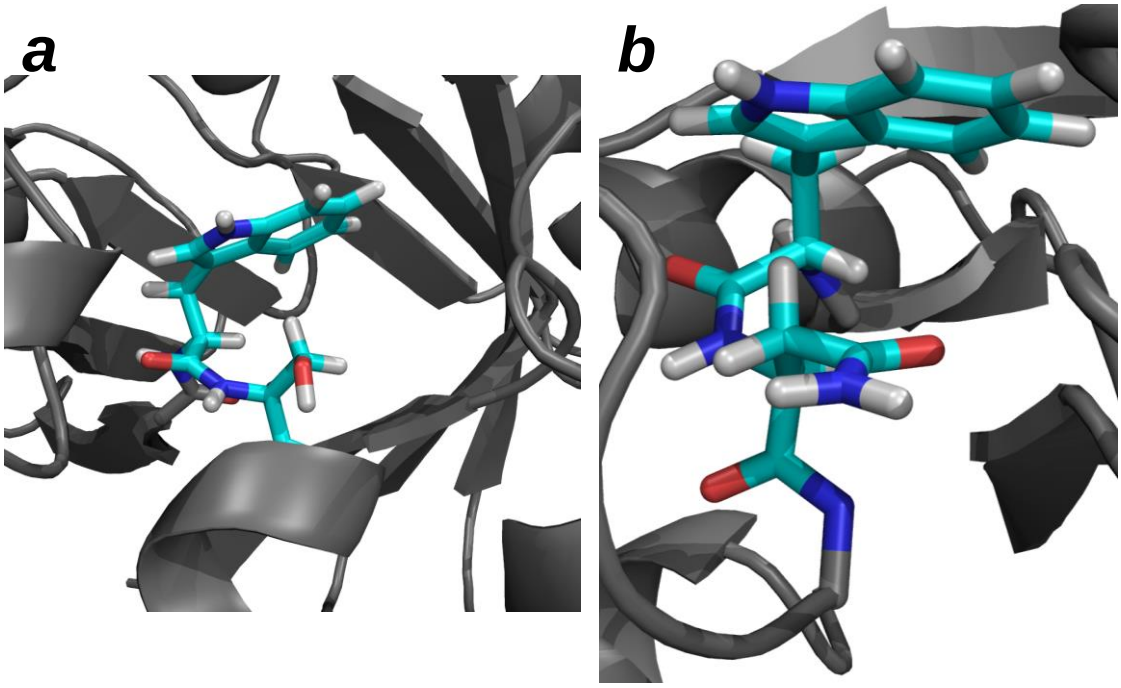


Figure 3.3 Non-proline cis amide bonds with a preceding aromatic amino acid
Examples of non-proline cis-amide bonds with a preceding aromatic amino acid. (a) Trp255-cis-Ser256 in hevamine enzyme, which exhibits lysozyme and chitinase activity (PDB ID: 2HVM);[299] (b) Trp313-cis-Ser314 in endoglucanase from *Clostridium thermocellum* (PDB ID: 1CEC).[300] Aromatic amino acids have a higher likelihood of preceding cis-amide bonds, either with proline or non-proline residues, suggesting that the aromatic ring may be involved in stabilizing the cis conformation. Note the Cβ–Hβ/π interactions in both structures.

Proline cis-trans isomerism can be a rate-determining step in protein folding, and amide conformation of prolines can be the difference between active and inactive forms of an enzyme or receptor.[282-284, 287] While proline cis-trans isomerism occurs spontaneously at room temperature within minutes, prolyl isomerases are enzymes that catalyze the transition between cis and trans isomerism in proline.[283] Proline isomerases stabilize the transition state between cis and trans conformers of proline, thus paying the energetic cost of the transition.[282, 301]

There are numerous biologically relevant examples of the importance of proline cis-trans isomerism, which have been reviewed previously,[282, 287, 295] including folding HIV-1 capsid protein and MS2 bacteriophage coat protein,[5] metal binding in thioredoxin[302] and mannose-binding proteins,[303] RNase A activity,[295] and amyloid fibril formation.[304, 305] While cis amide bonds are disruptive to α-helices, they are important in turn formation, specifically type VI β-turns,[306] which can dictate conformation of exposed recognition loops. In addition, Pin1 is a proline isomerase that is involved in numerous regulatory pathways, including transcription regulation, mitotic progression, and kinase and phosphatase activity.[287]

Brazin *et al.* identified via NMR shift perturbation that the two modes of substrate binding in interleukin-2 tyrosine kinase (Itk) SH2 domain are dictated by a proline within a flexible loop (Figure 3.4).[307] When Asn286-Pro287 is in the trans conformation, Itk SH2 favors binding to its phosphopeptide substrate.[308] When Asn286-Pro287 is in the cis conformation, catalyzed by the prolyl isomerase cyclophilin A, it favors binding to the SH3 domain.[307, 309] In this case, the conformation of a single proline residue affects the conformation of an exposed loop, which can interface between two different substrates, and proline cis-trans isomerism acts as a "switch" between the two binding modes.[282, 307-309]

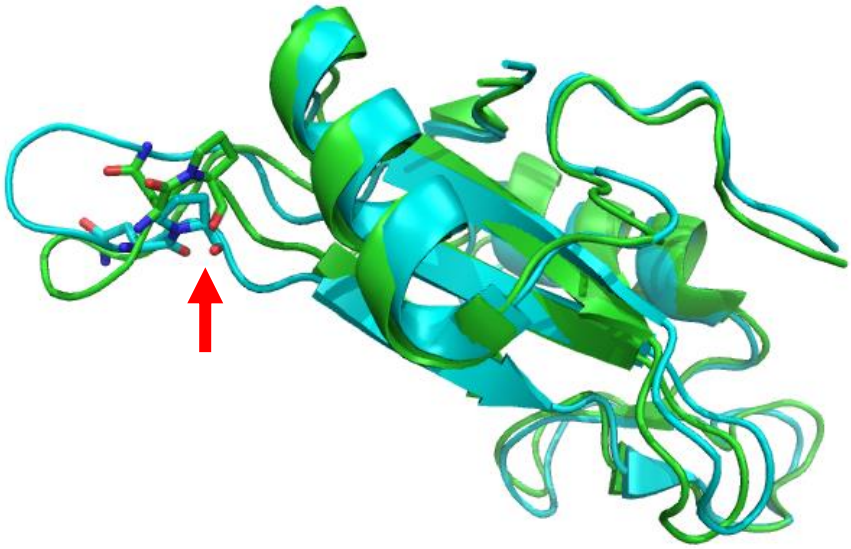


Figure 3.4 Cis-trans isomerism of Pro287 in Itk SH2 domain dictates substrate recognition
Itk SH2 domain with Pro 287 in cis (green, PDB ID: 1LUK) and trans (cyan, PDB ID: 1LUN) conformations. Pro287 (indicated) is part of a flexible loop region that is involved in substrate binding. Trans-Pro287 favors binding to the phosphopeptide substrate, and cis-Pro287 promotes docking with the SH3 domain.[308, 309]

Proline residues also play roles in the structure of transmembrane helices,[310] and it has been suggested conformations of proline can function as a means of regulation in these membrane proteins.[311] The 5-hydroxytryptamine type 3 (5-HT_3) receptor is a pentameric gated-ion channel, which has been proposed to undergo a conformational change that allows ion permeation upon binding its substrate, serotonin.[312-314] The nature of the conformational change upon substrate binding and channel "opening" is not completely clear, but Lummis *et al.* conducted voltage clamp studies on 5-HT_3 receptor protein, and found ion channel permeability was dependent on mutations of a specific proline in the transmembrane helix domain.[314] Pro308 is situated within a flexible loop that spans M2 and M3 helices in the transmembrane domain (Figure 3.5), which was thought to potentially act as a conformational

"hinge".[310, 313, 314] Mutations of Pro308 to cis-promoting non-natural proline analogues resulted in 5-HT_3 receptor protein mutants that were highly active, in that they allowed ion permeation across a membrane that was essentially irreversible.[314] In contrast, Pro308 mutations to trans-promoting analogues resulted in inactive mutants, where no ion permeation was observed across a membrane.[314]

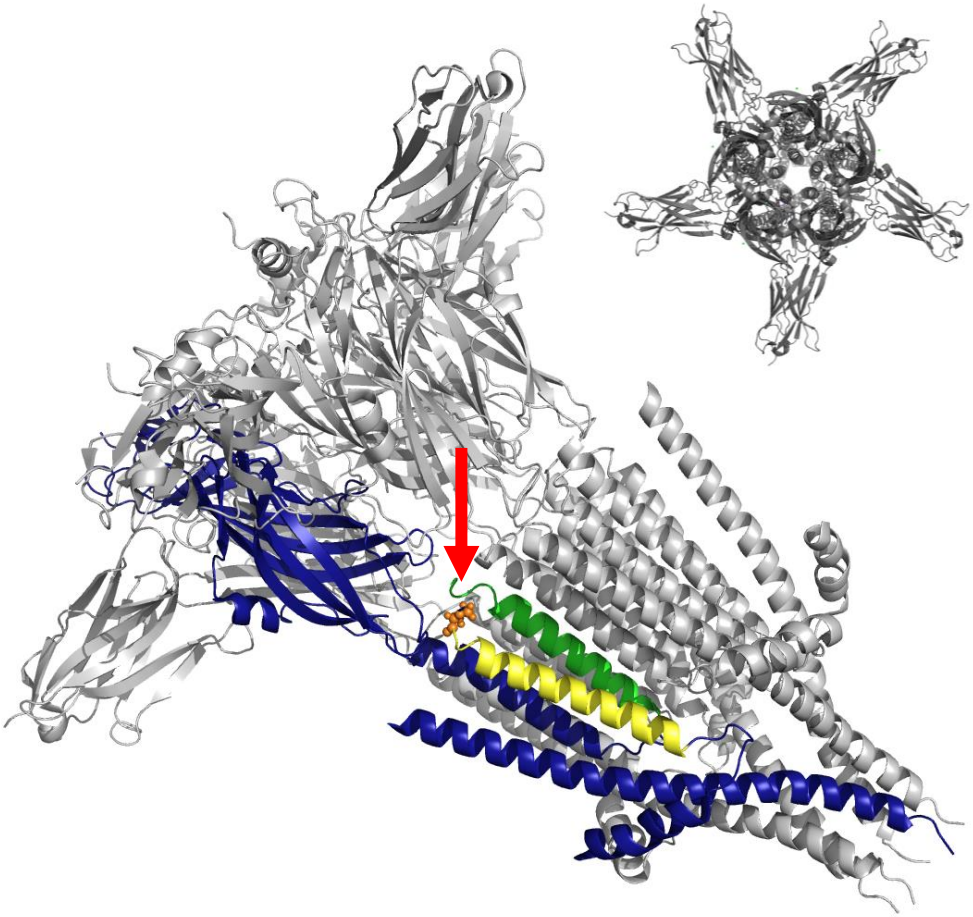


Figure 3.5 Serotonin receptor gated ion channel function is dependent on proline cis-trans isomerism[312-314]
The serotonin receptor protein is a pentameric (inset) gated ion channel (PDB ID: 4PIR).[315] One subunit of the pentamer is shown in blue. On binding to its substrates, the receptor undergoes a conformational change that allows for permeation of cations through the membrane. Pro308 (orange) is part of the flexible loop spanning helix M2 (yellow) and helix M3 (green), and has been implicated in the "opening" and "closing of the ion channel. When Pro308 was mutated to trans-promoting proline analogues, ion flow across a membrane was completely inhibited in the presence of its substrate; when it was mutated to cis-promoting analogues, ion permeation across a membrane was highly active and partially irreversible.[314]

With the enhanced ability of proline to undergo cis-trans isomerism, Nature has utilized this aspect as a conformational "switch" in numerous biological contexts.[282, 287] Understanding the interactions that influence cis-trans isomerism of

proline would further our ability to rationally design proteins and biological materials. This can be particularly useful in modifying substrate recognition loops, or designing metal coordination spheres where proline is known to play a role in local structure. The over-representation of aromatic amino acids near cis-proline residues suggests that cis-prolines may be influenced by non-covalent aromatic interactions; Pal & Chakrabarti specifically suggested that C–H/π interactions may play a role.[285, 286, 292, 316] Studies on proline cis-trans isomerism in model peptides has elucidated some of the interactions between cis-prolines and nearby amino acids.

3.1.2 Proline cis-trans isomerism in model peptides

Numerous studies in model peptides have identified that cis-proline residues have a special propensity with aromatic amino acids. Reimer *et al.* examined cis-proline propensity in model peptides Ac-AXPAK-NH_2 and found higher cis populations in peptides where X = Gly, Pro, Tyr, and Trp.[317] Yao, Dyson, & Wright examined β-turn propensities in a series of peptides based on the biologically relevant fragment from influenza virus hemagglutinin.[289, 318, 319] In this work, model peptides SXPYDV, SYPXDV, and SYPYXV were studied in aqueous conditions for β-turn formation, as measured by % cis-proline content via NMR.[289, 318] It was found that aromatic amino acids preceding or following proline increased the population of β-turn formation with a prolyl cis amide bond.[289, 318] Wu & Raleigh[288] examined GXPG model peptides, and observed the largest mole fractions of cis-prolines when X was tyrosine or tryptophan.

From these model peptide studies, it had become apparent that cis-proline conformations were promoted by an adjacent aromatic amino acid, although aromatic amino acids more than 3 residues from proline disfavor cis-proline. Meng *et al.*[90]

examined tetrapeptide motifs, Ac-XYPZ-NH_2, screening for cis propensity of peptides by varying the amino acids at *i* and *i*+3 positions across the β-turn motif YP. For the amino acid preceding tyrosine (X), small, polar, or β-branched amino acids (Thr, Cys, Val, and Gly) had greater cis populations, while aromatic amino acids at this position disfavored the cis-conformation.[90] Potentially, a competitive aromatic-aromatic interaction between the side-chains of XY (X = aromatic amino acid) destabilized any favorable interactions between tyrosine and proline.[90] Following proline, the opposite trend was apparent, where aromatic amino acids increased cis-populations, and small charged residues increased trans-populations (i.e. Gly, Glu).[90] Ganguly *et al.* studied PPX tripeptides, and found that aromatic amino acids following Pro-Pro motifs enhanced cis populations, due to a Cα–Hα/π interaction between the aromatic side-chain and Pro_{i-2}.[320]

Aromatic amino acids immediately preceding or following proline increases the likelihood of cis conformation, both in proteins[321, 322] and in model peptides.[90, 289, 318] However, these studies did not explore the nature of the aromatic-proline interaction itself. In order to investigate the underlying nature of the aromatic-cis-proline interaction, Thomas *et al.*[91] examined the cis-populations of tetrapeptides Ac-TXPN-NH_2, where X was a canonical or non-natural aromatic amino acid. Meng *et al.*[90] identified the tetrapeptide Ac-TYPN-NH_2 as having the highest cis-proline population, excluding proline or aromatic residues at the *i* or *i*+3 positions, and this peptide served as a basis for the study on the interaction in aromatic-cis-proline motifs. Thomas *et al.*[91] examined substituent effects on cis-proline populations with 4-substituted phenylalanine derivatives in the model peptide Ac-TXPN-NH_2. The aromatic substituents represented a range of electron-withdrawing or electron-

donating substituents on phenylalanine, and heterocyclic and non-aromatic amino acids. Electron-donating aromatic substituents increased cis-proline populations (i.e. 4-aminophenylalanine, tyrosine phenolate), and electron-withdrawing aromatic substituents increased trans-proline populations (i.e. 4-nitrophenylalanine, 4-trifluoromethylphenylalanine).[91] A model was proposed for how aromatic amino acids preceding proline effected cis-trans isomerism (Figure 3.6).[91]

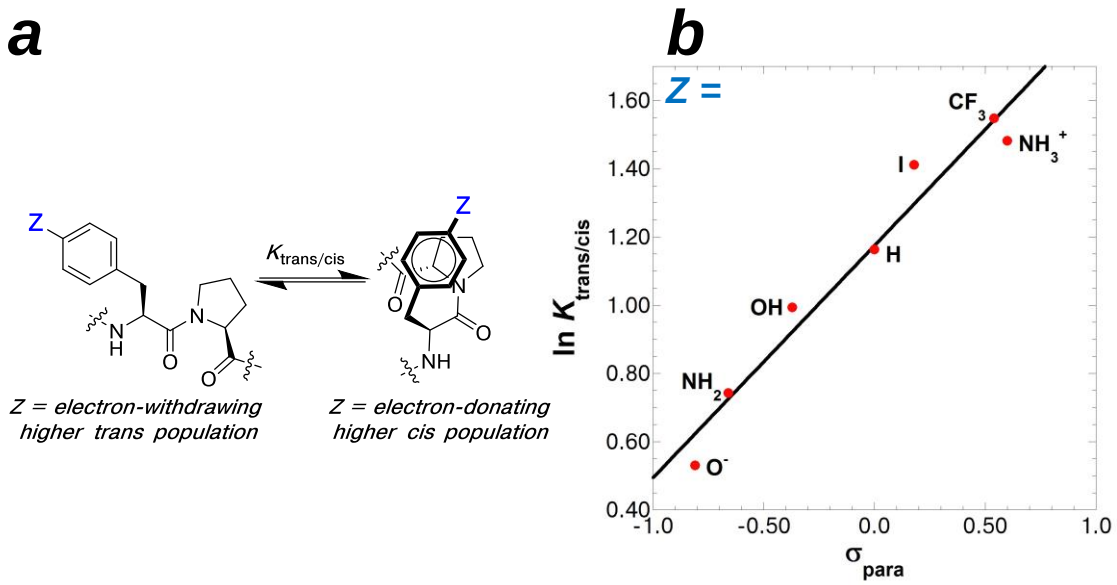


Figure 3.6 Electronic effects on proline cis-trans isomerism through aromatic substituents[91]
Proline cis-trans isomerism was measured in model peptides Ac-TXPN-NH_2 as $K_{trans/cis}$, where X was an aromatic amino acid with substituent Z. The $K_{trans/cis}$ (a) of the peptides were measured via NMR. A direct correlation was observed between the $K_{trans/cis}$ of the peptide and the electron-donating ability (σ_{para}) of the aromatic substituent. Electron-donating aromatic substituents had higher cis-proline populations (lower $K_{trans/cis}$), and electron-withdrawing substituents had higher trans populations (higher $K_{trans/cis}$).

Thomas *et al.*[91] observed a direct correlation between the cis-populations of peptides Ac-TXPN-NH_2 and the electron-donating ability of the aromatic substituent in the amino acid X. This study in model peptides established an electronic influence on cis-trans isomerism of proline through an adjacent aromatic amino acid.[91] Peptides containing non-aromatic amino acids, cyclohexylalanine and alanine, exhibited high populations of trans amide bonds, suggesting that the aromatic-cis-proline interaction was not stabilized by a purely hydrophobic effect.[91] Furthermore, these data suggested that the aromatic-cis-proline conformation could be tuned and controlled via aromatic

substituent effects. NMR data indicated close contact distances between the prolyl Hα and the aromatic ring in the peptide Ac-TWflpN-NH_2 (flp = 2*S*,4*S*-fluoroproline),[91] suggestive of an aromatic C–H/π interaction.

3.1.3 The nature of C–H/π interactions

In protein crystal structures, Pal & Chakrabarti observed a specific interaction in aromatic-cis-proline motifs between the prolyl Cα–Hα bond and the aromatic Cγ, suggesting a directionally defined C–H/π interaction.[285] In the model studies by Thomas *et al.*,[91] it was suggested that the β-turn motif was stabilized by a proline C–Hα/π aromatic interaction. With the sensitivity of the peptide structure to aromatic substituent effects, the peptide context Ac-TXPN-NH_2 provides a unique model to observe, identify, and characterize C–H/π interactions in the context of proteins in water. Similar to S–H/π interactions (Chapter 2), C–H/π aromatic interactions are generally regarded as weak hydrogen bonds that are driven by dispersive interactions.[190]

The C–H bonds are ubiquitous in proteins and small organic molecules, and can interact with π bonds in hydrogen bond-type interactions. Given the comparatively low electronegativity of carbon, C–H bonds are weakly polarized, and participate as a soft acid in non-covalent interactions.[106] Non-covalent C–H/π interactions can dictate stereoselectivity of some reactions, and can influence conformations of inter-ligand interactions in organometallic complexes (Figure 3.7).[106] The strength of an individual C–H/π interaction is typically 1 kcal mol^{-1} or less,[106] but cooperativity of multiple interactions can contribute substantially to the stability of a molecule or protein (Figure 3.8).[190] C–H/π interactions exhibit defined geometry and directionality, which is inconsistent with interactions that are purely hydrophobic or dispersive in nature.[106]

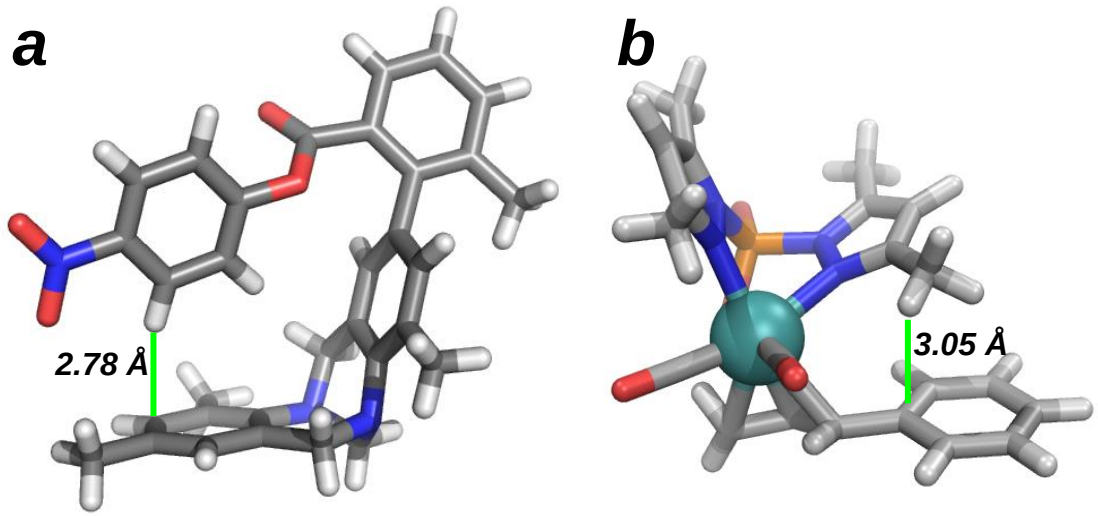


Figure 3.7 C–H/π aromatic interactions that stabilize conformations of complex small molecules[323]

(a) An "edge-to-face" aryl C–H/π interaction (CSD identifier: PIWYEZ). Substituent effects were examined within the context of aryl esters to determine the strength of this aromatic intermolecular C–H/π interaction;[324] (b) An aliphatic C–H/π interaction between ligands in a molybdenum complex. (CSD identifier: POMJUW).[325]

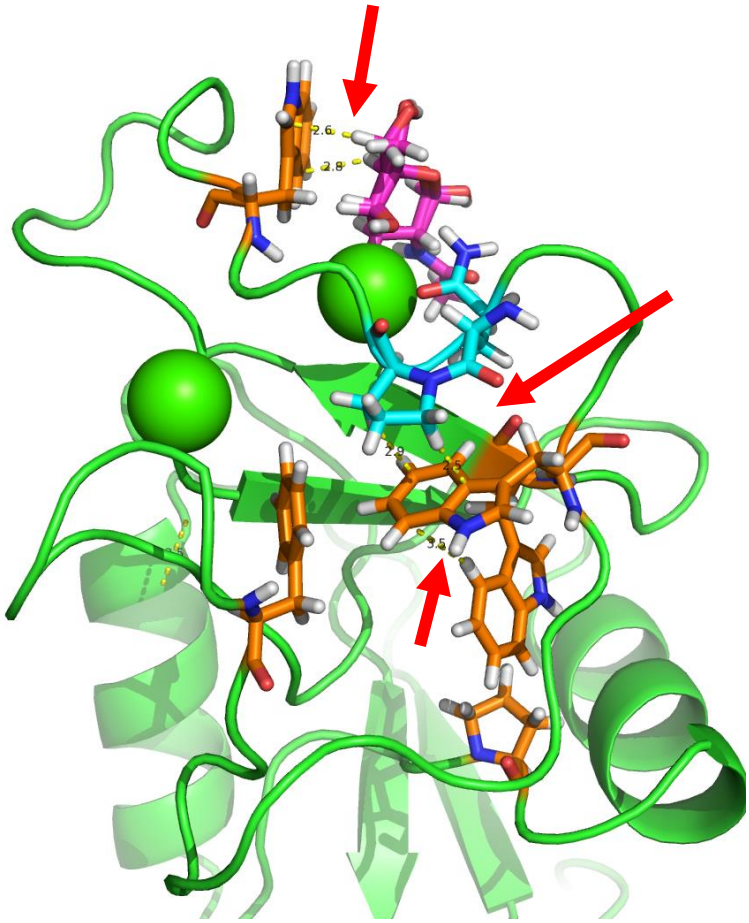


Figure 3.8 Extensive C–H/π aromatic interactions in mannose binding protein
Multiple aromatic residues (orange) are involved in a network of C–H/π aromatic interactions that stabilize mannose binding protein. C–H/π aromatic interactions can be seen between Trp181 and cis-Pro186, shown in cyan. The proximity of cis-Pro186 to the calcium (shown in spheres) and mannose (magenta) binding sites is also noteworthy, as the cis-Pro (stabilized by Trp181) is potentially requisite for binding (PDB ID: 1BCH).[326]

C–H bonds, as hydrogen bond donors, became recognized along with other X–H hydrogen bonds in the late 1930s, as the ^{1}H NMR chemical shift and IR frequency of chloroform were known to shift upon interactions with different solvents (also see Chapter 2.1.1 and Chapter 2.1.2).[190] Similar to S–H/π aromatic interactions, C–H/π

aromatic interactions were initially identified by IR, where chloroform had an observable frequency shift in the presence of benzene, suggesting that the aromatic ring interacted with the C–H bond.[327] Nature has harnessed C–H/π interactions for protein stability and substrate recognition in many proteins and biomolecules, and numerous examples have been reviewed in the literature.[105, 160, 161, 222, 290, 328] For example, C–H/π aromatic interactions have been implicated in saccharide binding to a chemoreceptor in *E. coli*, where three C–H bonds are directed at a tryptophan aromatic ring (Figure 3.9a).[105, 329] The heme group in horse hemoglobin is also coordinated by several C–H/π interactions (Figure 3.9b).[330] The trp cage miniprotein is stabilized by cross-helix C–H/π aromatic interactions, as described in Chapter 1 (Figure 1.27).

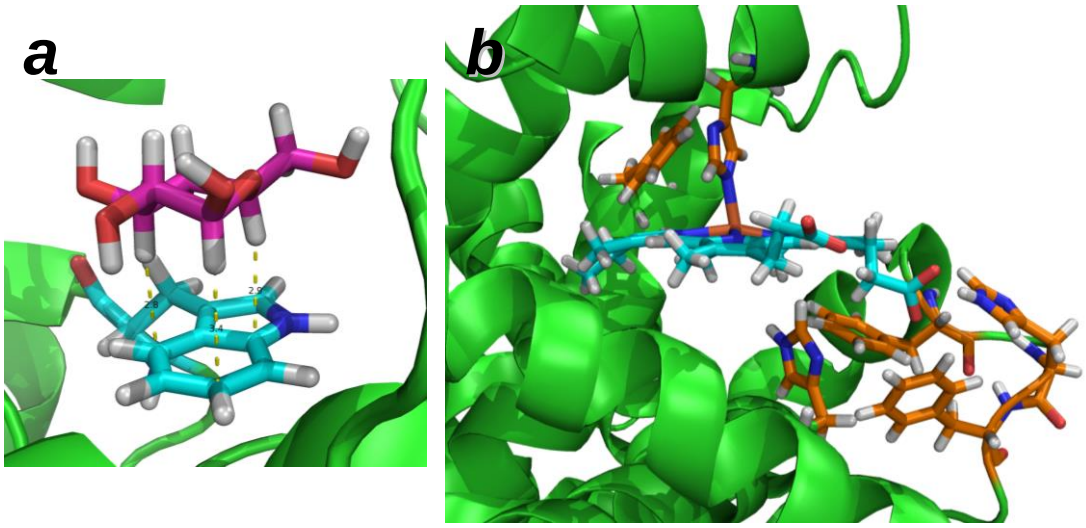


Figure 3.9 Multiple C–H/π aromatic interactions involved in coordinating substrates and cofactors
(a) β-D-Glucose interacts with Trp 183 in D-galactose binding protein from *E. coli* (PDB ID: 2GBP).[105, 329] The C–H··C_{aro} distances are near or below the sum of the van der Waals radii for hydrogen and carbon (2.90 Å); (b) numerous aromatic amino acids are involved (orange) in C–H/π interactions are involved in the coordination sphere around a heme group (cyan) in hemoglobin (PDB ID: 2DHB).[330] Several leucine residues in this region (not shown) have close contacts with the heme group, which may be argued as C–H/π aromatic interactions (or hydrophobic interactions).

Given the prevalence of aromatic rings in proteins and DNA, understanding the mechanisms of aromatic interactions is crucial for understanding ligand binding, for analysis of SAR studies, and for designing small molecule and peptide-based inhibitors. Brandl *et al.*[331] found that 50% of aromatic residues in crystal structures in the PDB were found to participate as "acceptors" in aromatic interactions, underscoring the importance of identifying underlying energetic features and structural consequences of C–H/π interactions. Based on the observed trends in proteins for aromatic-cis-proline motifs,[285] and based on work by Thomas *et al.*[91] we sought to further investigate the nature, strength, and scope of C–H/π interactions in peptide

structure. We aim to identify the fundamental nature of C–H/π interactions by closely examining the aromatic-cis-proline interaction by ^{1}H NMR, thermodynamic analysis, and crystallographic observations. Through this work, we will also evaluate the scope and breadth of electronic effects on cis-trans isomerism of proline, using an expanded set of aromatic amino acids with novel substituents.

3.2 Results

3.2.1 Synthesis of an expanded series of peptides Ac-TXPN-NH_2

The work by Thomas *et al.*[91] demonstrated that the electron-donating ability of aromatic substituents influenced cis-amide bond populations of Ac-TXPN-NH_2 model peptides (X = 4-substituted aromatic amino acid). In this work,[91] it was suggested that the aromatic-cis-proline motif was stabilized by an interaction between the aromatic ring and the prolyl C–H bonds. We sought to characterize the scope and limitations of this aromatic-cis-prolyl interaction using non-natural aromatic amino acids. Building from the work completed by Thomas *et al.*,[91] we expanded on the library of non-canonical amino acids used in peptides Ac-TXPN-NH_2 to encompass a broader range of aromatic substituents. In addition to electronic effects on cis-trans isomerism of proline, many of these non-natural aromatic amino acids introduce novel functionality or properties, potentially for bioconjugation or fluorescence applications. The electronic effects of non-natural aromatic substituents were quantified by measuring the equilibrium between cis and trans amide bond conformations ($K_{trans/cis}$) via ^{1}H NMR.[91] The following sections (3.2.1.1 and 3.2.1.2) describe the synthesis of aromatic amino acids in the expanded library of peptides Ac-TXPN-NH_2. NMR spectra of the

peptides, used for calculating the $K_{trans/cis}$, are incorporated within these sections; a summary of the $K_{trans/cis}$ values are included in a table in Chapter 3.2.2 (Table 3.1).

3.2.1.1 Synthesis of a series of peptides Ac-TXPN-NH_2 (X = 4-substituted aromatic amino acid)

3.2.1.1.1 Additional Ac-T(4-X-Phe)PN-NH_2 peptides generated from commercially available aromatic amino acids

In our expansion on the library of model peptides Ac-TXPN-NH_2,[91] additional commercially available aromatic amino acids were incorporated via solid-phase peptide synthesis (where X = His, 4-fluorophenylalanine, 4-chlorophenylalanine, 4-bromophenylalanine, and 4-cyanophenylalanine). The purification conditions for these peptides are described in the experimental section. The $K_{trans/cis}$ for each peptide was measured via ^{1}H NMR (Figures 3.10-3.14). The equilibrium between cis and trans conformations ($K_{trans/cis}$) was determined by measuring the ratios of 2 or 3 pairs of amide proton signals in the ^{1}H NMR spectra ($\Sigma\int$(trans signals)/$\Sigma\int$(cis signals)).[91]

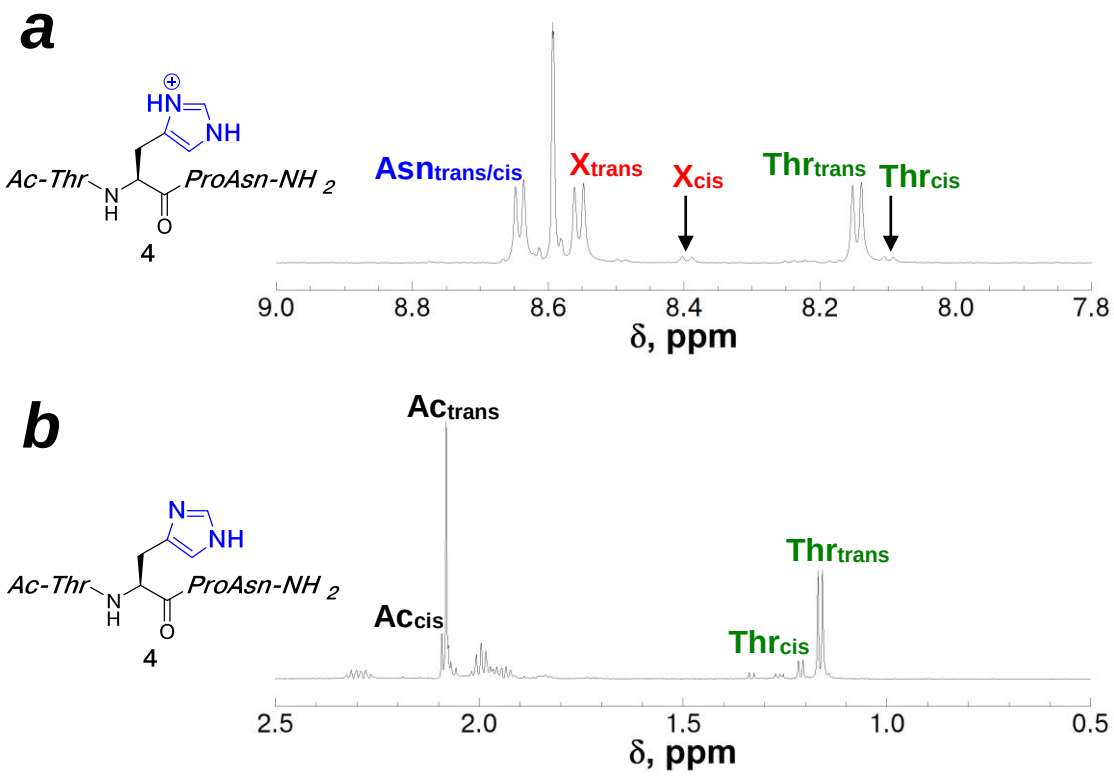


Figure 3.10 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = His)
NMR spectrum of amide region for the peptide Ac-THPN-NH_2 at (a) pH 4.0 and (b) pH 8.5. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

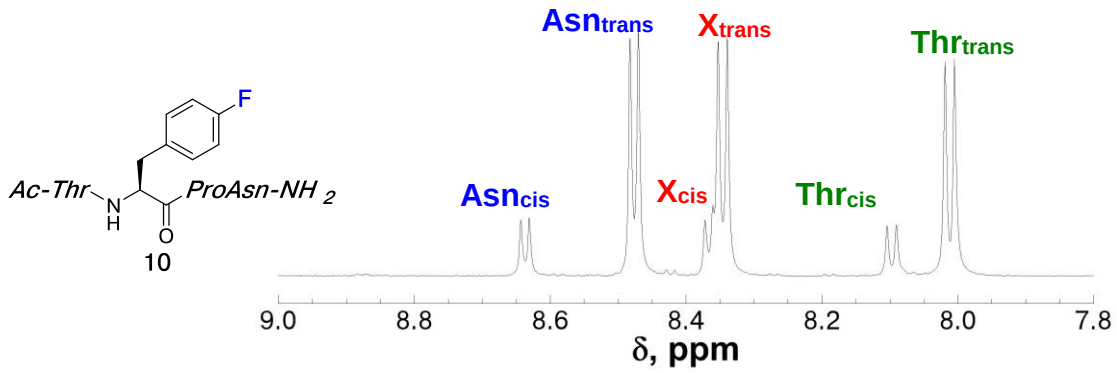


Figure 3.11 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-fluoro-phenylalanine)

NMR spectrum of amide region for the peptide Ac-T(4-fluoro-phenylalanine)PN-NH_2 pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

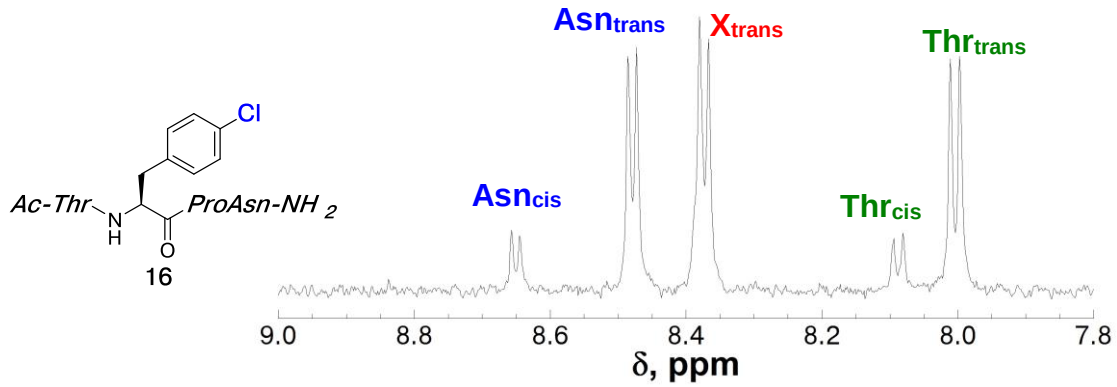


Figure 3.12 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-chloro-phenylalanine)

NMR spectrum of amide region for the peptide Ac-T(4-chloro-phenylalanine)PN-NH_2 pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

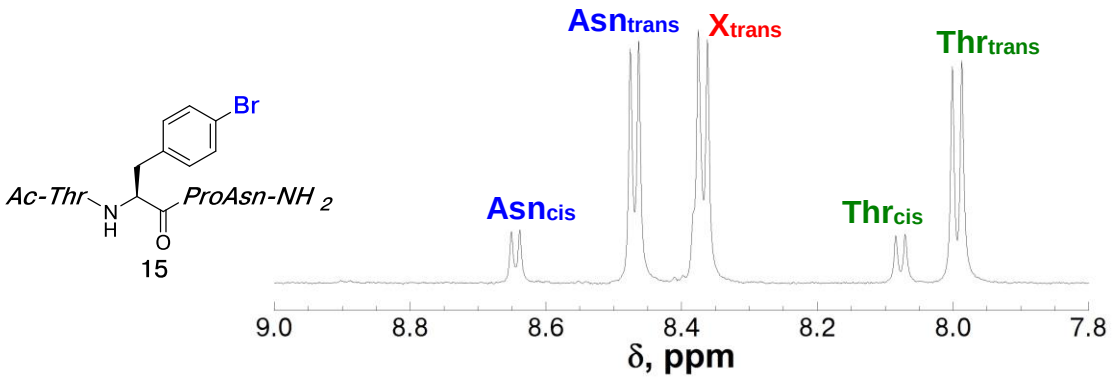


Figure 3.13 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-bromo-phenylalanine)
NMR spectrum of amide region for the peptide Ac-T(4-bromo-phenylalanine)PN-NH_2 pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

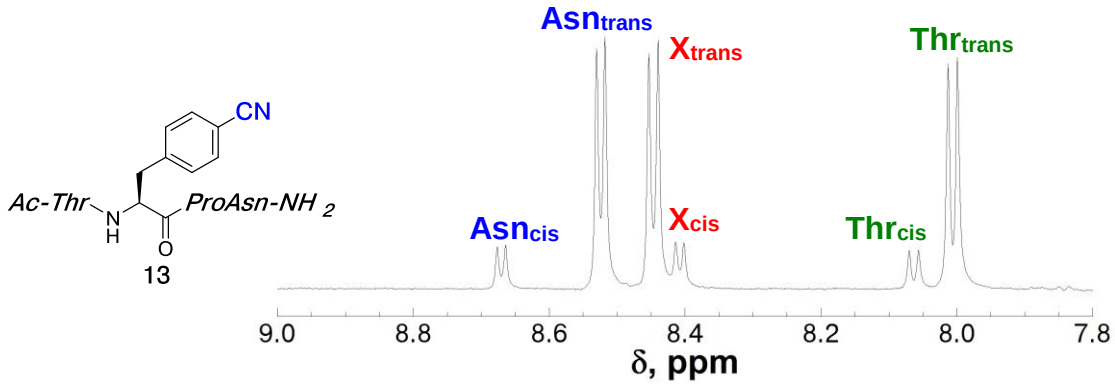


Figure 3.14 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-cyano-phenylalanine)
NMR spectrum of amide region for the peptide Ac-T(4-cyano-phenylalanine)PN-NH_2 pH 4.0. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

These additional aromatic amino acids, histidine, 4-fluorophenylalanine, 4-chlorophenylalanine, 4-bromophenylalanine, and 4-cyanophenylalanine, were not previously characterized and quantified for aromatic electronic effects on proline cis-trans isomerism. Notably, the peptide Ac-THPN-NH_2 exhibited different cis and trans

populations depending on the protonation state of histidine (8% cis at pH 4; 15% cis at pH 8.5), suggesting that histidine can act as a structural "switch" for prolyl-amide bond conformation. These additional data provided important comparisons of the structural and electronic consequences of halogen substituents on aromatic rings, and the structural and electronic implications of histidine at different protonation states.

3.2.1.1.2 Solid-phase synthesis of the peptide Ac-T(4-B(OH)$_2$-Phe)PN-NH$_2$

The development of novel amino acids for bioconjugation reactions is a constantly evolving area of research, specifically for site-selective protein tagging and modification.[51] As new organic reactions with broad substrate scope are developed, bioorganic chemists adapt these reactions and substrates to be compatible with proteins. Suzuki-Miyaura coupling reactions, palladium-catalyzed reactions between aryl boronates and aryl iodides to form biaryl linkages (Figure 3.15), have several examples in modifying organic molecules and peptides on solid phase,[332] and have been more recently adapted and demonstrated for site-selective modification in proteins.[333, 334] Schultz and coworkers[335, 336] have developed effective strategies for genetic incorporation of 4-boronic acid-phenylalanine (4-B(OH)$_2$-Phe) in Green Fluorescent Protein (GFP) as a "turn-on" sensor for oxidative conditions.[335] GFP containing 4-B(OH)$_2$-Phe was not fluorescent, but under oxidative conditions 4-B(OH)$_2$-Phe was converted to tyrosine, and GFP fluorescence was restored.[335] Aryl boronates have also been utilized for conjugating diols and carbohydrates,[337-339] and 4-B(OH)$_2$-Phe can have potential applications for designing peptide-based sensors for post-translational glyosylation reactions.

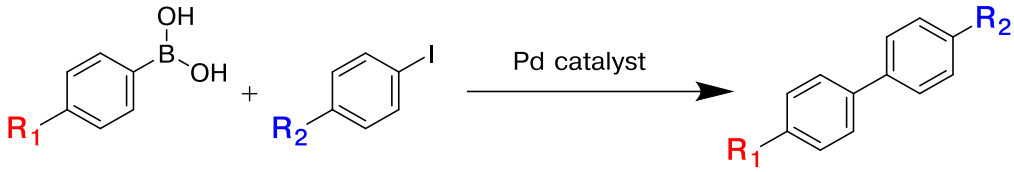


Figure 3.15 Generalized Suzuki-Miyaura coupling reaction
The palladium-catalyzed formation of biaryl linkages from aryl halides and aryl boronates have been widely utilized in protein modification, surface chemistry, sensing applications, and in organic synthesis.[332, 338-340]

Given the wide utility of aryl boronic acids for protein modification and for applications in biological sensors, we sought to characterize the aromatic electronic effects for charged and neutral 4-$B(OH)_2$-Phe using the model peptide Ac-TXPN-NH_2. In addition, the pK_a of phenylboronic acid is 8.9,[341] suggesting that 4-$B(OH)_2$-Phe can function as a structural switch in the context of proline. The change in fluorescence properties in mutated GFP on replacing tyrosine with 4-$B(OH)_2$-Phe is an indicator of significantly shifted aromatic electronics.[335] 4-$B(OH)_2$-Phe has been synthesized previously, both in solution[342] and via solid-phase peptide modification strategies.[343] We utilized a more practical solid-phase modification methodology that was previously developed,[343] where resin-bound peptides containing commercially available 4-iodophenylalanine were subjected to palladium-mediated cross-coupling reaction with bos(pinacolato)diboron in order to generate the peptide Ac-T(4-$B(OH)_2$-Phe)PN-NH_2 (Figure 3.16).

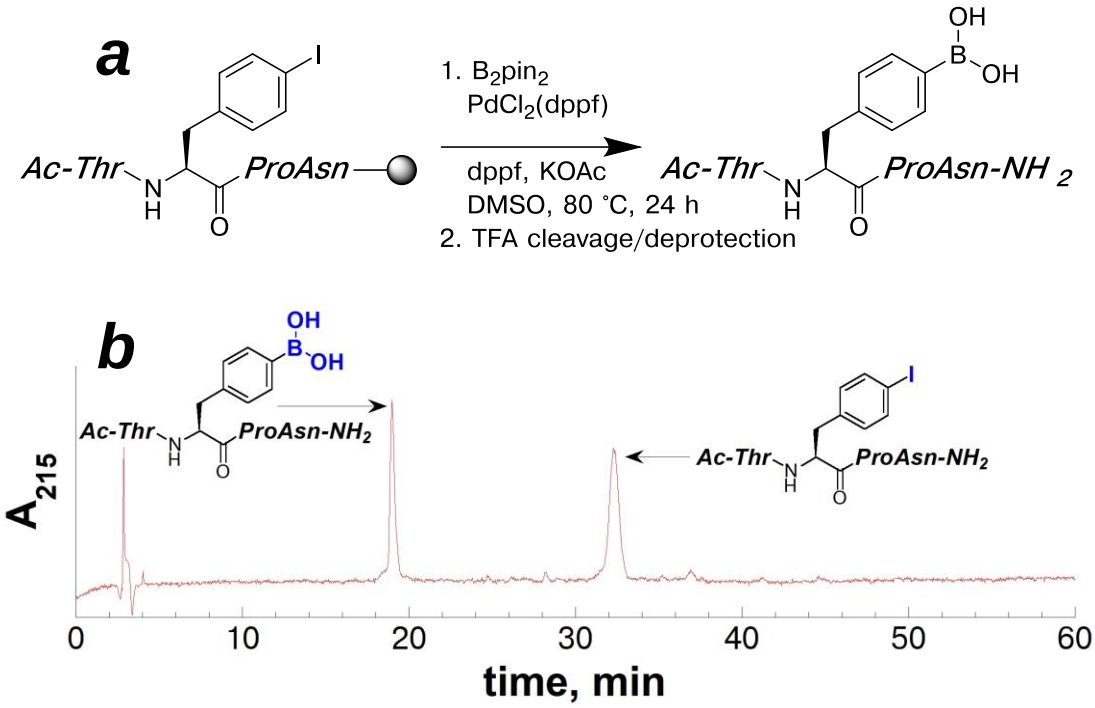


Figure 3.16 Synthesis of the peptide Ac-TXPN-NH_2 (X = 4-$B(OH)_2$-phenylalanine) on solid phase
(a) Scheme for the palladium-mediated cross-coupling reaction on Ac-T(4-iodophenylalanine)PN-NH_2 to generate the peptide containing 4-$B(OH)_2$-Phe; (b) HPLC chromatogram of the crude reaction products resulting from the palladium-mediated cross-coupling reaction on the peptide containing 4-iodophenylalanine using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

The model peptide containing 4-$B(OH)_2$-Phe was cleanly synthesized via modification of the peptide containing 4-iodophenylalanine on solid phase in 41% conversion with no side-products. The product and the starting material had identical molecular weights, and so the recovered product at 31 minutes (Figure 3.16b) was confirmed via NMR to be the peptide Ac-T(4-I-Phe)PN-NH_2 via superposition with authentic Ac-T(4-$B(OH)_2$-Phe)PN-NH_2.

The purified peptide Ac-T(4-B$(OH)_2$-Phe)PN-NH_2 was characterized via NMR under acidic and basic conditions to characterize both charged and neutral forms of this aromatic amino acid (Figure 3.17). The equilibrium between cis and trans conformations ($K_{trans/cis}$) was measured from the ^{1}H NMR spectra using amide signals. However, for the spectrum at pH 10, the Thr methyl signals, the acetyl methyl, and Thr Hα were used for calculations instead due to rapid exchange of the amide protons at pH > 6.5.

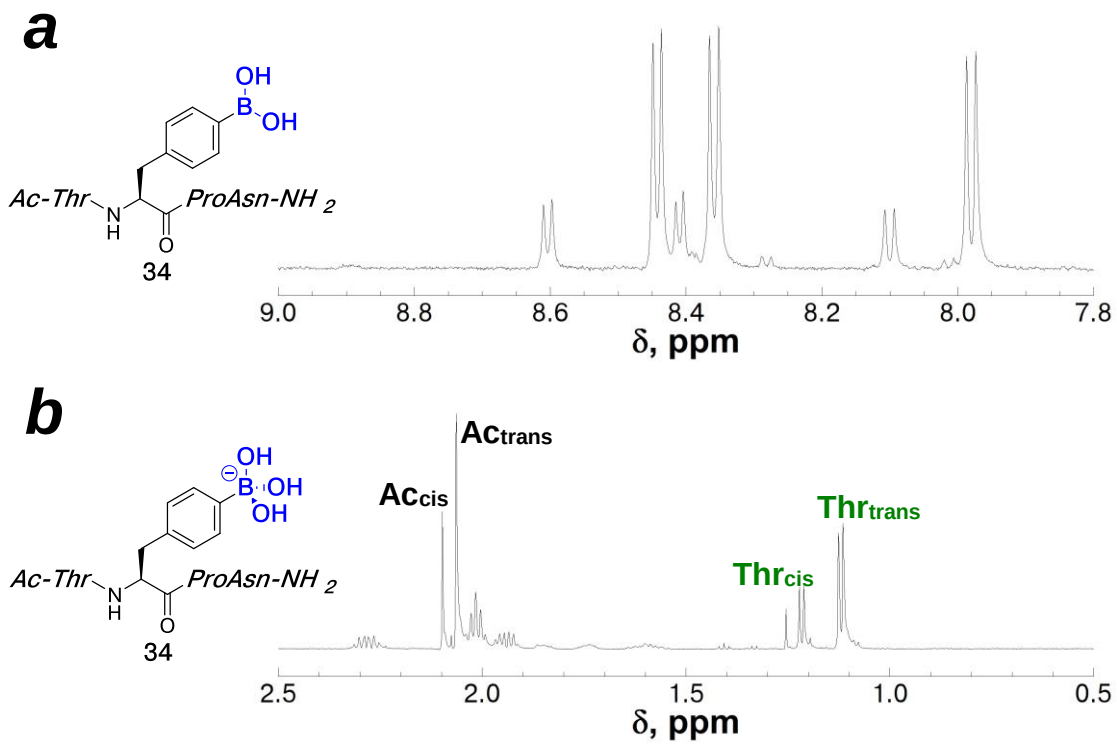


Figure 3.17 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-$B(OH)_2$-phenylalanine)
NMR spectrum of the amide region of the peptide Ac-T(4-$B(OH)_2$-Phe)PN-NH_2 at (a) pH 4 and (b) pH 10. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. The aliphatic region is shown for pH 10 (due to amide exchange at high pH, the amide proton signals are not observed).

The peptide containing the boronate form of 4-$B(OH)_3^-$-Phe had a slightly higher population of the cis conformation ($K_{trans/cis}$ = 2.4) over the boronic acid form of 4-$B(OH)_2$-Phe ($K_{trans/cis}$ = 3.5). It can be expected that boronate esters, such as those formed by boronates reacting with diols or carbohydrates, may affect electronic distribution of the aromatic ring, although other factors (changes in sterics and hydrophobic surface area) may dominate over the aromatic electronic effects. Peptides

containing 4-$B(OH)_2$-Phe can potentially function as structural switches in response to changes in pH, presence of carbohydrates, or oxidative conditions.

3.2.1.1.3 Solution phase synthesis of the peptide Ac-T(4-Pyridyl(N-oxide)-Ala)PN-NH_2

The protonation state of aromatic substituents affects aromatic ring electronics, which manifests as a change in the $K_{trans/cis}$ in model peptides Ac-TXPN-NH_2.[91] Aromatic amino acids that induced substantial changes in the $K_{trans/cis}$ with changes in protonation state included 4-aminophenylalanine, 4-pyridyl alanine, histidine, and tyrosine.[91] In the same sense that protonation state of aromatic substituents can influence the cis-population of peptides Ac-TXPN-NH_2, the oxidation state of aromatic substituents can be expected to have significant structural consequences as well. Derivatives of 4-thiophenylalanine representing different oxidation states of sulfur exhibited substantially different UV spectra, indicating changes in the aromatic electronic properties (Chapter 1, Figure 1.55).[169] 4-Pyridyl alanine has different populations of cis-proline for protonated ($K_{trans/cis}$ = 5.7,) and neutral forms ($K_{trans/cis}$ = 7.6).[91] Oxidation of 4-pyridyl alanine to form the N-oxide can significantly alter the electron distribution in the pyridyl ring, in turn affecting the relative population of cis-prolines. The aromatic ring in pyridyl N-oxides is relatively electron deficient, and the 4-pyridyl(N-oxide) alanine derivative was expected destabilize to the cis amide bond conformation, exhibited by a high $K_{trans/cis}$ in the model peptide Ac-TXPN-NH_2.

Pyridine rings are readily oxidized to form N-oxides,[344-346] the peptide containing 4-pyridyl(N-oxide) alanine was readily generated within the synthesized peptide in solution following a previously established protocol.[346] The peptide Ac-T(4-pyridyl alanine)PN-NH_2 was subjected to reaction with hydrogen peroxide in

acetonitrile to generate the peptide containing 4-pyridyl(N-oxide) alanine (Figure 3.18). The product was verified via ESI-MS, and the $K_{trans/cis}$ for the peptide Ac-T(4-pyridyl(N-oxide)-Ala)PN-NH_2 was obtained via NMR (Figure 3.19).

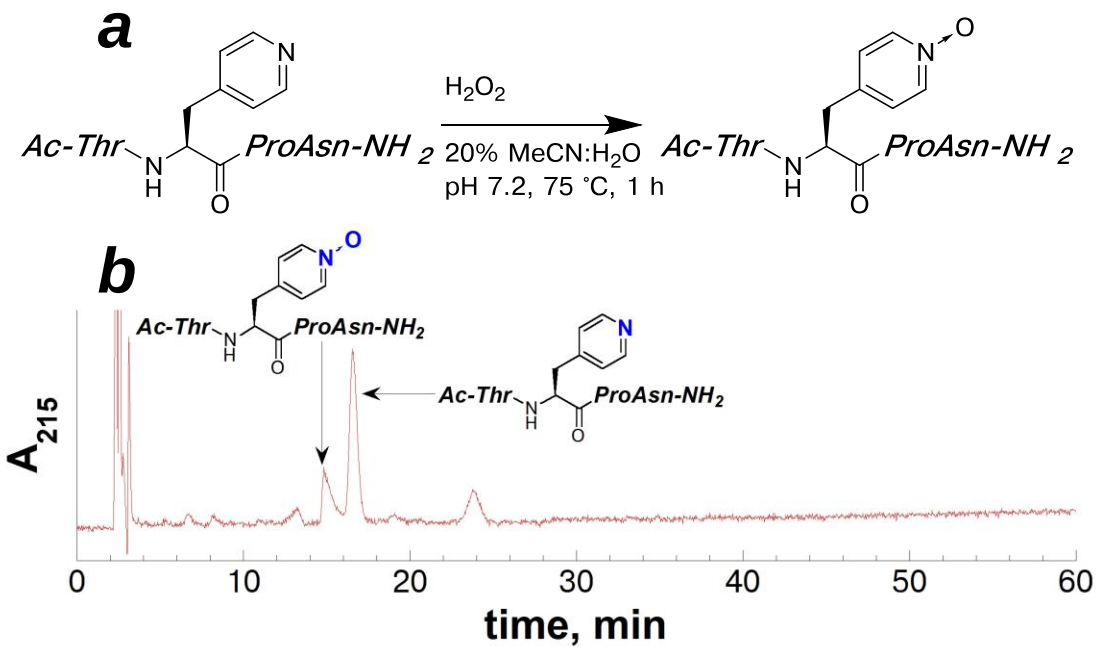


Figure 3.18 Synthesis of the peptide Ac-TXPN-NH_2 (X = 4-pyridyl(N-oxide)-alanine) in solution phase
(a) Scheme for the reaction on the peptide Ac-T(4-pyridyl alanine)PN-NH_2 to generate the peptide containing 4-pyridyl(N-oxide) alanine; (b) HPLC chromatogram of the products resulting from the crude reaction to generate the peptide Ac-T(4-pyridyl(N-oxide) alanine)PN-NH_2 using isocratic 100% buffer A (98% water, 2% MeCN, 0.06% TFA) for 20 minutes, followed by a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A over 40 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

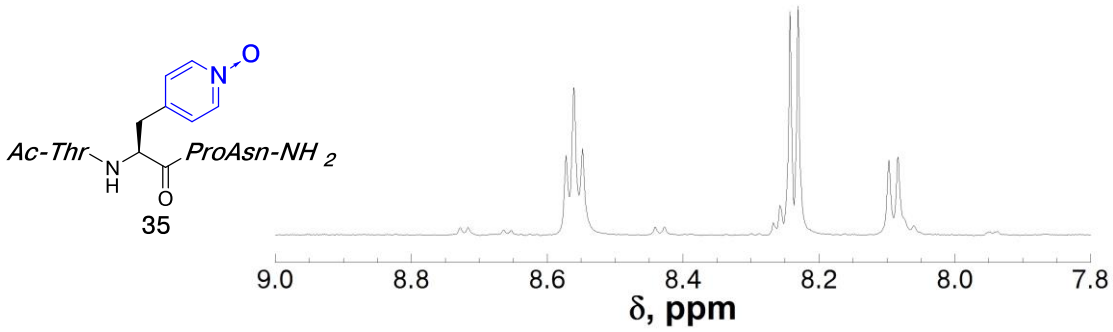


Figure 3.19 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-pyridyl(N-oxide) alanine).
NMR spectrum of the amide region of the peptide Ac-T(4-Pyridyl(N-oxide)-Ala)PN-NH_2 at pH 4. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

The 4-pyridyl alanine amino acid can tune cis-trans isomerism of proline via two different modes of modification: via protonation or oxidation (N-oxide). Cis-trans isomerism of proline can be tightly controlled or inducibly switched upon changing pH conditions or upon oxidizing the peptide containing 4-pyridyl alanine. 4-Pyridyl-alanine can potentially be utilized for designing peptide-based sensors for oxidative conditions, or designing peptide-based biomaterials that expand in morphology upon oxidation or protonation.

3.2.1.1.4 Solid-phase synthesis of the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2

Heterocycles containing nitrogen and sulfur are broadly employed in pharmaceutical compounds,[347] and examination of heterocyclic systems can help provide insights into C–H/π interactions in medicinal chemistry applications. Given that nitrogen heterocycles were shown to have significant effects on aromatic rings through changes in protonation or oxidation state, we sought to characterize an

aromatic amino acid functionalized with a nitrogen a pyrrolidyl functionality, the most common 5-membered non-aromatic nitrogen heterocycle utilized in pharmaceuticals.[347] Utilizing Buchwald-Hartwig methodologies in palladium-catalyzed cross-coupling reactions with aryl halides and amines,[24, 348-350] we developed a synthesis for peptides containing 4-pyrrolidyl-phenylalanine via cross-coupling reaction on solid-phase. Similar to the strategy employed for the synthesis of 4-thiophenylalanine (described in Chapter 1), the resin-bound peptide Ac-TXPN-NH_2 containing a halogenated aromatic amino acid was subjected to cross-coupling reactions with pyrrolidine. Pyrrolidine was utilized for reaction development, but other secondary amines, such as pyrroles, could also potentially be incorporated as aromatic substituents within peptides using these approaches.

Initial attempts to synthesize the peptide containing 4-pyrrolidyl-phenylalanine from the peptide containing 4-chlorophenylalanine or 4-iodophenylalanine were unsuccessful (using copper- and palladium-mediated methodologies).[351, 352] However, the peptide containing 4-bromophenylalanine successfully generated 4-pyrrolidyl-phenylalanine using pyrrolidine via a palladium-mediated cross-coupling approach on solid-phase.[348] Epimerization was observed via HPLC for both the product peptide and reactant peptide when sodium *tert*-butoxide was used as a base in these cross-coupling reactions, as had been employed by Buchwald and coworkers.[348] Other bases were screened in the cross-coupling reaction conditions in order to reduce epimerization of the peptide. Little or no product was observed when basic tertiary amines (DIPEA and DBU) were utilized in the pyrrolidine cross-coupling reaction. Cesium carbonate reactions generated the highest conversion to 4-pyrrolidyl-phenylalanine, but significant epimerization was observed when the reaction was conducted at

temperatures above 60 °C (Figure 3.20). The highest reaction conversion with the lowest epimerization was observed when the cross-coupling reaction was conducted at 50 °C for 12 hours, and then at 60 °C for the final 8 hours of the reaction (Figure 3.21).

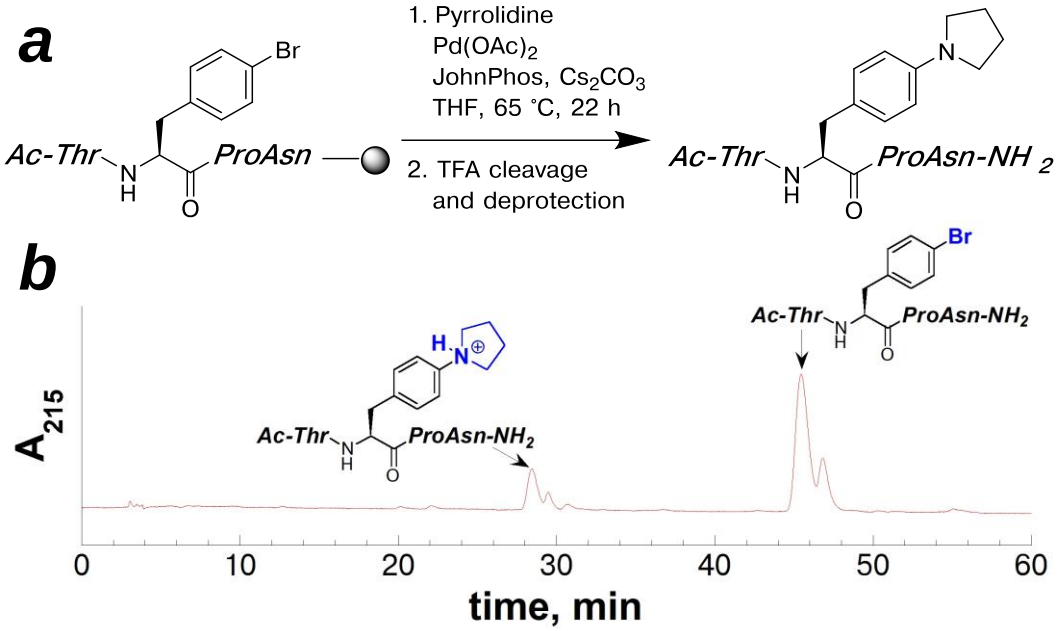


Figure 3.20 Initial synthesis of the peptide Ac-TXPN-NH_2 (X = 4-pyrrolidyl-phenylalanine) on solid phase
(a) Scheme for the initial solid-phase palladium-mediated cross-coupling reaction on Ac-T(4-Br-Phe)PN-NH_2 to generate the peptide containing 4-pyrrolidyl-phenylalanine, where the reaction was conducted at 65 °C for 22 hours; (b) HPLC chromatogram of the crude reaction products resulting from the conditions shown in (a). The epimerization of the products and reactants is apparent as two peaks. HPLC chromatogram was obtained using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.
"JohnPhos" refers to the phosphine ligand, (2-biphenyl)di-*tert*-butylphosphine[348]

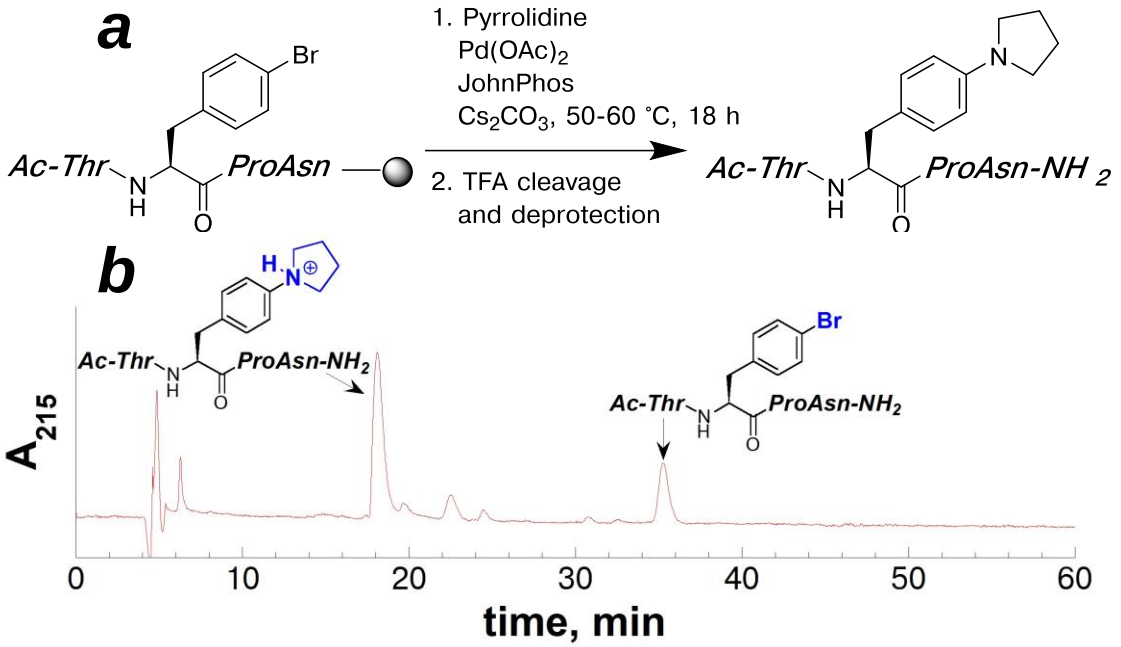


Figure 3.21 Optimized synthesis of the peptide Ac-TXPN-NH_2 (X = 4-pyrrolidyl-phenylalanine) on solid phase
(a) Scheme for the solid-phase palladium-mediated cross-coupling reaction on Ac-T(4-Br-Phe)PN-NH_2 to generate the peptide containing 4-pyrrolidyl-phenylalanine, where the reaction was conducted at 50 °C for the first 12 hours of the reaction, and then at 60 °C for the final 6 hours of the reaction; (b) HPLC chromatogram of the crude reaction products resulting from the reaction shown in (a) using a linear gradient of 0-35% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). HPLC chromatogram was obtained using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).
Products identified via ESI-MS are indicated.
"JohnPhos" refers to the phosphine ligand, (2-biphenyl)di-*tert*-butylphosphine[348]

The model peptide Ac-TXPN-NH_2 containing 4-pyrrolidyl-phenylalanine was generated in 73% conversion using the optimized palladium-mediated cross-coupling reaction on the resin-bound peptide containing 4-bromophenylalanine (Figure 3.21). Similar to the copper-mediated cross-coupling reaction described in Chapter 1, this

palladium-mediated cross-coupling reaction generates a novel amino acid using no solution-phase chemistry or column purification. From one reaction on the synthesized peptide containing 4-bromophenylalanine, and one HPLC purification, the peptide containing 4-pyrrolidyl-phenylalanine was generated in a rapid, practical manner. With practical access to peptides containing the novel amino acid 4-pyrrolidyl-phenylalanine, the spectral and fluorescent properties were characterized within the model peptide Ac-TXPN-NH_2 (Figures 3.22 and 3.23).

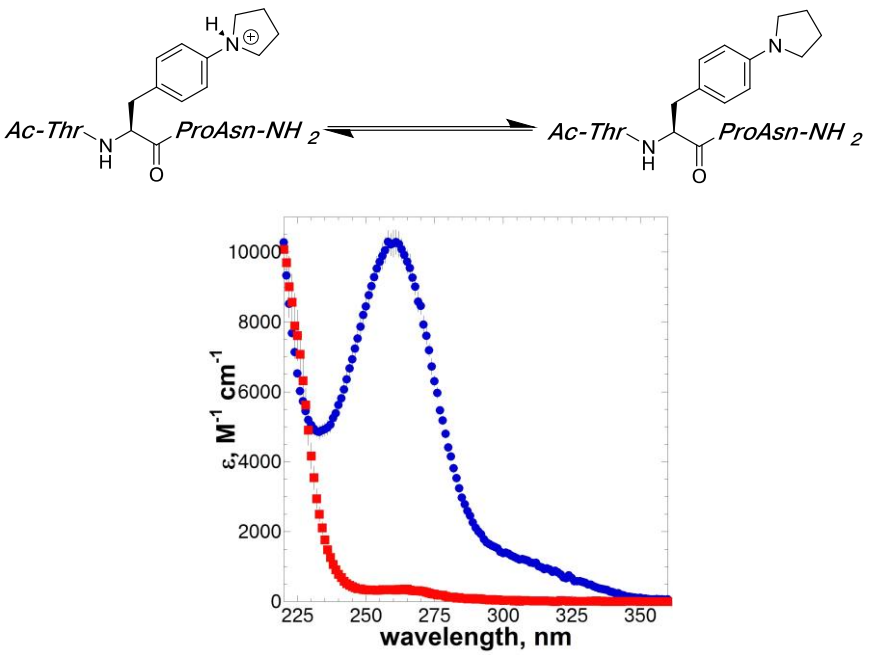


Figure 3.22 UV-Vis spectra of the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2
UV-Vis spectrum of the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2 at pH 7.2 (blue circles) and pH 2.5 (red squares). Error bars indicate standard error of at least three independent trials.

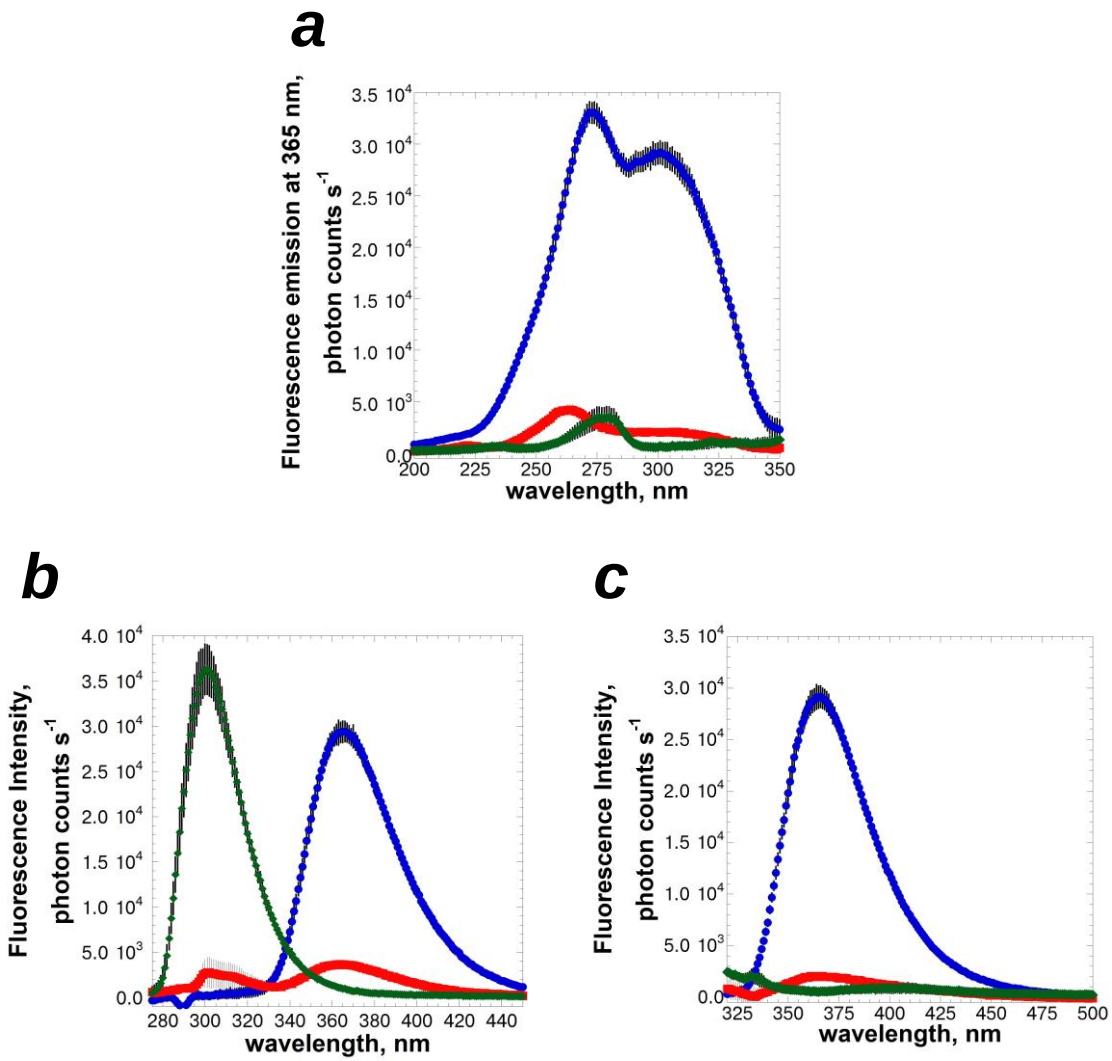


Figure 3.23 Fluorescence excitation and emission spectra of the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2

Fluorescence intensity of the peptides Ac-T(4-Pyrrolidyl-Phe)PN-NH_2 at pH 7.2 (blue circles) and pH 2.5 (red squares) was compared to the peptide Ac-TYPN-NH_2 (2) at pH 7.2 (green diamonds). All fluorescence spectra were obtained with 0.1 mM peptide and 5 mM phosphate buffer at the indicated pH. (a) Excitation spectra with fluorescence detection at 365 nm; (b) fluorescence emission with excitation at 265 nm; (c) fluorescence emission with excitation at 300 nm. Error bars indicate standard error of the mean of at least three independent trials.

4-Pyrrolidyl-phenylalanine exhibited a strong UV absorbance at 260 nm at pH 7.2 which was not observed under acidic conditions (Figure 3.22). The significant shift in UV absorbance indicates that the aromatic electronics are distinctive depending on the protonation state of the nitrogen heterocycle. The fluorescence emission was also distinctive depending on the protonation state of the pyrrolidyl substituent. The fluorescent properties of the peptide containing 4-pyrrolidyl-phenylalanine were compared to the peptide containing tyrosine (Figure 3.23). When using 265 nm as the excitation wavelength (near the UV absorbance λ_{max} for neutral 4-pyrrolidyl-phenylalanine), a strong fluorescence emission was observed at 365 nm for the neutral species of 4-pyrrolidyl-phenylalanine, which was not observed in the protonated species of this amino acid. Excitation at 265 nm also resulted in a strong emission from the peptide containing tyrosine at 300 nm (Figure 3.23b). However, upon excitation at 300 nm, no fluorescence emission was observed from the peptide containing tyrosine, and a strong fluorescence emission was observed for the neutral species of 4-pyrrolidyl-phenylalanine only (Figure 3.23c). With these fluorescent properties, 4-pyrrolidyl-phenylalanine can potentially be used as a fluorescent label that is orthogonal to tyrosine under physiologically relevant conditions. In addition, 4-pyrrolidyl-phenylalanine exhibited a strong, broad excitation bandwidth with emission detection at 365 nm. Excitation with a broader bandwidth can potentially generate a significantly brighter, pH-dependent fluorescence emission from this novel aromatic amino acid.

The UV-Vis and fluorescence properties of 4-pyrrolidyl-phenylalanine indicate significant changes in the aromatic electronics that are depending on protonation state of the pyrrolidyl substituent. In addition to the changes in the optical properties, the

aromatic electronic effects in 4-pyrrolidyl-phenylalanine were characterized through cis-trans isomerism in Ac-TXPN-NH_2 peptides (Figure 3.24).

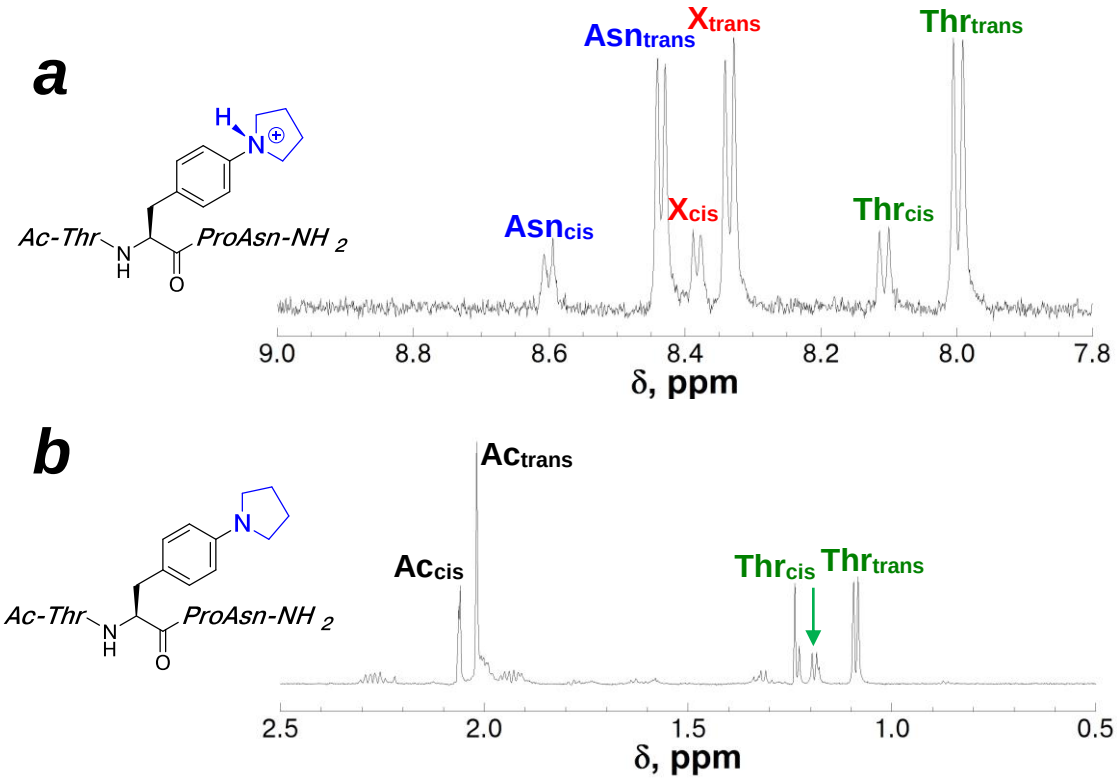


Figure 3.24 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-pyrrolidyl-phenylalanine)
NMR spectra for the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2 at (a) pH 2.5 of the amide region, and (b) pH 7.3 of the aliphatic region. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. The aliphatic region is shown for pH 7.3 (due to amide exchange at high pH, the amide proton signals are not observed).

The pyrrolidyl heterocycle is electron donating to the aromatic ring, with cis-proline populations between 22%-29% depending on the protonation state of the

pyrrolidyl substituent. The $K_{trans/cis}$ values (protonated, 3.5; neutral, 2.5) are comparable to those observed for neutral tyrosine (3.2).

3.2.1.1.5 Solid-phase synthesis of the peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2

In the synthesis of the peptides containing 4-pyrrolidyl-phenylalanine and 4-thiophenylalanine, transition-metal-mediated cross-coupling methodologies were applied to peptides containing a halogenated amino acid. However, it is not always possible to synthesize peptides containing halogenated amino acids for cross-coupling reactions. Aryl triflates are known to have similar reactivity to aryl halides, and they have been informally termed "pseudo-halides."[348] Aryl triflates can be prepared from phenols using triflic anhydride, which can be translated to convert a tyrosine hydroxyl to a triflate (4-OSO_2CF_3-Phe). An aryl triflate would functionalize tyrosine so that it would have similar reactivity to 4-bromophenylalanine or 4-iodophenylalanine,[348] which would allow for cross-coupling methodologies in peptides where halogenated amino acids can not be incorporated. The optimized reaction conditions for synthesizing peptides containing 4-OSO_2CF_3-Phe are shown in Figure 3.25.

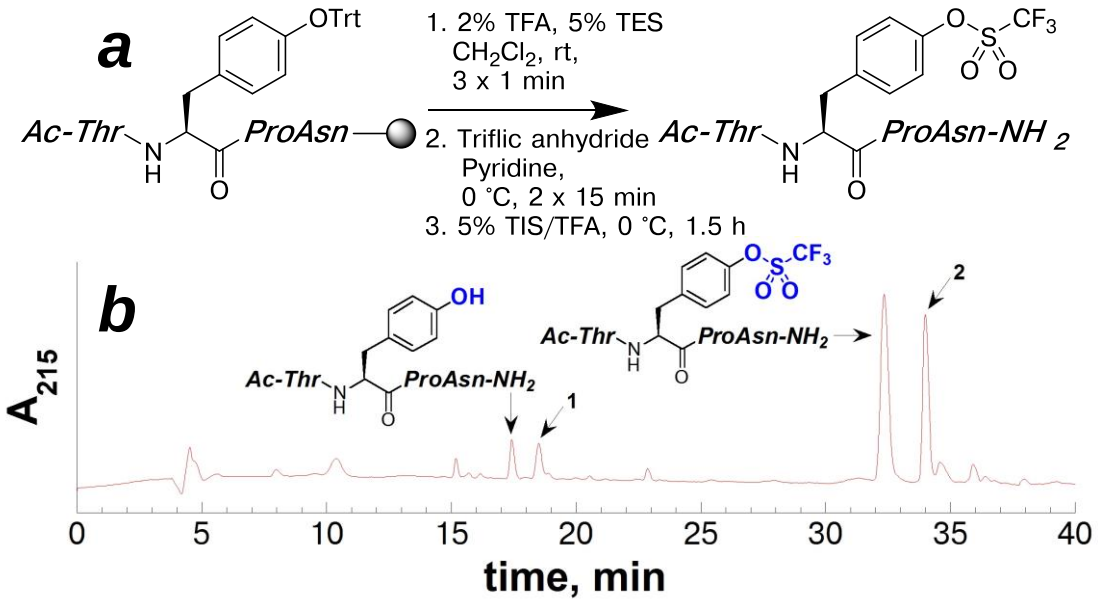


Figure 3.25 Synthesis of the peptide Ac-TXPN-NH_2 (X = 4-OSO_2CF_3-phenylalanine) on solid phase
(a) Scheme for the reaction on the protected peptide Ac-T(4-O(trityl)-tyrosine)PN-NH_2 to generate the peptide containing 4-OSO_2CF_3-phenylalanine; (b) HPLC chromatogram of the crude reaction products resulting from the reaction to generate the peptide containing 4-OSO_2CF_3-phenylalanine using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 30 minutes, followed by a linear gradient to 100% buffer B in buffer A over 5 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.
1 indicates a product identified by ESI-MS that is consistent with the peptide Ac-TYPN-NH_2 with loss of a water molecule (mass – 18 MW).
2 indicates a product identified by ESI-MS that is consistent with the peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 with loss of a water molecule (mass – 18 MW).

Synthesis of the peptide containing 4-OSO_2CF_3-Phe was successfully achieved through reaction of the peptide containing tyrosine with triflic anhydride, followed by mild cleavage and deprotection conditions in the absence of water. Utilizing typical cleavage and deprotection conditions (room temperature for 3 hours in 90% TFA in the presence of scavengers) apparently yielded only starting material, as Ac-TYPN-

NH_2. When the cleavage conditions were modified, by decreasing the reaction temperature and duration and excluding water, the desired product was observed by HPLC (Figure 3.25). However, two additional side-products were observed among the reaction products: identified by ESI-MS, species consistent with elimination on the starting material (Ac-TYPN-NH_2 – 18MW) and on the product (Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 – 18) were observed. The loss of a water molecule from a peptide containing a serine or threonine can generally be attributed to β-elimination. This β-elimination can potentially be due to triflic acid formation as the tyrosine hydroxyl reacts with triflic anhydride: the resulting acidic conditions can potentially hydrolyze the *tert*-butyl ester protection group on the threonine hydroxyl group. Triflic acid is able to catalyze E2 elimination in alcohols, and it possible that E2 elimination on threonine occurs in both the starting material Ac-TYPN-NH_2 and the product Ac-T(4-OSO_2CF_3-Phe))PN-NH_2. Therefore, care should be exercised in application of this reaction chemistry to modifying peptides on solid phase. The $K_{trans/cis}$ for the product peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 was characterized via NMR (Figures 3.26) in order to evaluate the aromatic electronic effect of this substituent.

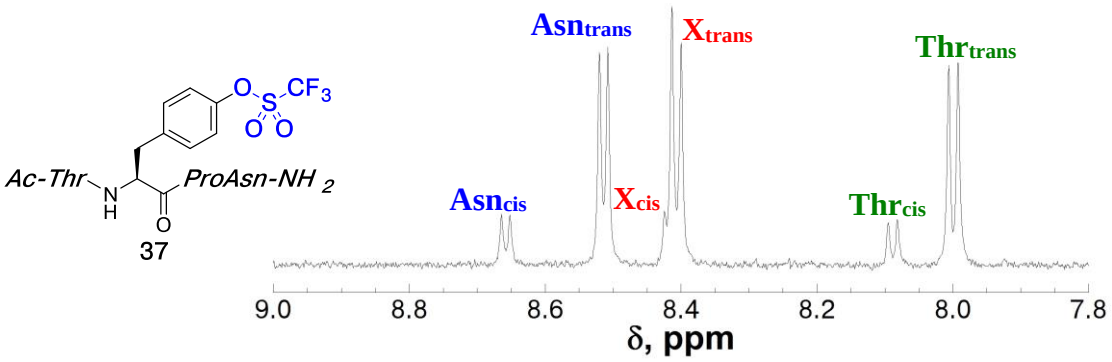


Figure 3.26 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 4-OSO_2CF_3-Phe).
NMR spectrum of amide region for the peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 pH 4. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

The $K_{trans/cis}$ for Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 was measured to be 4.5, which was consistent with the $K_{trans/cis}$ values that were identified for the peptides containing aryl halides, with 4-chlorophenylalanine, 4-bromophenylalanine, and 4-iodophenylalanine (4.2, 4.6, and 4.3, respectively).[91] These data demonstrate that aryl triflates and aryl halides have similar aromatic electronic properties. This synthesis demonstrates that a tyrosine amino acid can be converted on peptides on solid phase into 4-OSO_2CF_3-Phe, which furnishes the aryl ring with a reactive functional group that can potentially act as a substrate for transition-metal-catalyzed cross-coupling reactions.

3.2.1.2 Synthesis of a series of peptides Ac-TXPN-NH_2 (X = 3,4-disubstituted aromatic amino acid)

Di-substituted aromatic amino acids occur naturally in proteins, including 3-nitrotyrosine, 3-iodotyrosine, and 3,4-dihydroxyphenylalanine (DOPA), as a result of post-translational modifications.[353, 354] These post-translational modifications of

tyrosine can impact tyrosine reactivity and interactions, which can have significant impacts in protein structure, and in physiology. For example, tyrosine nitration is a biomarker for cardiovascular diseases and neurodegenerative disorders, and occurs as a result of a reaction between tyrosine and reactive nitrogen species.[141, 353, 355] Tyrosine nitration impacts the acidity, hydrogen bonding ability, spectral properties, metal-coordination ability, and redox properties of tyrosine.[353, 355] 3-Iodotyrosine is a precursor to thyroid hormones,[356] and the introduction of a halogen on an aromatic ring can have significant consequences in potential non-covalent interactions (i.e. halogen bonding).[194, 357] However, the aromatic electronic consequences of these post-translational modifications on tyrosine have not been characterized quantitatively.

In the model peptide Ac-TXPN-NH_2 (X = para-substituted phenylalanine), a direct correlation was observed between the $K_{trans/cis}$ of the peptide and the σ_{para} for the aromatic substituent (Figure 3.6).[91] Due to resonance and inductive effects in aromatic rings, the electron-donating ability of an aromatic substituent depends on the position of the substitution. Meta-substituents and para-substituents in aromatic rings have different contributions due to different resonance effects, and therefore have slightly different σ values denoted as σ_{meta} and σ_{para}, respectively.[358] The σ_{meta} and σ_{para} values for aromatic ring substituents are measured based on the change in the pK_a for substituted benzoic acid derivatives, and have been reviewed extensively.[358] In our expanded library of model peptides Ac-TXPN-NH_2, we sought to characterize the combined aromatic electronic effects of di-substituted aromatic amino acids, which were not previously explored by Thomas *et al.*[91]

In addition to natural modifications of tyrosine to 3,4-disubtituted aromatic amino acid derivatives, novel di-substituted aromatic amino acids were also

synthesized within model peptides Ac-TXPN-NH_2. In particular, syntheses for 3-amino-tyrosine and 3-mercapto-tyrosine were developed via modification of synthesized peptides (described in later sections).

3.2.1.2.1 Synthesis and characterization of peptides Ac-TXPN-NH_2 (X = commercially available 3,4-disubstituted aromatic amino acid)

Tyrosine is subject to numerous post-translational modifications in proteins, and these modifications to the aromatic ring can impact the non-covalent interactions involving tyrosine. In addition to sulfation or phosphorylation on the phenolic -OH, tyrosine can be nitrated to form 3-nitrotyrosine (3-NO_2-Tyr), iodinated to form 3-iodo-tyrosine, or oxidized to form 3,4-dihydroxyphenylalanine (3-OH-Tyr).

3-Nitrotyrosine (3-NO_2-Tyr) and 3-iodotyrosine (3-I-Tyr) are both commercially available as Fmoc-protected amino acids. The model peptides Ac-TXPN-NH_2 containing 3-NO_2-Tyr and 3-I-Tyr were synthesized using standard Fmoc-solid phase peptide synthetic approaches, protecting the tyrosine hydroxyl with an acetyl group via reaction with acetic anhydride during peptide synthesis on solid-phase (see experimental section). Following cleavage and deprotection reactions on the resin containing the peptides with the 3,4-disubstituted aromatic amino acids, the peptides were subjected to hydrolysis reaction conditions with LiOH in solution phase, generating the peptides Ac-T(3-NO_2-Tyr)PN-NH_2 and Ac-T(3-I-Tyr)PN-NH_2. These peptides were characterized via ^{1}H NMR to determine the structural implications of disubstituted aromatic amino acids on cis-trans isomerism of proline (Figures 3.27 and 3.28). Both of the protonation states for the tyrosine hydroxyl were examined.

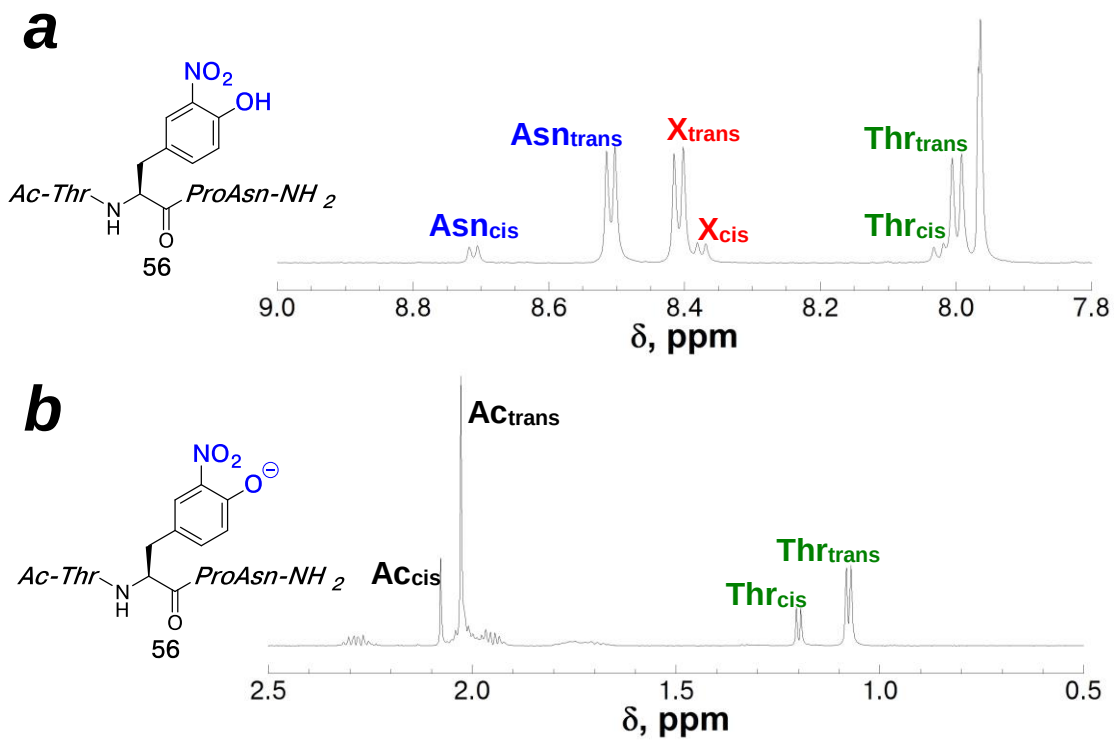


Figure 3.27 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 3-nitrotyrosine).
NMR spectra of the peptide Ac-T(3-NO_2-Tyr)PN-NH_2 at (a) pH 4 for the amide region, and (b) pH 11 for the aliphatic region. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. The aliphatic region is shown for pH 11 (due to rapid amide exchange at high pH, the amide proton signals are not observed).

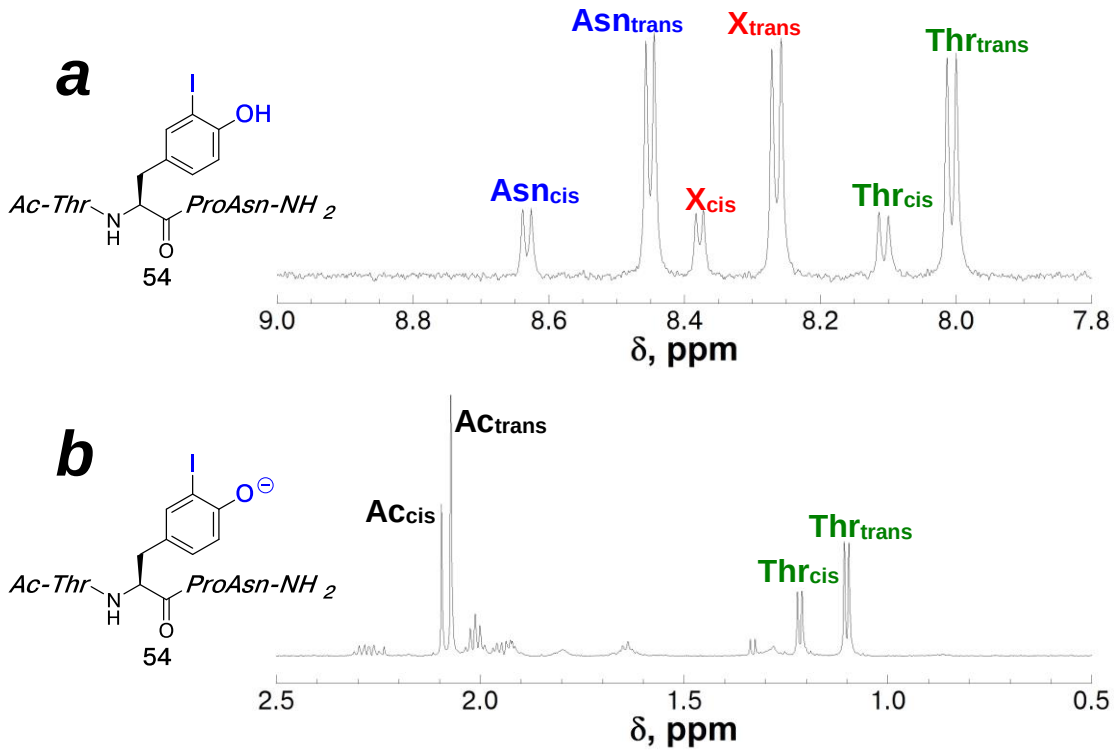


Figure 3.28 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 3-iodotyrosine).
NMR spectra of the peptide Ac-T(3-I-Tyr)PN-NH_2 at (a) pH 4 for the amide region, and (b) pH 11 for the aliphatic region. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. The aliphatic region is shown for pH 11 (due to rapid amide exchange at high pH, the amide proton signals are not observed).

In both of the peptides, Ac-T(3-NO_2-Tyr)PN-NH_2 and Ac-T(3-I-Tyr)PN-NH_2, the cis population of the peptides increased with pH, and upon deprotonation of the para-substituent (tyrosine –OH). For the peptide Ac-T(3-NO_2-Tyr)PN-NH_2, the $K_{trans/cis}$ decreased from 6.5 to 3.1 on increasing the pH from 4 to pH 11, respectively. For the peptide Ac-T(3-I-Tyr)PN-NH_2, the $K_{trans/cis}$ decreased from 3.7 to 1.9 on increasing the from pH 4 to pH 11, respectively. These trends are consistent with the

model proposed by Thomas *et al.*, and the prior $K_{trans/cis}$ data for tyrosine and tyrosinate,[91] where increased electron-donating ability of an aromatic substituent increases the cis-propensity of the peptide. The peptide containing 3-iodo-tyrosine exhibited increased cis-proline population in comparison to the peptide containing 3-nitrotyrosine, consistent with the enhanced ability of halogens to activate an aromatic ring.

3.2.1.2.2 Solution phase synthesis of the peptide Ac-T(3-OH-Tyr)PN-NH_2

In addition to its uses in treatment of Parkinson's disease, 3,4-dihydroxyphenylalanine (DOPA) has been found to be an important component in the proteins that mollusks produce for surface adhesion.[354, 359, 360] Resulting from an oxidative post-translational modification of tyrosine, DOPA can undergo cross-linking, and peptides containing DOPA have been explored in potential applications in materials science.[354, 359, 360] However, the aromatic electronic effects of DOPA have not been fully characterized. Utilizing a previously established protocol,[361] the purified peptide Ac-TYPN-NH_2 was subjected to reaction with 2-iodoxybenzoic acid (IBX) to generate the peptide containing DOPA (Figure 3.29). With practical access to the model peptide containing DOPA, and the electronic and structural consequences of this oxidative modification were characterized via ^{1}H NMR (Figure 3.30).

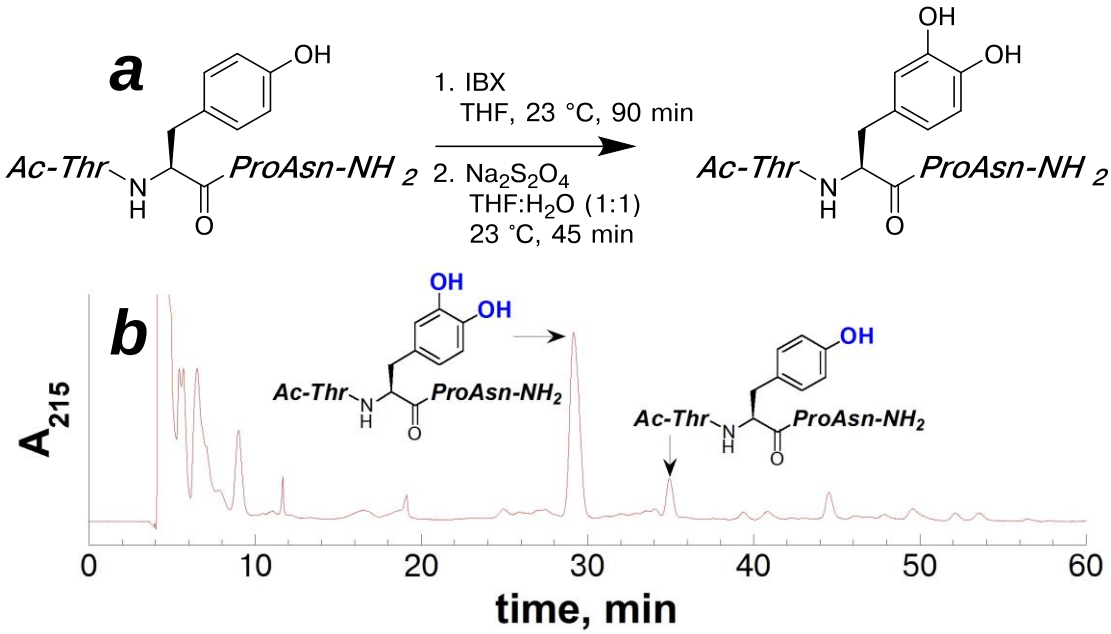


Figure 3.29 Synthesis of the peptide Ac-TXPN-NH_2 (X = 3,4-dihydroxyphenylalanine) in solution phase

(a) Scheme for the reaction on Ac-TYPN-NH_2 with 2-iodoxybenzoic acid in solution to generate the peptide containing 3,4-dihydroxyphenylalanine; (b) HPLC chromatogram of the resultant crude reaction products that generated the peptide Ac-T(3-OH-Tyr)PN-NH_2 using a linear gradient of 0-15% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

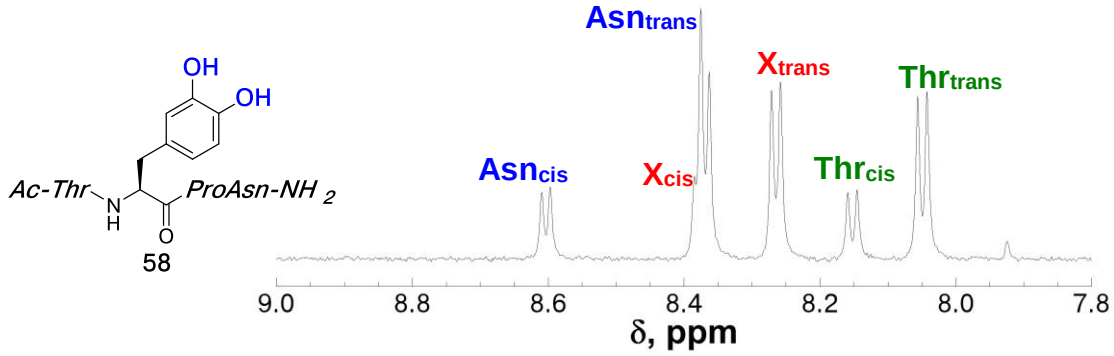


Figure 3.30 NMR Characterization of the peptide Ac-TXPN-NH_2 (X = 3,4-dihydroxyphenylalanine)
NMR spectrum of amide region for the peptide Ac-T(3-OH-Tyr)PN-NH_2 pH 4. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

The model peptide containing DOPA was synthesized in one step from the synthesized peptide Ac-TYPN-NH_2, with only one HPLC purification required. The $K_{trans/cis}$ of the peptide Ac-T(3-OH-Tyr)PN-NH_2 was measured to be 2.5, with a higher cis population compared to tyrosine or other aromatic amino acids. Characterization of the peptide structure was attempted at higher pH, but the peptide containing DOPA degraded into multiple species, potentially due to oxidation of the peptide, metal chelation, or cross-linking and aggregation. The increased population of cis-proline in the peptide containing DOPA is suggestive of an activated aromatic ring.

3.2.1.2.3 Solid-phase synthesis of the peptide Ac-T(3-NH_2-Tyr)PN-NH_2

The peptide Ac-TXPN-NH_2 containing 4-amino-phenylalanine was characterized via NMR previously, and this peptide exhibited significant structural and electronic changes that correlated with the protonation state of the aromatic amino substituent.[91, 172] We sought to investigate the aromatic electronic effects of an amino substituent in concert with an aromatic hydroxyl substituent in the context of the

peptide Ac-TXPN-NH_2. The amino acid 3-amino-tyrosine could potentially function as a structural "switch" with proline, with three distinct protonation states on this aromatic amino acid (Figure 3.33a). Hooker *et al*.[362] demonstrated that 3-amino-tyrosine can be site-selectively generated from a surface-exposed tyrosine residue within viral capsid proteins using nitrobenzenediazonium tetrafluoroborate, following reduction with sodium dithionite. In this work, the surface-exposed 3-amino-tyrosine residues were, upon oxidation, further able to react with dienophiles, yielding Diels-Alder reaction products under mild oxidative conditions. This Diels-Alder reaction on oxidized 3-amino-tyrosine effectively functionalized the interior surface of a viral capsid protein, with potential applications in drug delivery.[362] Wisastra *et al.* demonstrated that 3-nitrotyrosine could be reduced to 3-amino-tyrosine in proteins, and reacted with $AlCl_3$ and salicylaldehyde to generate a fluorescent probe of tyrosine nitration.[363] Potentially, the reactivity of 3-amino-tyrosine is permitted by the unique aromatic electron distribution, which allows for its reactivity with electron-rich groups. In order to further study the aromatic electronic effects of aryl amines, 3-amino-tyrosine was synthesized within the model peptide Ac-TXPN-NH_2 using two different synthetic approaches from purified peptides in solution: one from the peptide Ac-TYPN-NH_2 using nitrobenzediazonium tetrafluoroborate and subsequent reduction reaction (Figure 3.31), and one from the peptide Ac-T(3-NO_2-Tyr)PN-NH_2 using sodium dithionite (Figure 3.32).

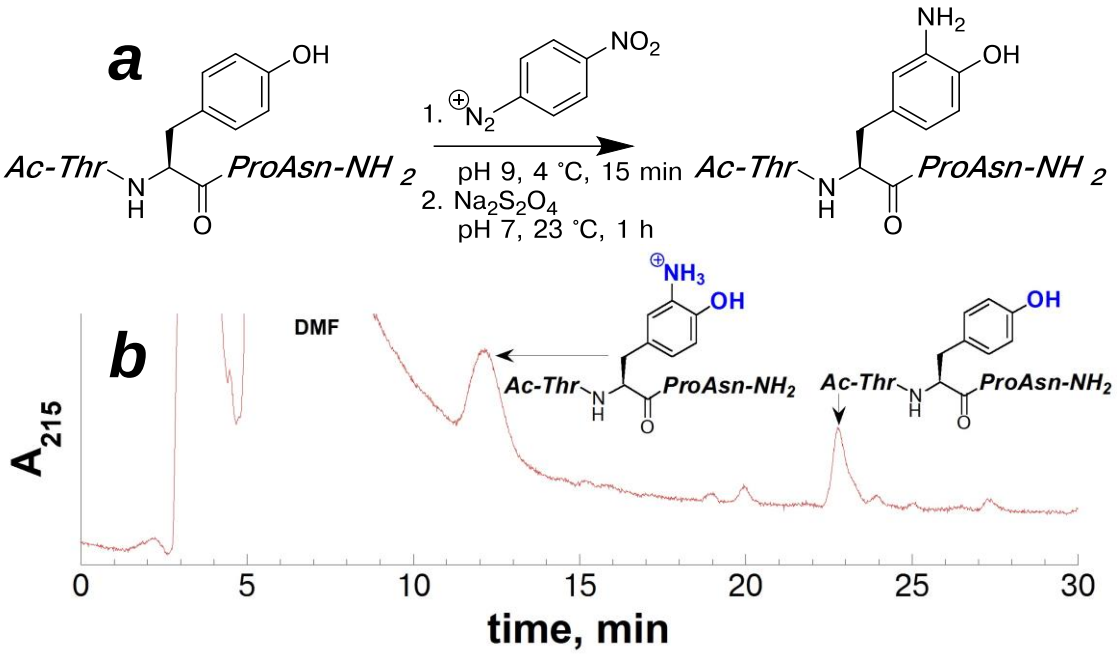


Figure 3.31 Synthesis of the peptide Ac-TXPN-NH_2 (X = 3-aminotyrosine) in solution phase from a peptide containing tyrosine
(a) Scheme for the reaction on Ac-TYPN-NH_2 with 4-nitrobenzenediazonium tetrafluoroborate to generate the peptide containing 3-aminotyrosine; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with Ac-TYPN-NH_2 and 4-nitrobenzenediazonium tetrafluoroborate using a linear gradient of 0-15% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 30 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

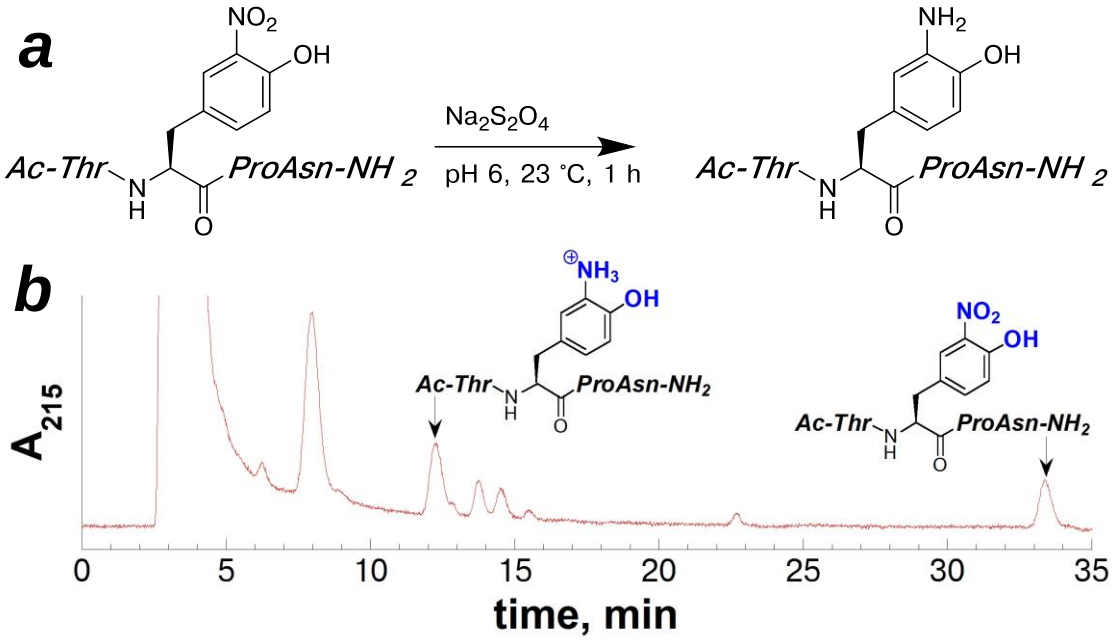


Figure 3.32 Synthesis of the peptide Ac-TXPN-NH_2 (X = 3-aminotyrosine) in solution phase from the peptide containing 3-nitrotyrosine
(a) Scheme for the reaction on the peptide Ac-T(3-NO_2-Tyr)PN-NH_2 to generate the peptide containing 3-aminotyrosine; (b) HPLC chromatogram of the crude reaction products resulting from the reaction with the peptide Ac-T(3-NO_2-Tyr)PN-NH_2 and sodium dithionite using a linear gradient of 0-15% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 30 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

Peptides containing 3-amino-tyrosine was successfully generated from purified peptides Ac-TXPN-NH_2 in solution from two different synthetic strategies, requiring only one reaction and one HPLC purification (Figures 3.31-3.32). Utilizing a protocol similar to that employed by Hooker *et al.*,[362] the peptide containing tyrosine was subjected to reaction with 4-nitrobenzenediazonium tetrafluoroborate to generate the diazo product. The peptide was subjected to sodium dithionite in solution in order to generate the peptide containing 3-amino-tyrosine.[362] Alternatively, the peptide containing 3-amino-tyrosine was also generated in solution from the purified model

peptide containing 3-nitrotyrosine, by subjecting the peptide to sodium dithionite (Figure 3.32).[363] Using these two approaches, 3-amino-tyrosine was successfully generated within synthesized peptides only one or two synthesis and purification steps. With the modified peptide Ac-T(3-NH_2-Tyr)PN-NH_2 in hand, the aromatic electronic effects of 3-amino-tyrosine were probed by examining cis-trans isomerism of proline (Figure 3.33).

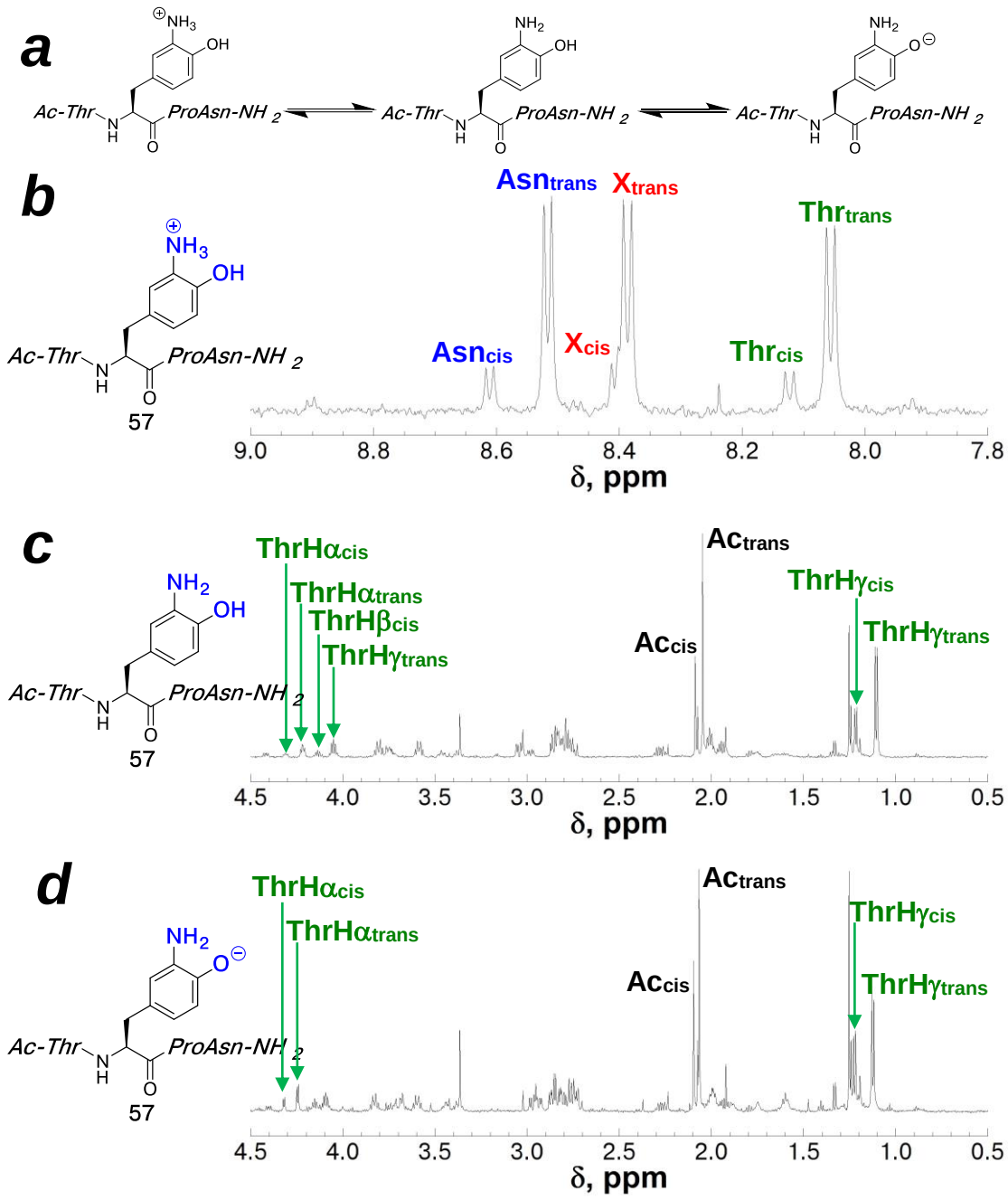


Figure 3.33 NMR Characterization of the peptide Ac-TXPN-NH$_2$ (X = 3-aminotyrosine)

NMR spectra of the amide region for (a) the peptide Ac-T(3-NH_2-Tyr)PN-NH_2 at (b) pH 2.5, (c) pH 7.2, and (d) 11.8. The samples contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. The aliphatic region is shown for pH 7.2 and 11.8, because rapid amide exchange at high pH obscures the amide peaks. Due to spectral overlap, the Thr Hα and Thr Hβ peaks were also used to determine the $K_{trans/cis}$.

The $K_{trans/cis}$ of the peptide containing 3-amino-tyrosine decreased from 3.3 to 1.6 with increasing pH (pH 2.5-11.8), indicating an increasing structural preference for the cis amide bond conformation. These data suggest that as the electron-richness of the aromatic ring increases (where both of the aromatic substituents become more electron-donating with increased pH), the intermolecular interaction with proline increases in stability. This stabilizing effect on the cis amide bond conformation is potentially due to a strengthened C–H/π aromatic interaction. However, over multiple trials with 3-amino-tyrosine, significant degradation of the peptide with increasing pH was observed, and the values of $K_{trans/cis}$ varied slightly. Multiple species are observed in the aliphatic region at pH 7.2 and 11.8, particularly near the acetyl and threonine Hγ peaks, which were used for determining the $K_{trans/cis}$ when amide signals were too broad for reliable integration (due to rapid amide proton exchange under neutral and basic conditions). The $K_{trans/cis}$ was estimated at elevated pH based on the threonine Hα and Hβ protons, identified by the TOCSY spectra, which were reasonably well resolved. However, the impurities from the degraded peptide may have affected the structural conformation in an unknown manner. This observed degradation may have been a result of oxidation of the aromatic ring, similar to that observed with DOPA and 3-mercaptotyrosine.

3.2.1.2.4 Solid-phase synthesis of the peptide Ac-T(3-SH-Tyr)PN-NH_2

With the utility and versatility of 4-thiophenylalanine described in Chapter 1, and with its unique structural effects discussed in Chapter 2, we envisioned that 3-mercaptotyrosine could also provide special functional properties that would merit further investigation. DOPA is known to chelate metals, and 3-mercaptotyrosine is expected to act in a similar manner. Sulfur-containing amino acids bind to soft metals,[1, 2] which are crucial for electron transfer and energy conversion mechanisms in metalloenzymes. 3-Mercaptotyrosine could potentially be used in designed peptides for modulating the coordination sphere around metals, such as tungsten or molybdenum, for performing difficult redox reactions, such as reduction of CO_2.[364-366]

The novel amino acid 3-mercaptotyrosine was synthesized within the model peptide Ac-TXPN-NH_2 using a similar solid-phase approach as described in Chapter 1. 3-Iodotyrosine was incorporated into the model peptide Ac-TXPN-NH_2 using standard Fmoc- solid-phase peptide synthesis, and the tyrosine hydroxyl was protected during synthesis using acetic anhydride (see experimental methods). The protected peptide on solid phase was subjected to the copper-mediated cross-coupling conditions that were optimized previously.[169] The resin containing the peptide was subjected to cleavage and deprotection reactions using TFA. The peptide was subjected to DTT and LiOH in solution following the TFA cleavage and deprotection reactions, in order to reduce disulfides and hydrolyze the acetyl protecting groups (Figure 3.34). The UV-Vis spectral properties of this novel aromatic amino acid were characterized, and the pK_a of the thiol was measured via UV (Figure 3.35).

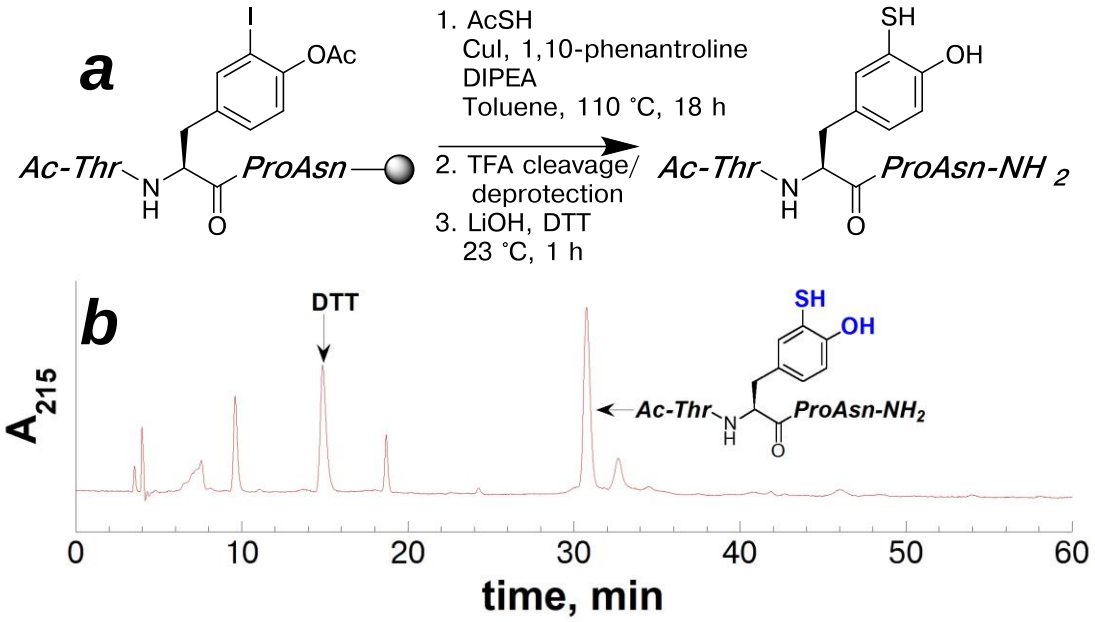
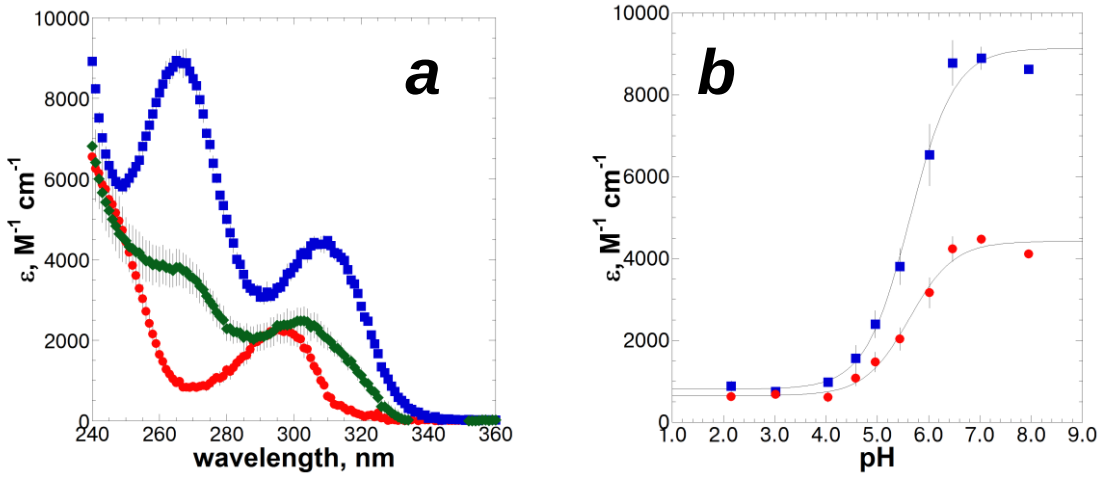


Figure 3.34 Synthesis of the peptide Ac-TXPN-NH_2 (X = 3-mercaptotyrosine) on solid phase

(a) Scheme for the solid-phase copper-mediated cross-coupling reaction on the peptide Ac-T(4-O(acetyl)-3-iodo-phenylalanine)PN-NH_2 to generate the peptide containing 3-mercaptotyrosine; (b) HPLC chromatogram of the crude reaction products resulting from the solid-phase cross-coupling reaction and subsequent reduction and hydrolysis reactions in solution to generate the peptide Ac-T(3-SH-Tyr)PN-NH_2 using a linear gradient of 0-35% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated. "DTT" indicates oxidized dithiothreitol.



$$\varepsilon_{266} = \frac{\varepsilon_{266,\text{thiolate}} 10^{(\text{pH} - \text{p}Ka)} + \varepsilon_{266,\text{thiol}}}{1 + 10^{(\text{pH} - \text{p}Ka)}}$$ Equation (2) $\text{p}K_a = 5.62 \pm 0.09$

$$\varepsilon_{310} = \frac{\varepsilon_{310,\text{thiolate}} 10^{(\text{pH} - \text{p}Ka)} + \varepsilon_{310,\text{thiol}}}{1 + 10^{(\text{pH} - \text{p}Ka)}}$$ Equation (3) $\text{p}K_a = 5.63 \pm 0.07$

Figure 3.35 UV-Vis spectroscopy of the peptide Ac-T(3-SH-Tyr)PN-NH$_2$
(a) Absorbance spectra of the peptide Ac-T(3-SH-Tyr)PN-NH$_2$ at pH 4.0 (thiol, red circles), pH 5.5 (green diamonds), and pH 7.0 (thiolate, blue squares). Data represent an average of at least 3 independent trials; (b) pH dependence of the molar extinction coefficient for the peptide Ac-T(3-SH-Tyr)PN-NH$_2$ at 266 nm (blue squares) and 310 nm (red circles). Peptide samples contained 100 mM phosphate at the indicated pH, and contained 1 mM TCEP to reduce disulfides. Error bars indicate standard error. The pK_a was determined using Equations (2) and (3).

The model peptide containing 3-mercaptotyrosine was synthesized in 77% conversion from 3-iodotyrosine on the protected peptide on solid-phase. Distinctive changes were observed in the UV spectrum of 3-mercaptotyrosoine with respect to the pH, which allowed for determination of the acid dissociation constant for this novel amino acid. The pK_a was calculated to be 5.6, based on the UV absorbance at two different wavelengths with respect to pH (Figure 3.35b, and Equations 2 and 3). The

pK_a of 3-mercaptotyrosine was less than that of 4-thiophenylalanine (pK_a = 6.4),[169] potentially due to intramolecular hydrogen bonding with the tyrosine hydroxyl that can stabilize the thiolate.

With substantial changes observed in the UV absorbance spectrum with changes in protonation state, the peptide containing 3-mercaptotyrosine was characterized via NMR (Figure 3.36). Upon increasing the pH to pH >9.0, significant degradation of the peptide was observed as multiple products formed, similar to observations with DOPA and 3-amino-tyrosine.

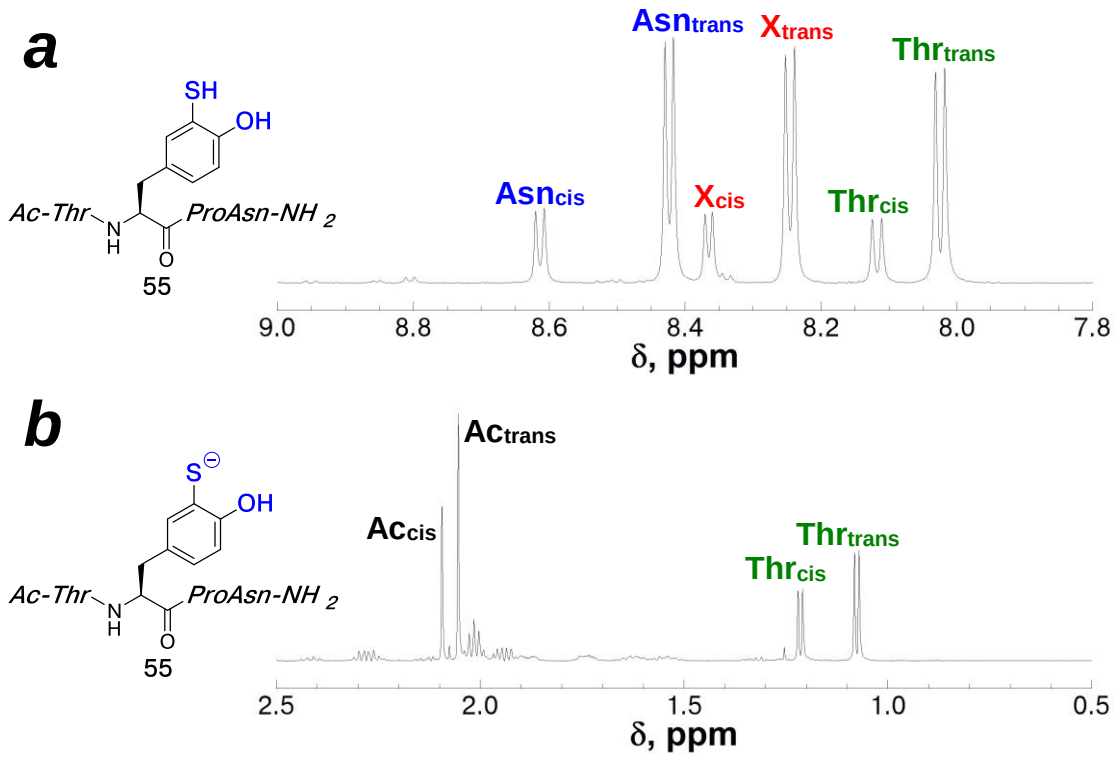


Figure 3.36 NMR Characterization of the peptide Ac-TXPN-NH$_2$ (X = 3-mercaptotyrosine)
NMR spectra of amide region for the peptide Ac-T(3-SH-Tyr)PN-NH$_2$ at (a) pH 2.5 and (b) pH 7.2. The sample contained 5 mM phosphate and 25 mM NaCl, and 100 μM TCEP, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum.

Similar to the observations in the peptide Ac-T(3-NH$_2$-Tyr)PN-NH$_2$, 3-mercaptotyrosine exhibited differing cis-populations depending on the protonation state of the thiol group. With increasing the pH, the $K_{trans/cis}$ of the peptide Ac-T(3-SH-Tyr)PN-NH$_2$ decreased from 3.3 to 1.6, with higher cis population as more of the thiolate because deprotonated. Potentially, metal-binding on 3-mercaptotyrosine can affect the UV spectra and noncovalent interactions, due to a shift in the electron distribution in the aromatic ring upon interaction binding to a soft metal.

3.2.2 Hammett correlations of the model peptides Ac-TXPN-NH_2

For all of the aromatic amino acids examined in the model peptide Ac-TXPN-NH_2, including those previously published,[91] those discussed in Chapter 1,[169] and those described in the previous sections, the relative populations of cis and trans amide bond conformations ($K_{trans/cis}$) were determined via NMR. Thus, the aromatic electronic properties of these amino acids were probed and quantified using cis-trans isomerism of proline. It has been shown that the cis conformation of aromatic-proline motifs can be stabilized by an intermolecular proline C–H/π aromatic interaction, where electron-rich aromatic rings stabilize the cis conformation through an interaction with the proline Hα.[91] Based on this model, the equilibrium between cis and trans conformations ($K_{trans/cis}$) in aromatic-proline motifs is indicative of the nature of aromatic electronics of the amino acid.

In addition to the peptides described in the previous sections, model peptides that were synthesized and characterized by Krista Thomas and Dr. Devan Naduthambi were included in these studies of comparisons and analyses of aromatic-cis-proline interactions, although the original data has been reanalyzed.[91] A model peptide synthesized and characterized by Michael Scheuermann, Ac-T(4-OSO_3^--Phe)PN-NH_2, was also included within these comparisons of aromatic-cis-proline interactions. All of the $K_{trans/cis}$ values measured for each peptide containing a 4-substituted aromatic amino acid are compiled in Table 3.1. The correlation of $K_{trans/cis}$ and Hammett parameters for each 4-substituted aromatic amino acid within the model peptide Ac-TXPN-NH_2 are shown in Figure 3.37.

Table 3.1. $K_{trans/cis}$ values for the peptides Ac-TXPN-NH_2, where X = 4-substituted amino acid

The equilibrium between cis and trans conformations ($K_{trans/cis}$) in the peptide Ac-TXPN-NH_2 was determined by measuring the ratios of 2 or 3 pairs of proton signals in the ^{1}H NMR spectra ($\Sigma\int$(trans signals)/$\Sigma\int$(cis signals)). When possible, amide proton signals were used, but other peaks, including the acetyl methyl, threonine methyl, and threonine Hα, were also used for calculation in the event that amide signals could not be used due to spectral overlap or rapid amide exchange. The σ_{para} and σ_{meta}, when available, are included in the table.[358]

[a]Reported previously.[139]

n.d. refers to "not determined" due to spectral overlap or broadened amide signals.

Ac-TXPN-NH_2, X =	σ_{para}	σ_{meta}	$K_{trans/cis}$	$\Delta G_{trans/cis}$, kcal mol^{-1}	$^3J_{\alpha N}$ X-cis, Hz	%cis	pH
His(H^+)			12.3	-1.49	8.2	8%	4
4-Pyridyl(N-oxide)-Ala			11.2	-1.43	8.1	8%	4
4-Pyridyl(H^+)-Ala			11.2	-1.43	n.d.	8%	2.5
F_5-Phe			6.7	-1.13	8.4	13%	4
4-Pyridyl-Ala			6.1	-1.07	n.d.	14%	7.2
4-SO_2Me-Phe	0.72	0.60	5.8	-1.04	7.2	15%	4
His			5.7	-1.03	n.d.	15%	8.5
4-NO_2-Phe	0.78	0.71	5.6	-1.02	7.4	15%	4
4-NH_3^+-Phe	0.60	0.86	5.5	-1.01	n.d.	15%	2.5
4-S(*N*-ethyl succinimide)-Phe			5.4	-1.00	n.d.	16%	4.0
4-CN-Phe	0.66	0.56	5.3	-0.99	7.3	16%	4
4-CF_3-Phe	0.54	0.43	4.8	-0.93	n.d.	17%	4
4-S(O)Me-Phe	0.49	0.52	4.7	-0.92	n.d.	18%	4
4-Br-Phe	0.23	0.39	4.6	-0.90	n.d.	18%	4
4-SO_2-propargyl-Phe[a]			4.5	-0.89	6.7	18%	4
4-OSO_2CF_3-Phe	0.53	0.56	4.5	-0.89	n.d.	18%	4
4-SO_3^--Phe	0.35	0.30	4.4	-0.88	n.d.	19%	4
4-SBz-Phe			4.3	-0.86	6.7	19%	4
4-S(O)propargyl-Phe[a]			4.3	-0.86	n.d.	19%	4
4-I-Phe	0.18	0.35	4.3	-0.86	n.d.	19%	4
4-Cl-Phe	0.23	0.37	4.2	-0.85	n.d.	19%	4
4-S(2-nitrobenzyl)-Phe			4.2	-0.85	6.9	19%	4
4-SAc-Phe	0.44	0.39	4.2	-0.85	6.9	19%	4
4-F-Phe	0.06	0.34	4.2	-0.85	7.0	19%	4
4-OAc-Phe	0.44	0.39	3.9	-0.81	6.8	20%	4
4-S-S-Glutathione-Phe			3.9	-0.81	n.d.	20%	4

Table 3.1 continued

4-SH-Phe	0.15	0.25	3.8	-0.79	6.6	21%	4
4-SO_2^--Phe	-0.05	-0.02	3.8	-0.79	n.d.	21%	4
4-SPh-Phe	0.07	0.23	3.7	-0.77	6.4	21%	4
4-SMe-Phe	0.00	0.15	3.6	-0.76	6.3	22%	4
Phe	0.00	0.00	3.6	-0.76	6.3	22%	4
4-Pyrrolidyl(H^+)-Phe			3.5	-0.74	6.6	22%	2.5
4-$SCH_2CH{=}CHCH_2OH$-Phe			3.5	-0.74	n.d.	22%	4
4-$B(OH)_2$-Phe	0.12	-0.01	3.5	-0.74	6.6	22%	4
4-S-allyl-Phe	0.12	0.19	3.5	-0.74	6.5	22%	4
4-S-propargyl-Phe			3.2	-0.69	6.4	24%	4
Tyr	-0.37	0.12	3.2	-0.69	6.2	24%	4
4-OPO_3H^--Phe	0.00	0.29	3.1	-0.67	6.7	24%	4
4-OPO_3^{3-}-Phe			3.0	-0.65	n.d.	25%	8.0
4-OSO_3^--Phe			2.9	-0.63	6.8	26%	4
4-NH_2-Phe	-0.66	-0.16	2.9	-0.63	n.d.	26%	6.5
4-S-S(2-thiopyridyl)-Phe			2.7	-0.59	n.d.	27%	4
4-Pyrrolidyl-Phe			2.5	-0.54	n.d.	29%	7.3
4-$B(OH)_3^-$-Phe	-0.44	-0.48	2.4	-0.52	n.d.	29%	10
4-S^--Phe	-1.21	-0.36	2.3	-0.49	n.d.	30%	8.5
Trp			2.0	-0.41	5.6	33%	4
4-O^--Phe	-0.81	-0.47	1.9	-0.38	n.d.	34%	12
Ala			11.5	-1.45	5.4	8%	4
Cha			8.6	-1.27	6.9	10%	4

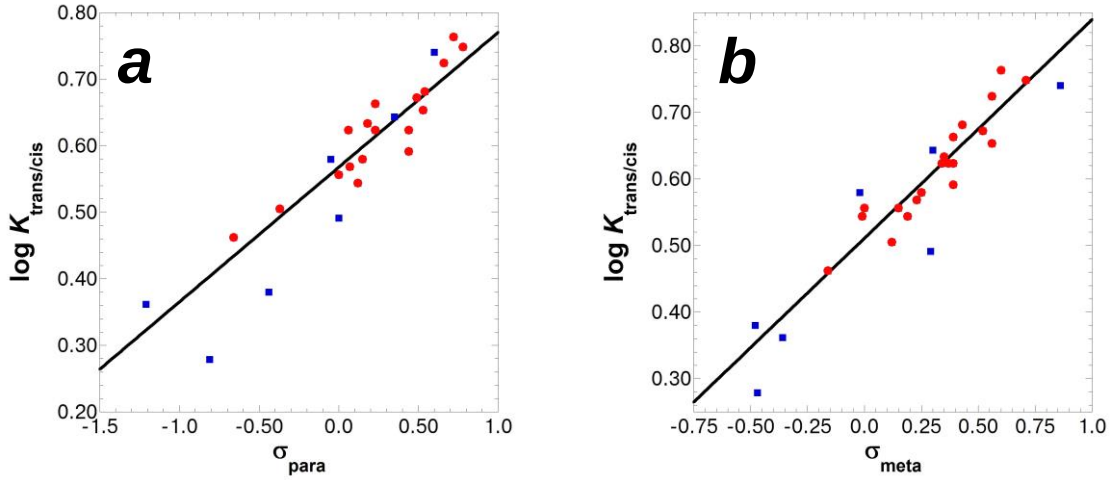


Figure 3.37 Hammett correlation with $K_{trans/cis}$ in model peptides Ac-TXPN-NH_2

The $K_{trans/cis}$ values for neutral (red circles) and charged (blue squares) 4-substituted aromatic amino acids are shown. The log($K_{trans/cis}$) of each peptide is correlated with the Hammett constant (a, σ_{para}; b, σ_{meta}) for each substituent. Linear regressions are fitted to neutral substituents only.

σ_{para} correlation (a): $\rho = 0.202 \pm 0.022$, $R = 0.908$

σ_{meta} correlation (b): $\rho = 0.328 \pm 0.031$, $R = 0.928$

It was previously observed by Thomas *et al.*[91] that the cis population was increased in the model peptide Ac-TXPN-NH_2 containing cyclohexylalanine (Cha) over the peptide containing canonical alanine (Ala $K_{trans/cis}$ = 11.5 versus Cha $K_{trans/cis}$ = 8.5). The stabilization of the cis amide bond conformation in the peptide Ac-TXPN-NH_2 with Cha is potentially due to the increased hydrophobic surface area of cyclohexylalanine over alanine. However, the peptide Ac-TXPN-NH_2 containing phenylalanine (Phe $K_{trans/cis}$ = 3.6) had a lower $K_{trans/cis}$ over cyclohexylalanine, even though the hydrophobic surface area is similar. The lower $K_{trans/cis}$ with phenylalanine indicates that the aromaticity of the 6-membered ring is more stabilizing to the cis conformation.[91] The comparison between phenylalanine and cyclohexylalanine in aromatic-proline motifs demonstrates that the increased cis propensity of these

aromatic-proline motifs is not due to a simple hydrophobic effect or van der Waals interaction.

A direct correlation was observed between the Hammett value of each aromatic substituent (σ) and the $K_{trans/cis}$ of each model peptide Ac-TXPN-NH_2 (Figure 3.37), similar to that described by Thomas *et al.*[91] The Hammett value for a given substituent is a measure of the "electron-donating ability" or activation of given substituent on an aromatic ring. Due to resonance effects within the aromatic ring, a given substituent activates the aromatic ring depending on the location of the substituent. To account for these differences, due to differences in resonance and inductive effects, Hammett values are given as σ_{para} (for para substituents) or σ_{meta} (for meta substituents).[358] For example, the acidity of benzoic acid is affected depending on the location of a fluorine substituent relative to the carboxylic acid (3-fluorobenzoic acid, pK_a = 3.87, 4-fluorobenzoic acid pK_a = 4.14).[367] In a typical Hammett analyses, an equilibrium constant for a reaction is correlated to the σ_{para} or σ_{meta} for different aromatic substituents, in order to determine if the reaction is influenced by substituent effects. If a correlation is observed, the linear correlation is fit to equation 4, where ρ is a measure for how sensitive the reaction is to aromatic substituent effects.[368] For the 4-substituted phenylalanine derivatives, both σ_{para} and σ_{meta} Hammett values exhibited excellent correlation ($R > 0.90$) for the $K_{trans/cis}$ of the model Ac-TXPN-NH_2 peptides across a broad range of aromatic substituents. Outliers in the Hammett correlation tended to be charged aromatic substituents, such as protonated 4-aminophenylalanine or anionic 4-thiophenylalanine, indicating that electrostatic contributions can have additional effects on peptide structure outside of aromatic substituent effects. These correlations provide further evidence that a broad

range of aromatic amino acid substituents can exert concerted structural influence on cis-trans isomerism in the context of aromatic-proline motifs.

$$\log \frac{K}{K_0} = \sigma\rho \qquad (4)$$

In comparing the Hammett correlations between the $K_{trans/cis}$ and the σ_{para} or σ_{meta} in the model peptides Ac-TXPN-NH_2, we observed better fit and a greater ρ value when the $K_{trans/cis}$ data were fit to σ_{meta} values (Figure 3.37b). Both the σ_{para} and σ_{meta} Hammett values represent inductive effects for a given substituent on an aromatic ring. However, σ_{para} Hammett values have greater resonance effects than σ_{meta} Hammett values.[369] The greater correlation and greater sensitivity of the $K_{trans/cis}$ for the model peptide Ac-TXPN-NH_2 on the σ_{meta} value for a given aromatic substituent suggests that the inductive effects are more important than the resonance effects.

The spin-spin coupling constant between the amide proton and the Hα proton can be determined via NMR, as the $^3J_{\alpha N}$. The $^3J_{\alpha N}$ is directly related to the ϕ torsion angle for amino acids in peptides and proteins, generally following the Karplus relationship in Equation 5:

$$^3J_{\alpha N} = 6.51 \cos^2\theta - 1.76 \cos\theta + 1.60 \qquad (5)$$

where $^3J_{\alpha N}$ is the measured coupling constant in hertz derived from the proton NMR spectrum, and $\theta = |\phi - 60°|$.[91, 370] The $^3J_{\alpha N}$ values for the aromatic amino acid in the peptides Ac-TXPN-NH_2 are shown in Table 3.1. With increasing cis-population of the

peptides and with more electron-donating aromatic substituents, the $^3J_{\alpha N}$ coupling constant decreases. The largest $^3J_{\alpha N}$ in this series of peptides Ac-TXPN-NH_2 was measured for protonated histidine, at 8.2 Hz, which corresponds to a ϕ torsion angle of –92°. The smallest $^3J_{\alpha N}$ was measured for the peptide Ac-TWPN-NH_2 at 5.6 Hz, corresponding to a Trp ϕ torsion angle of –71°. The idealized ϕ torsion angle for residues preceding proline in type VIa1 β-turns is –60°.[289, 306] With increasingly electron-donating aromatic substituents, the smaller $^3J_{\alpha N}$ coupling constants for the residue preceding proline indicate that the peptide Ac-TXPN-NH_2 approaches an idealized type VIa1 β-turn with more electron-rich aromatic rings.

With an established model that correlates the σ_{meta} and σ_{para} for given substituents to the measured $K_{trans/cis}$ within model peptides Ac-TXPN-NH_2, the approximate Hammett values can be determined from the $K_{trans/cis}$ for substituents which have no established value. The aromatic effects from heteroaromatic and polycyclic rings can also be approximated from the measured $K_{trans/cis}$. Table 3.2 provides the calculated values for σ_{meta} and σ_{para} for all aromatic rings based on the model peptide Ac-TXPN-NH_2.

Table 3.2. Calculated Hammett values for aromatic substituents and heteroaromatic rings based on the measured $K_{trans/cis}$ for model peptides Ac-TXPN-NH_2

The calculated σ_{para} and σ_{meta} values were determined based on the Hammett correlations and ρ values shown in Figure 3.37.

Ac-TXPN-NH_2, X =	σ_{para}	σ_{meta}	$K_{trans/cis}$
His(H^+)	2.58	1.76	12.3
4-Pyridyl(N-oxide)-Ala	2.38	1.64	11.2
4-Pyridyl(H^+)-Ala	2.38	1.64	11.2
F_5-Phe	1.28	0.96	6.7
4-Pyridyl-Ala	1.08	0.84	6.1
His	0.93	0.75	5.7
4-S-NEM-Phe	0.81	0.67	5.4
4-SO_2-propargyl-Phe	0.42	0.43	4.5
4-SBz-Phe	0.32	0.37	4.3
4-SO-propargyl-Phe	0.32	0.37	4.3
4-S(2-nitrobenzyl)-Phe	0.27	0.34	4.2
4-SSGlutathione-Phe	0.11	0.24	3.9
4-Pyrrolidyl(H^+)-Phe	-0.12	0.10	3.5
4-$SCH_2CH{=}CHCH_2OH$-Phe	-0.12	0.10	3.5
4-Spropargyl-Phe	-0.31	-0.02	3.2
4-OPO_3^{3-}-Phe	-0.45	-0.10	3.0
4-OSO_3^--Phe	-0.52	-0.15	2.9
4-SSPyridyl-Phe	-0.68	-0.24	2.7
4-Pyrrolidyl-Phe	-0.84	-0.34	2.5
Trp	-1.32	-0.64	2.0

In order to examine the scope and limitations of the aromatic-cis-proline interaction, we quantified the aromatic electronic effects of di-substituted aromatic amino acids on cis-trans isomerism of proline. The $K_{trans/cis}$ values measured for each peptide Ac-TXPN-NH_2 containing disubstituted aromatic amino acids are compiled in Table 3.3. The Hammett correlations for 3,4-substituted aromatic amino acids within the model peptide Ac-TXPN-NH_2 are shown in Figure 3.38.

Table 3.3. Combined $K_{trans/cis}$ values for the peptide Ac-TXPN-NH_2 where X = 3,4-disubstituted phenylalanine

The equilibrium between cis and trans conformations ($K_{trans/cis}$) was determined by measuring the ratios of 2 or 3 pairs of proton signals in the ^{1}H NMR spectra ($\Sigma\int$(trans signals)/$\Sigma\int$(cis signals)). When possible, amide proton signals were used, but other peaks, including the acetyl methyl, threonine methyl, and threonine Hα, were also used for calculation in the event that amide signals could not be used due to spectral overlap or rapid amide exchange. The σ_{para} and σ_{meta}, when available, are included in the table.[358] Chemical shift data for the proline Hα were derived from TOCSY data. The σ_{meta} for the 4-substituent.

Ac-TXPN-NH_2, 3,4-Phe		4-substitutent	3-substitutent					
4-X	**3-X**	σ_{meta}	σ_{meta}	σ_{para}	$K_{trans/cis}$	$\Delta G_{trans/cis}$, kcal mol^{-1}	δ ProHα cis	δ ProHα trans
OH	NO_2	0.12	0.71	0.78	6.5	-1.11	4.24	4.41
OH	NH_3^+	0.12	0.86	0.60	5.0	-0.95	3.95	4.41
OH	I	0.12	0.35	0.18	3.7	-0.77	3.87	4.42
OH	SH	0.12	0.25	0.15	3.3	-0.71	3.82	4.41
OH	H	0.12	0.00	0.00	3.2	-0.69	3.82	4.42
OH	NH_2	0.12	-0.16	-0.66	2.9	-0.63	3.76	4.42
OH	OH	0.12	0.12	-0.37	2.5	-0.54	3.75	4.42
OH	S^-	0.12	-0.36	-1.21	1.6	-0.28	3.73	4.42
O^-	NO_2	-0.47	0.71	0.78	3.1	-0.67	3.97	4.41
O^-	I	-0.47	0.35	0.18	1.9	-0.38	3.72	4.42
O^-	H	-0.47	0.00	0.00	1.9	-0.38		
O^-	NH_2	-0.47	-0.16	-0.66	1.6	-0.28		

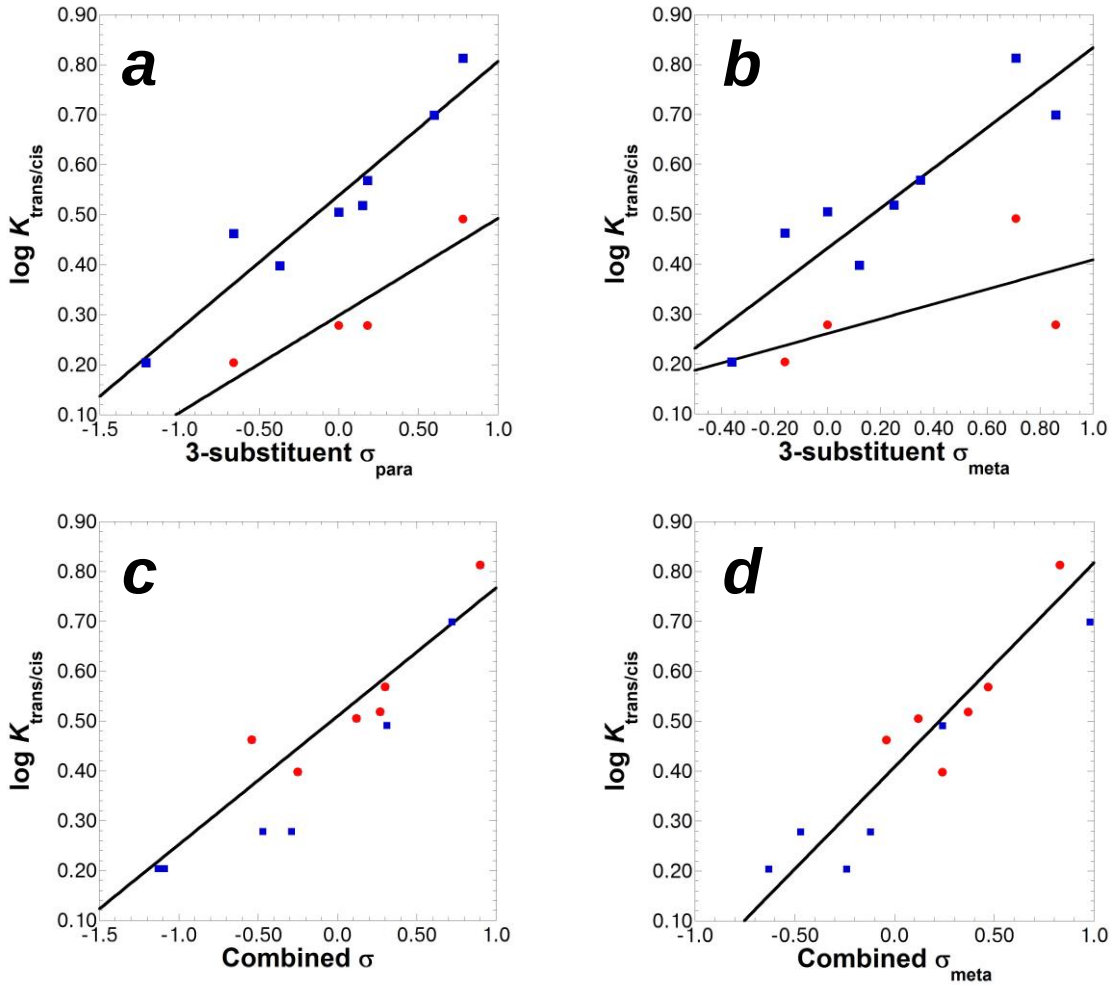


Figure 3.38 Hammett correlations of the $K_{trans/cis}$ in model peptides Ac-TXPN-NH_2, where X = 3,4-disubstituted aromatic amino acid

(a) The log($K_{trans/cis}$) of each peptide was correlated with the σ_{para} Hammett constant for the 3-substituent in 3-substituted tyrosine derivatives within Ac-TXPN-NH_2 model peptides. The tyrosine hydroxyl is either neutral (blue squares) or anionic (red circles); (b) The log($K_{trans/cis}$) of each peptide was correlated with the σ_{meta} Hammett constant for the 3-substituent in 3-substituted tyrosine derivatives within Ac-TXPN-NH_2 model peptides. The tyrosine hydroxyl is either neutral (blue squares) or anionic (red circles); (c) The log($K_{trans/cis}$) of each peptide is correlated with the combined Hammett constant σ in 3,4-disubstituted aromatic amino acids within Ac-TXPN-NH_2 model peptides. The combined Hammett constant σ is additive between the σ_{para} value of the 3-substituent and the σ_{meta} value of the 4-substituent. (d) The log($K_{trans/cis}$) of each peptide is correlated with the combined Hammett constant σ in 3,4-disubstituted aromatic amino acids within Ac-TXPN-NH_2 model peptides. The combined Hammett constant σ is additive between the σ_{meta} value of both the 3-substituent and the 4-substituent.
(a) neutral tyrosine correlation: $\rho = 0.268 \pm 0.034$, $R = 0.954$
(a) anionic tyrosine correlation: $\rho = 0.194 \pm 0.054$, $R = 0.930$
(b) neutral tyrosine correlation: $\rho = 0.432 \pm 0.035$, $R = 0.904$
(b) anionic tyrosine correlation: $\rho = 0.261 \pm 0.073$, $R = 0.606$
(c) Combined σ correlation: $\rho = 0.258 \pm 0.065$, $R = 0.892$
(d) Combined σ_{meta} correlation: $\rho = 0.366 \pm 0.046$, $R = 0.929$

With disubstituted aromatic amino acids, a correlation was observed for the $K_{trans/cis}$ of the model peptide Ac-TXPN-NH_2 with the σ_{para} value of the 3-substitutent of tyrosine (Figure 3.38a). In contrast, the σ_{meta} value of the 3-substitutent of tyrosine was not as well correlated to the $K_{trans/cis}$, particularly under conditions were tyrosine was anionic (Figure 3.38b). Substituent effects are known to be approximately additive,[371] an approximate combined σ value for 3,4-phenylalanine derivatives can be determined by the σ_{para} and σ_{meta} values of the respective substituents. The combined σ values (determined as 4-σ_{meta} + 3-σ_{para} in Figure 3.38c, or 4-σ_{meta} + 3-σ_{meta} in Figure 3.38d) were found to have direct correlations with the $K_{trans/cis}$ of the model Ac-TXPN-NH_2 peptides, with both anionic and neutral forms of the tyrosine hydroxyl group. These correlations suggest that the resonance and inductive effects from the 3-

substituents in disubstituted tyrosine derivatives are important for stabilizing the cis conformation of proline in model peptides Ac-TXPN-NH_2.

In addition to the observed aromatic substituent effects on on the cis population for the peptides Ac-TXPN-NH_2, the proline Hα chemical shift between cis and trans conformations exhibited significant dependence on the aromatic substituents as well. By ^{1}H-^{1}H-TOCSY NMR spectra, a substantial upfield chemical shift was observed for the δ cis-Pro Hα for peptides Ac-TXPN-NH_2 containing electron-rich aromatic amino acids. In contrast, the chemical shift for trans-Pro Hα within peptides Ac-TXPN-NH_2 was unchanged for all aromatic amino acids. Fundamentally, an upfield chemical shift in the proline Hα indicates a change in the local environment, potentially as an aromatic ring current effect.[106, 370] The proline Hα chemical shifts for cis and trans conformations for the peptides Ac-TXPN-NH_2 are compiled in Table 3.4. A direct correlation was observed between the chemical shift of the cis-proline Hα and the $K_{trans/cis}$ of the peptide (Figure 3.39). Furthermore, a direct correlation was also observed between the proline Hα and the σ_{meta} value for 4-substituted aromatic amino acids (Figure 3.40).

Table 3.4. Proton chemical shifts for ProHα for the peptides Ac-TXPN-NH_2

Proton chemical shifts for the proline Hα (as cis and trans conformations) were obtained via TOCSY spectra for some of the peptides Ac-TXPN-NH_2. The σ_{para} and σ_{meta}, when available, are included in the table.[358]

Ac-TXPN-NH_2, X =	σ_{para}	σ_{meta}	$K_{trans/cis}$	**δ $Pro_{H\alpha cis}$**	**δ $Pro_{H\alpha trans}$**
His(H^+)			12.3	4.64	4.43
4-Pyridyl(N-oxide)-Ala			11.2	4.52	4.42
4-Pyridyl(H^+)-Ala			11.2		4.42
4-Pyridyl-Ala			6.1	4.26	4.44
His			5.7	4.14	4.42
4-NO_2-Phe	0.78	0.71	5.6	4.18	4.43
4-NH_3^+-Phe	0.60	0.86	5.5	3.98	4.42
4-CN-Phe	0.66	0.56	5.3	4.10	4.42
4-CF_3-Phe	0.54	0.43	4.8	4.01	4.42
4-S(O)Me-Phe	0.49	0.52	4.7	3.98	4.43
4-Br-Phe	0.23	0.39	4.6	4.00	4.42
4-OSO_2CF_3-Phe	0.53	0.56	4.5	3.97	4.42
4-SBz-Phe			4.3	3.96	4.42
4-I-Phe	0.18	0.35	4.3	3.97	4.41
4-F-Phe	0.06	0.34	4.2	3.94	4.41
4-Cl-Phe	0.23	0.37	4.2	3.99	4.42
4-OAc-Phe	0.44	0.39	3.9	3.89	4.41
4-SH-Phe	0.15	0.25	3.8	3.87	4.42
4-SO_3^--Phe	0.35	0.30	4.4	3.90	4.42
Phe	0.00	0.00	3.6	3.82	4.42
4-SMe-Phe	0.00	0.15	3.6	3.87	4.41
4-Pyrrolidyl(H^+)-Phe			3.5	3.82	4.41
4-$B(OH)_2$-Phe	0.12	-0.01	3.5	3.82	4.42
4-S(propargyl)-Phe			3.2	3.85	4.41
Tyr	-0.37	0.12	3.2	3.82	4.42
4-OPO_3H^--Phe			3.1	3.82	4.41
4-OPO_3^{3-}-Phe			3.0	3.84	4.42
4-OSO_3^--Phe			2.9	3.81	4.42
4-NH_2-Phe	-0.66	-0.16	2.9	3.81	4.43
4-Pyrrolidyl-Phe			2.5	3.82	4.44
4-$B(OH)_3^-$-Phe	-0.44	-0.48	2.4	3.64	4.42
4-S^--Phe	-1.21	-0.36	2.3	3.75	4.41
Trp			2.0	3.45	4.43
Ala			11.5	4.65	4.43
Cha			8.6	4.52	4.41

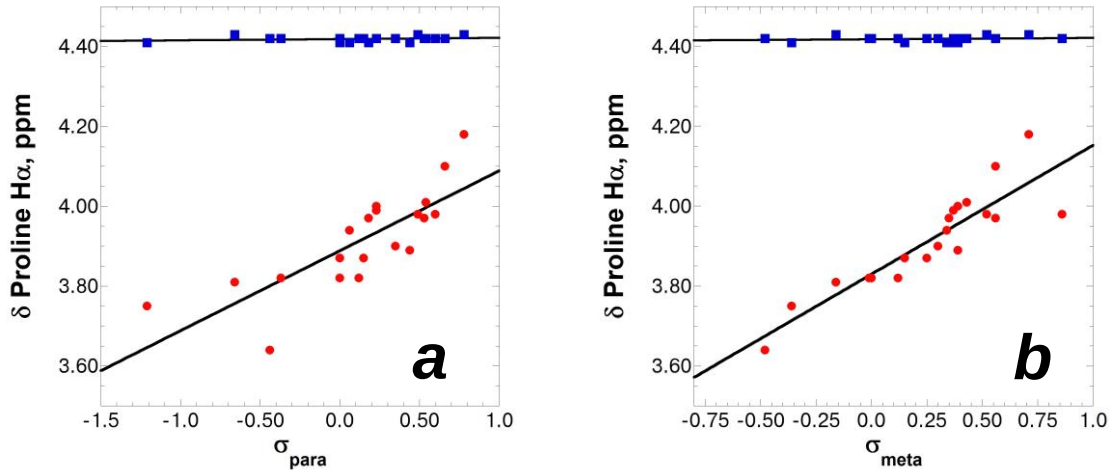


Figure 3.39 Hammett correlation for the proton chemical shift of ProHα in peptides Ac-TXPN-NH_2 (where X = 4-substituted amino acid)

The chemical shift (δ) for proline Hα in cis (red circles) and trans (blue squares) conformations are compared against the Hammett constant (a, σ_{para}; b, σ_{meta}) for each aromatic substituent.

σ_{para} correlation with cis-ProHα (a): $\rho = 0.200 \pm 0.035$, $R = 0.800$

σ_{para} correlation with trans-ProHα (a): $\rho = 0.003 \pm 0.004$, $R = 0.177$

σ_{meta} correlation with cis-ProHα (b): $\rho = 0.323 \pm 0.039$, $R = 0.890$

σ_{meta} correlation with trans-ProHα (b): $\rho = 0.003 \pm 0.003$, $R = 0.236$

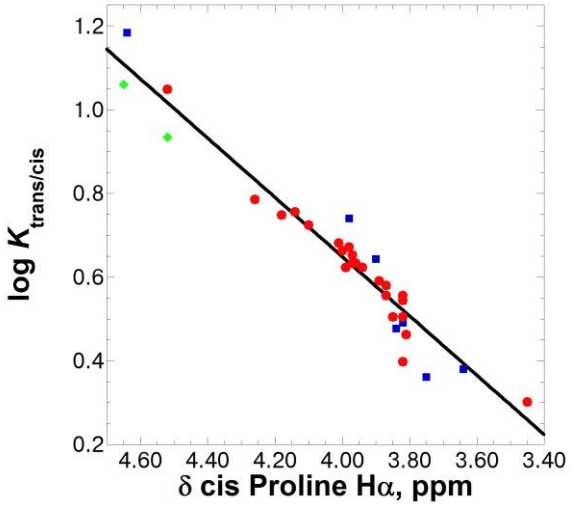


Figure 3.40 Proton chemical shift for cis-ProHα correlation with $K_{trans/cis}$ of Ac-TXPN-NH_2 model peptides (where X = amino acid)
The chemical shift (δ) for cis proline Hα for neutral (red circles) and charged (blue squares) aromatic substituents in model peptides Ac-TXPN-NH_2 are correlated to the cis-population of the peptide. Non-aromatic amino acids (Ala or Cha) are shown as green diamonds.
R = 0.967 for neutral substituents only.

The aromatic substituent effects in the peptides Ac-TXPN-NH_2 directly influence the local magnetic environment around the cis-Pro Hα, but *not* the trans-Pro Hα (Figure 3.39). Furthermore, the correlation between the $K_{trans/cis}$ for each peptide and the cis-proline Hα chemical shift demonstrates that the local magnetic environment around the proline Hα is directly related with the cis-propensity of the model peptides (Figure 3.40). For the Pro Hα in the cis conformation, the range of chemical shifts varied from protonated histidine, at 4.64 ppm, to tryptophan, at 3.45 ppm. In contrast, Pro Hα in the trans conformation was consistently at 4.42 ± 0.01 ppm for all peptides examined. These observations are consistent with prior examples of model peptides involving aromatic-cis-proline motifs.[288, 318, 372] Combined, these data indicate that the backbone conformational preference of these aromatic-proline

motifs is directly influenced by the aromatic substituent effects in the aromatic amino acid, and a shielding interaction with the proline Hα is correlated with these conformational preferences.

With the observed increase in cis-populations of the model peptides Ac-TXPN-NH_2 with electron-donating aromatic substituents, and the upfield chemical shifts in the cis-Pro Hα, the nature of this aromatic-cis-proline interaction demanded further insight. The Hammett correlation on the $K_{trans/cis}$ of the model peptides Ac-TXPN-NH_2 suggested that the cis conformation was stabilized by an electrostatic effect. The greater correlation $K_{trans/cis}$ with the aromatic substituent σ_{meta} values suggests that the cis-conformation is influenced more by inductive effects rather than resonance effects, but this can not be established based on Hammett correlations alone. The role of hydrophobic effects, electrostatic effects, and potential orbital-overlap interactions were addressed in further studies on these aromatic-cis-proline motifs, via solvent dependent studies, determination of thermodynamic parameters, and detailed examination of minimized aromatic-cis-proline dipeptides via x-ray crystallography, NBO analysis, and NMR.

3.2.3 Solvent effects on proline cis-trans isomerism in the peptides Ac-TXPN-NH_2

In comparing the model peptides Ac-TXPN-NH_2 containing either alanine or cyclohexylalanine, the peptide containing cyclohexylalanine had increased cis-Pro population, suggesting that an increase in the hydrophobic surface area can play a role in stabilizing the cis-proline within aromatic-proline motifs. The face of the pyrrolidyl ring in proline is hydrophobic, and it can be expected to interact favorably with other hydrophobic surfaces. In order to determine the degree to which hydrophobic

contributions stabilize the cis-proline conformation, a selection of model peptides Ac-TXPN-NH_2 was examined in organic solvents, methanol and acetonitrile, and the $K_{trans/cis}$ was determined via NMR.

If the aromatic-cis-proline interaction is driven by the hydrophobic effect, then the interaction should be weaker in the presence of organic solvents, resulting in a higher $K_{trans/cis}$. In the context of a hydrophobic interaction, aromatic substituent effects would have little or no correlation with the cis-population of the peptides. In contrast, an electrostatic interaction will be enhanced in the presence of organic solvents, as competing interactions with bulk polar solvents decrease and intramolecular association of charged groups is favored. To examine these possibilities, and to elucidate the underlying nature of the aromatic-cis-proline interaction, a series of five model peptides Ac-TXPN-NH_2 was selected, with 4-substituted aromatic amino acids representing a range of σ values. The $K_{trans/cis}$ for each model peptide was measured via NMR in either 100% MeCN-d_3 or 90% MeOH-d_3/10% MeOD-d_4, and the data was compared with previous results in 90% H_2O/10% D_2O (Table 3.5). MeOH-d_3 was used so that the amide proton resonances could be used for calculating the $K_{trans/cis}$ of the model peptides. Correlations of the measured $K_{trans/cis}$ values with the Hammett values, σ_{para} or σ_{meta}, are shown in Figure 3.41.

Table 3.5. $K_{trans/cis}$ values for model peptides Ac-TXPN-NH_2 in organic solvents

The equilibrium between cis and trans conformations ($K_{trans/cis}$) was determined by measuring the ratios of 2 or 3 pairs of proton signals in the ^{1}H NMR spectra ($\Sigma\int$(trans signals)/$\Sigma\int$(cis signals)) in the indicated solvent. When possible, amide proton resonances were used, but other peaks, including threonine methyl protons, and threonine Hα or Hβ, or aromatic protons were also used for calculation, where amide signals could not be used due to spectral overlap. The σ_{para} and σ_{meta}, when available, are included in the table.[358]

			90% H_2O/ 10% D_2O		90% MeOH-d_3/ 10% MeOD-d_4		100% MeCN-d_3	
Ac-TXPN-NH_2, X =	σ_{para}	σ_{meta}	$K_{trans/cis}$	ΔG, kcal mol^{-1}	$K_{trans/cis}$	ΔG, kcal mol^{-1}	$K_{trans/cis}$	ΔG, kcal mol^{-1}
4-CN-Phe	0.66	0.56	5.3	-0.99	5.4	-1.00	4.1	-0.83
4-SOMe-Phe	0.49	0.52	4.7	-0.92	3.2	-0.69	4.2	-0.85
4-I-Phe	0.18	0.35	4.3	-0.86	3.0	-0.65	2.9	-0.63
4-SMe-Phe	0.00	0.15	3.6	-0.76	2.6	-0.57	2.6	-0.57
Phe	0.00	0.00	3.6	-0.76	2.2	-0.45	2.2	-0.47
Tyr	-0.37	0.12	3.2	-0.69	1.8	-0.33	2.2	-0.45
Ala			11.5	-1.45	7.9	-1.22	5.8	-1.04

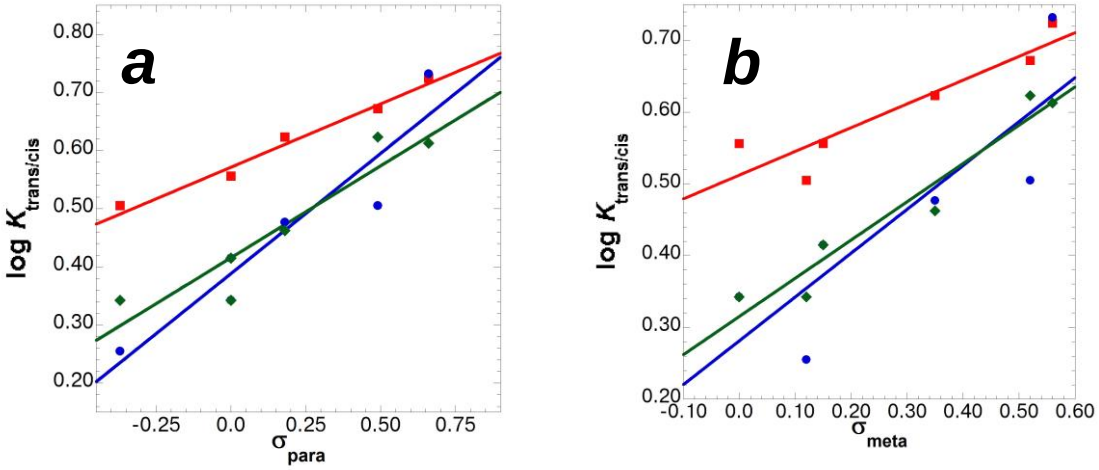


Figure 3.41 Solvent effects on the $K_{trans/cis}$ of model peptides Ac-TXPN-NH_2 (where X = 4-substituted amino acid)

The $K_{trans/cis}$ was measured for selected model peptides in different solvents: 90% H_2O/10% D_2O (red circles), 90% MeOH-d_3/10% MeOD-d_4 (blue squares), or 100% MeCN-d_3 (green diamonds). The log($K_{trans/cis}$) of each peptide is correlated with the Hammett constant (a, σ_{para}; b, σ_{meta}) for each aromatic substituent.
σ_{para} correlation in H_2O/D_2O (a): $\rho = 0.218 \pm 0.018$, $R = 0.986$
σ_{para} correlation in MeOH/MeOD (a): $\rho = 0.414 \pm 0.076$, $R = 0.939$
σ_{para} correlation in MeCN (a): $\rho = 0.316 \pm 0.061$, $R = 0.933$
σ_{meta} correlation in H_2O/D_2O (b): $\rho = 0.331 \pm 0.069$, $R = 0.922$
σ_{meta} correlation in MeOH/MeOD (b): $\rho = 0.611 \pm 0.186$, $R = 0.854$
σ_{meta} correlation in MeCN (b): $\rho = 0.533 \pm 0.069$, $R = 0.968$

For most of the model peptides Ac-TXPN-NH_2 examined, the aromatic-cis-proline interaction was stabilized in the presence of organic solvent; the only exception was with 4-cyanophenylalanine, which was observed to have a slightly higher $K_{trans/cis}$ in methanol than in water. The aromatic-cis-proline interaction had a greater ρ value in both acetonitrile and methanol compared to water, indicating that aromatic substituent effects are more important in organic solvents. These data strongly suggest that the energetic stability of the aromatic-cis-proline interaction has a minimal contribution due to the hydrophobic effect.

However, it can be seen that the peptide Ac-TAPN-NH_2 has an increased cis-population in the presence of organic solvents as well, even in the absence of a stabilizing aromatic-cis-proline interaction. This increased cis population is potentially due to an enhancement of an *i,i*+3 intermolecular hydrogen bond between the carbonyl oxygen and the amide proton, typical of type VIa1 β-turns.[285] In organic solvents, conventional hydrogen bonding interactions that are stabilized by electrostatic effects are enhanced,[203] for the same reasons outlined above.

3.2.4 Effect of modified prolines: combined aromatic electronic and stereoelectronic effects in Ac-TXPro$_x$N-NH_2 model peptides

These studies on cis-trans isomerism of proline in model tetrapeptides Ac-TXPN-NH_2 (where X = aromatic amino acid) demonstrated that increasingly electron-donating aromatic substituents stabilize the cis-proline conformation. Modification on prolines can also influence cis-trans isomerism via stereoelectronic effects.[93, 98, 373-375] Proline is a saturated amino acid that is uniquely cyclic through its own amide nitrogen, and the ring must "pucker" to accommodate its saturated cyclic structure (Figure 3.42). Control of the proline ring-pucker can influence the amide bond (ω) and other main chain (φ, ϕ) conformations.[98, 285, 374, 375] Functional groups on proline, such as 4*R*-hydroxyproline, can dictate the ring "pucker" via hyperconjugative effects between the C_δ–N σ* and Cγ–Hγ σ orbitals, C_β–H_β σ and Cγ–Xγ σ* orbitals, or C_δ–H_δ σ and Cγ–Xγ σ* orbitals (X = 4-substituent on proline).[373] The stereoelectronic control of proline ring-pucker through 4*R*- or 4*S*-substituents is known to stabilize polyproline helices or collagen.[98] In general, when proline adopts an *endo* ring-pucker, it has a tendency for the cis amide bond conformation, due to restrictions about the ϕ and ψ torsion angles.[98, 285] 4*S*-fluoroproline is known to promote the cis amide bond

conformation while 4*R*-hydroxyproline promotes the trans amide bond conformation.[98, 374] We hypothesized that a tetrapeptide could be designed to have exceptional stability for the cis-conformation, based on the guiding principles of aromatic-cis-proline interactions and stereoelectronic control of proline ring pucker. With such a designed peptide, the aromatic-cis-proline can be stabilized and the nature of the prolyl C–H/π aromatic interaction could be studied in greater detail.

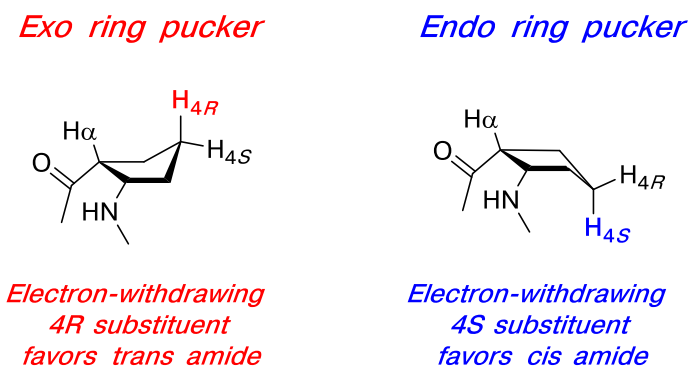


Figure 3.42 Prolyl ring-pucker dictates influences proline amide bond conformation
In prior work,[373, 374] stereoelectronic control of the proline ring pucker, as either endo or exo, was demonstrated as a means for promoting cis or trans amide bond conformations in peptides. The ring pucker influences the allowed and disallowed amide bond conformations due to restrictions on the ϕ and ψ torsion angles.[98, 285]

We synthesized a series of model peptides containing various aromatic amino acids and a modified proline (Pro$_x$), either 2*S*,4*S*-fluoroproline (*4S*-fluoroproline or flp) or 2*S*,4*R*-hydroxyproline (4*R*-hydroxyproline or Hyp). Using established protocols,[374] 4*R*-hydroxyproline was incorporated into peptides via Fmoc-solid-phase peptide synthesis, and 4*S*-fluoroproline was incorporated via solid-phase modification of peptides containing 4*R*-hydroxyproline. Peptides Ac-TXPro$_x$N-NH$_2$ (where X = 4-

substituted aromatic amino acid; Pro_x = Hyp or flp) were synthesized and characterized by Krista Thomas and Dr. Devan Naduthambi,[91, 374] in addition to the peptides Ac-T(4-I-Phe)Pro_xN-NH_2 and Ac-T(4-SH-Phe)Pro_xN-NH_2 described below (Figures 3.43-3.45). The $K_{trans/cis}$ was measured for all Ac-TXPro_xN-NH_2 peptides via NMR (Table 3.6). The peptides Ac-TXPro_xN-H were compared via Hammett correlation (Figure 3.46) between the peptides where Pro_x = 4*R*-hydroxyproline, proline, or 4*S*-fluoroproline.

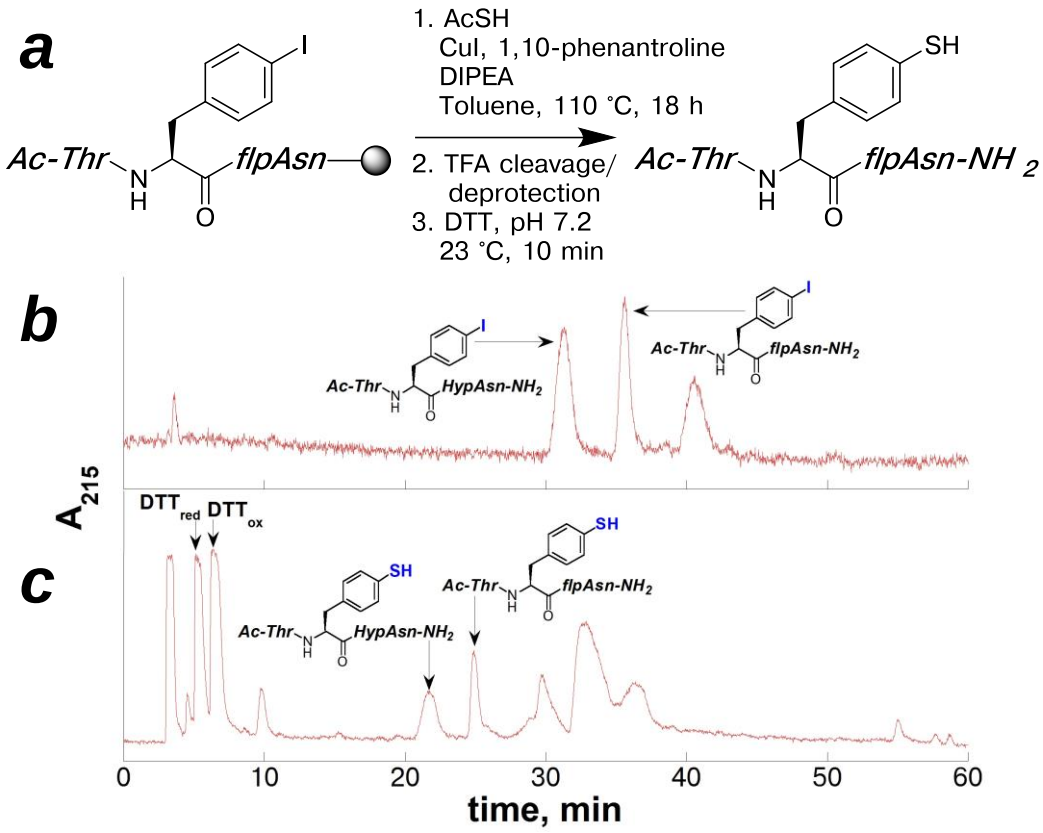


Figure 3.43 Synthesis of the peptide Ac-T(4-SH-Phe)Pro$_x$Asn-NH$_2$ (Pro$_x$ = 4*R*-hydroxyproline or 4*S*-fluoroproline) on solid-phase

(a) Scheme for the copper-mediated cross-coupling reaction on the peptide Ac-T(4-I-Phe)flpN-NH$_2$ (using a previously established protocol[374]) to generate the peptide Ac-T(4-SH-Phe)flpN-NH$_2$; (b) HPLC chromatogram of the crude peptide Ac-T(4-I-Phe)flpN-NH$_2$ that resulted from the solid-phase modification reaction;[374] (c) HPLC chromatogram of the crude peptide Ac-T(4-SH-Phe)flpN-NH$_2$ that resulted from the copper-mediated cross-coupling reaction on solid phase. The reaction products were analyzed via HPLC using a linear gradient of 0-35% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products identified via ESI-MS are indicated.

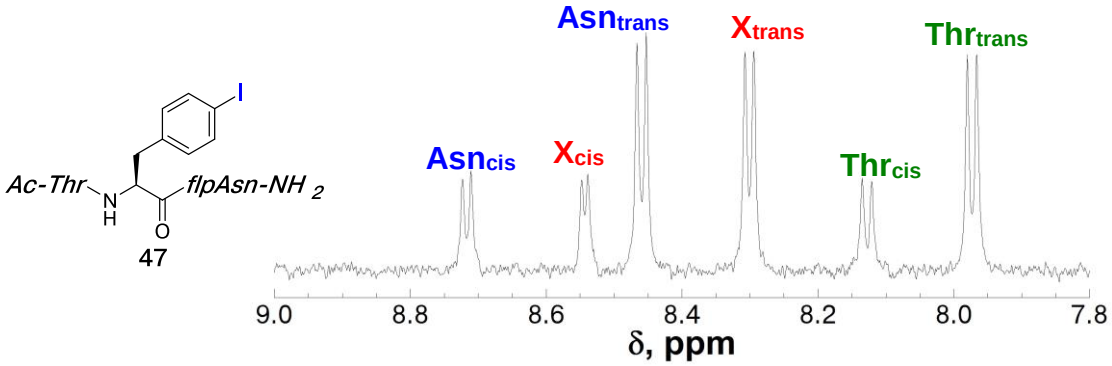


Figure 3.44 NMR Characterization of the peptide Ac-TXPro$_x$Asn-NH$_2$ (X = 4-iodophenylalanine)

NMR spectrum of the amide region of the peptide Ac-T(4-I-Phe)flpAsn-NH$_2$ at pH 4. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured 90% H$_2$O/10% D$_2$O at 298 K. Amide peaks were identified by the TOCSY spectrum.

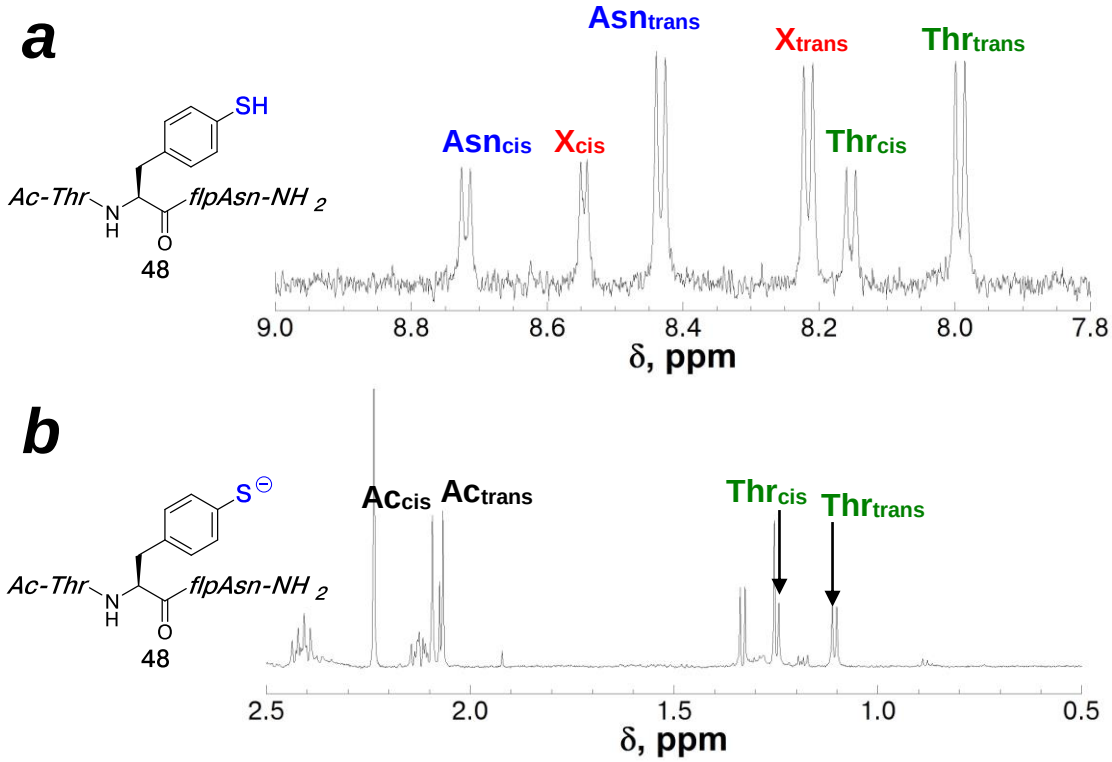


Figure 3.45 NMR Characterization of the peptide Ac-TXflpAsn-NH$_2$ (X = 4-thiolphenylalanine)

NMR spectra of the peptide Ac-T(4-SH-Phe)flpAsn-NH_2 at (a) pH 4.0 (amide region shown) and (b) pH 8.5 (aliphatic region shown). The sample contained 5 mM phosphate and 25 mM NaCl, and 100 μM TCEP, and was measured in 90% H_2O/10% D_2O at 298 K. Amide peaks were identified by the TOCSY spectrum. Due to spectral overlap at pH 8.5, the integrations were taken from the TOCSY spectrum as volumes of proton resonances.

Table 3.6. ^{1}H NMR derived data for the peptides Ac-TXPro$_x$N-NH_2 (Pro$_x$ = Hyp or flp)

The equilibrium between cis and trans conformations ($K_{trans/cis}$) was determined by measuring the ratios of 2 or 3 pairs of proton signals in the ^{1}H NMR spectra ($\Sigma\int$(trans signals)/$\Sigma\int$(cis signals)). When possible, amide proton signals were used, but other peaks, including the acetyl methyl, threonine methyl, and carboxamides were also used for calculation, in the event that amide signals could not be used due to spectral overlap or rapid amide exchange. Available σ_{meta} are included in the table.[358]
n.d. indicates "not determined" due to spectral overlap

Ac-TXZN-NH_2 X(Z) =	σ_{para}	σ_{meta}	$K_{cis/trans}$	$K_{trans/cis}$	$^3J_{\alpha N}$ X_{cis}, Hz	$\Delta G_{cis/trans}$, kcal mol^{-1}	% cis
Ala(Hyp)			0.07	15.5	n.d.	1.62	6%
4-NO_2-Phe(Hyp)	0.78	0.71	0.11	9.5	n.d.	1.33	10%
Tyr(Hyp)	-0.37	0.12	0.12	8.5	7.6	1.27	11%
Phe(Hyp)	0.00	0.00	0.15	6.6	7.3	1.12	13%
Trp(Hyp)			0.22	4.5	6.7	0.89	18%
Ala(flp)			0.18	5.5	n.d.	1.01	15%
4-I-Phe(flp)	0.18	0.35	0.40	2.5	5.6	0.55	28%
4-NH_3^+-Phe(flp)	0.60	0.86	0.46	2.2	5.2	0.46	32%
4-SH-Phe(flp)	0.15	0.25	0.49	2.1	5.5	0.42	33%
Phe(flp)	0.00	0.00	0.54	1.9	5.0	0.37	35%
Tyr(flp)	-0.37	0.12	0.68	1.5	4.7	0.22	41%
4-NH_2-Phe(flp)	-0.66	-0.16	0.91	1.2	n.d.	0.06	48%
4-S^--Phe(flp)	-1.21	-0.36	0.93	1.1	n.d.	0.05	48%
Trp(flp)			1.41	0.7	4.1	-0.20	58%

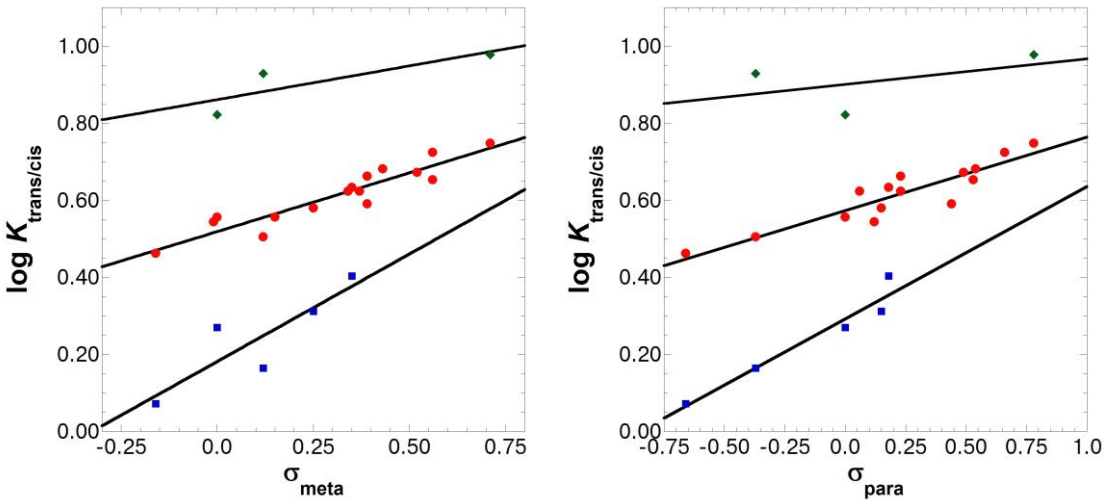


Figure 3.46 Hammett correlation for model peptides Ac-TXPro$_x$N-NH$_2$
Cis-trans isomerism of the model peptides Ac-TXProxN-NH$_2$ was measured via NMR, where Pro$_x$ was either 4*R*-hydroxyproline (Hyp, green diamonds), proline (red circles), or 4*S*-fluoroproline (blue squares). Hammett correlations are with respect to the (a) σ_{meta} or (b) σ_{para} of the aromatic substituent. Only aromatic amino acids with neutral substituents are shown.
(a) 4*R*-hydroxyproline: $\rho = 0.175$; $R = 0.836$
(a) proline: $\rho = 0.305$; $R = 0.933$
(a) 4*S*-fluoroproline: $\rho = 0.558$; $R = 0.871$
(b) 4*R*-hydroxyproline: $\rho = 0.066$; $R = 0.489$
(b) proline: $\rho = 0.191$; $R = 0.917$
(b) 4*S*-fluoroproline: $\rho = 0.344$; $R = 0.970$

Consistent with prior observations in model tetrapeptides containing modified proline residues, lower $K_{trans/cis}$ values were observed for peptides containing 4*S*-fluoroproline, and higher $K_{trans/cis}$ values for peptides containing 4*R*-hydroxyproline.[374] Within peptides Ac-TXPro$_x$N-NH$_2$ containing either 4*S*-fluoroproline or 4*R*-hydroxyproline, an increase in cis-population was observed with increasingly electron-donating aromatic substituents, consistent with our studies in model tetrapeptides containing native proline (Table 3.1). These data demonstrate that modified proline residues provide additional ability to fine-tune proline conformation in designed

motifs, and can increase the range of conformational control over cis-trans isomerism of proline. The peptides exhibited different sensitivities to aromatic substituent effects, where the value ρ increased as Hyp < Pro < flp. The electron-donating ability of 4-substituents on proline (such as in 4*R*-hydroxyproline or 4*S*-fluoroproline) is known to influence the "ring pucker" of the pyrrolidyl ring,[373, 375] and the conformation of the pyrrolidyl ring can enhance the stability of the aromatic C–H/π interaction. The endo ring-pucker is favored for 4*S*-fluoroproline, which promotes the C–H/π interaction, and the peptides containing flp exhibit enhanced sensitivity to aromatic substituent effects. For peptides containing 4*R*-hydroxyproline, the exo ring pucker disfavors the C–H/π interaction, and the peptides containing Hyp are less sensitive to aromatic substituent effects. In addition, the $^{3}J_{\alpha N}$ coupling constants were generally smaller for peptides containing flp in comparison to Pro or Hyp, indicating ϕ torsion angles that are closer to idealized type VIa1 β-turns (Ac-TWflpN-NH_2 $^{3}J_{\alpha N}$ = 4.1 Hz, Trp ϕ = –59°).[91, 289, 306, 370] Notably, with the use of 4*S*-fluoroproline in place of proline, the cis population of the peptide Ac-TWPro$_x$N-NH_2 was increased to 58% from 33%, which became the basis for designing a stabilized aromatic-cis-proline motif.

In order to more closely examine the fundamental nature of the aromatic-cis-proline interaction, we sought to design a peptide that had exceptionally high population and stability for the cis-proline conformation. By decreasing the population of trans-proline conformation within model peptides, the NMR resonances become more enhanced for cis-proline conformation. The model tetrapeptide Ac-TWflpN-NH_2, with almost 60% cis population, allowed for examination of the solution structure of the aromatic-cis-proline motif by ROESY spectrum; the ROESY spectrum indicated a through-space interaction between the aromatic ring and the prolyl Hα.[91]

Based on this work and prior literature examples, we established the following guidelines for designing stabilized aromatic-cis-proline tetrapeptides: (1) electron-rich aromatic amino acids preceding proline have the highest populations of cis conformations, with tryptophan showing the lowest $K_{trans/cis}$ values;[91] (2) modified prolines can further stabilize the cis conformation, when electron-withdrawing substituents are incorporated at the *4S* position, such as 4*S*-fluoroproline (alternatively, sterically-demanding substituents at the 4*R* position will also increase the cis-population);[373, 374] and (3) aromatic amino acids *following* proline can also increase the cis population of a model tetrapeptide.[90, 376] Based on these guidelines, we synthesized the model peptide Ac-TWflpF-NH_2, which was expected to have a highly stabilized cis conformation, which would allow for further study on the nature of the aromatic-cis-proline interaction. The ^{1}H NMR characterization for this peptide is shown in Figures 3.47-3.50 and key resonance assignments are shown in Table 3.7. The synthetic strategy for this peptide is included in the experimental section.

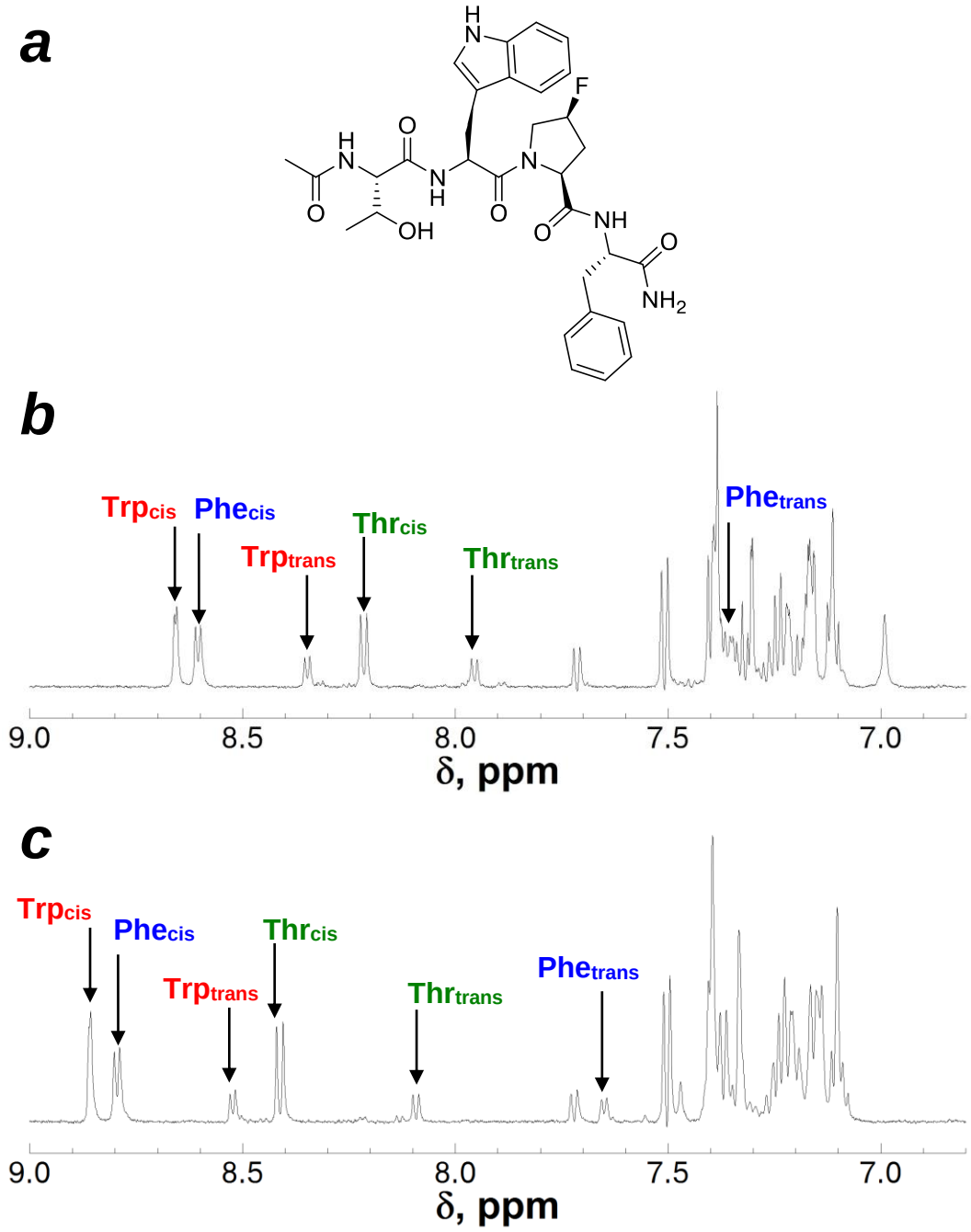


Figure 3.47 ^{1}H NMR Characterization of the designed peptide Ac-TWflpF-NH$_2$
^{1}H NMR spectrum of the aromatic/amide region of the peptide (a) Ac-TWflpF-NH$_2$ at (b) 298 K and (c) 277 K. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H$_2$O/10% D$_2$O. Amide peaks were identified by the TOCSY.

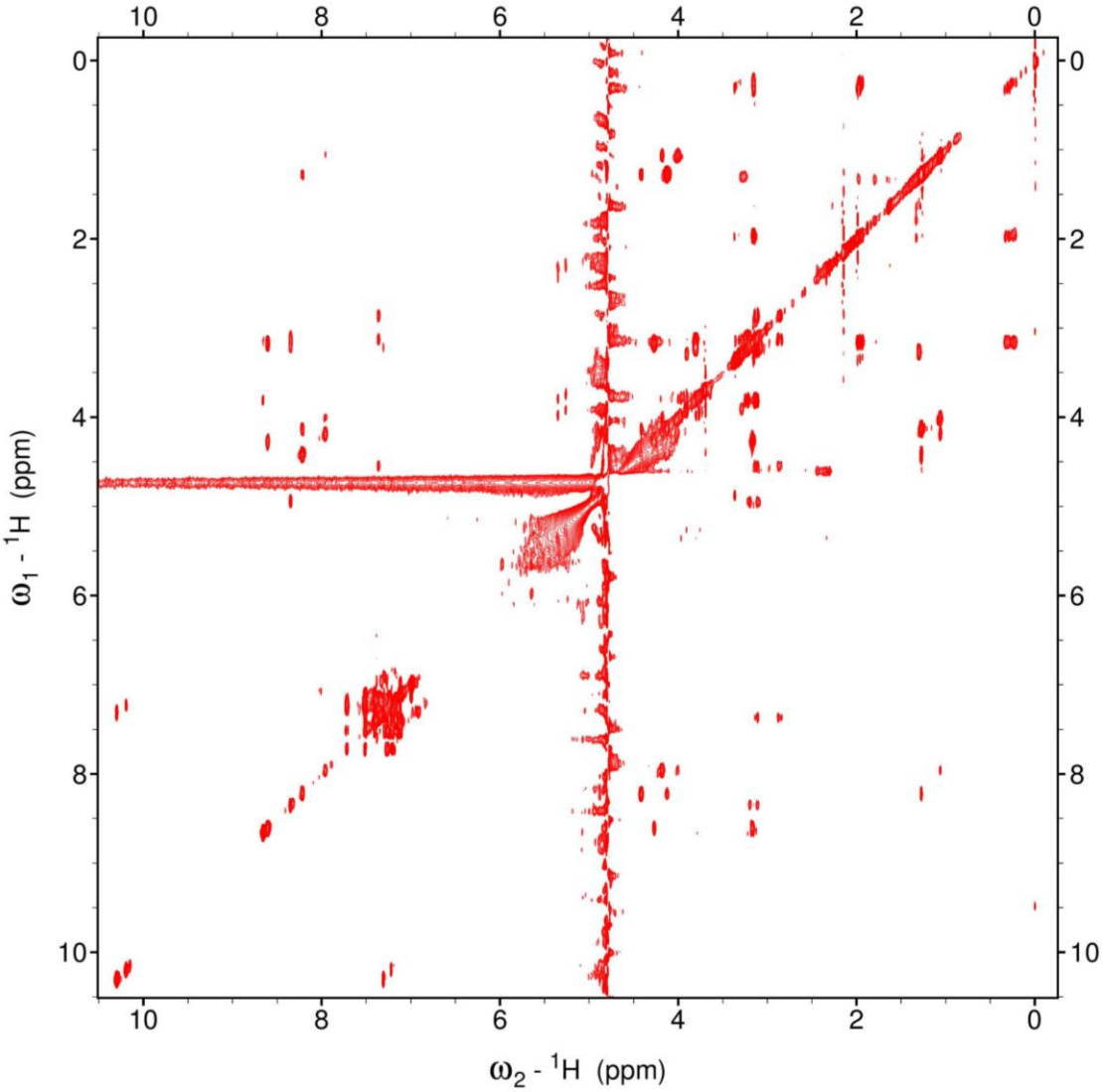


Figure 3.48 TOCSY spectrum of the designed peptide Ac-TWflpF-NH_2 at 298 K

^1H-^1H TOCSY NMR spectrum of the peptide Ac-TWflpF-NH_2 at 298 K. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O.

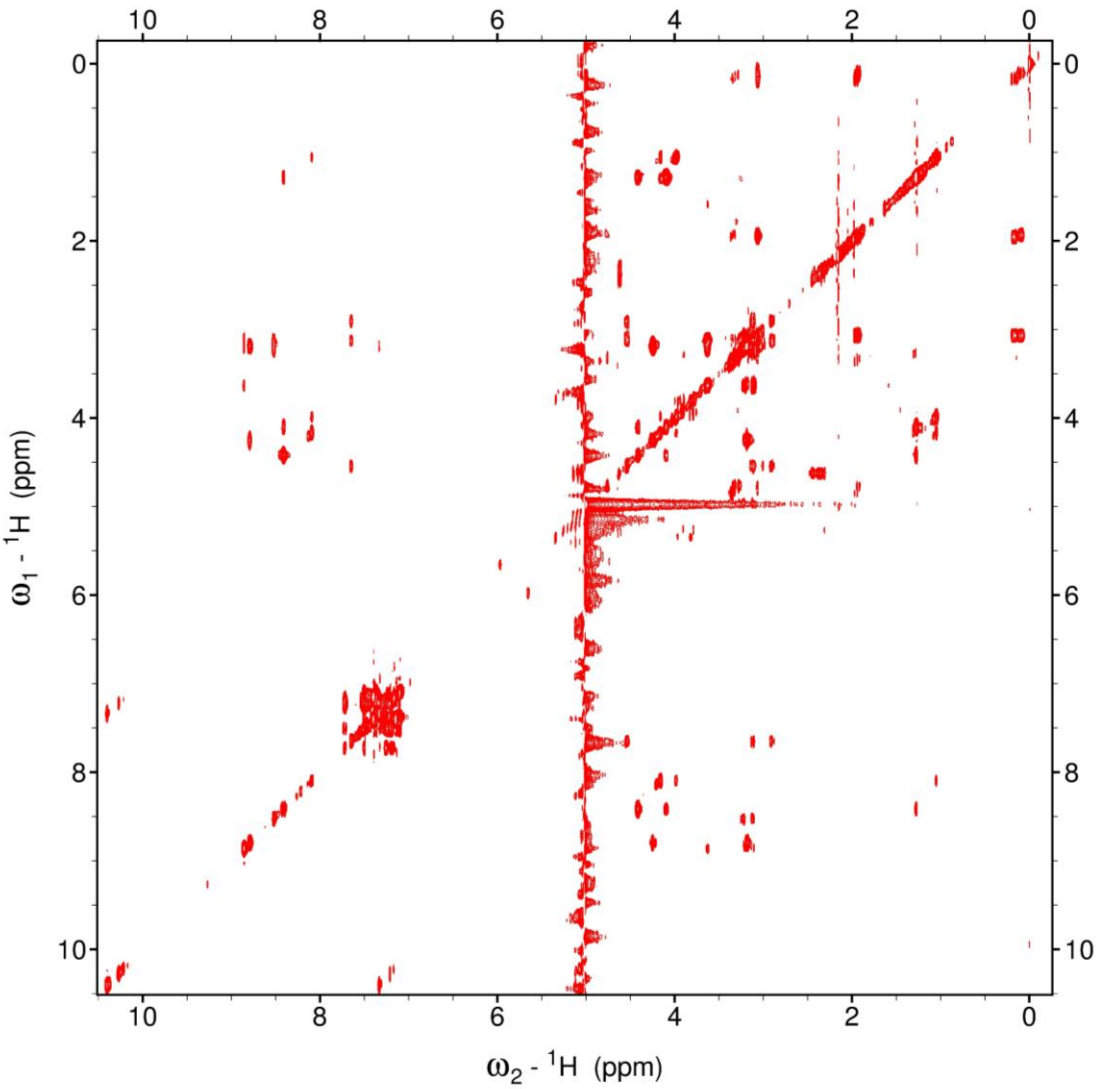


Figure 3.49 TOCSY spectrum of the designed peptide Ac-TWflpF-NH_2 at 277 K
^1H-^1H TOCSY NMR spectrum of the peptide Ac-TWflpF-NH_2 at 277 K. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O.

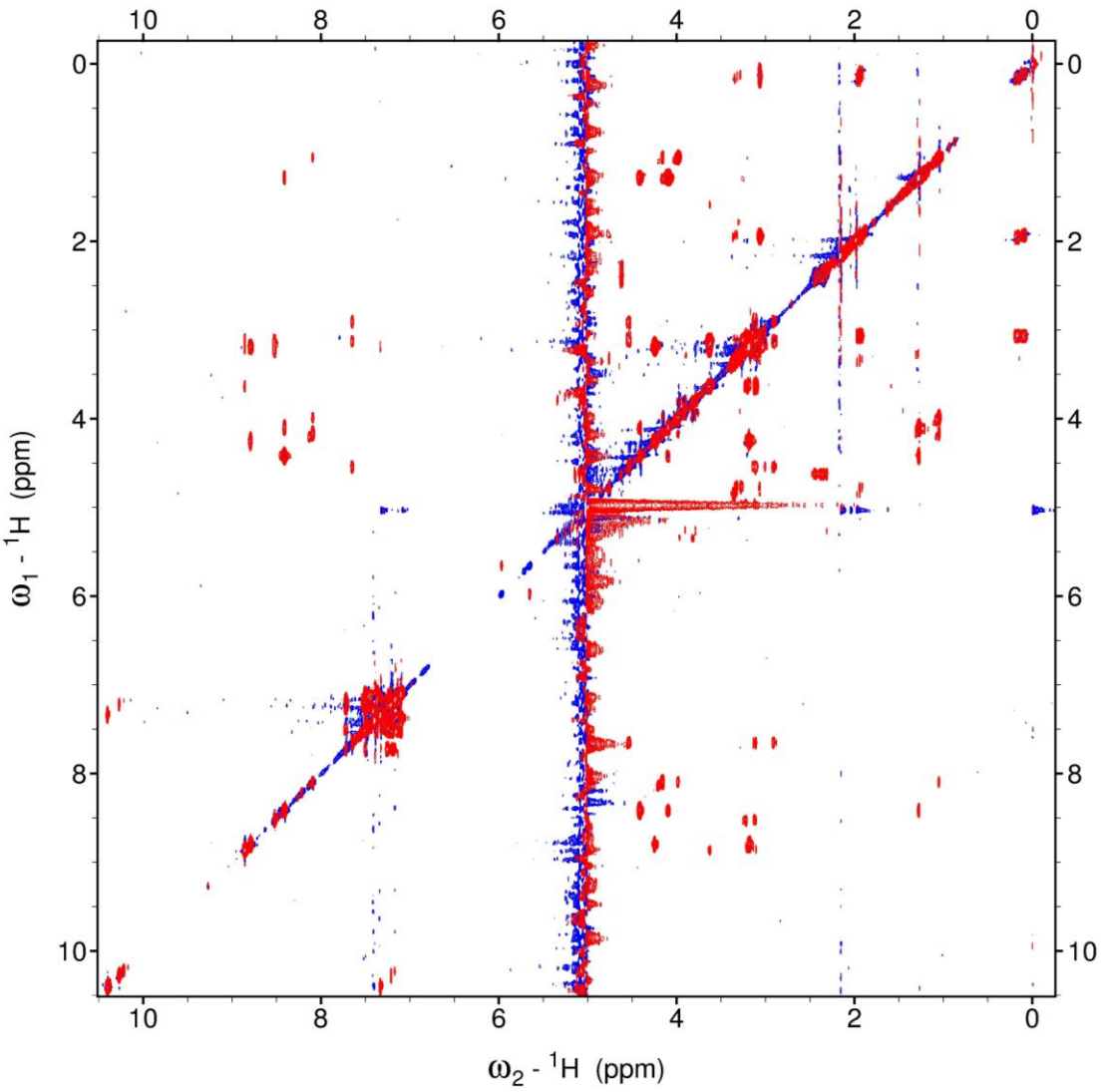


Figure 3.50 Superposition of the TOCSY and ROESY spectra of the designed peptide Ac-TWflpF-NH_2 at 277 K

Superposition of the ^{1}H-^{1}H TOCSY NMR spectrum (red) and the ROESY spectrum (blue) of the peptide Ac-TWflpF-NH_2 at 277 K. The sample contained 5 mM phosphate and 25 mM NaCl, and was measured in 90% H_2O/10% D_2O.

Table 3.7. ^{1}H NMR resonance assignments for the peptide Ac-TWflpF-NH_2

Resonance assignments were obtained from the ^{1}H NMR and TOCSY spectra at the indicated temperatures. The first table (a) shows key, non-prolyl resonance assignments, and the second table (b) shows 4*S*-fluoroproline resonance assignments. n.d. indicates "not determined" due to spectral overlap.

a

	Hα		**Hβ**		**H_N**			
	trans	**cis**	**trans**	**cis**	**trans**		**cis**	
298 K	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**$^{3}J_{\alpha N}$, Hz**	**δ, ppm**	**$^{3}J_{\alpha N}$, Hz**
Thr	4.19	4.42	4.02	4.13	7.97	8.2	8.23	8.8
Trp	4.95	3.82	3.21, 3.11	3.24, 3.14	8.36	6.9	8.67	3.4
Phe	4.55	4.28	3.13, 2.87	3.18	7.35	7.7	8.62	7.1
277 K	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**$^{3}J_{\alpha N}$, Hz**	**δ, ppm**	**$^{3}J_{\alpha N}$, Hz**
Thr	4.16	4.42	3.99	4.10	8.09	7.9	8.41	9.0
Trp	4.94	3.81	3.23, 3.12	3.63, 3.12	8.52	7.4	8.86	2.2
Phe	4.54	3.19	3.12, 2.90	3.19	7.65	7.0	8.80	7.4

b

	Hα		**Hβ**		**Hγ**		**Hδ**	
	trans	**cis**	**trans**	**cis**	**trans**	**cis**	**trans**	**cis**
flp	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**	**δ, ppm**
298 K	4.61	3.17	2.40, 2.30	1.96, 0.27	5.30	4.94	3.91, 3.74	n.d.
277 K	4.63	3.06	2.45, 2.32	1.94, 0.13	5.34	4.77	3.97, 3.81	3.35

By utilizing the design principles that were established through this work and others,[90, 91, 93, 373, 374] we synthesized and designed a tetrapeptide that had a higher population of cis-proline over trans-proline at 298 K (72% cis population, $K_{trans/cis}$ = 0.39). The $^{3}J_{\alpha N}$ coupling constant can be used to approximate the dihedral angle ϕ using a parameterized Karplus equation for proteins.[377] The measured Trp $^{3}J_{\alpha N}$ of 3.4 Hz corresponded to a ϕ torsion angle of approximately –54° at room temperature, and –41° at 277 K, nearly an idealized ϕ torsion angle for a type VIa1 β-turn.[90, 285, 306] The Trp $^{3}J_{\alpha N}$ observed in Ac-TWflpF-NH_2 is even smaller than Ac-TWflpN-NH_2 (3.4 Hz vs. 4.2 Hz at 298 K),[91] suggesting enhancement of the ordering and structure of the type VIa1 β-turn.

Between the cis and trans conformations in Ac-TWflpF-NH_2, the amide proton resonances for phenylalanine were significantly different ($\Delta\delta$ < 1.1 ppm). The trans amide proton for phenylalanine was significantly shifted upfield. This upfield chemical shift in the trans amide proton was consistent with other aromatic-proline-aromatic motifs previously reported.[90, 289] Potentially, the Phe amide proton may participate in roles that stabilize the trans conformation, such as interacting with the tryptophan aromatic ring.[90]

An upfield chemical shift of more than 1.4 ppm was observed for the 4*S*-fluoroproline Hα in the cis conformation over the trans conformation, at both 298 K and 277 K. This upfield chemical shift of the cis-prolyl Hα was also observed in the model tetrapeptides Ac-TXPN-NH_2, and suggests an interaction with π-orbitals from one of the aromatic rings. The flp Hβ resonances between the cis and trans conformations were also significantly different. The chemical shifts for the diastereotopic flp Hβ resonances in the trans conformation were similar ($\Delta\delta$ < 0.2

ppm), at 2.40 ppm and 2.30 ppm. In contrast, the diastereotopic flp Hβ resonances in the cis conformation were highly divergent and shifted upfield, separated by more than 1.0 ppm (1.96 ppm and 0.27 ppm). These chemical shifts suggest that both the Hβ and Hα in 4*S*-fluoroproline are involved in the aromatic-cis-proline interaction, potentially through an interaction with an aromatic ring. ROESY spectra were obtained for the peptide Ac-TWflpF-NH_2 in order to identify which aromatic ring was interacting with flp Hα and Hβ. Although the ROESY spectrum did not indicate which aromatic ring was interacting with the Hβ and Hα in 4*S*-fluoroproline in the peptide Ac-TWflpF-NH_2, prior data on the peptide Ac-TWflpN-NH_2 indicated an interaction with the tryptophan aromatic ring.[91]

However, the designed peptide Ac-TWflpF-NH_2 had substantially greater hydrophobic surface area than the model tetrapeptides Ac-TXPN-NH_2, and the cis-conformation could potentially be stabilized by a hydrophobic effect rather than a prolyl C–H/π aromatic interaction. In this regard, the upfield chemical shifts in the prolyl Hα and Hβ could potentially be an artifact of a hydrophobic interaction (where an aromatic ring still shields these protons, but is not a stabilizing interaction), rather than contributing to the stability of the cis-proline conformation. In order to definitively identify the contributions of hydrophobic interactions, backbone hydrogen bonding interactions, and C–H/π interactions to the overall stability of the aromatic-cis-proline motif, the thermodynamic parameters of these model peptides would need to be determined.

3.2.5 Thermodynamics of stability of the aromatic-cis-proline motif in model tetrapeptides

The enhanced stability of the aromatic-cis-proline motif in organic solvents suggested that the aromatic-proline interaction was not primarily driven by a hydrophobic effect. Interactions that are stabilized by a classical hydrophobic effect are entropically driven, while hydrogen-bond-like interactions are driven by a greater enthalphic component.[203, 378] Therefore, determination of the enthalphic and entropic contributions in the aromatic-proline interaction can establish the fundamental nature of the noncovalent aromatic interactions.

van′t Hoff analysis can be applied to measure the enthalpic and entropic contributions to an interaction.[378] The equilibrium for an interaction is measured with respect to temperature, and the data are correlated on a van′t Hoff plot. The van′t Hoff plot correlates ln$K$ with inverse temperature (1/T), and a linear fit of the data to the van′t Hoff equation determines the enthalpy (ΔH) and entropy (ΔS) of the interaction, assuming a constant heat capacity. The van′t Hoff equation is derived from free energy relationships, and applied here for cis-trans isomerism of proline in Equation 6:

$$\Delta G = -RT\ln K_{eq}$$
$$\Delta G = \Delta H - T\Delta S$$
$$-RT\ln K_{eq} = \Delta H - T\Delta S$$
$$\ln K_{eq} = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$
$$\ln(K_{trans/cis}) = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \qquad (6)$$

where $K_{trans/cis}$ represents the equilibrium between the cis and trans conformations of the model peptide; ΔH and ΔS represent the enthalpy and entropy of the interaction,

respectively; *T* represents a measured temperature, and *R* represents the universal gas constant.

In this study, the temperature dependence of the $K_{trans/cis}$ within a selection of model tetrapeptides was examined. The selected peptides represented Ac-TXPN-NH_2 peptides containing non-aromatic and electron-rich and electron-poor aromatic amino acids. In addition, peptides containing 4*S*-fluoroproline were included to examine the thermodynamics of stabilized aromatic-cis-proline motifs (Ac-TWflpN-NH_2 and Ac-TWflpF-NH_2). The van′t Hoff plots and summary of derived data are shown in Figure 3.51 and Table 3.8.

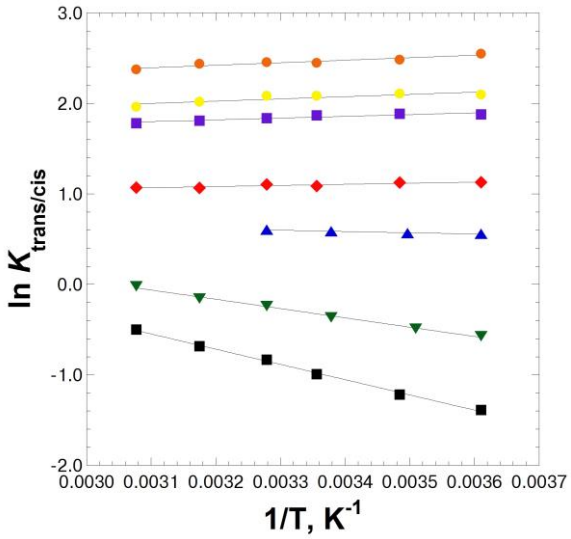


Figure 3.51 van′t Hoff plots for selected peptides
The temperature dependence of the equilibrium $K_{trans/cis}$, measured via NMR, with respect to temperature in 90% H_2O/10% D_2O for selected peptides: Ac-TAPN-NH_2 (orange circles), Ac-TChaPN-NH_2 (yellow circles), Ac-T(4-NO_2-Phe)PN-NH_2 (purple squares), Ac-TYPN-NH_2 (red diamonds), Ac-TWPN-NH_2 (blue triangles),[91] Ac-TWflpN-NH_2 (green triangles),[91] and Ac-TWflpF-NH_2 (black squares). The data represent an average of at least 2 independent trials at each temperature. Error bars indicate standard error.

Table 3.8. Thermodynamic parameters for stability of cis-proline conformation in selected model peptides

The data shown were derived from the van′t Hoff analyses shown in Figure 3.51, based on a linear fit to the temperature dependence of the $K_{trans/cis}$ for the given peptides. The data were derived from averages of at least 2 independent trials at each temperature. The data are shown as $K_{cis/trans}$, where greater values indicate higher populations of cis conformation (the inverse of $K_{trans/cis}$).
$\Delta G_{cis/trans}$[a], was calculated as $\Delta G_{cis/trans,\ 277\ K} = -R\mathrm{Tln}(K_{cis/trans,\ 277\ K})$
$\Delta G_{cis/trans}$[b], was calculated as $\Delta G_{cis/trans,\ 277\ K} = \Delta H_{cis/trans} - (277\ K)\cdot\Delta S_{cis/trans}$
[c]Data obtained at 298 K.[90]
Data for the peptides Ac-AYPN-NH_2 and Ac-YYPN-NH_2 were reported previously and are shown for comparison.[90] Data for the peptides Ac-TWPN-NH_2 and Ac-TWflpN-NH_2 were reported previously.[91]

Peptide	$K_{cis/trans}$ at 277 K	% cis	$\Delta G_{cis/trans}$,[a] kcal mol^{-1}	$\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}	$\Delta G_{cis/trans}$,[b] kcal mol^{-1}
YYPN[90]	0.15[c]	13%[c]	+1.12[c]	+1.01	–0.4	+1.13[c]
AYPN[90]	0.27[c]	21%[c]	+0.76[c]	+0.22	–1.8	+0.72[c]
TAPN	0.078	7%	+1.40	+0.55 ±0.093	–3.1 ±0.309	+1.40
TChaPN	0.13	11%	+1.14	+0.49 ±0.139	–2.4 ±0.465	+1.17
T(4-NO_2-Phe)PN	0.15	13%	+1.03	+0.38 ±0.075	–2.4 ±0.251	+1.04
TYPN	0.32	24%	+0.62	+0.24 ±0.054	–1.4 ±0.178	+0.62
TWPN[91]	0.57	36%	+0.31	–0.28 ±0.044	–2.1 ±0.150	+0.31
TWflpN[91]	1.8	64%	–0.31	–2.04 ±0.079	–6.2 ±0.263	–0.32
TWflpF	4.0	80%	–0.76	–3.35 ±0.084	–9.3 ±0.280	–0.85

The enthalpy and entropy of the aromatic-cis-proline interaction in water for each selected peptide was calculated from the linear van′t Hoff plot (Figure 3.51). An increasingly favorable enthalpic component was observed with increasing cis-population of the selected peptides. Prior studies examining the enthalpic and entropic components of cis-trans isomerism in models containing proline have also shown favorable enthalpy for interaction, such as the peptides GYPG (ΔH = –0.65 kcal mol^{-1}), GWPG (ΔH = –0.75 kcal mol^{-1}), Ac-FP-OMe (ΔH = –0.25 kcal mol^{-1}), or Ac-Pro-OMe (ΔH = –1.04 kcal mol^{-1}).[288, 372, 379] Peptides containing 4*S*-fluoroproline had

the greatest stability and the most favorable enthalpy for the cis conformation. The highly stabilized designed peptide Ac-TWflpF-NH_2 exhibited a high population of cis-proline and a favorable enthalpy for the interaction (ΔH = –3.35 kcal mol^{-1}). In contrast, the calculated enthalpy for peptides containing alanine or cyclohexylalanine were +0.55 kcal mol^{-1} and +0.49 kcal mol^{-1}, respectively. In general, the enthalpy was increasingly favorable with greater cis-population, and the entropy became more unfavorable. This increase in unfavorable entropy can potentially be attributed to a weak enthalpy-entropy compensation,[380] where the higher favorable enthalpy for a rigid conformation is offset by the loss of degrees of freedom and a greater entropic cost.

The ΔG for the cis-conformation was directly correlated with the $K_{cis/trans}$ of the aromatic-cis-proline interaction. The ΔG for this interaction can also be calculated from the measured ΔH and ΔS, which should agree with the ΔG calculated from the $K_{cis/trans}$ at a given temperature. The $\Delta G_{cis/trans}$ at 277 K calculated via $-RT\ln K_{cis/trans}$ and via ($\Delta H - T\Delta S$) are both shown in Table 3.8; the calculated $\Delta G_{cis/trans}$ are nearly identical from both calculations for all peptides, which provides excellent validation of the measured values of ΔH and ΔS. These thermodynamic data directly establish that the aromatic-cis-proline interaction in these tetrapeptides is not stabilized by an entropy-driven, classical hydrophobic effect. An enthalpically driven aromatic-cis-proline interaction is consistent with the electrostatic nature of this interaction, corroborating the previously discussed Hammett correlations and solvent effects.

3.2.6 NMR characterization of minimized aromatic-cis-proline motifs

The model tetrapeptides Ac-TXPN-NH_2 were previously established as short peptides with a high population for cis-proline conformation, when X was an aromatic

amino acid.[90, 91] These model tetrapeptides served as a basis for understanding the underlying stabilizing features of aromatic-cis-proline motifs. Increasingly electron-donating aromatic substituents on the aromatic amino acid increased the population of cis amide bond conformation. With increasing cis population, a significant upfield chemical shift in the cis-proline Hα was observed. Based on the van′t Hoff analysis of these model tetrapeptides, it was established that the aromatic-cis-proline interaction was enthalpically driven, suggesting an electronic aromatic interaction. In this model tetrapeptide, the stability of the cis-proline conformation may also be due to an *i*,*i*+3 backbone hydrogen bond between Thr C=O and Asn N–H, which is a defining feature in type VIa1 β-turns involving cis-proline.[285] The upfield chemical shift in the amide proton in phenylalanine in the peptide Ac-TWflpF-NH_2 suggests interactions outside of the central C–H/π aromatic interactions. These additional interactions may influence or compete with the central aromatic-cis-proline C–H/π interaction. In order to study the fundamental nature of the core aromatic-cis-proline interaction, additional stabilizing effects from other amino acids would need to be excluded, and the central dipeptide motif would need to be more closely examined.

In order to address these questions, minimal aromatic-proline motifs were synthesized: Ac-XPro-OMe and Ac-XPro-NHMe, where X was either Trp or Ala, shown in Figure 3.52. The $K_{trans/cis}$ of these dipeptides were measured via NMR to directly examine the structural influence of the tryptophan ring on proline cis-trans isomerism. The comparisons of the dipeptides containing a C-terminal methyl amide or methyl ester allowed for comparison of the aromatic-cis-proline interaction in the presence or absence of the *i*,*i*+3 backbone hydrogen bond (Figure 3.52a). van′t Hoff analysis was conducted on the four dipeptides in water, methanol, and chloroform

(Figures 3.53 and 3.54), in order to identify the enthalpic and entropic contributions to the stability of the cis conformation. The enthalpy and entropy of the aromatic-cis-proline interaction was expected to vary as a function of solvent, depending on the relative hydrophobic contributions and the strength of hydrogen bonds as a function of solvent. The summary of derived ΔH and ΔS parameters for the series of dipeptides in various solvents are shown in Table 3.9. The ΔH and ΔS parameters for the tetrapeptides previously discussed are included in Table 3.9 for comparison.

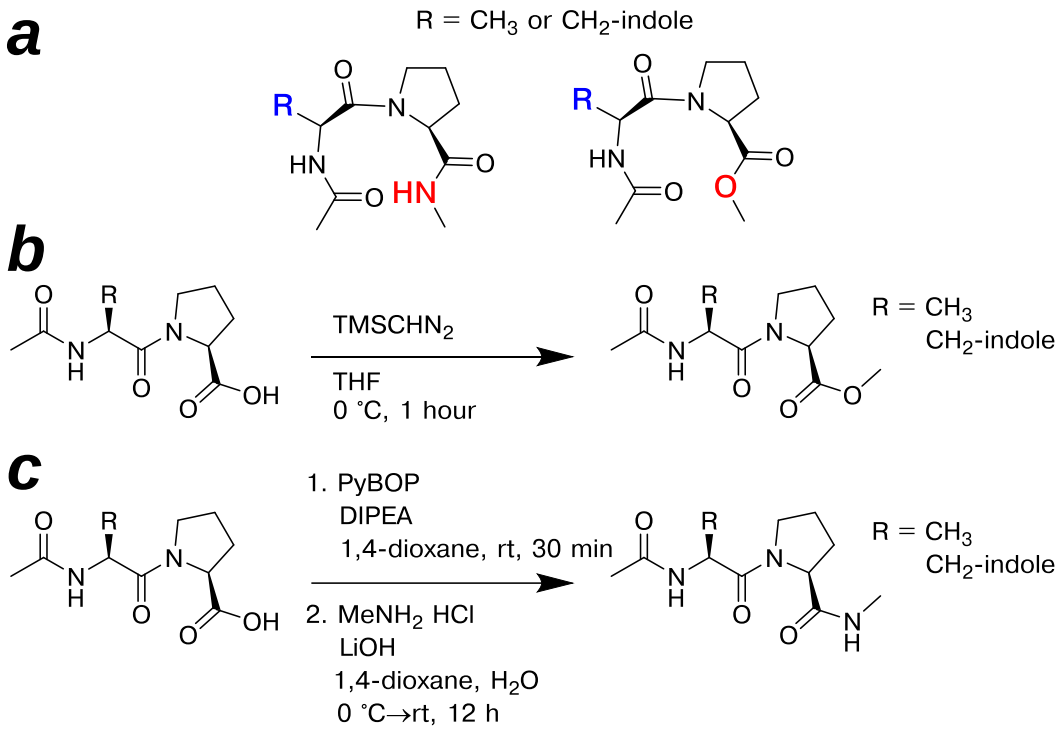


Figure 3.52 Synthesis of minimized dipeptides Ac-AP-XMe and Ac-WP-XMe (where X = O or NH)

(a) In order to examine the fundamental nature of the aromatic-cis-proline interaction, and to determine the role of the *i*,*i*+3 backbone hydrogen bond in stabilizing the cis-proline conformation, a series of synthesized dipeptides were modified in solution using these protocols: (b) modification of the synthesized peptides Ac-AlaPro-OH or Ac-TrpPro-OH to generate the peptides containing a C-terminal methyl ester; (c) modification of the synthesized peptides Ac-AlaPro-OH or Ac-TrpPro-OH to generate the peptides containing a C-terminal methyl amide. The dipeptides containing a methyl amide are capable of forming a type VIa1 β-turn with a backbone *i*,*i*+3 hydrogen bond, while the dipeptides containing methyl esters are unable to form this intramolecular hydrogen bond.

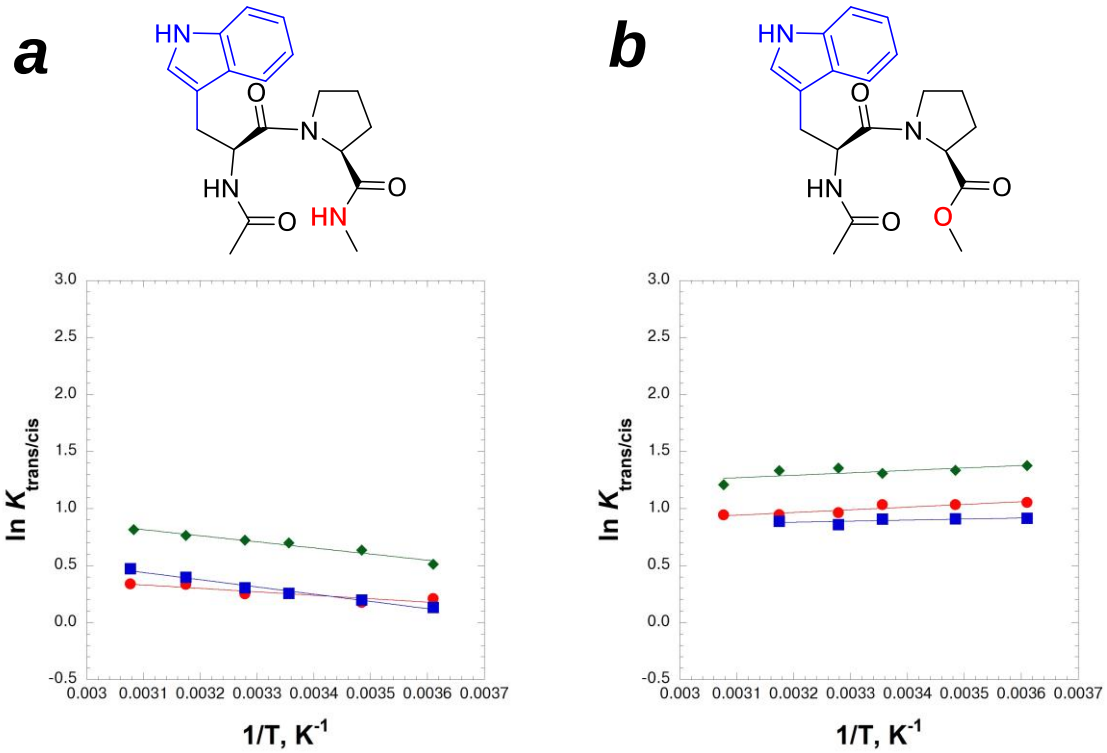


Figure 3.53 van′t Hoff plots for Ac-WP-XMe in various solvents (X = O or NH) $K_{trans/cis}$ was measured for the Ac-WP-XMe dipeptides via NMR as a function of temperature in various solvents: 10% D_2O/90% H_2O (red circles), 10% MeOD-d_4/90% MeOH-d_3 (blue squares), and $CDCl_3$ (green diamonds). Error bars indicate standard error of the mean from at least two independent trials. (a) van′t Hoff plot for Ac-WP-NHMe and (b) Ac-WP-OMe. This study examined the enthalpic and entropic contributions to stabilize the cis-proline conformation in the presence and absence of an intramolecular hydrogen bond (with an intramolecular aromatic-proline interaction).

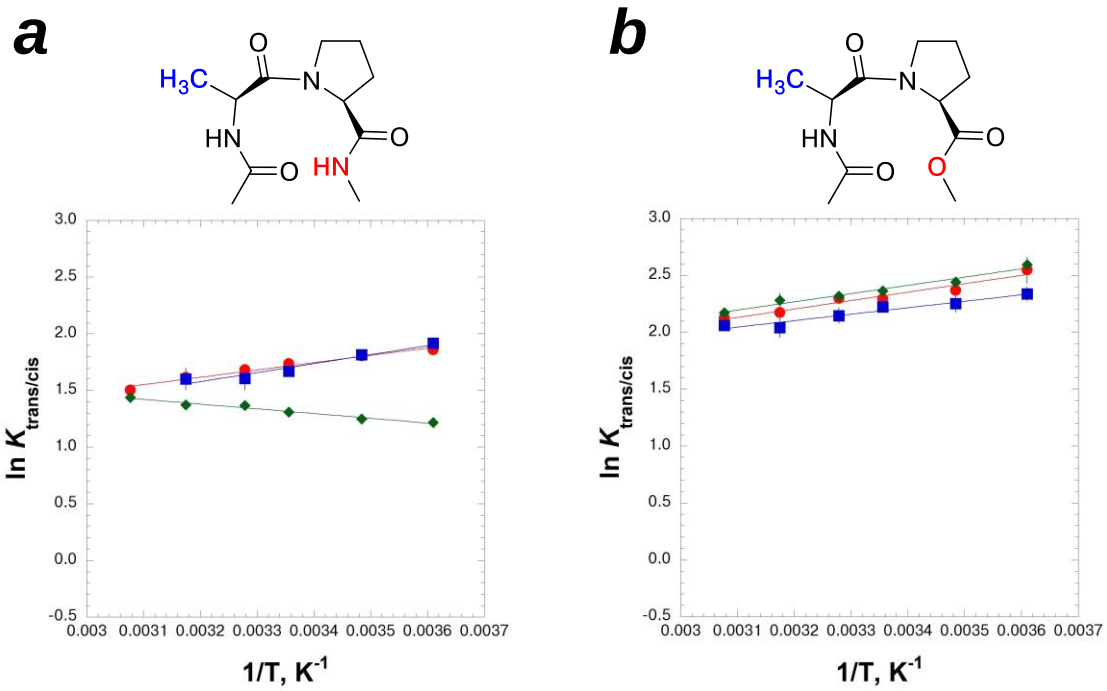


Figure 3.54 van′t Hoff plots for Ac-AP-XMe in various solvents (X = O or NH) $K_{trans/cis}$ was measured for Ac-AP-XMe dipeptides via NMR as a function of temperature in various solvents: 10% D_2O/90% H_2O (red circles), 10% MeOD-d_4/90% MeOH-d_3 (blue squares), and $CDCl_3$ (green diamonds). Error bars indicate standard error of the mean from at least two independent trials. (a) van′t Hoff plot for Ac-AP-NHMe and (b) Ac-AP-OMe. This study examined the enthalpic and entropic contributions to stabilize the cis-proline conformation in the presence and absence of an intramolecular hydrogen bond (without an intramolecular aromatic-proline interaction).

Table 3.9. Summary of van′t Hoff analysis of dipeptides Ac-AP-XMe and Ac-WP-XMe (X = O or NH) in water, methanol, and chloroform

The data shown were derived from the van′t Hoff analyses shown in Figures 3.53 and 3.54, based on a linear fit to the temperature dependence of the $\ln(K_{trans/cis})$ for the given peptides. The data are shown as $K_{cis/trans}$, where greater values indicate higher populations of cis conformation (the inverse of $K_{trans/cis}$).
$\Delta G_{cis/trans}{}^{a}$, was calculated as $\Delta G_{cis/trans,\ 277\ K} = -R\mathrm{T}\ln(K_{cis/trans,\ 277\ K})$

	Peptide	$K_{cis/trans}$ 277 K	% cis	$\Delta G_{cis/trans}$,[a] kcal mol^{-1}	$\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}
Water	TAPN	0.078	7%	+1.40	+0.55 ±0.093	-3.05 ±0.309
	TChaPN	0.13	11%	+1.14	+0.49 ±0.075	-2.45 ±0.251
	T(4-NO_2-Phe)PN	0.15	13%	+1.03	+0.38 ±0.075	-2.38 ±0.251
	TYPN	0.32	24%	+0.62	+0.24 ±0.054	-1.39 ±0.178
	TWPN	0.57	36%	+0.31	-0.28 ±0.044	-2.11 ±0.150
	TWflpN	1.8	64%	-0.31	-2.04 ±0.079	-6.20 ±0.263
	TWflpF	4.0	80%	-0.76	-3.35 ±0.084	-9.29 ±0.280
	Ac-AP-OMe	0.077	7%	+1.41	+1.47 ±0.174	+0.33 ±0.580
	Ac-WP-OMe	0.35	26%	+0.58	+0.47 ±0.093	-0.42 ±0.310
	Ac-AP-NHMe	0.16	13%	+1.02	+1.28 ±0.114	+0.89 ±0.380
	Ac-WP-NHMe	0.81	45%	+0.11	-0.60 ±0.141	-2.52 ±0.470
Methanol	Ac-AP-OMe	0.10	9%	+1.29	+1.13 ±0.146	-0.57 ±0.487
	Ac-WP-OMe	0.40	29%	+0.50	+0.18 ±0.112	-1.16 ±0.378
	Ac-AP-NHMe	0.15	13%	+1.05	+1.58 ±0.223	+1.92 ±0.754
	Ac-WP-NHMe	0.88	47%	+0.07	-1.26 ±0.092	-4.79 ±0.307
$CDCl_3$	Ac-AP-OMe	0.075	7%	+1.43	+1.44 ±0.131	+0.10 ±0.436
	Ac-WP-OMe	0.25	20%	+0.76	+0.43 ±0.202	-1.20 ±0.672
	Ac-AP-NHMe	0.30	23%	+0.67	-0.83 ±0.074	-5.40 ±0.246
	Ac-WP-NHMe	0.60	37%	+0.28	-1.06 ±0.106	-4.92 ±0.354

The dipeptides containing the C-terminal methyl amide modification had consistently greater populations of cis-proline over the peptides with C-terminal methyl esters, indicating, as expected, that the *i*,*i*+3 backbone hydrogen bond significantly stabilizes the type VIa1 β-turn (by definition[285]). In the organic solvent,

electrostatic interactions are enhanced because competing interactions with bulk solvent are reduced (in comparison to water). An enhancement of the enthalpic contribution for the cis-conformation was observed in chloroform, consistent with stabilizing the *i*,*i*+3 hydrogen bond in methyl amide dipeptides in chloroform. This stabilizing effect due to the backbone hydrogen bond is clearly observable in the dipeptide Ac-AP-NHMe, where there are few other driving interactions that could stabilize the cis conformation (note that favorable enthalpy with this dipeptide is *only* observed in chloroform).

In general, the dipeptides containing tryptophan had consistently higher $K_{cis/trans}$ values over peptides containing alanine, with more favorable enthalpy, indicating that the aromatic interaction plays a significant role in stabilizing the cis conformation. In all solvents examined, the ΔH for cis-proline was more favorable by approximately 1 kcal mol^{-1} for the dipeptides WP over AP, *regardless of the ability to form the* i,i+*3 backbone hydrogen bond*. These thermodynamic parameters across multiple solvents demonstrates that the aromatic-proline interaction *alone* contributes approximately 1 kcal mol^{-1} of stability for the cis conformation. Similar to the model tetrapeptides in water, the dipeptides with higher $K_{cis/trans}$ values also had the most favorable ΔH, often with unfavorable ΔS, regardless of solvent. This unfavorable entropy with more favorable enthalpy of interaction is consistent with the enthalpy-entropy compensation phenomenon observed in tetrapeptides, discussed in section 3.2.5.

Having determined that the nature of the aromatic-cis-proline interaction in minimized dipeptides was enthalpy-driven, we sought to further explore the strength and limitations of this aromatic-cis-proline interaction in minimized dipeptides

containing 4*S*-fluoroproline. The model tetrapeptide studies showed that 4*S*-fluoroproline increased the population of the cis conformation through stereoelectronic control of the proline ring-pucker.[373] We synthesized a series of peptides containing the Trpflp motif with different C- and N-terminal functional groups (Figure 3.55), in order to study the aromatic-cis-proline interaction in highly stabilized, minimized sequences. It was expected that these highly stabilized aromatic-cis-proline structures could potentially be crystallized (discussed in detail in Chapter 3.2.7), given exceptional stability of the cis conformation in different solvents. The $K_{trans/cis}$ and chemical shift for the flp Hα in these Trpflp dipeptides were measured in different solvents, shown in Table 3.10. Comparisons to the dipeptides Ac-XP-OMe and Ac-XP-NHMe and tetrapeptides containing tryptophan are included in Table 3.10.

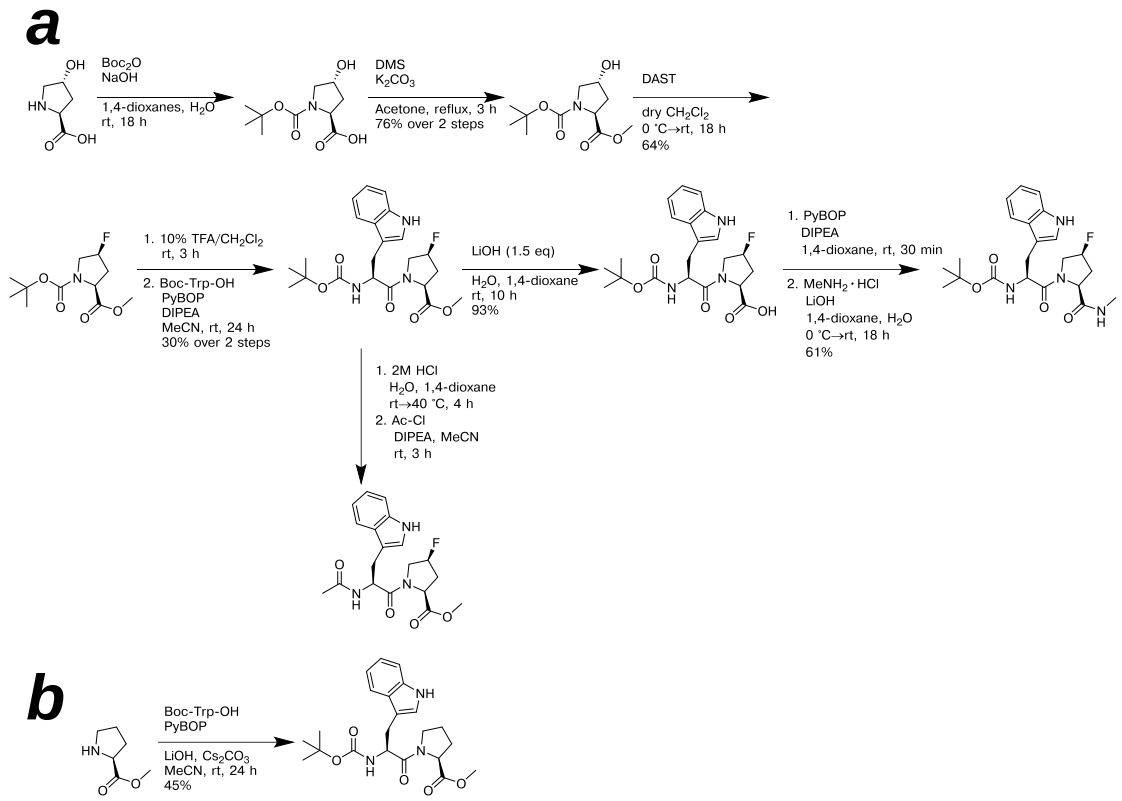


Figure 3.55 Synthesis of minimized Trp-4*S*-fluoroproline dipeptides in solution In order to examine the fundamental nature of the aromatic-cis-proline interaction, a series of minimized dipeptides were synthesized in solution using these synthesis protocols. The dipeptide series included: (a) Boc-Trpflp-OMe, Boc-Trpflp-OH, Boc-Trpflp-NHMe, Ac-Trpflp-OMe, and (b) Boc-TrpPro-OMe.

Table 3.10. NMR characterization of stabilized aromatic-cis-proline motifs in water, methanol, and chloroform

The data were derived from solution NMR data of each peptide in the indicated solvent. Chemical shifts were identified via the TOCSY spectrum. All NMR data were obtained at 298 K.
n.d. indicates "not determined" due to spectral overlap or broad signals. n.s. indicates "not soluble."flp indicates 4*S*-fluoroproline

	90% H_2O/10% D_2O							90% MeOH-d_3/10% MeOD-d_4							$CDCl_3$						
Peptide	$K_{cis/trans}$	$K_{trans/cis}$	Z Hα_{cis}, ppm	Z Hα_{trans}, ppm	$^3J_{\alpha N}X_{cis}$, Hz	$\Delta G_{cis/trans}$, kcal mol^{-1}	% cis	$K_{cis/trans}$	$K_{trans/cis}$	Z Hα_{cis}, ppm	Z Hα_{trans}, ppm	$^3J_{\alpha N}$ X_{cis}, Hz	$\Delta G_{cis/trans}$, kcal mol^{-1}	% cis	$K_{cis/trans}$	$K_{trans/cis}$	Z Hα_{cis}, ppm	Z Hα_{trans}, ppm	$^3J_{\alpha N}$ X_{cis}, Hz	$\Delta G_{cis/trans}$, kcal mol^{-1}	% cis
Ac-TWPN-NH_2	0.50	2.0	3.45	4.43	5.6	0.41	33%	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Ac-TWflpN-NH_2	1.4	0.71	3.50	4.67	4.1	-0.20	58%	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Ac-TWflpF-NH_2	2.6	0.39	3.17	4.61	3.4	-0.56	72%	3.2	0.31	3.28	n.d.	2.9	-0.69	76%	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Ac-AP-OMe	0.10	9.9	4.46	4.47	n.d.	1.36	9%	0.11	9.2	4.63	4.45	n.d.	1.32	10%	0.094	10.6	4.46	4.53	n.d.	1.40	9%
Ac-AP-NHMe	0.18	5.7	4.51	4.34	n.d.	1.03	15%	0.19	5.3	n.d.	n.d.	n.d.	0.99	16%	0.26	3.9	4.53	4.29	n.d.	0.80	20%
Ac-WP-OMe	0.36	2.8	3.65	4.43	7.8	0.61	26%	0.40	2.5	3.51	4.42	7.8	0.54	29%	0.27	3.7	3.43	4.49	7.0	0.77	21%
Ac-WP-NHMe	0.78	1.3	3.32	4.33	n.d.	0.15	44%	0.78	1.3	3.31	4.34	4.1	0.15	44%	0.50	2.0	3.41	4.48	n.d.	0.41	33%
Ac-Wflp-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.74	1.4	3.61	4.71	8.2	0.18	43%	0.82	1.2	3.42	4.81	7.6	0.12	45%
Boc-WP-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.38	2.6	3.81	4.43	8.5	0.57	28%	0.29	3.4	3.45	4.50	8.7	0.72	23%
Boc-Wflp-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.67	1.5	3.60	4.74	8.6	0.24	40%	0.84	1.2	3.42	4.80	8.3	0.10	46%
Boc-Wflp-OH	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.48	2.1	3.45	4.66	n.d.	0.43	32%	0.67	1.5	3.34	4.78	7.9	0.24	40%
Boc-Wflp-NHMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	1.9	0.52	3.29	4.54	n.d.	-0.38	66%	1.8	0.56	3.25	4.68	6.1	-0.35	64%

In general, the dipeptides containing 4*S*-fluoroproline had lower $K_{trans/cis}$ values than comparable peptides containing proline, consistent with observations in the model tetrapeptide series and prior work with 4*S*-fluoroprolines.[91, 373, 374] Boc-Trpflp-NHMe dipeptide showed a greater population for the cis conformation over trans, which was the only dipeptide to attain this degree of stability (65% cis in chloroform and methanol). This highly stabilized conformation is due to a combination of a type VIa1 β-turn intramolecular *i*,*i*+3 hydrogen bond, an intramolecular aromatic-proline interaction with an electron-rich indole ring, and the promotion of the cis conformation by stereoelectronic control of the prolyl ring pucker. While the hydrophobic surface area of the indole in tryptophan is greater than the benzene ring in phenylalanine, the substantial population of cis-proline in chloroform in all dipeptides (excluding Ac-AP-OMe) indicated that this aromatic-cis-proline interaction was not primarily driven by a hydrophobic interaction, consistent with our prior data in tetrapeptides and canonical dipeptides. A substantial upfield chemical shift in the cis-flp Hα was observed with increasing cis populations. The the dipeptide Boc-Trpflp-NHMe, the flp Hα chemical shift difference between the cis and trans conformations was greater than 1.2 ppm at room temperature, independent of solvent. Furthermore, the NMR data for the dipeptides showed that the individual cis-prolyl Hβ protons exhibit divergent chemical environment, and also demonstrated substantial upfield chemical shifts (Table 3.11).

Table 3.11. Upfield chemical shifts in prolyl Hβ protons in dipeptides in water, methanol, and chloroform

The data were derived from solution NMR data of each peptide in the indicated solvent. Chemical shifts were identified via TOCSY spectrum. All NMR data were taken at 298 K. “Pro Hβ1” and “Pro Hβ2” can refer to proline or 4*S*-fluoroproline chemical shifts, respectively.
n.d. indicates “not determined” due to spectral overlap or broad signals.
n.s. indicates “not soluble”

	90% H_2O/10% D_2O							90% MeOH-d_3/10% MeOD-d_4							$CDCl_3$						
Peptide	$K_{cis/trans}$	**Pro Hβ1 $_{trans}$, ppm**	**Pro Hβ2 $_{trans}$, ppm**	**Pro Hβ1 $_{cis}$, ppm**	**Pro Hβ2 $_{cis}$, ppm**	**Δ Pro Hβ$_{trans}$, ppm**	**Δ Pro Hβ$_{cis}$, ppm**	$K_{cis/trans}$	**Pro Hβ1 $_{trans}$, ppm**	**Pro Hβ2 $_{trans}$, ppm**	**Pro Hβ1 $_{cis}$, ppm**	**Pro Hβ2 $_{cis}$, ppm**	**Δ Pro Hβ$_{trans}$, ppm**	**Δ Pro Hβ$_{cis}$, ppm**	$K_{cis/trans}$	**Pro Hβ1 $_{trans}$, ppm**	**Pro Hβ2 $_{trans}$, ppm**	**Pro Hβ1 $_{cis}$, ppm**	**Pro Hβ2 $_{cis}$, ppm**	**Δ Pro Hβ$_{trans}$, ppm**	**Δ Pro Hβ$_{cis}$, ppm**
Ac-AP-OMe	0.10	2.29	2.02	2.22	1.90	0.27	0.32	0.11	2.03	1.96	2.29	2.17	0.07	0.12	0.094	2.24	2.03	1.24	0.84	0.21	0.40
Ac-AP-NHMe	0.18	2.04	1.90	1.98	1.83	0.14	0.15	0.19	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.34	2.16	1.91	1.92	1.79	0.25	0.13
Ac-WP-OMe	0.36	2.22	1.89	1.44	1.15	0.33	0.29	0.40	2.18	1.87	1.34	0.97	0.31	0.37	0.27	2.17	1.88	1.33	0.88	0.29	0.45
Ac-WP-NHMe	0.78	2.18	1.87	1.47	0.92	0.31	0.55	0.78	2.06	1.83	1.30	0.51	0.23	0.79	0.50	2.24	1.84	1.44	0.54	0.40	0.90
Ac-Wflp-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.74	2.39	2.34	2.03	0.87	0.05	1.16	0.82	2.49	2.26	2.05	0.47	0.23	1.58
Boc-WP-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.38	2.18	1.89	1.33	1.00	0.29	0.33	0.29	2.14	1.88	1.31	0.87	0.26	0.44
Boc-Wflp-OMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.67	2.41	2.37	2.02	0.84	0.04	1.18	0.84	2.44	2.22	2.04	0.50	0.22	1.54
Boc-Wflp-OH	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.48	2.46	2.37	2.11	0.85	0.09	1.26	0.67	2.56	2.24	2.07	0.51	0.32	1.56
Boc-Wflp-NHMe	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	1.9	2.24	2.16	2.10	0.88	0.08	1.22	1.8	2.65	2.19	2.04	0.16	0.46	1.88

Similar to the prolyl Hα protons, the prolyl Hβ protons were substantially shifted upfield with increasing cis-population, indicating that the Hβ protons in the proline ring were also subject to a shielding effect as the cis conformation became more stable. In addition, the individual prolyl Hβ protons exhibited divergent chemical shifts, indicative that each Hβ proton in proline was subject to different chemical environments. In all Trpflp dipeptides, the Δδ was more than 1.0 ppm between the two diastereotopic Hβ protons in 4*S*-fluoroproline. The most divergent chemical shifts for flpHβ was observed for the peptide Boc-Trpflp-NHMe in chloroform, with a Δδ of 1.88 ppm. In contrast, the Δδ between the two Hβ protons was less than 0.5 ppm in the trans conformation for all dipeptides examined.

The non-degeneracy of the diastereotopic prolyl Hβ protons supports the hypothesis that the aromatic ring stacks over one face of the prolyl ring, where one prolyl Hβ is more shielded than the other. Based on this model, the greater population of cis conformation in all peptides containing tryptophan is potentially due to an increased *electronic* surface area of interaction, rather than an increased *hydrophobic* surface area. With a larger area for aromatic interaction, multiple prolyl C–H bonds can potentially interact with the tryptophan ring, rather than only one C–Hα. The more favorable enthalpies of interaction that were observed in the tetrapeptides and dipeptides containing tryptophan may be due to one or multiple proline C–H/π aromatic interactions. The non-degeneracy of the cis-Hβ protons in proline could also be partially due to the proximal a fluorine atom, although neither the trans-Hβ protons nor the Hδ protons exhibited this degree of divergent chemical shifts. The NMR data strongly indicate a favorable proline C–H/π aromatic interaction that stabilizes the cis amide bond conformation.

3.2.7 Crystallization of an aromatic-cis-proline motif: Direct observation of a proline C–H/π aromatic interaction

With the NMR characterization of the minimized peptides showing a high stability of the aromatic-cis-proline motifs, broadly independent of solvent, we were confident that the fundamental nature of the proline C–H/π aromatic interaction could be more closely examined within these model peptides. The van′t Hoff analyses indicated that the interaction was enthalpy-driven, consistent with a hydrogen-bond-type interaction. Furthermore, the NMR data for the Trp-cis-flp motifs indicated that the interaction was not due to a backbone hydrogen bond, and that multiple proline C–H bonds were involved in stabilizing the cis conformation. The nature of C–H/π aromatic interactions has been reviewed previously,[105, 106, 190, 191, 260, 290] but the geometry and details of stabilizing components have not been completely understood. Specifically, the results obtained in Chapter 2 suggested that stabilizing aromatic π→ $\sigma^*_{X–H}$ molecular orbital interactions could be general to X–H/π interactions. With the enhanced stability of the Trp-cis-flp motifs, we were encouraged that one or several of these minimal peptides could crystallize using appropriate conditions. Obtaining a crystal structure of these aromatic-cis-proline motifs could provide unique and detailed insights into the nature and geometry of C–H/π aromatic interactions.

Initially, a series of tripeptides based on Trp-(2*S*,4*S*-(*p*-nitrobenzoyl)-hydroxyproline)-X was synthesized. A 2*S*,4*S*-*p*-nitrobenzoyl substituent on proline was known to promote the endo ring pucker (promoting cis conformations), and could potentially aid in crystallization.[381] In this peptide series, the *i*+1 amino acid was varied as phenylalanine, 4-bromophenylalanine, or 4-iodophenylalanine, and contained either a free acid C-terminus or C-terminal amine. The tripeptides that were based on Trp-(2*S*,4*S*-(*p*-nitrobenzoyl)-hydroxyproline)-Phe contained an N-terminal 4-

bromobenzylamide. The use of aryl halides in the peptides can potentially aid in crystal packing interactions through halogen bonding interactions.[382] Unfortunately, we observed limited crystallization from this initial series of peptides. Only one tripeptide from this series formed a diffractable crystal, Ac-Trp(2S,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-bromophenylalanine)-NH_2, from 60% methanol and chloroform, although only a low resolution structure was solved (Figure 3.56).

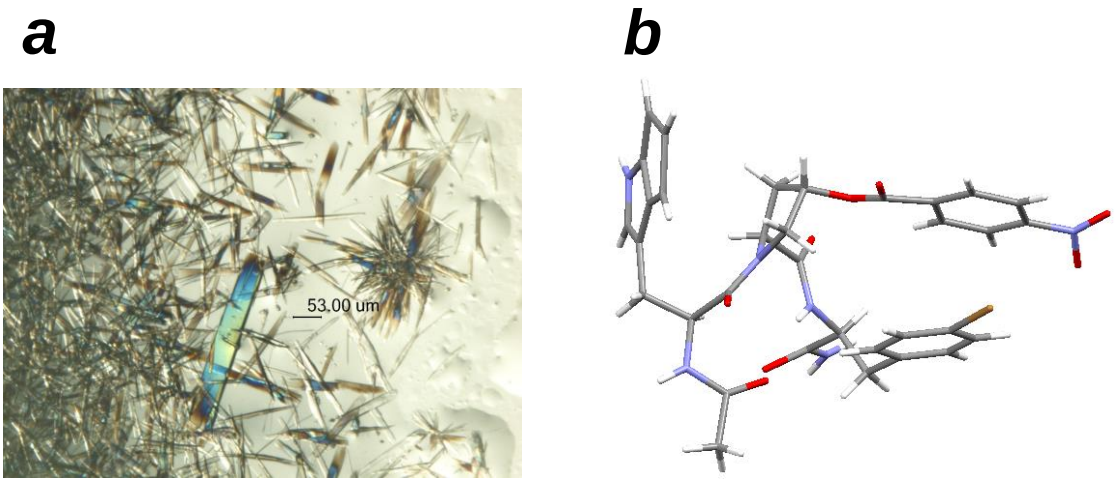


Figure 3.56 X-ray crystal structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-bromophenylalanine)-NH_2
(a) Crystals were obtained via slow evaporation at room temperature over 3 weeks from a solution of 60% methanol in chloroform. (b) Crystal structure of the peptide Ac-Trp(2S,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-bromophenylalanine)-NH_2. The structure was solved to 1.1 Å resolution.

Although the crystal structure did not provide reliable measurements of an aromatic-cis-proline C–H/π interaction, the crystal structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-bromophenylalanine)-NH_2 did provide guidance for designing more stabilized minimized aromatic-cis-proline motifs. The Trp-Pro amide

bond adopted an expected cis conformation (ω = 4.7°), with a typical *i*,*i*+3 backbone hydrogen bond between the acetyl carbonyl and the 4-bromophenylalanine N–H bond, characteristic of type VIa1 β-turns.[90, 91] However, an aromatic-aromatic stacking interaction was observed between the *p*-nitrobenzoate group and 4-bromophenylalanine. This interaction could potentially compete with the proline C–H/π aromatic interaction with tryptophan. Although the *p*-nitrobenzoate was utilized both to promote the cis conformation via stereoelectronic control of the prolyl ring-puckers, and to promote crystallization, this bulky group introduced undesired, competitive aromatic interactions.

In place of a bulky aromatic substituent on hydroxyproline, a series of Trp-flp dipeptides was screened for crystallization. Similar to 2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline, 4*S*-fluoroproline also provides stereoelectronic control over the prolyl ring-pucker and can promote the cis amide bond conformation.[98] The minimal aromatic-cis-proline motifs containing 4*S*-fluoroproline exhibited stabilized cis-conformations in solution as observed via NMR (described in Chapter 3.2.6). In addition, the tetrapeptides containing the Trp-flp motif also had especially high cis-population (Chapter 3.2.4). In fact, all of the dipeptides containing the Trp-flp motif were successfully crystallized, which allowed for detailed examination of multiple perspectives of the proline C–H/π aromatic interaction via x-ray crystallography. The crystal structure for the peptide Ac-Trpflp-OMe is shown in Figure 3.57; the peptide Boc-Trpflp-OMe crystallized from a solution of 50% methanol and water at room temperature (Figure 3.58); the peptide Boc-Trpflp-OH crystallized from an NMR sample in $CDCl_3$, where large crystals were observed within a few hours at room temperature (Figure 3.59); the peptide Boc-TrpPro-OMe crystallized from methanol

over 3 days at room temperature (Figure 3.60); the peptide Boc-Trpflp-NHMe crystallized via slow evaporation from $CDCl_3$ (Figure 3.61). A summary of the data obtained from x-ray crystallography of this series of Trp-flp dipeptides, including measurements from the crystal structure and NMR characterization, are included in Table 3.12.

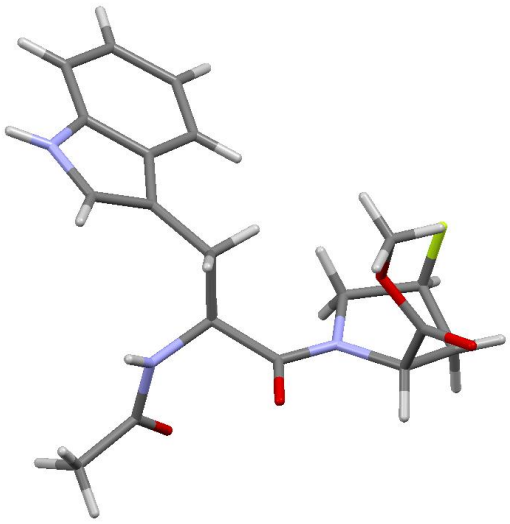


Figure 3.57 X-ray crystal structure of the dipeptide Ac-Trpflp-OMe
The crystal structure of the peptide Boc-Trpflp-OMe was solved to 0.77 Å resolution.

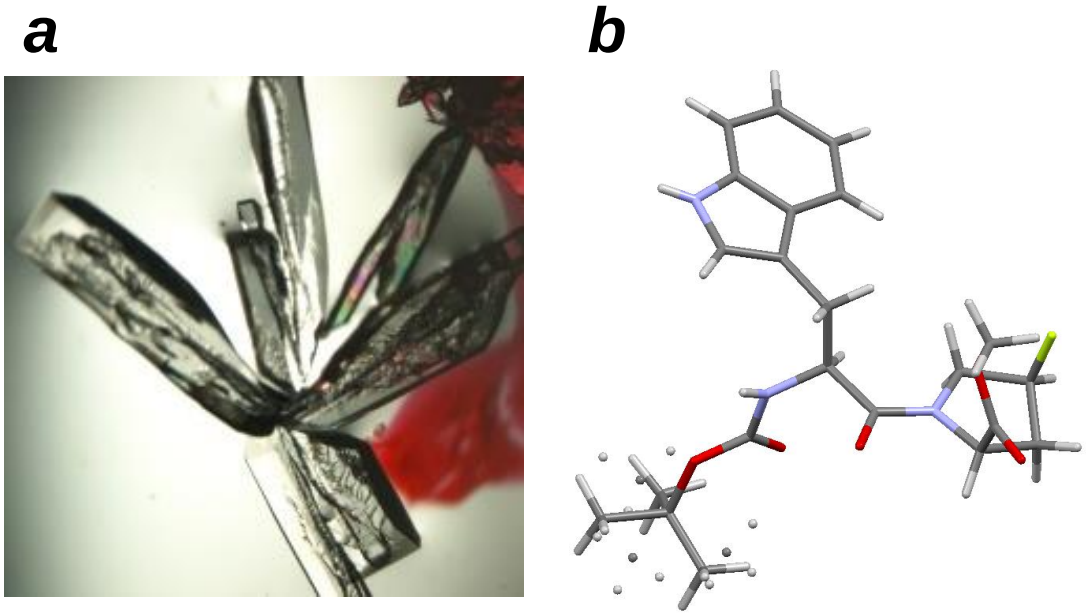


Figure 3.58 X-ray crystal structure of the dipeptide Boc-Trpflp-OMe
(a) Crystals were obtained via slow evaporation at room temperature from a solution of 50% methanol in water (crystallization of the peptide was conducted by Dana Reigner); (b) crystal structure of the peptide Boc-Trpflp-OMe. The structure was solved to 0.80 Å resolution. This dipeptide exhibited a close intermolecular contact, with the form shown in Figure 3.64.

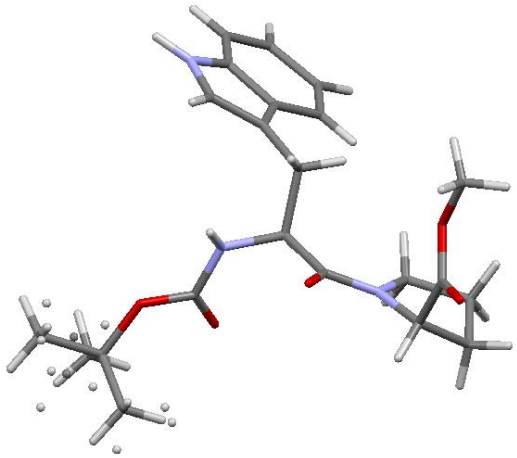


Figure 3.59 X-ray crystal structure of the dipeptide Boc-TrpPro-OMe
Crystals were obtained via slow evaporation at room temperature from a sample dissolved in methanol. The crystal structure of the peptide Boc-TrpPro-OMe was solved to 0.77 Å resolution. This dipeptide exhibited a close intermolecular contact, with the form shown in Figure 3.64.

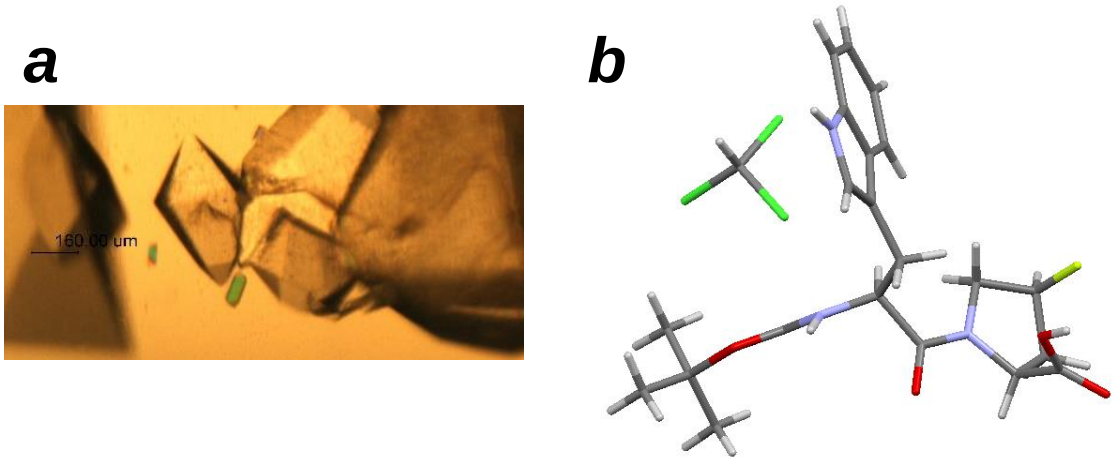


Figure 3.60 X-ray crystal structure of the dipeptide Boc-Trpflp-OH
(a) Crystals were obtained via slow evaporation at room temperature from an NMR sample in $CDCl_3$; (b) crystal structure of the peptide Boc-Trpflp-OMe. The structure was solved to 0.77 Å resolution. This dipeptide exhibited a close intermolecular contact, with the form shown in Figure 3.64.

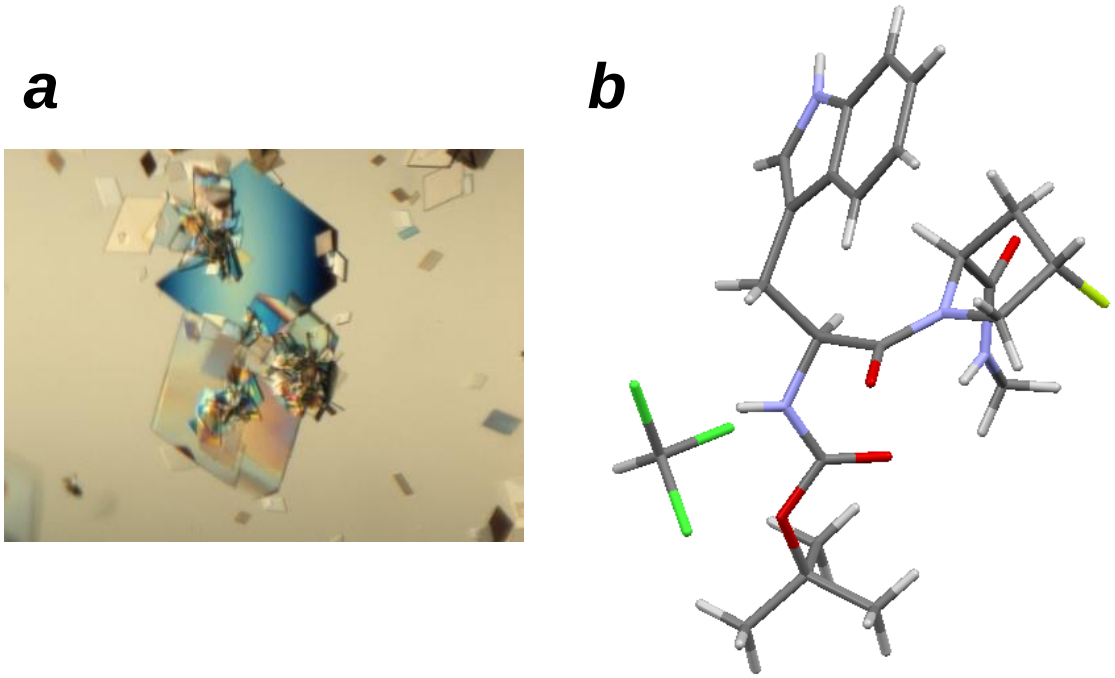


Figure 3.61 X-ray crystal structure of the dipeptide Boc-Trpflp-NHMe
(a) Crystals were obtained via slow evaporation at room temperature from a dissolved in $CDCl_3$; (b) crystal structure of the peptide Boc-Trpflp-NHMe. The structure was solved to 0.75 Å resolution. This dipeptide exhibited close C–H/π aromatic intramolecular interactions.

Table 3.12. Summary of minimal aromatic-proline motifs: structural data from x-ray crystallography and NMR in solution.

All of the peptides in the series of dipeptides containing the Trp-flp motif successfully crystallized and diffracted, which allowed for unique insights into the nature of C–H/π aromatic interactions that stabilize the cis amide bond conformation. Detailed NMR data for these dipeptides are described in Section 3.2.6, and the structures obtained via x-ray crystallography are shown in Figures 3.56-3.61.

[a]The $^3J_{\alpha N}$ coupling constant was difficult to discern due to spectral overlap, and the value shown is only an estimation.

flp indicates 4*S*-fluoroproline

4-PNB-hyp indicates 2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline

NMR Measurements MeOH, 298 K	Ac-Trp(4-PNB-hyp)(4-Br-Phe)-NH_2	Boc-TrpPro-OMe	Ac-Trpflp-OMe	Boc-Trpflp-OMe	Boc-Trpflp-OH	Boc-Trpflp-NHMe
$K_{trans/cis}$	0.35	2.9	0.8	0.75	2.1	0.49
$K_{cis/trans}$	2.8	0.35	1.3	1.3	0.48	2.1
% cis	74%	26%	56%	57%	32%	67%
$^3J_{\alpha N}$, Hz	2.6	8.6	7.0	7.7	8.1	4.4[a]
$\Delta G_{cis/trans}$ kcal mol^{-1}	-0.61	0.63	-0.13	-0.17	0.43	-0.42
X-ray crystal measurements						
Resolution	1.10	0.77	0.77	0.80	0.77	0.75
C–H/π contact inter- or intramolecular	Intra-	Inter-	Inter-	Inter-	Inter-	Intra-
Trp ϕ	-85.3	-82.3	-71.1	-78.7	-83.4	-62.8
Trp ψ	143.0	147.1	150.8	147.3	151.7	137.2
ω	4.7	173.8	164.2	173.4	170.2	11.1
flp ϕ	-89.8	-66.6	-67.0	-67.0	-82.2	-92.5
flp ψ	15.4	-25.8	-10.8	-18.0	9.9	12.4

In spite of the high populations of cis-proline that were observed in solution with most of these dipeptides, only two peptides crystallized in the cis conformation, Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-bromophenylalanine)-NH_2 (Figure 3.56) and Boc-Trpflp-NHMe (Figure 3.61). The crystal structure of the peptide Boc-Trpflp-NHMe revealed an exceptionally close contact between the 4*S*-fluoroproline Hα and the Trp indole ring (Figure 3.62). The 4*S*-fluoroproline Hα was located 2.62 Å from the 5-membered aromatic ring centroid with reasonable alignment of interaction[190] (angle C–Hα··centroid = 147.1°). In light of the insights into the nature of S-H/π aromatic interactions described in Chapter 2, it is potentially more important to examine the distances to individual aromatic ring carbons (Figure 3.62). In fact, the 4*S*-fluoroproline Hα participated in two intramolecular interactions with aromatic ring carbons at distances that were less than the sum of the van der Waals radii for carbon and hydrogen (2.90 Å): one distance at 2.58 Å (angle C–Hα··C = 167.8°), and another at 2.80 Å (angle C–Hα··C = 139.0°). In addition, one 4*S*-fluoroproline Hβ proton and one Hδ proton also interacted with the aromatic ring carbons at distances near 3.0 Å (Figure 3.62, indicated in right panel). Indeed, the upfield chemical shift in the cis-proline Hβ protons observed via solution NMR is consistent with the crystallographic observations, due to ring-current effects from the aromatic ring. The position of the interacting 4*S*-fluoroproline Hβ proton is directly over the aromatic centroid of the 6-membered ring in tryptophan (angle C–Hβ··ring centroid = 149.6°), but the alignment of the Cβ–Hβ bond is toward a C–C bond in the ring (angle C–Hβ··bond centroid = 171.9°). The favorable alignment and close contact distances between the prolyl C–H bonds and the tryptophan aromatic ring carbons strongly suggests a stabilizing effect due to C–H/π aromatic interactions.

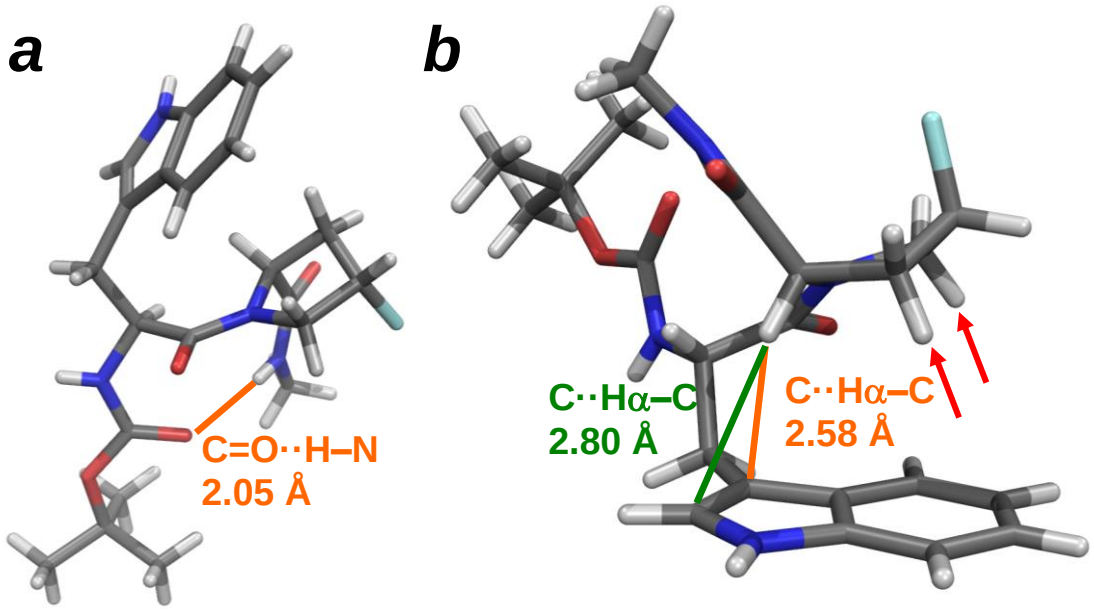


Figure 3.62 Intramolecular contacts in Boc-Trpflp-NHMe: Direct observation of an intramolecular C–H/π interaction between flp Hα and the tryptophan aromatic ring carbons
Two perspectives on the crystal structure of Boc-Trpflp-NHMe. The C–H bond in 4*S*-fluoroproline interacts closely with the aromatic ring carbons in tryptophan: 2.58 Å (angle C–Hα··C = 167.8°) and 2.80 Å (angle C–Hα··C = 139.0°). The distances from the flp Hα to the 5-membered ring centroid is 2.62 Å (angle C–Hα··centroid = 147.1°) and to the 6-membered ring centroid is 3.48 Å (angle C–Hα··centroid = 110.4°). In addition to the flp Hα, one flp Hβ proton and one flp Hδ proton also engage in close intramolecular contacts (C–H··C ~ 3.0 Å) with aromatic ring carbons.

The exceptionally close contacts between prolyl hydrogen atoms and aromatic carbon atoms in the crystal structure of Boc-TrpflpNHMe merited further examination via van′t Hoff analysis. Temperature dependence of equilibrium of an interaction can be used to determine the enthalpic and entropic contributions to the interaction (Chapter 3.2.5 and 3.2.6). The $K_{trans/cis}$ was measured for the peptide Boc-Trpflp-NHMe in methanol via ^{1}H NMR (Figure 3.63), and these data were compared to a similar peptide in methanol, Ac-TrpPro-NHMe.

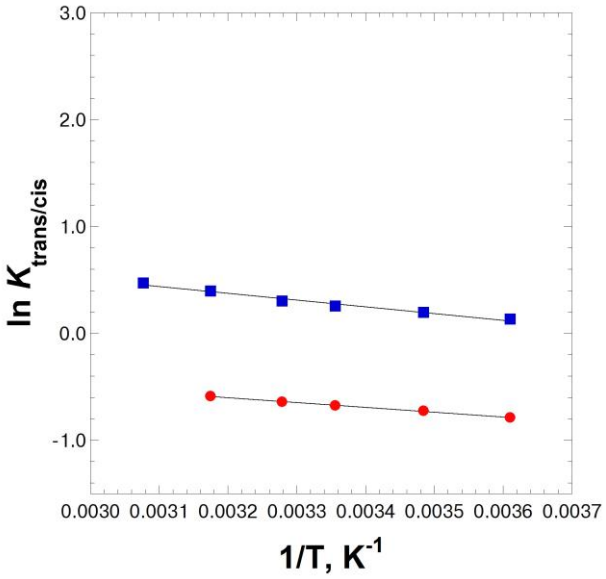


Figure 3.63 van′t Hoff plots for Boc-Trpflp-NHMe and Ac-WP-NHMe in methanol

The temperature dependence of the equilibrium $K_{trans/cis}$, measured via NMR, with respect to temperature in 90% MeOH-d_3/10% MeOD-d_4 for Boc-Wflp-NHMe (red circles) and Ac-WP-NHMe (blue squares). The data represent an average of at least 2 independent trials at each temperature. Error bars indicate standard error.
ΔH for Ac-WP-NHMe = –1.26 ± 0.092 kcal mol^{-1}
ΔS Ac-WP-NHMe = –4.79 ± 0.307 cal mol^{-1} K^{-1}
ΔH for Boc-Wflp-NHMe = –0.90 ± 0.092 kcal mol^{-1}
ΔS Boc-Wflp-NHMe = –1.68 ± 0.307 cal mol^{-1} K^{-1}

In examining the temperature-dependent NMR for these dipeptides in methanol, the entropy is more favorable with Boc-Trpflp-NHMe than for Ac-TrpPro-NHMe, while the enthalpy is slightly less favorable. However, the overall energy for the cis conformation is more favorable for Boc-Trpflp-NHMe ($\Delta G_{277\ K}$ = –0.43 kcal mol^{-1}) than for Ac-TrpPro-NHMe ($\Delta G_{277\ K}$ = +0.07 kcal mol^{-1}). Based on the van′t Hoff analysis, the cis conformation is more stabilized in the peptide Boc-Trpflp-NHMe in methanol due to a more favorable enthalpy of interaction, with less unfavorable entropy.

While only one dipeptide crystallized in the cis conformation, the dipeptides that crystallized in the trans conformation also revealed close contacts for *inter*molecular C–H/π aromatic interactions (Figure 3.64). The Trp-trans-proline motifs all crystallized with the same general structure of intermolecular interactions: the tryptophan indole N–H participated in an hydrogen bond with a carbonyl oxygen, and the prolyl C–H bonds interacted with the aromatic face of the tryptophan, with two or more C–H$\cdot\cdot$C$_{aro}$ distances that were *less* than the sum of the van der Waals radii. For some of these intermolecular contacts between Trp-trans-proline motifs, C–H$\cdot\cdot$N$_{aro}$ interactions were also observed with contact distances that were below the sum of the van der Waals radii. Similarly, intermolecular C–H/π aromatic interactions were also observed in the dipeptide Boc-Trpflp-NHMe between the Boc-methyl C–H bonds and the tryptophan aromatic ring (Figure 3.64c). Not only are these C–H/π aromatic interactions important for stabilizing aromatic-cis-proline turn conformations, via intramolecular interactions, but they can also stabilize supramolecular structures via *inter*molecular interactions. These intermolecular interactions can potentially provide guiding principles for designing supramolecular assembled structures based on C–H/π aromatic interactions.

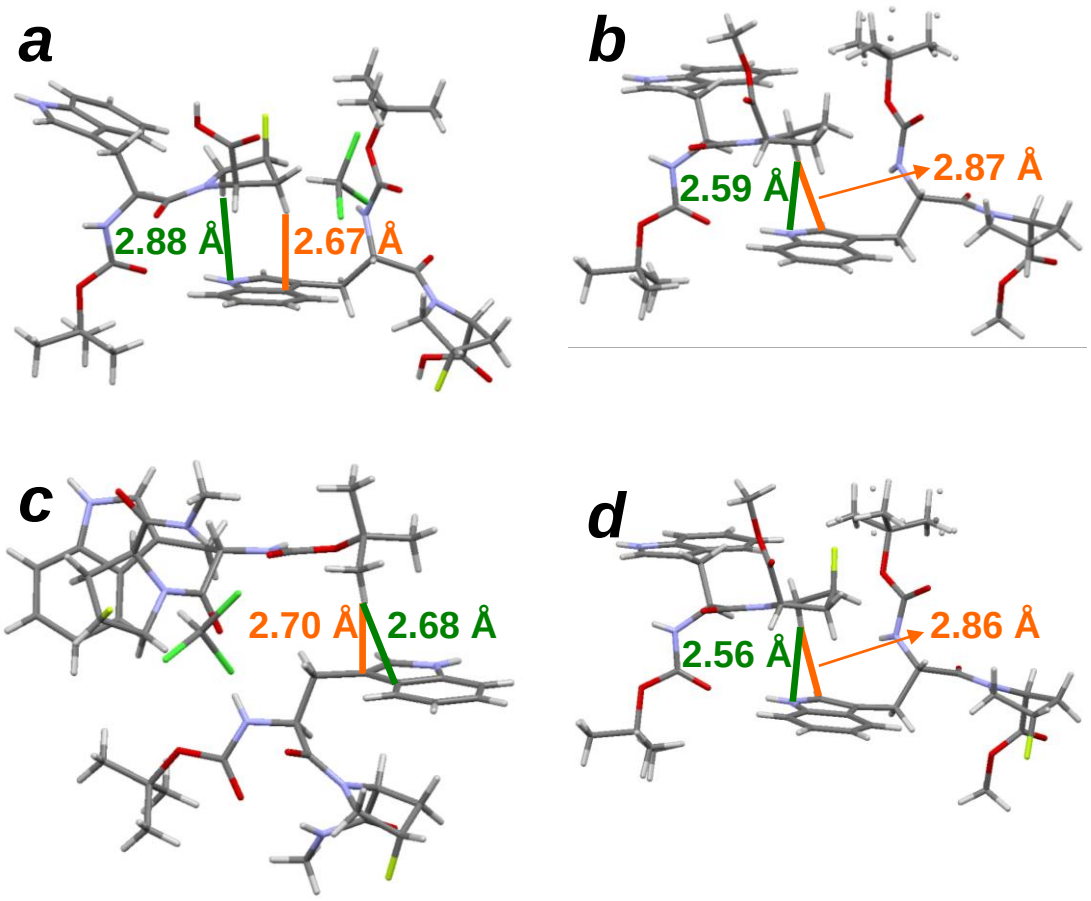


Figure 3.64 Extensive intermolecular C–H/π aromatic interactions at sub-van der Waals distances within minimal dipeptides
Both cis and trans conformations of various Trp-flp and Trp-Pro motifs engaged in extensive intermolecular C–H/π aromatic interactions. The distances indicated represent the C–H··C_{aro} distance measured from the solved crystal structures. (a) Crystal structure of Boc-Trpflp-OH. The intermolecular contact is between the prolyl ring and the tryptophan ring; (b) crystal structure of Boc-TrpPro-OMe. The intermolecular interaction is between the prolyl ring and tryptophan ring; (c) crystal structure of Boc-Trpflp-NHMe. One side of the aromatic ring interacts with the prolyl protons, and the other side of the aromatic ring interacts with a Boc-*tert*-butyl methyl group; (d) crystal structure of Boc-Trpflp-OMe. The intermolecular contact is between the prolyl ring and the tryptophan ring.

The two peptides that crystallized in the cis conformation formed the requisite *i*,*i*+3 backbone hydrogen bond that is typical for type VIa1 β-turns, indicating that this

hydrogen bond contributes substantially to the stability of the aromatic-cis-proline conformation. These observations in the crystal structures are consistent with the solution NMR data of WP and AP dipeptides. It can be hypothesized that the intramolecular proline C–H/π aromatic interaction provides favorable enthalpy of forming the β-turn motif, compensating for the unfavorable entropy of the cis conformation, and stabilizing the backbone hydrogen bond. Alternatively, the hydrogen bond may be requisite to formation, but the intramolecular proline C–H/π aromatic interaction may provide additional enthalpic stability to lock the turn motif in place.

In all of these crystal structures with the dipeptide Trp-flp motif, extensive inter- and intramolecular C–H/π aromatic interactions were observed at close contact distances between individual hydrogen and carbon atoms. Notably, in the closest interactions, the C–H bonds were directed concertedly towards individual aromatic ring carbons rather than toward either of the centroids of the tryptophan ring (Figures 3.62 and 3.64). Indeed, many of the observed C–H$\cdot\cdot$C$_{aro}$ distances were less than the sum of the van der Waals radii for carbon and hydrogen (2.90 Å). The geometry and distances observed between the flp Hα and tryptophan aromatic ring carbons (near linearity with aromatic ring carbons at sub-van der Waals distances) were reminiscent of the observations in the Boc-4-thiol-L-phenylalanine-*tert*-butyl ester crystal structure, described in detail in Chapter 2. Potentially, the orbital interaction $\pi_{aromatic} \rightarrow \sigma^*_{S–H}$ that was observed previously is a general stabilizing feature in X–H/π aromatic interactions.

3.3 Discussion

In this work, proline cis-trans isomerism was utilized as a means to characterize aromatic interactions using the model peptide Ac-TXPN-NH_2. Prior work had established the foundation for model peptide Ac-TXPN-NH_2 to act as a direct probe of aromatic electronics through the influence of aromatic substituents on proline cis-trans isomerism.[90, 91] This work encompassed a broader range of aromatic amino acids to identify the breadth and limitations of the aromatic-prolyl interaction, and provided further insights into the stabilizing contributions in the aromatic-cis-proline C–H/π interaction. Overall, more than 80 peptides were characterized in this analysis of aromatic-cis-proline interactions, through NMR characterization, thermodynamic analysis, and x-ray crystallography.[91, 169, 373] Many of the aromatic amino acids used in this study were not commercially available or were novel, and practical synthetic methods to obtain the desired non-natural amino acids within synthesized peptides.

3.3.1 Practical synthesis of novel aromatic amino acids with unique functionality and reactivity

In addition to the structural control over cis-trans isomerism, the non-natural aromatic amino acids explored in this work also had unique reactive or spectral properties. While some of these amino acids were synthesized and characterized for their unique ability to control cis-trans isomerism of proline via aromatic electronic effects, these aromatic amino acids could be utilized in other applications, such as metal-coordination or fluorescent labeling. Synthetic methodologies were developed in order to obtain these novel aromatic amino acids within peptides in a practical manner.

The pyrrolidyl group is commonly utilized in pharmaceutical compounds,[347] and we sought to quantitatively characterize the electronic properties of an aromatic

ring that contained a pyrrolidyl group. However, the novel amino acid 4-pyrrolidyl-phenylalanine exhibited strong fluorescent properties, and has potential utility for incorporation into proteins as a pH-dependent fluorescent label. The reaction methodology that was developed for synthesizing 4-pyrrolidyl-phenylalanine could potentially be extended to other secondary amines, such as pyrroles, which further extends the utility of the palladium-mediated cross-coupling approach on solid phase.

We expanded the application of the solid-phase copper-mediated cross-coupling reaction to generate peptides containing 3-mercaptotyrosine, a novel amino acid with can potentially be used for metal binding. 4-Thiophenylalanine was shown to have unique reactivity, as described in Chapter 1, such as the ability to react with alkylation reagents orthogonal to cysteine under acidic conditions. For different oxidation states, 4-thiophenylalanine exhibited distinctive UV spectroscopic signatures, and ability to control the side-chain conformation (i.e. orientation of the propargyl thioether moiety through sulfur oxidation states). 3-Mercaptotyrosine is currently under additional investigation for coordinating molybdenum or other metals in order to perform difficult redox reactions, such as nitrogen reduction or carbon dioxide reduction.[383]

The aryl triflate derivative of tyrosine was synthesized for electronic control of cis-trans isomerism of proline, and can potentially be used as an aryl halide for subsequent reactions. This practical methodology for synthesizing 4-OSO_2CF_3-phenylalanine within peptides allows for tyrosine to be converted to a pseudohalide, which can act as a substrate in metal-mediated cross-coupling reactions. This synthetic approach inspired a synthetic approach to generate a disubstituted amino acid, 4-(trifluoromethylsulfonyl),3-iodophenylalanine, which can be subjected to the copper-

mediated cross-coupling reaction with thiolacetic acid to generate 3,4-dithiolphenylalanine (Figure 3.65). We anticipate that this dithiolated aromatic amino acid can potentially coordinate soft metals for redox reactions, similar to bis-dithiol cofactors in nitrogenase enzymes.[383-385]

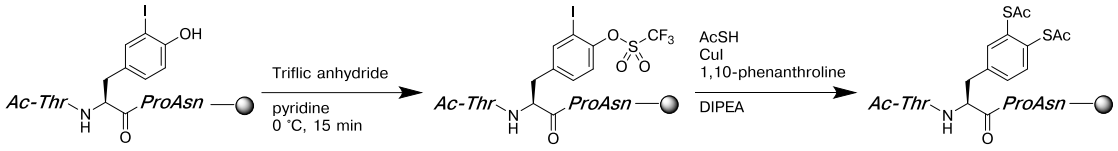


Figure 3.65 Proposed synthesis of a bis-dithiolated amino acid using practical synthetic approaches
The practical synthetic methodologies developed through this work can be utilized for generating novel di-substituted aromatic amino acids. In this potential synthesis, a bis-dithiolated aromatic amino acid can be generated in 2 steps from synthesized peptides on solid-phase. Aromatic bis-dithiols can be utilized to coordinate metals to catalyze difficult redox reactions.[364, 383-385]

3.3.2 Aromatic amino acids as a structural "switch" of proline conformation

The nature of some of these aromatic amino acids that were examined in model peptides Ac-TXPN-NH_2 allowed for inducible "switching" of the peptide structure. For example, the different protonation states of 4-amino-phenylalanine, tyrosine, and 4-thiophenylalanine exhibited substantially different preferences for cis and trans conformations. Structural "switches" based on oxidation state of the aromatic substituent, such as 4-thiophenylalanine and derivatives such as the sulfinic acid and sulfonic acid, also exhibited unique populations of cis and trans conformations. These aromatic amino acids can be used as a "switch" between cis or trans proline conformations, inducing or disrupting β-turn formation.

Several of these potential structural "switches" were examined during this study, including aromatic amino acids that can induce different cis- or trans-amide

bond populations based on protonation state or oxidation state (Tabe 3.13). Changes in protonation state induced changes in the cis populations for the peptide containing 4-amino-phenylalanine, 4-boronic acid-phenylalanine, and 4-pyrrolidyl-phenylalanine. The cis population can be shifted for the peptide containing 4-pyridyl-alanine upon either protonation or oxidation of the nitrogen atom. Canonical amino acids can also be used as structural switches that are sensitive to protonation state, such as tyrosine or histidine. Tyrosine acts as a substrate for post-translational modification in proteins, including phosphorylation, sulfation, nitration, iodination, and oxidation, and these modifications on tyrosine were characterized within the context of the model peptide Ac-TXPN-NH_2 for changes in the cis-proline population. 4-Thiophenylalanine and its related derivatives induced changes in proline cis-trans isomerism based on a variety of different modifications on the sulfur atom. Upon deprotonation of the thiol group, 4-thiophenylalanine increased the population of cis-proline in the model peptide Ac-TXPN-NH_2, and exhibited decreased cis-populations with higher sulfur oxidation states. Many of the oxidative modifications examined with 4-thiophenylalanine represent natural post-translational modifications on cysteine, including the sulfinic acid, sulfonic acid, S-glutathionylated disulfide. The thioether derivatives of 4-thiophenylalanine, and their sulfoxides and sulfones, also induced changes in the cis-proline populations of the model peptide Ac-TXPN-NH_2. 3-Mercaptotyrosine also induced significant, pH-dependent changes in the cis-proline population within the model peptide. These inducible “switches” on proline cis-trans isomerism could be utilized for designing reactive biomaterials or peptide-based sensors.

Table 3.13. Selected peptides Ac-TXPN-NH_2 with potential ability to act as a structural "switch"

Complete list of peptides characterized in this study is included in Table 3.1. $\Delta G_{trans/cis}$ and %cis populations were derived from ^{1}H NMR data for peptides Ac-TXPN-NH_2, where X = aromatic amino acid.
[a]indicates an amino acid that can modulate structure based on protonation state
[b]indicates an amino acid that can modulate structure based on oxidation state
[c]indicates an amino acid that can modulate structure based a post-translational modification
[d]indicates an amino acid that can modulate structure based a photochemical reaction

	Ac-TXPN-NH_2, X =	$K_{trans/cis}$	$\Delta G_{trans/cis}$, kcal mol^{-1}	%cis	pH
Canonical amino acids	His(H^+)[a]	12.3	-1.49	8%	4
	His[a]	5.7	-1.03	15%	8.5
	Tyr[a,c]	3.2	-0.69	24%	4
	4-O^--Phe[a]	1.9	-0.38	34%	12
	4-OPO_3H^--Phe[a,c]	3.1	-0.67	24%	4
	4-OPO_3^{2-}-Phe[a,c]	3.0	-0.65	25%	8.0
	4-OSO_3^--Phe[c]	2.9	-0.63	26%	4
	3-NO_2-Tyr[a,c]	6.5	-1.11	13%	4
	3-NO_2-Tyr[a,c]	3.1	-0.67	24%	9.5
	3-I-Tyr[c]	3.7	-0.77	21%	4
	3-OH-Tyr[c]	2.5	-0.54	29%	4
Non-natural amino acids	4-Pyridyl(N-oxide)-Ala[b]	11.2	-1.43	8%	4
	4-Pyridyl(H^+)-Ala[a]	11.2	-1.43	8%	2.5
	4-Pyridyl-Ala[a,b]	6.1	-1.07	14%	7.2
	4-Pyrrolidyl(H^+)-Phe[a]	3.5	-0.74	22%	2.5
	4-Pyrrolidyl-Phe[a]	2.5	-0.54	29%	7.3
	4-NH_3^+-Phe[a]	5.5	-1.01	15%	2.5
	4-NH_2-Phe[a]	2.9	-0.63	26%	6.5
	4-$B(OH)_2$-Phe[a]	3.5	-0.74	22%	4
	4-$B(OH)_3^-$-Phe[a]	2.4	-0.52	29%	10
Thiophenylalanine derivatives	4-SH-Phe[a,b,c]	3.8	-0.79	21%	4
	4-S^--Phe[a,b,c]	2.3	-0.49	30%	8.5
	4-S(2-nitrobenzyl)-Phe[d]	4.2	-0.85	19%	4
	4-S-S-Glutathione-Phe[b,c]	3.9	-0.81	20%	4
	4-SO_2^--Phe[b,c]	3.8	-0.79	21%	4
	4-SO_3^--Phe[b,c]	4.4	-0.88	19%	4

Table 3.13 continued

4-SMe-Phe[c]	3.6	-0.76	22%	4
4-S(O)Me-Phe[c]	4.7	-0.92	18%	4
4-SO_2Me-Phe[c]	5.8	-1.04	15%	4
4-S-propargyl-Phe[c]	3.2	-0.69	24%	4
4-S(O)propargyl-Phe[c]	4.3	-0.86	19%	4
4-SO_2-propargyl-Phe[c]	4.5	-0.89	18%	4
3-SH-Tyr[a]	3.3	-0.71	23%	3
3-S^--Tyr[a]	1.6	-0.28	39%	7.5

These inducible "switches" for cis-trans isomerism of proline were applied to proline-rich peptides that could adopt compact or extended conformations. In this work, model peptides Ac-GPPXPPGY-NH_2 were synthesized containing different aromatic amino acids, and the conformational populations were examined via NMR. With electron-rich aromatic amino acids, multiple proline residues could interact with the aromatic ring to invoke a mixture of cis-populations, an overall compact structure.[172] In contrast, with electron-poor aromatic amino acids, the trans conformation was stabilized, invoking an overall extended, polyproline helix II structure in these proline-rich peptides.[172] In this manner, polyproline helix II formation was favored or disfavored based on the aromatic substituent effects. This work demonstrated that aromatic-cis-proline interactions could be used for designing responsive peptide materials, for example, proline-rich fibril polymers that can expand or contract under acidic or basic conditions.

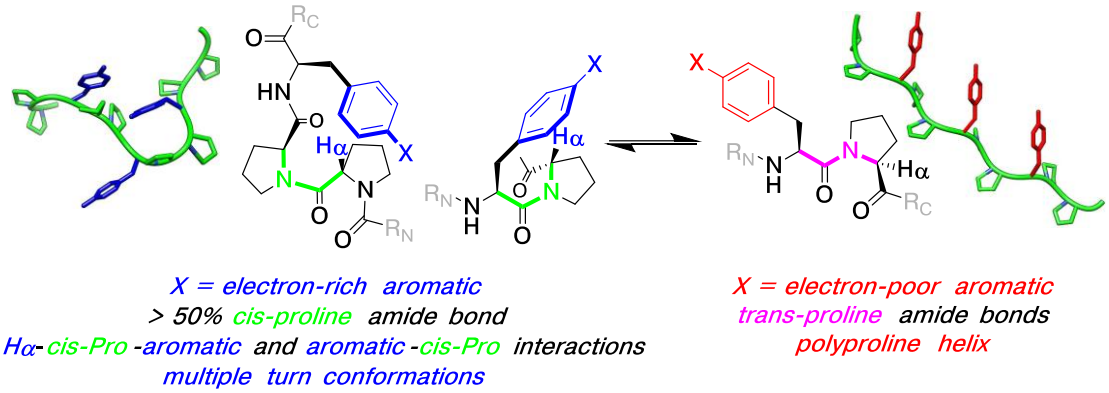


Figure 3.66 Electronic control of polyproline helix stability through aromatic substituent effects[172]
Electron-rich aromatic amino acids stabilize cis-proline conformations, which is unfavorable in polyproline II helices (left). In contrast, electron-poor aromatic amino acids stabilize trans-proline conformations, which is favored in polyproline II helices (right). Aromatic amino acids such as 4-amino-phenylalanine can act as inducible switches between rigid, ordered conformations and disordered conformations, due to differences in electron-donating ability of aromatic substituents based on protonation state.

3.3.3 Aromatic substituent effects on cis-trans isomerism of proline: Linear-free energy relationships of the aromatic-cis-proline interaction

In this work, model peptides Ac-TXPN-NH_2 were synthesized with more than 50 different amino acids at position X in order to examine the influence of 4-substituted and 3,4-disubstituted aromatic amino acids on cis-trans isomerism of proline. A Hammett correlation was observed with the electron-donating ability of the aromatic substituents and the $K_{trans/cis}$ for each peptide.[91] In addition to gaining further insights into the nature of C–H/π aromatic interactions in peptides and proteins, these Hammett correlations allowed for approximation of the σ values in water for given aromatic substituents which do not have a reported σ value. The σ_{meta} and σ_{para} values were calculated for aromatic substituents including 4-pyrrolidine, 4-sulfonate, and oxidized derivatives of thioethers based on the Hammett correlations for the peptides

Ac-TXPN-NH_2 (Table 3.2). The pseudo-σ_{meta} and σ_{para} values for heteroaromatic rings was also possible with these Hammett correlations, and so σ_{meta} and σ_{para} values were calculated for both protonation states of histidine, 4-pyridine, and for tryptophan in water (Table 3.2). These calculated σ and pseudo-σ values could potentially be utilized in practice where aromatic electronic effects are important in water, such as in designing peptide or small-molecule therapeutics.

Improved fit and a greater correlation constant (ρ) was observed for when the σ_{meta} values were used for Hammett correlation (ρ = 0.33 for neutral substituents), rather than the σ_{para} values (ρ = 0.20 for neutral substituents). In structure-activity relationship analysis, identifying a greater magnitude in the ρ value implies a greater sensitivity due to substituent effects.[369] Fundamentally, the values for σ_{para} and σ_{meta} for an aromatic substituent are different: a para-substituent (σ_{para} values) exerts both inductive and resonance effects, while a meta-substituent (σ_{meta} values) exerts primarily inductive or field effects.[358, 369] In essence, σ_{meta} values for a given substituent represent an entirely inductive, through-space effect on the 'reaction center' in the context of aromatic rings.[369, 386] In the Hammett correlations with the model peptides Ac-TXPN-NH_2 (Figure 3.37), the greater ρ value when using σ_{meta} values suggests that cis-trans isomerism of proline is more sensitive to inductive field effects rather than resonance effects.

All of the model peptides Ac-TXPN-NH_2, both using mono- and di-substituted phenylalanine derivatives, and regardless of charge or number of substituents, were fit to a common Hammett correlation based on this model of primarily inductive effects (Figure 3.67). The Hammett constant for the disubstituted aromatic amino acids were calculated by adding the σ_{meta} for the 4-substituent to either the σ_{meta} or σ_{para} of the 3-

substituent (Figure 3.67a and 3.67b, respectively). Both of these combined Hammett correlations exhibited excellent fit ($R > 0.88$).

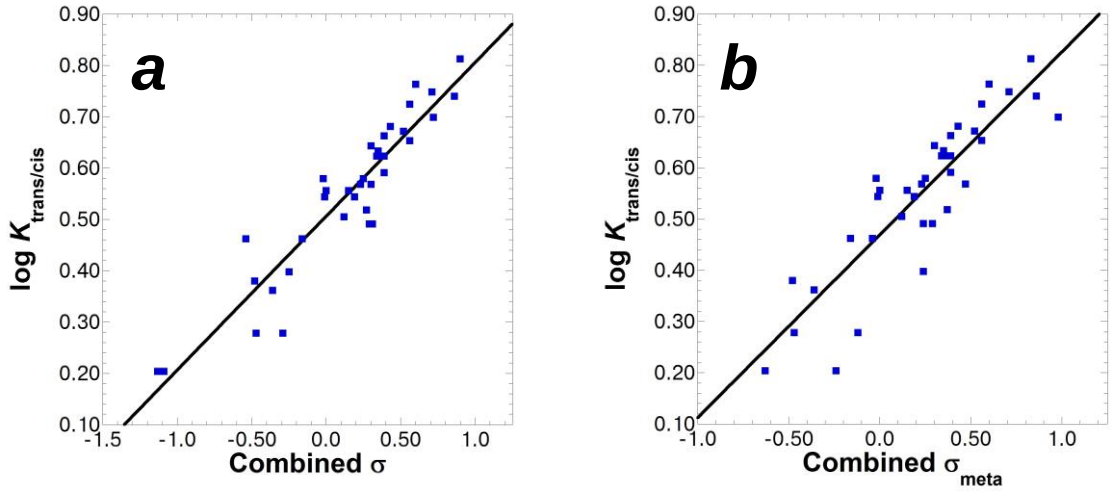


Figure 3.67 Combined Hammett correlation of $K_{trans/cis}$ in model peptides Ac-TXPN-NH_2, where X = aromatic amino acid
The log($K_{trans/cis}$) of each model peptide is correlated with the σ Hammett constant for the aromatic substituent within Ac-TXPN-NH_2 model peptides. Both 4- and 3,4-substituted aromatic amino acids are included, and both charged and neutral species are fit to the Hammett correlation.
For 4-substituted amino acids, the σ_{meta} was used
(a) For 3,4-disubstituted amino acids, the combined σ was computed from $\sigma_{meta,4\text{-substituent}} + \sigma_{para,3\text{-substituent}}$; combined σ correlation: $\rho = 0.300 \pm 0.019$, $R = 0.934$
(b) For 3,4-disubstituted amino acids, the combined σ was computed from $\sigma_{meta,4\text{-substituent}} + \sigma_{meta,3\text{-substituent}}$; combined σ correlation: $\rho = 0.357 \pm 0.032$, $R = 0.882$

The combined Hammett correlation plots showed excellent fitness between the σ value and the $K_{trans/cis}$ across a broad scope of aromatic substituents, regardless of charge or position on the aromatic ring. The σ values for these plots (Figure 3.67) were σ_{meta} for 4-substituted aromatic amino acids and $\sigma_{meta,4\text{-substituent}} + \sigma_{para,3\text{-substituent}}$ for disubstituted aromatic amino acids, based on the concept that the "reaction center"

was focused at the aromatic ortho carbon. Greater influence with meta substituents implies a greater dependence on field induction, rather than resonance effects, although both play roles.[387] While the proline C–Hα interaction geometry may have preferences towards the adjacent aromatic ortho carbons for steric reasons, the fundamental nature of C–H/π aromatic interactions that stabilize cis-proline demanded further explanation. The directed nature of proline C–Hα for specific carbons was reminiscent of the results obtained in Chapter 2, where S–H bonds were "directed" towards aromatic ring carbons due to favorable molecular orbital interactions. Utilizing cis-trans isomerism of proline as a probe, we aimed to study the nature of the prolyl C–H/π aromatic interaction and determine if the favorable molecular orbital interactions identified in S–H/π aromatic interactions are generalized.

In addition to the inductive or resonance effects on an aromatic ring, aromatic substituents also perturb the frontier molecular orbitals within the aromatic ring. Perturbation of the π orbitals of the aromatic ring can have significant consequences in reactivity and noncovalent interactions.[388] Molecular orbital calculations were conducted on phenol and thiophenol, and the two highest occupied molecular orbitals and calculated energies are shown (Figure 3.68). The sizes of the molecular orbitals were slightly different between these two benzene derivatives, which is particularly noticeable in comparing the localization of the molecular orbital at the carbon atom located opposite of the aromatic substituent (Figure 3.68, top panel, indicated). In addition, the energies of the orbitals differ between thiophenol and phenol (Figure 3.68). In the context of noncovalent interactions involving π orbitals as electron-donors, the substituent effects on the molecular orbitals of the aromatic ring can have

significant consequences on modulating the strengths of these noncovalent interactions.

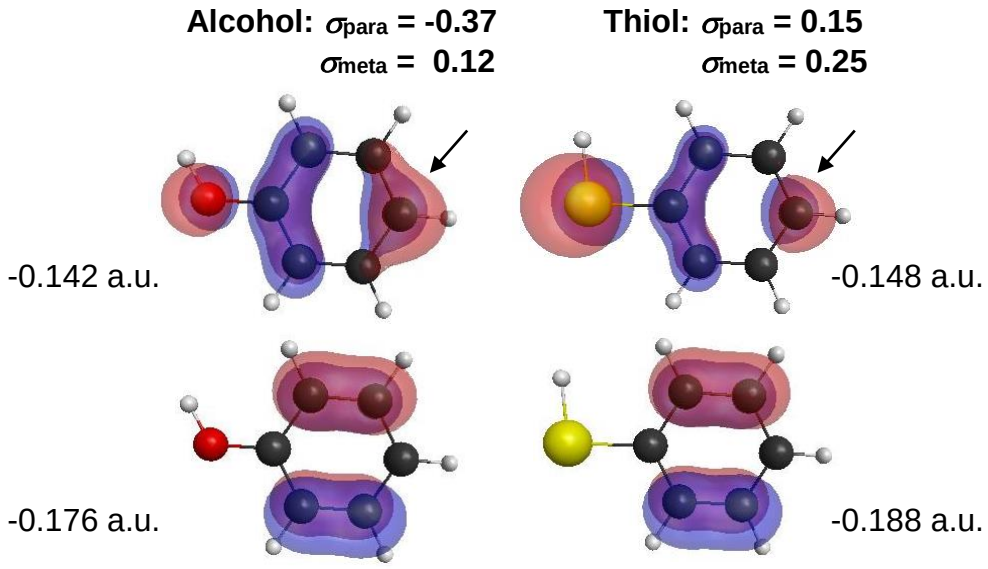


Figure 3.68 Calculated molecular orbitals for phenol and thiophenol
Calculations for the two highest occupied molecular orbitals (HOMOs) were conducted using GAMESS (RHF 6-31G),[195] and visualized with MacMolPlt (contoured at 0.05).[196] The different aromatic substituents influence the size and energy of the molecular orbitals. The size and energy of the molecular orbitals can influence the noncovalent interactions that involve the benzene ring.

In many of the peptides examined, a TOCSY spectrum was obtained during the NMR experiments, which allowed for more accurate determination of the chemical shift for the proline Hα in cis and trans conformations. With higher cis-proline populations and increasingly electron-donating aromatic substituents, the chemical shift for the cis-Pro Hα was shifted upfield, while the trans-Pro Hα chemical shift remained at 4.42 ppm (Table 3.4, Figure 3.39). In the peptide Ac-TWPN-NH_2, with the greatest cis population for the tetrapeptide series Ac-TWPN-NH_2, the cis-Pro Hα

resonance was 0.98 ppm upfield from the trans-Pro Hα resonance. A similar effect was observed for the cis-Pro Hβ resonances, shown for the dipeptide series in water, methanol, and chloroform in Table 3.11. An upfield chemical shift implies a shielding effect from a local, induced magnetic field, which can result from a proximal electron-rich atom or π-orbital due to ring current effects.[248, 370, 389] The observed upfield chemical shifts of the cis-prolyl Hα and Hβ with increasing population of cis-proline suggested that these protons engage in hydrogen bonding behavior that stabilizes the cis conformation. Other studies in model peptides have also observed this upfield chemical shift in prolyl Hα, where aromatic amino acids were immediately adjacent to proline.[289, 318] In studies on Ac-GXPG-NH_2 model peptides by Wu & Raleigh,[288] the trans-proline Hα resonance in GWPG was at 4.40 ppm, while cis-proline Hα was at 3.47 ppm,[288] consistent with our observations for the peptide Ac-TWPN-NH_2.

While the NMR data is strongly suggestive of a C–H/π aromatic interaction, the upfield chemical shift in the prolyl Hα could potentially be an artifact from another interaction (such as hydrophobic effects). In other words, if the aromatic-cis-proline interaction is stabilized by a hydrophobic effect between proline and the aromatic ring, and the prolyl Hα will still be positioned near the aromatic ring, and an upfield chemical shift can still be observed due to aromatic ring current effects. In order to determine if the aromatic-cis-proline interaction was due to a hydrophobic effect or an aromatic interaction, we examined selected peptides in organic solvents. By examining the aromatic-cis-proline interaction in organic solvents, the intramolecular hydrophobic interactions are weakened due to competing interactions with bulk solvent. However, electrostatically-driven interactions such as hydrogen bonds are enhanced in organic solvents because of fewer competing interactions with

solvent. In the organic solvents acetonitrile and methanol, the tetrapeptides Ac-TXPN-NH_2 exhibited generally higher cis populations than observed in water, strongly suggesting that the stability of the cis conformation is not driven by a hydrophobic effect. The Hammett correlation constants (ρ) were greater in magnitude in organic solvents compared to water (Figure 3.41), suggesting that proline cis-trans isomerism is even more sensitive to aromatic substituent effects in organic solvents than in water.[369]

The hydrophobic effect is a largely-entropically driven process with a small component of enthalpy compensation.[18] Exclusion of water in non-polar media allows for more degrees of freedom, since inclusion of water requires entropically unfavorable ordered structures.[18] In contrast, hydrogen bonds are largely enthalpy-driven with a modestly unfavorable entropy, which is due to the loss of degrees of freedom for a favorable electrostatic interaction.[203] Identifying the entropy and enthalpy of interaction can elucidate the underlying nature of interaction. Through van′t Hoff analyses examine an interaction as a function of temperature, which can determine the enthalpy and entropy of the interaction.[369] Through van′t Hoff analyses of the model tetrapeptides in water, we determined the enthalpic and entropic contributions for the cis-proline conformation. The calculated ΔH and ΔS thermodynamic parameters were calculated for the model tetrapeptides, and an increasingly favorable enthalpy was correlated with increasing cis population and increasingly electron-donating aromatic substituents (Chapter 3.2.5). Wu & Raleigh[288] observed that more favorable enthalpies were observed for increasing cis populations of GXPG model peptides, particularly for tryptophan and tyrosine in water and in DMSO. In the tetrapeptides containing aromatic amino acids, a modest enthalpy-

entropy compensation is observed (Figure 3.69), with a β value of 474 K.[203] The tetrapeptides containing alanine and cyclohexylalanine appeared to be outliers in the context of the plot $\Delta H = \beta \Delta S$, potentially suggesting a different mechanism for stabilization of the cis conformation.

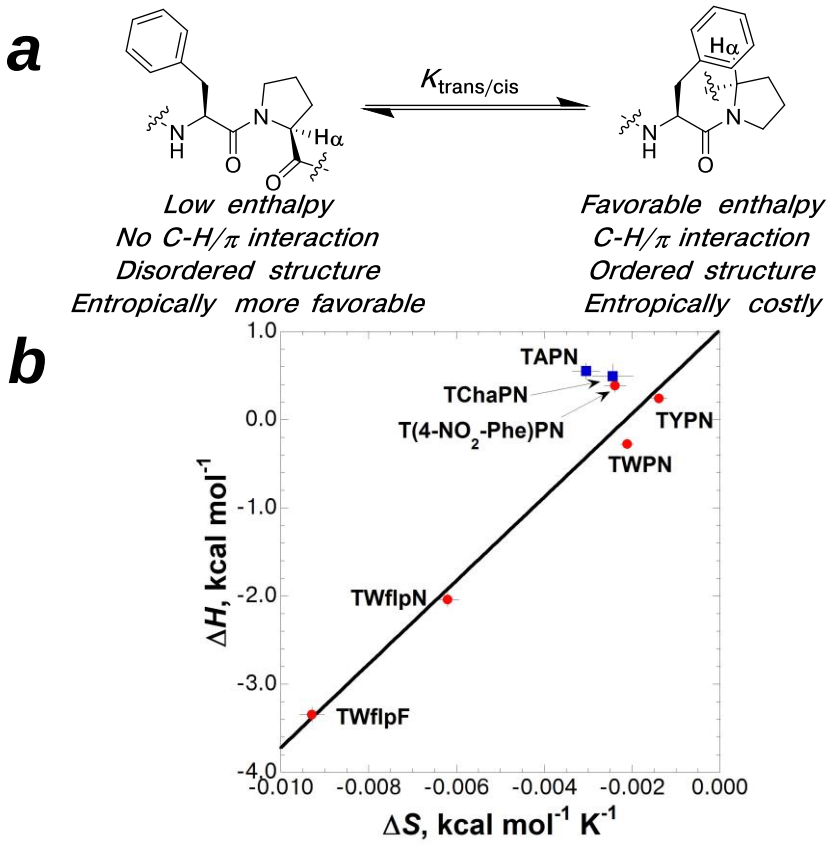


Figure 3.69 Enthalpy-entropy compensation in model tetrapeptides
(a) The peptides designed to promote the cis-conformation have a strong enthalpic component that stabilizes the aromatic-cis-proline interaction. However, the entropy becomes more unfavorable with loss of degrees of freedom and increasingly ordered structures. (b) The peptides containing aromatic amino acids (red circles) exhibited typical enthalpy-entropy compensation, while non-aromatic peptides (blue squares) appeared to be outliers. The error bars are shown from linear fit of the van′t Hoff analysis (Table 3.8).

3.3.4 Designed peptides with increased stability for the cis conformation: A closer examination of the nature of C–H/π aromatic interactions

With the design of the stabilized peptide, Ac-TWflpF-NH_2, as part of the van′t Hoff analyses, we sought additional characterization of the tetrapeptide with the highest population of cis-proline in our studies (72% cis conformation in water at room temperature). NMR analysis revealed unique features in the peptide Ac-TWflpF-NH_2 that were distinctive from the other model tetrapeptides. By NMR, the prolyl Hβ protons in the cis conformation exhibited divergent resonances, with more than a 1.5 ppm chemical shift difference between the geminally coupled protons. In contrast, the chemical shift difference for the trans prolyl Hβ protons was less than 0.15 ppm. The differences between the cis-prolyl Hβ and trans-prolyl Hβ protons indicates distinctive electronic environments surrounding the prolyl Hβ protons between the cis and trans conformations. The prolyl Hα was also shifted substantially upfield in the peptide Ac-TWflpF-NH_2, consistent with the observations in the model peptides TXPN.

The peptides containing tryptophan, both with natural and modified prolines, generally had the highest populations for cis-proline conformation, with the greatest enthalpic driving force for this interaction. The high population for cis conformation involving tryptophan-proline sequences has been observed in model peptides previously.[90, 91, 288, 289, 318] In database analyses of the Protein Data Bank, the occurrence of tryptophan preceding cis-proline is often over-represented, when normalized for the relatively low abundance of tryptophan in proteins.[285, 376] It can potentially be argued that proline interacts more favorably with tryptophan due to the increased hydrophobic surface area of the indole ring. However, we propose that the broader surface area for electronic interactions plays a substantial role.

The solution NMR data on the designed peptide Ac-TWflpF-NH_2 indicated that both the prolyl Hα and Hβ protons participated in interactions that stabilize the cis conformation. Interactions with both of these prolyl protons increases the overall energy of enthalpic stability, while being only moderately entropically costly. In essence, the larger electronic surface area of tryptophan allows for "polyvalency" of the participating prolyl protons.[291, 390] Similar to glucose that can interact with tryptophan through three C–H/π aromatic interactions,[105, 329] proline can also interact favorably with tryptophan through multiple C–H/π aromatic interactions.

These observations of upfield chemical shifts in the trans-amide proton following proline were consistent with observations by Wu & Raleigh[288] in GXPG model peptides: the Gly4 amide proton chemical shifts were at 7.91 ppm and 8.31 ppm for trans and cis conformations of the peptide GWPG, respectively. In model studies on dipeptides and modified prolines, Taylor *et al.*[372] suggested that the trans conformation is stabilized by an hydrogen bonding interaction between the amide proton following proline and the carbonyl oxygen preceding proline. This hydrogen bonding interaction can cause a significant upfield chemical shift in the trans-Phe amide proton in Ac-TWflpF-NH_2 (δ Phe $NH\alpha_{trans}$ = 7.65 ppm, Phe $NH\alpha_{cis}$ = 8.80 ppm). This upfield chemical shift in the amide proton was also noted by Yao *et al.* in model peptides SYPXDV,[289, 318] where aromatic amino acids at position X produced the greatest upfield chemical shifts in trans-amide protons. Aromatic amino acids following proline can promote the cis conformation, potentially by interacting with the prolyl face opposite of the aromatic-cis-proline interaction (Figure 3.70).[90] However, aromatic amino acids may also stabilize the trans conformation via aromatic-stacking interactions or C–H/π interactions between the aromatic amino acids (Figure 3.70c).

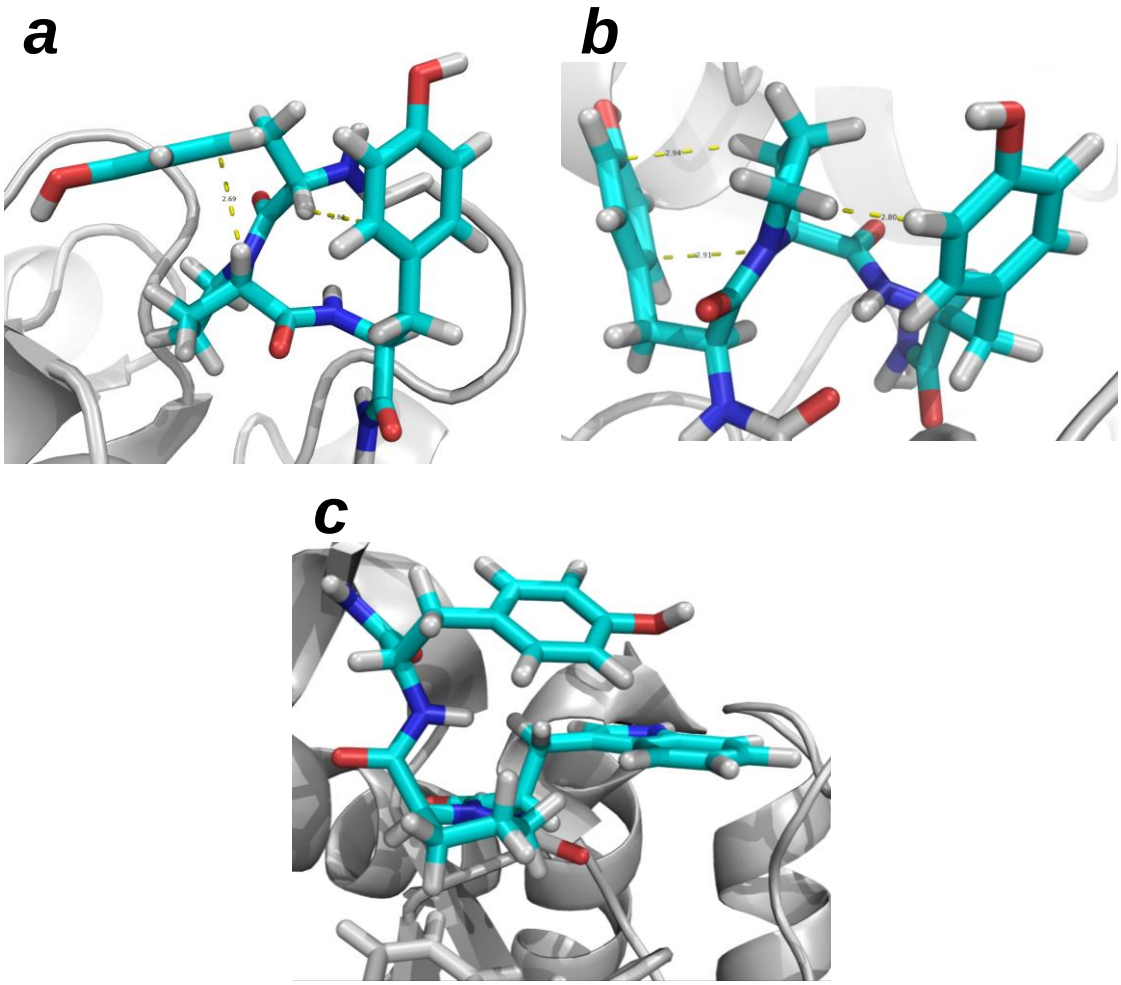


Figure 3.70 Conformations of aromatic-proline-aromatic motifs
(a) Tyr-cis-Pro-Tyr motif, with two C–H/π aromatic interactions that stabilize the turn motif: Pro$_i$Cα–Hα/C$_{aro}$ Tyr$_{i-1}$ at 2.69 Å and Tyr$_{i-1}$Cα–Hα/C$_{aro}$ Tyr$_{i+1}$ at 2.84 Å (PDB ID: 2RAU); (b) Tyr-cis-Pro-Tyr motif, with both aromatic amino acids participating in C–H/π interactions: Pro$_i$Cα–Hα/C$_{aro}$ Tyr$_{i-1}$ at 2.91 Å, Pro$_i$Cβ–Hβ/C$_{aro}$ Tyr$_{i-1}$ at 2.94 Å, and Pro$_i$Cδ–Hδ/C$_{aro}$ Tyr$_{i-1}$ at 2.80 Å (PDB ID: 1ADE); (c) Tyr-Pro-Trp motif (all trans amide bonds), stabilized by interactions between aromatic amino acids: Trp$_{i+1}$Cβ–Hβ/C$_{aro}$ Tyr$_{i-1}$ at 3.39 Å and Tyr$_{i-1}$C$_{aro}$/C$_{aro}$ Trp$_{i+1}$ at 3.76 Å (PDB ID: 3FX4).

Incorporating Phe in the position following proline was designed to increase the cis-population[90, 288] in order to study this conformation more thoroughly. While the Ac-TWflpF-NH$_2$ peptide had the highest cis population in all tetrapeptides studied

($\Delta G_{cis/trans}$ = –0.76 at 299 K in water), the phenylalanine residue clearly introduced competing interactions that stabilized the trans conformation. However, the *i*,*i*+3 backbone hydrogen bond is still crucial for formation of type VIa1 β-turns, and contributes a substantial enthalpic driving force for aromatic-cis-proline motifs. In order to avoid introducing interfering interactions to the central aromatic-cis-proline motif, and to separate the energetic contributions of the backbone hydrogen bond from the aromatic-proline interaction, minimized dipeptides were synthesized and characterized in order to interrogate the fundamental nature of the prolyl C–H/π aromatic interaction. Dipeptides Ac-AP and Ac-WP were synthesized containing either C-terminal methyl esters or C-terminal methyl amides. In order to determine the role of the backbone hydrogen bond, the peptides containing C-terminal methyl esters were compared against the peptides containing methyl amides. In order to determine the role of the prolyl C–H/π aromatic interaction, the peptides containing tryptophan were compared against the peptides containing alanine. This series of dipeptides was studied via van′t Hoff analysis in different solvents, water, methanol, and chloroform, in order to identify the relative contribution of the hydrophobic effect on the cis-proline conformation. As expected, in all dipeptides, the cis population was increased with formation of the backbone hydrogen bond (Table 3.14, XP-OMe vs. XP-NHMe), regardless of the solvent, as an enthalpically driven process. In contrast, a substantial increase in cis population was observed on formation of the aromatic-proline interaction, even in the absence of the backbone hydrogen bond Table 3.14, (AP-ZMe vs. WP-ZMe). In comparing the relative energies of the dipeptides AP vs. WP, the cis conformation was stabilized by 0.7-0.8 kcal mol^{-1} regardless of solvent *in the absence of a backbone hydrogen bond* (Table 3.14, AP-OMe vs. WP-OMe). Furthermore, the

enthalpy for the cis conformation is more favorable by 1 kcal mol^{-1}, even though no backbone hydrogen bond is formed. These data strongly indicate that the aromatic-proline C–H/π interaction significantly stabilizes the cis conformation, even when the *i*,*i*+3 backbone hydrogen bond is not formed.

Table 3.14. Measured relative energies of minimized dipeptides: Energetic contributions on stabilizing the cis-proline conformation from the backbone hydrogen bond and the proline C–H/π aromatic interaction

Data were derived from van′t Hoff plots and associated calculations, described in section 3.2.6, Table 3.9. $\Delta G_{cis/trans}$ was calculated from $-RT\ln K_{cis/trans}$ at 277 K. Direct comparisons between dipeptides, examining relative contributions of the backbone hydrogen bond and the prolyl C–H/π aromatic interaction, are shown.

	AP-OMe vs. AP-NHMe			**WP-OMe vs. WP-NHMe**		
	$\Delta\Delta G_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}	$\Delta\Delta G_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}
H_2O	–0.39	–0.19	+0.56	–0.47	–1.07	–2.10
MeOH	–0.24	+0.45	+2.49	–0.43	–1.44	–3.63
$CDCl_3$	–0.76	–2.27	–5.50	–0.48	–1.49	–3.72

	AP-OMe vs. WP-OMe			**AP-NHMe vs. WP-NHMe**		
	$\Delta\Delta G_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}	$\Delta\Delta G_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta H_{cis/trans}$, kcal mol^{-1}	$\Delta\Delta S_{cis/trans}$, cal mol^{-1} K^{-1}
H_2O	–0.83	–1.00	–0.75	–0.91	–1.88	–3.41
MeOH	–0.79	–0.95	–0.59	–0.98	–2.84	–6.71
$CDCl_3$	–0.67	–1.01	–1.30	–0.39	–0.23	+0.48

Having established the enthalpic nature of the aromatic-cis-proline interaction with model tetrapeptides, we sought to design highly stabilized peptide motifs to directly study the geometry of the prolyl C–H/π aromatic interaction. By solution NMR, the enthalpy and entropy of interaction were studied, but examination of the geometry of the prolyl C–H/π aromatic interaction required crystallization of the

stabilized aromatic-cis-proline motif. Our initial attempts focused on tripeptides Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-X-Phe)-NH_2 as the aromatic-Pro-aromatic motif had shown enhanced cis-populations based on our studies on TWflpF and other work.[90, 289, 318, 320, 376] (2*S*,4*R*-*p*-nitrobenzoyl)hydroxyproline is known to crystallize with an exo ring pucker on proline (promoting trans amide conformation),[381] and so (2*S*,4*S*-*p*-nitrobenzoyl)hydroxyproline should crystallize with an endo ring pucker on proline (promoting cis amide conformation).[373] The aromatic amino acid following proline was varied to include halogenated aromatic amino acids, with the expectation that bromine or iodine atoms could promote crystallization and aid in solution of the crystal structure. In this initial tripeptide series, only one peptide formed a diffractable crystal, Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-Br-Phe)-NH_2. However, the structure Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-Br-Phe)-NH_2 of was solved with numerous errors and relatively low resolution. While this tripeptide crystallized in the cis conformation, the aromatic interaction between the tryptophan indole ring and proline was weakened by a competing aromatic-aromatic stacking interaction between 4-nitrobenzoate and 4-bromophenylalanine (Figure 3.56). This structure did not exhibit either a cis-proline-aromatic interaction (as described by Yao *et al*.[289, 318]) or an Hα-cis-proline aromatic interaction that is known to be the primary basis of increased cis amide bond in this motif (Figure 3.70). Based on the crystal structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoyl-hydroxyproline)(4-Br-Phe)-NH_2, we redesigned the peptide series to a more concise motif, which more closely matched the dipeptide motifs that were studied via NMR.

A series of dipeptides based on Trp-(4*S*-fluoroproline) was synthesized in solution, in an effort to crystallize the aromatic-cis-proline motif and directly observe

the geometry of the proline C–H/π aromatic interaction. 4*S*-Fluoroproline promotes cis amide conformations through stereoelectronic control of the proline ring-pucker, based on numerous prior studies.[91, 98, 373-375] In addition, in order to reduce flexibility and increase rigidity, the tripeptide model was truncated to a simplified dipeptide. We successfully obtained several crystals from the series of dipeptides based on Boc-Trpflp motifs. The only dipeptide that crystallized as the cis amide conformation contained a C-terminal methyl amide, which allowed for formation of the intramolecular hydrogen bond, a hallmark of type VIa1 β-turns (N–H··O = 2.05 Å). The dipeptide Boc-Trpflp-NHMe exhibited spectacular intramolecular contacts between proline and the tryptophan indole ring that were consistent with NMR interpretations, specifically involving the prolyl C–Hα and one C–Hβ and one C–Hδ proton. The prolyl Hα was located near *two* aromatic indole carbons at distances less than the sum of the van der Waals radii for carbon and hydrogen (flpHα··C_{aro} = 2.80 Å and 2.56 Å). We also observed excellent alignment of the angle flpCα–Hα··C_{aro}, with one angle at 167° and the other at 139°. The trajectory of the flpCα–Hα bond was directed nearest to an aromatic ring carbon, and slightly outside of the aromatic ring face (Figure 3.71).

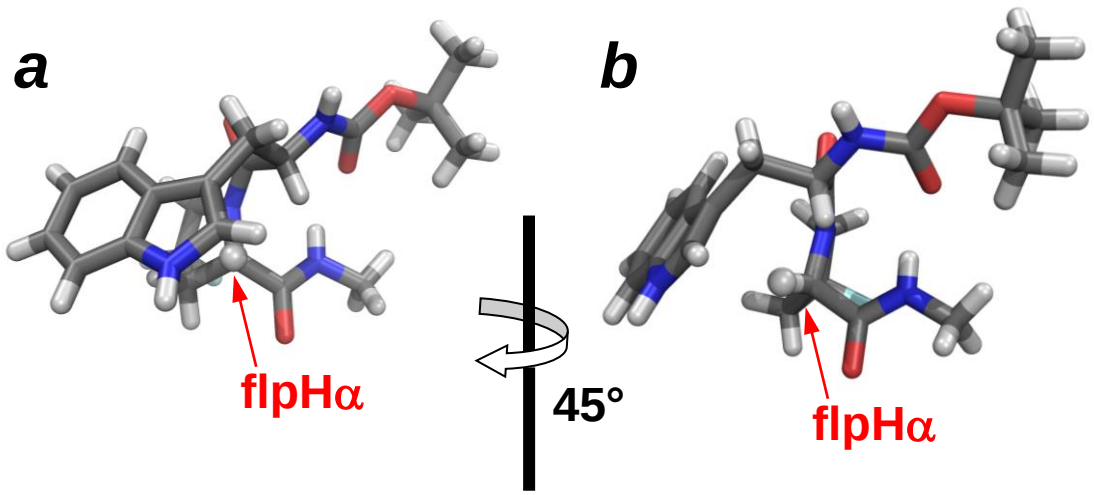


Figure 3.71 Prolyl Hα is directed towards the edge of the indole ring face. Viewing down the flpCα–Hα bond from the crystal structure of Boc-Trpflp-NHMe (flpHα is shown as a sphere), the near linear alignment to the aromatic ring carbons is apparent. The distances between the Hβ hydrogens and the aromatic carbons are less than the sum of van der Waals radii for hydrogen and carbon. This is consistent with our observations for S–H/π aromatic interactions, discussed in Chapter 2, suggesting that prolyl C–H/π aromatic interactions may also be motivated by orbital interactions.

Several other crystal structure were obtained within the Trpflp series, although all crystallized in the trans amide conformation (Boc-Trpflp-OH, Boc-Trpflp-OMe, Ac-Trpflp-OMe, and Boc-TrpPro-OMe). However, even the crystal structures of dipeptides in the trans conformation were also stabilized by prolyl C–H/π aromatic interactions, and we observed numerous *inter*molecular interactions that stabilized the overall crystal packing. A common feature in these dipeptide crystal structures with intermolecular interactions was the hydrogen bond between the indole N–H and the Boc carbonyl oxygen. Notably, this same Boc carbonyl oxygen acts as the electron donor in the intramolecular hydrogen bond with the methyl amide in the Boc-Trpflp-NHMe crystal structure. These intermolecular flpCα–H··C_{aro} contact distances were near or less than the sum of the van der Waals radii for hydrogen and carbon atoms,

involving prolyl Hα, Hβ or Hδ hydrogens (Figure 3.72). The alignment for all of these C–H/π aromatic interactions exhibited carbon-oriented geometry rather than centroid-oriented geometry. Both the short contact distances and the directionality of the interacting C–H bonds towards the edge of the aromatic ring face that were observed in these crystal structures are strongly suggestive of an orbital interaction.

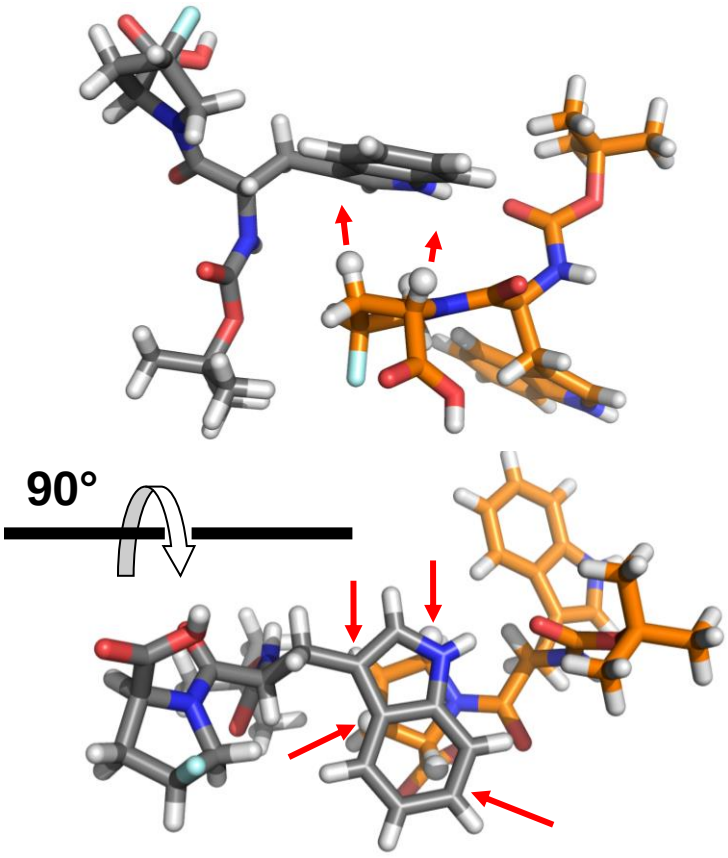


Figure 3.72 Prolyl C-H/π aromatic interactions with the tryptophan indole stabilize intermolecular interactions
Boc-Trpflp-OH is shown with intermolecular C–H/π aromatic interactions. All of the dipeptides showed crystal packing similar to that in this structure. The prolyl Hα, one Hβ, and one Hδ all interact with the aromatic ring face of proline to stabilize this interaction. The prolyl Hα generally interacted with indole ring carbons at distances *that were less than the sum of the van der Waals radii between carbon and hydrogen*, and with *carbon-oriented* geometry (indicated with red arrows).

The geometry of these intra- and inter-molecular prolyl C–H/π aromatic interactions, and the sub-van der Waals contact distances, were strikingly similar to the observations from the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester, described in Chapter 2. In our studies of S–H/π aromatic interactions, we discovered a significant stabilizing energy due to the orbital overlap between the S–H

σ^* and the π aromatic orbitals, similar to a very weak covalent bond. Given that both of these X–H/π aromatic interactions are considered non-conventional hydrogen bonds (particularly C–H, with lower bond polarization), there is the possibility that they are both driven by similar molecular orbital contributions. In this case, a $\pi \rightarrow \sigma^*_{C-H}$ interaction may stabilize the observed inter- and intra-molecular aromatic-proline interactions (Figure 3.73).

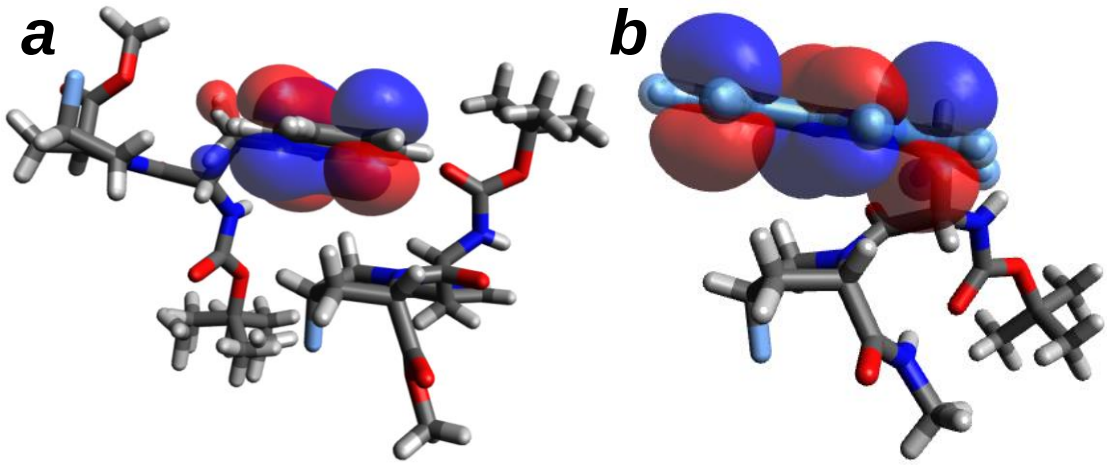


Figure 3.73 Potential orbital overlap interactions that stabilize prolyl C–H/π aromatic interactions
Prolyl C–H/π aromatic interactions observed in the Trp-flp crystal structures suggest stabilizing contributions from $\pi \rightarrow \sigma^*_{C-H}$ interactions, similar to the interactions discovered in Chapter 2. The HOMO π orbitals for an individual methyl-indole ring was calculated in GAMESS (RHF 6-31G),[195] and aligned to crystal structures of (a) Boc-Trpflp-OMe and (b) Boc-Trpflp-NHMe in Avogadro (contoured at 0.02).

The possibility of orbital overlap component to C–H/π interactions has been very briefly discussed previously in the literature, but the nomenclature has been somewhat inconsistent. In general, C–H/π interactions have been largely attributed to electrostatic contributions.[190, 225] In only one paper out of over 200 of his papers on this topic, Nishio discusses the orbital overlap contributions in the stability of C–H/π interactions,[106] describing these contributions as "delocalization or charge transfer interactions." In comparing C–H/π interactions to related hydrogen bonds, Nishio points out that "electrostatic stabilization decreases abruptly in the order OH/O > CH/O > C–H/π. In contrast, the delocalization energy decreases only moderately in this order, and becomes most important in the C–H/π interaction."[106] Due to the increasing relative contribution of orbital overlap interactions in C–H/π interactions, it was noted that maximizing the integral orbital overlap between occupied (π HOMO) and unoccupied (C–H σ* LUMO) will play a role in the geometry of these interactions.[106] Desiraju & Steiner[190] use similar terminology to describe orbital contributions to weak hydrogen bonds: "charge transfer involves transfer of electrons from an occupied orbital of one molecule to the unoccupied orbitals of the other, and is therefore conceptually similar to covalency." They describe strong and weak hydrogen bonds as a continuum between electrostatics forces and dispersive forces, where electrostatic, charge transfer interactions are stronger and more directional while dispersive/van der Waals forces are weaker and non-directional.[190] Specifically commenting on C–H/π interactions in this same work,[190] they state: "with falling C–H acidity, the hydrogen bond nature of C–H··Ph interactions becomes increasingly questionable. Although the interactions are still directional, that is they do not

represent van der Waals interactions, one could argue that a degree of long-range nature is not reached that would allow their classification as hydrogen bonds."

In our work, we observed an aromatic-cis-proline interaction with defined geometry that indicates an interaction that is more concerted than a dispersive van der Waals interaction. Biswal & Wategoankar[164] specifically noted significant energetic contributions from orbital overlap $\pi \rightarrow \sigma^*_{S-H}$, describing this interaction as "charge delocalization," and later called attention to this interaction as being distinct from classical hydrogen bonds. Further complicating discussion of these orbital interactions, they pose a concluding question: "what does one call these complexes which are stabilized entirely due to the dispersion interaction but do involve a hydrogen atom covalently bound to one atom and that interacts with another atom or group of atoms which are rich in electron density."[164] They tentatively suggest a new class of hydrogen bonds termed "dispersion-stabilized hydrogen bonds,"[164] although we believe this is as much of a misnomer as "charge-transfer interaction" in reference to $\pi \rightarrow \sigma^*_{X-H}$ orbital contributions.

Deciphering the relative energetic contributions of these ubiquitious X–H/π interactions has unrealized importance. The nature and relative contributions that stabilize noncovalent interactions can have significant implications in the geometric preferences of two interacting molecules, as described in Chapter 2 in comparing cation/π and S–H/π interactions. Misidentification of the relative energetic contributions in non-covalent interactions can skew results obtained from database analyses. Malone *et al.*[225] conducted a comprehensive database analysis on X–H/π interactions in crystal structures from the CSD, and identified many trends and geometries that furthered our understanding of X–H/π interactions in general. The

restrictive search parameters employed by Malone *et al.*[225] focused on X–H bond approaches to aromatic *centroids,* where structures with angles X–H··centroid lower than 150° were excluded.[225] These search parameters examined only centroid-oriented X–H/π interactions, either for practicality or with the assumption that X–H/π interactions are driven by electrostatic interactions (similar to cation/π). It is noteworthy that the crystal structure Boc-Trpflp-NHMe would have been excluded from this study using these search parameters, and perhaps countless other related structures have been excluded from similar database analyses. In addition, inappropriate parameters for interactions would impact molecular modeling or calculations, by biasing against certain geometries which are actually stabilizing, with preference for geometries that represent electrostatically driven interactions.

3.3.5 Aromatic-cis-proline interactions in proteins: Stabilization of the cis conformation due to a molecular orbital overlap

With the understanding that prolyl C–H/π aromatic interactions are potentially motivated by an orbital overlap interaction with aromatic ring carbons (where degenerate π orbitals would be located), we conducted a search of the Protein Data Bank (PDB) to gain insights into the prevalence of these interactions in proteins. The analysis of the PDB was conducted and analyzed by Dr. Himal Ganguly as part of this study. In non-redundant chains from protein crystal structures (< 2.0 Å resolution), the distance from the Pro Hα to the aromatic centroid and to the nearest aromatic ring carbon was obtained. Both ring centroids were considered in the case of tryptophan. His-cis-Pro motifs were excluded because the protonation state of histidine, which is unknown in specific proteins, can have significant consequences on the strength of the C–H/π interaction. The results from Dr. Ganguly's analysis are shown in Figure 3.74.

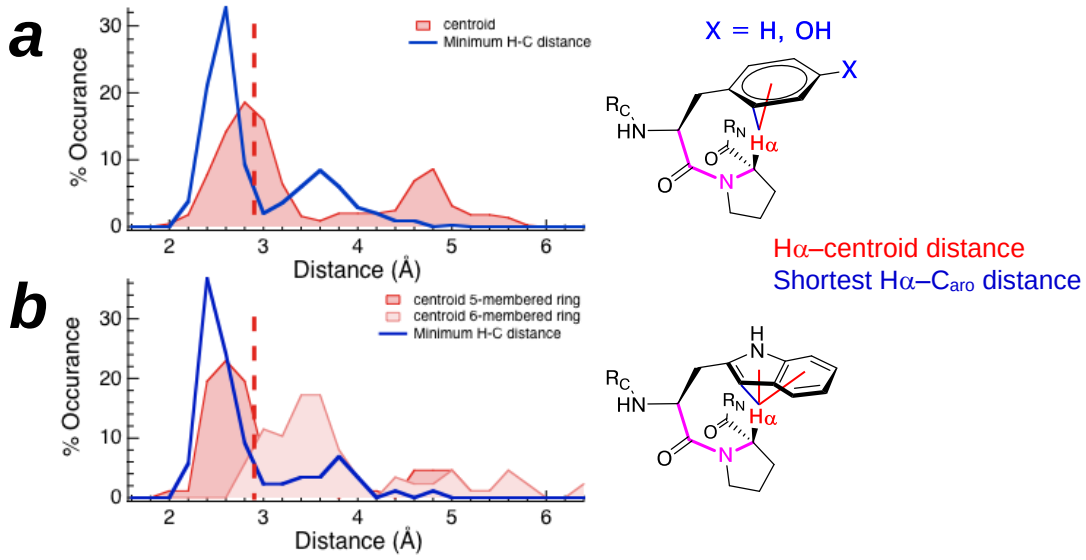


Figure 3.74 PDB analysis of aromatic-cis-proline motifs: Distances from Pro Hα to aromatic rings

Occurrences of Hα-aromatic distances in aromatic-cis-proline motifs in proteins from the PDB. The red line indicates the distance Hα–centroid and the blue line indicates the shortest distance Hα–C_{aro} (see description of d_{H-Cmin} in Chapter 2). (a) Phe-cis-Pro and Tyr-cis-Pro motifs in proteins; (b) Trp-cis-Pro motifs in proteins. The distance to the 5-membered ring centroid or the 6-membered ring centroid are indicated. The vertical dashed line is at 2.90 Å, the sum of the van der Waals radii for carbon and hydrogen atoms.

In this analysis of the PDB, 480 Phe-cis-Pro and Tyr-cis-Pro motifs were identified (combined) and 97 Trp-cis-Pro motifs were identified. In the majority of cases, the measured distances from ProHα to individual aromatic carbons were *less* than the distance to the aromatic centroid, for both Phe/Tyr-cis-Pro and Trp-cis-Pro. More than 60% of structures showed a distance ProHα–C_{aro} that was *less* than the sum of the van der Waals radii for carbon and hydrogen. These data also suggest that the observed geometries within our crystallographic studies of Boc-Trpflp dipeptides are

more general to peptide and protein structures, implying that C–H/π orbital overlap interactions are important in protein secondary structure, folding, and function.

The geometry of the aromatic-cis-proline interaction in the crystal structure of Boc-Trpflp-NHMe and the crystal structures from the PDB indicated that the proline-Hα/aromatic distances were often less than the sum of the van der Waals radii for carbon and hydrogen. The close distances of these C-H/π aromatic interactions are suggestive of a molecular orbital interaction, similar to the one observed in the crystal structure of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester that was described in detail in Chapter 2. In order to investigate the possibility of a molecular orbital interaction that stabilizes the aromatic-cis-proline conformation, calculations were performed on dipeptide motifs from protein crystal structures (Figure 3.75). In all of the examples shown, with Phe-, Tyr-, or Trp-cis-proline interactions, an overlap between the proline C-Hα σ* and an aromatic π orbital was evident.

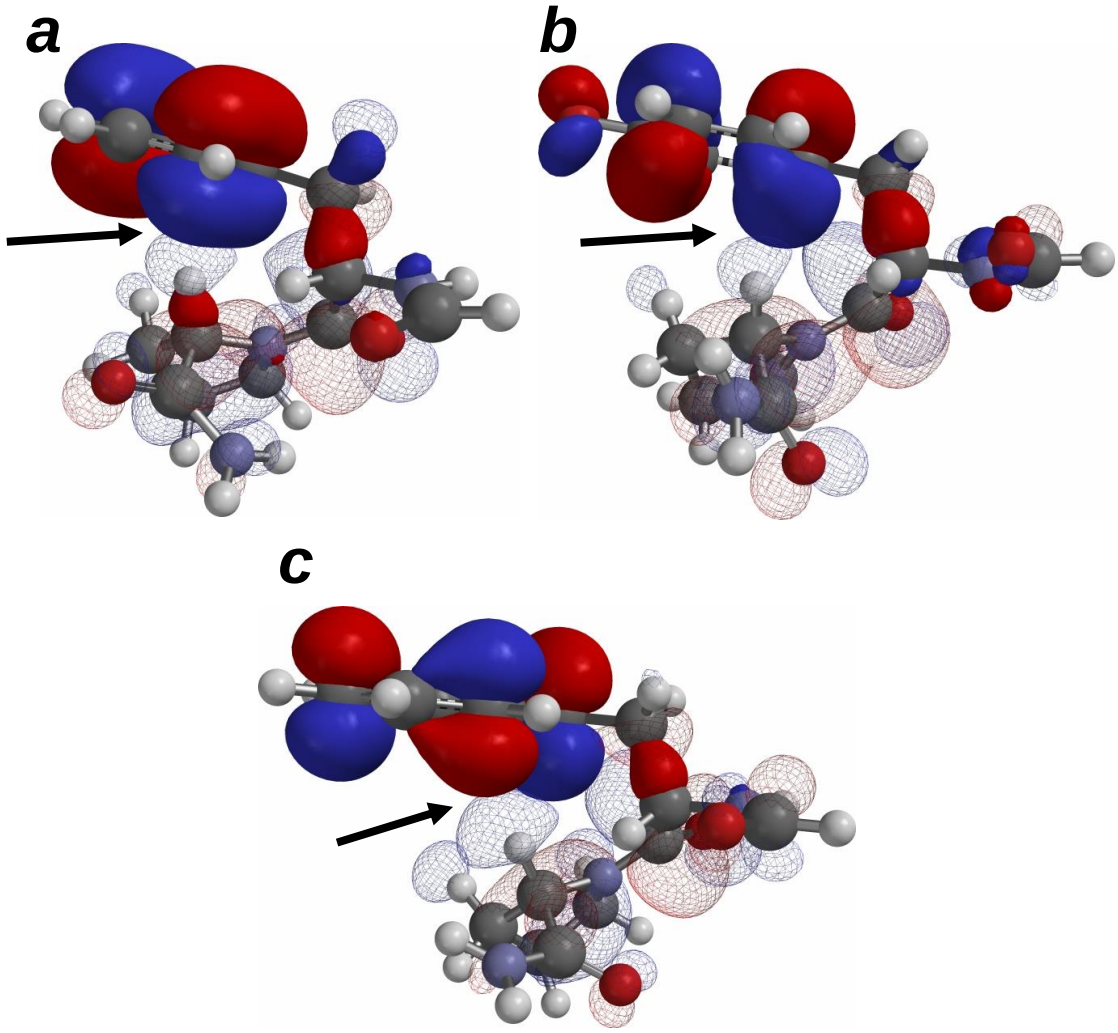


Figure 3.75 Calculated molecular orbitals in aromatic-cis-proline motifs from protein crystal structures from the PDB
Calculations for the highest occupied molecular orbitals (HOMOs) were conducted using Spartan 06 (6-31G). (a) Phe631-Pro632, PDB ID: 3SXX; (b) Tyr206-Pro207, PDB ID: 4DB8; (c) Trp207-Pro208, PDB ID: 4KB1.
In all three of these examples, an overlap between the C–H σ* orbital and the aromatic π orbital is apparent (indicated).

Calculations conducted on aromatic-cis-proline motifs from protein crystal structures from the PDB suggest that a molecular orbital interaction $\pi \rightarrow \sigma^*_{C-H}$ potentially contributes to the stability of the motif (indicated, Figure 3.75). The

generality of carbon-directed geometry of C–H/π interactions in aromatic-cis-proline motifs in crystal structures from the PDB, similar to the carbon-directed geometry in the S–H/π interaction described in Chapter 2, suggests that aromatic interactions in proteins are potentially stabilized by molecular orbital interactions. The energy and geometry of aromatic interactions based on molecular orbital interactions can potentially affect design principles in the context of designing therapeutic protein inhibitors, or modeling protein-protein interactions. In structure-activity relationship analysis of small molecule libraries containing aromatic rings, aromatic susbtituent effects must be considered in the context of electrostatic interactions *and* molecular orbital interactions.

In this work, we have determined that the nature of the C–H/π interaction which stabilizes aromatic-cis-proline type VIa1 β-turns, is an enthalpically-driven interaction with close contact distances between the proline hydrogens and the aromatic ring carbons. In this work, several novel synthetic methods were developed in order to generate non-natural aromatic amino acids, many with unique reactive or spectroscopic properties. Crystallographic analysis of the peptide Boc-Trpflp-NHMe revealed the preference for carbon-oriented geometry in C–H/π interactions, and these geometric preferences were also found in the majority of aromatic-cis-proline motifs from protein crystal structures. Combined with our conclusions from Chapter 2, we propse that the molecular orbital interaction $\pi \rightarrow \sigma^*_{X-H}$ is general and significant in X–H/π aromatic interactions.

3.4 Experimental

3.4.1 Materials

Natural and non-natural Fmoc-L-amino acids, Boc-L-amino acids, and L-amino acids were purchased from Novabiochem (San Diego, CA), Bachem (San Carlos, CA), or Chem-Impex (Wood Dale, IL). Rink amide MBHA resin, diisopropylethylamine (DIPEA), (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP), and O-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Chem-Impex. Acetic anhydride (Ac_2O), trifluoroacetic acid (TFA), phenol, thioanisole, triethylsilane (TES), triisopropylsilane (TIS), 5,5'-dithiobis(2-nitrobenzoic acid), bis(pinacolato)diboron, (2-biphenyl)di-*tert*-butylphosphine (JohnPhos), cesium carbonate (Cs_2CO_3), sodium hypochlorite (NaOCl), lithium hydroxide, N-ethylmaleimide (NEM), sodium hydrosulfite ($Na_2S_2O_4$), triphenylmethyl chloride (Trt-Cl), diisopropyl azodicarboxylate (DIAD), trimethylsilyldiazomethane, copper(I) iodide, 1,10-phenanthroline, thiolacetic acid, and 4-nitrobenzoic acid were purchased from Acros. Triflic anhydride, dithiothreitol (DTT), 1,1′-bis(diphenylphosphino)ferrocene, palladium(II) acetate ($Pd(OAc)_2$), pyrrolidine, (diethyamino)sulfur trifluoride (DAST), acyl chloride, triphenylphosphine, 2-iodoxybenzoic acid (IBX), methylamine hydrochloride, and formic acid were purchased from Aldrich. All deuterated solvents were purchased either from Acros or Cambridge Isotopes (Tewksbury, MA). Acetonitrile (MeCN), dimethylformamide (DMF), methylene chloride (CH_2Cl_2), methanol (MeOH), ether, pyridine, hydrogen peroxide (30% solution in water), potassium acetate, and ethylenediaminetetraacetic acid (EDTA) were purchased from Fisher. Ethanedithiol (EDT) was purchased from Pfaltz & Bauer (Waterbury, CT).

Tri(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Hampton Research (Aliso Viejo, CA). 4-Nitrobenzenediazonium tetrafluoroborate was purchased from TCI America (Portland, OR). Deionized water was purified by a Millipore Synergy 185 water purification system with a Simpak2 cartridge. Solid-phase post-synthetic modification reactions were performed in capped disposable fritted columns (Image Molding), or in glass vials (2 mL). All materials were used as purchased with no additional purification. $PdCl_2$(dppf) was prepared as reported previously.[391]

3.4.2 Peptide Synthesis and Characterization

Peptides (0.1 or 0.25 mmol) were synthesized manually or on a Rainin PS3 peptide synthesizer on Rink amide resin via standard Fmoc solid-phase peptide synthesis using HBTU as a coupling reagent. 60 minute coupling reactions were performed using 4 equivalents of Fmoc amino acid and HBTU. 3 equivalents were used for coupling reactions with non-canonical amino acids. All TXPN peptides contained N-terminal amides and C-terminal amides.

The peptides were subjected to cleavage and deprotection reaction for 2-4 hours under standard conditions (90% TFA/5% TIS/5% H_2O or 84% TFA/4% each of H_2O/phenol/thioanisole/ethanedithiol), unless otherwise indicated. The TFA was removed by evaporation under nitrogen. Peptides were precipitated from the mixture with cold ether, and the precipitate was dried. The peptides were dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptides were purified and the conversion and purity were determined using reverse phase HPLC on a Vydac C18 semi-preparative column (250 × 10 mm, 5-10 μm particle, 300 Å pore) or on a Varian Microsorb MV C18 analytical column (250 × 4.6 mm, 3-5 μm particle,

100 Å pore) using a linear gradient of buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA), unless otherwise noted. Peptide purity was verified via reinjection on an analytical HPLC column. Peptides were characterized by ESI-MS (positive ion mode) on an LCQ Advantage (Finnigan) mass spectrometer, or on an LCMS 2020 (Shimadzu) mass spectrometer (positive or negative ion mode). Synthesis and NMR characterization of all peptides containing 4-thiophenylalanine and related derivatives were described in Chapter 1. The peptides Ac-TWPN-NH_2, Ac-T(4-CF_3-Phe)PN-NH_2, Ac-T(pentafluoro-Phe)PN-NH_2, Ac-TFflpN-NH_2, Ac-TYflpN-NH_2, Ac-TWflpN-NH_2, Ac-TAflpN-NH_2, Ac-T(4-NH_2-Phe)flpN-NH_2, Ac-TFHypN-NH_2, Ac-TYHypN-NH_2, Ac-TWHypN-NH_2, Ac-TAHypN-NH_2, and Ac-T(4-NO_2-Phe)HypN-NH_2 were purified by Krista Thomas or Dr. Devan Naduthambi, and conditions are reported elsewhere.[91] The peptide Ac-T(4-OSO_3^--Phe)PN-NH_2 was prepared and purified by Michael Scheuermann, and the conditions are reported elsewhere.[392]

Table 3.15. Purification and ESI-MS of peptides made from commercially available amino acids

[a]Peptides were described previously,[91] but were re-synthesized for these studies.

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-TFPN-NH_2[a]	60 minutes 0-20% buffer B	30.2	518.3	541.1 $(M+Na)^+$
Ac-TYPN-NH_2[a]	60 minutes 0-20% buffer B	22.8	534.2	557.3 $(M+H)^+$
Ac-THPN-NH_2[a]	60 minutes 0-35% buffer B	10.2	508.2	509.3 $(M+H)^+$
Ac-TAPN-NH_2[a]	60 minutes 0-10% buffer B	11.5	442.2	443.1 $(M+H)^+$
Ac-TChaPN-NH_2[a]	60 minutes 0-20% buffer B	30.7	524.3	525.3 $(M+H)^+$
Ac-T(4-Pyridyl-Ala)PN-NH_2[a]	60 minutes 0-20% buffer B	13.8	519.2	542.4 $(M+Na)^+$

Table 3.15 continued

Ac-T(4-NH_2-Phe)PN-NH_2[a]	60 minutes 0-20% buffer B	6.3	533.3	556.4 $(M+Na)^+$
Ac-T(4-NO_2-Phe)PN-NH_2[a]	60 minutes 0-35% buffer B	33.8	563.2	564.1 $(M+H)^+$
Ac-T(4-F-Phe)PN-NH_2	60 minutes 0-20% buffer B	31.5	536.2	559.2 $(M+H)^+$
Ac-T(4-CN-Phe)PN-NH_2	60 minutes 0-20% buffer B	40.7	543.2	544.3 $(M+H)^+$
Ac-T(4-Br-Phe)PN-NH_2	60 minutes 0-35% buffer B	43.5	596.2	598.9 $(M+H)^+$
Ac-T(4-Cl-Phe)PN-NH_2	60 minutes 0-45% buffer B	36.6	552.2	575.3 $(M+Na)^+$

3.4.3 Synthesis of novel aromatic amino acids via modification within peptides Ac-TXPN-NH_2

Peptide modification reactions were performed either on the peptides Ac-TXPN on Rink amide resin (X = indicated commercially available amino acid), or in solution on the purified peptides Ac-TXPN-NH_2. Peptides Ac-TXPN-NH_2 containing 4-thiophenylalanine and related derivatives were described in Chapter 1. The HPLC conditions, retention times, and ESI-MS data are summarized in Table 3.16.

Table 3.16. Purification and ESI-MS of peptides with modified aromatic amino acids

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-T(4-$B(OH)_2$-Phe)PN-NH_2	60 minutes 0-45% buffer B	19.1	562.3	585.3 $(M+Na)^+$
Ac-T(4-Pyridyl(N-oxide)-Ala)PN-NH_2	60 minutes, isocratic A for 20 minutes then 0-20% buffer B over 40 minutes	15.0	535.2	536.2 $(M+H)^+$
Ac-T(4-pyrrolidyl-Phe)PN-NH_2	60 minutes 0-35% buffer B	18.1	587.3	588.1 $(M+H)^+$
Ac-T(4-OSO_2CF_3-Phe)PN-NH_2	30 minutes 0-45% buffer B	36.3	596.2	619.1 $(M+Na)^+$
Ac-T(4-OAc-Phe)PN-NH_2	60 minutes 0-15% buffer B	38.4	576.3	577 $(M+H)^+$

Table 3.16 continued

Ac-T(3-I-Tyr)PN-NH_2	60 minutes 0-35% buffer B	36.6	659.1	683.1 $(M+Na)^+$
Ac-T(3-NO_2-Tyr)PN-NH_2	60 minutes 0-15% buffer B	49.0	579.2	580.1 $(M+H)^+$
Ac-T(3-SH-Tyr)PN-NH_2	60 minutes 0-35% buffer B	30.7	565.2	589.3 $(M+Na)^+$
Ac-T(3-NH_2-Tyr)PN-NH_2	30 minutes 0-15% buffer B	12.1	549.3	550 $(M+H)^+$
Ac-T(3-OH-Tyr)PN-NH_2	60 minutes 0-15% buffer B	29.2	550.2	551 $(M+H)^+$
Ac-T(4-SH-Phe)flpN-NH_2	60 minutes 0-35% buffer B	24.9	568.2	591.4 $(M+Na)^+$
Ac-TWflpF-NH_2	60 minutes 0-70% buffer B	37.6	608.2	609.4 $(M+H)^+$

3.4.3.2 Solid-phase synthesis of the peptide Ac-T(4-B$(OH)_2$-Phe)PN-NH_2

Potassium acetate (7.7 mg, 78 μmol) was dissolved in DMSO (460 μL) via sonication. The resin containing the peptide Ac-TXPN (where X = 4-iodophenylalanine, 15-25 mg, 14-23 μmol) was placed in a 2 mL glass vial. $PdCl_2$(dppf) (1.7 mg, 2.3 μmol), dppf (1.1 mg, 1.3 μmol), and bis(pinacolato) diboron (13.2 mg, 52 μmol) were added to the resin. The mixture of potassium acetate in DMSO was added to the resin. The mixture was sparged briefly with nitrogen gas. The vial was sealed and the mixture was subjected to stirring in an oil bath at 80 °C for 24 h. The reagents were removed from the resin by filtration. The resin was washed with DMF (4 mL × 3) and CH_2Cl_2 (4 mL × 2), and dried with ether. The resin containing the peptide was subjected to cleavage and deprotection reactions using standard conditions to generate the peptide Ac-T(4-B$(OH)_2$-Phe)PN-NH_2. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.3 Solution phase synthesis of the peptide Ac-T(4-Pyridyl(N-oxide)-Ala)PN-NH_2

The purified, lyophilized peptide Ac-TXPN-NH_2 peptide (X = 4-pyridylalanine, 0.3-0.8 μmol) was dissolved in 20% acetonitrile in water (100 μL) containing phosphate buffer (pH 7.2, 100 mM). The resultant solution was subjected to incubation at 75 ˚C on a heating block, and then H_2O_2 was added (30% in H_2O, 30 μL). The resultant reaction solution was allowed to incubate at 75 ˚C for 1 h to generate the peptide Ac-T(4-pyridyl(N-oxide)-Ala)PN-NH_2. After the reaction was allowed to cool to room temperature, the peptide was purified directly from the resultant reaction solution via HPLC using the conditions described in Table 3.16.

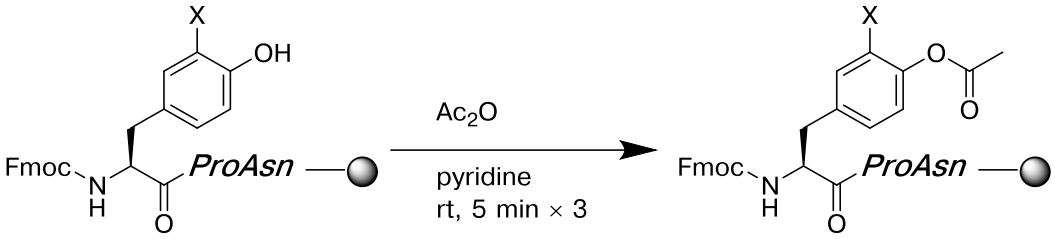
3.4.3.4 Solid-phase synthesis of the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2

The resin containing the peptide Ac-TXPN (X = 4-bromophenylalanine, 15-25 mg, 14-23 μmol) was placed in a 2 mL glass vial. Palladium(II) acetate (1.1 mg, 5 μmol), JohnPhos (6.0 mg, 20 μmol), cesium carbonate (456 mg, 1.4 mmol), and THF (1 mL) were added to the resin in the vial. The mixture was sparged briefly with nitrogen gas, and pyrrolidine (99 μL, 1.2 mmol) was added to the resultant mixture. The vial was sealed and the mixture was allowed to stir in an oil bath at 60 °C for 24 h. The reagents were removed from the resin via filtration. The resin was washed with DMF (4 mL × 3), CH_2Cl_2 (4 mL × 2), and dried with ether. The resin was subjected to cleavage and deprotection reaction under standard conditions to generate the peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.5 Solid-phase synthesis of the peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2

The resin containing the peptide Ac-TYPN, (trityl-protected tyrosine, 0.3-0.8 μmol) was allowed to swell in DMF (4 mL) for 20 minutes, and then washed with DCM (4 mL × 4). The trityl protection on tyrosine was selectively removed from the peptide using a solution of TFA (2%) and TES (5%) in CH_2Cl_2 (3 mL × 2 min × 3), and then the resin was washed with CH_2Cl_2 (4 mL × 4). Pyridine (1.5 mL) was added to the resin and the resultant mixture was placed in an ice bath, and the mixture was allowed to chill to 4 °C. Triflic anhydride (75 μL, 45 μmol) was added dropwise to the resin, and the mixture was allowed to stir at 4 °C for 15 minutes. The reagents were removed from resin by filtration, and the resin was allowed to react with triflic anhydride in pyridine at 4 °C for an additional 15 minutes as described. The resin was then washed with CH_2Cl_2 (4 mL × 4) and then dried with ether. The resultant peptide Ac-T(4-OSO_2CF_3-Phe)PN-NH_2 was subjected to cleavage and deprotection reaction on ice at 4 °C using 5% TIS in TFA for 1.5 h to generate the peptide. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16. A major side product was observed via HPLC: t_R 34.0 min, exp. 648.2, obs. 649.0 $(M + H)^+$, which was consistent with a β-elimination product of the peptide.

3.4.3.6 General solid phase acetylation reaction of Ac-T(3-Z-Tyr)PN-NH_2 (X = H, I, NO_2)



The peptide Fmoc-(3-X-Tyr)PN-NH_2 (X = H, I, or NO_2) was synthesized on Rink amide resin via SPPS. Immediately following the coupling reaction with Fmoc-(3-X-Tyr) and the peptide on resin, the resin was washed with DMF (3 mL × 3). The resin-bound peptide was subjected to reaction with a solution of 5% acetic anhydride in pyridine (3 mL × 5 min × 3). After this protection reaction, the peptide was synthesized using standard reaction conditions. The resin containing the synthesized peptide was subjected to cleavage and deprotection reaction using standard TFA conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in 10 mM LiOH in phosphate buffer (pH 4.0, 50 mM) allowed to incubate at room temperature for 30 minutes to generate the peptides containing 3-substituted tyrosine. The solution containing the crude peptide was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.15, 3.16, or Table 3.17.

Table 3.17. Purification and ESI-MS of peptides with modified aromatic amino acids: 4-O(acetyl)-Tyr and derivatives

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-T(4-OAc-Tyr)PN-NH_2	60 minutes 0-15% buffer B	38.4	576.3	577 $(M+H)^+$
Ac-T(3-I-Phe)PN-NH_2	60 minutes 0-35% buffer B	36.6	659.1	683.1 $(M+Na)^+$
Ac-T(4-OAc, 3-NO_2-Phe)PN-NH_2	60 minutes 0-15% buffer B	49.0	579.2	580.1 $(M+Na)^+$

3.4.3.7 Solution phase synthesis of the peptide Ac-T(3-OH-Tyr)PN-NH_2

The purified, lyophilized peptide Ac-TYPN-NH_2 (0.1 mmol) and 2-iodoxybenzoic acid (IBX, 11.2 mg, 45% wt, 18 μmol) were dissolved in THF (125 μL) while protected from light. The resultant solution was allowed to stir in the dark at room temperature for 1.5 h. $Na_2S_2O_4$ (5.5 mg, 32 μmol) was dissolved in water (125 μL) and the resultant solution was added to the solution containing the peptide. The resultant solution was allowed to stir at room temperature in the dark for 45 min to generate the peptide Ac-T(3-OH-Tyr)PN-NH_2. The reaction was quenched with MeOH (10 μL), the reaction solution was washed with ether (1 mL × 2), and the aqueous layer was filtered on a 0.45 μm nylon syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.8 Solution phase synthesis of the peptide Ac-T(3-NH_2-Tyr)PN-NH_2

The purified, lyophilized peptide Ac-TYPN-NH_2 was dissolved in phosphate buffer in water (100 μL, pH 9.0, 100 mM) and placed on ice. To the peptide, 6.0 μL of a solution of 4-nitrobenzenediazonium tetrafluoroborate solution (130 mM in DMF) was added, and the solution was allowed to stir on ice for 15 minutes. The solution was neutralized to pH 7 using 1 M HCl, and then solid $Na_2S_2O_4$ (10 mg, 57 μmol) was

added to the resultant solution. The solution was allowed to stir at room temperature for 1 hour to generate the peptide Ac-T(3-NH_2-Tyr)PN-NH_2. The solution containing the crude peptide was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

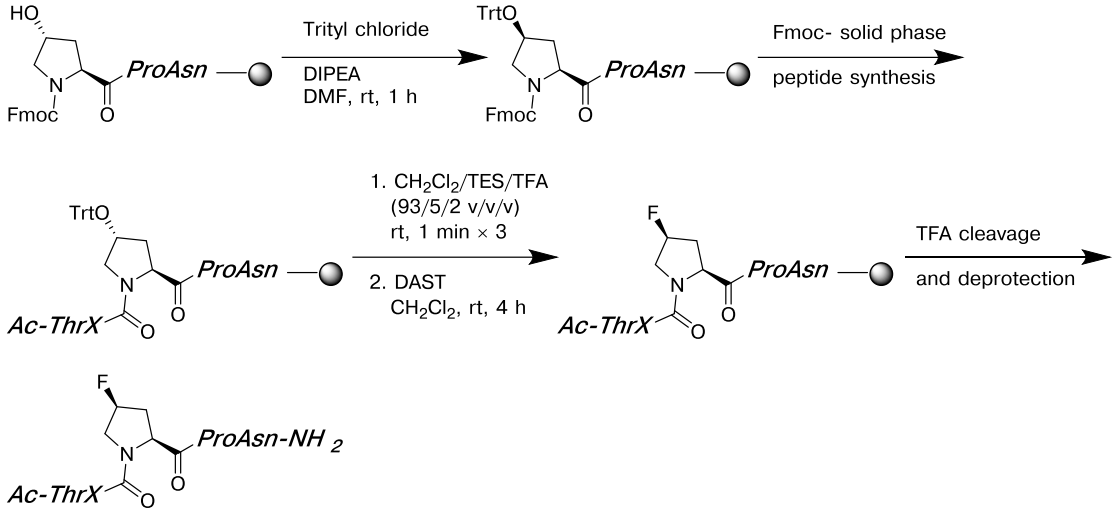
Alternately, the purified, lyophilized peptide Ac-T(3-NO_2-Tyr)PN-NH_2 was dissolved in phosphate buffer (100 μL, pH 6.0, 100 mM). Solid $Na_2S_2O_4$ (10 mg, 57 μmol) was added to the solution containing the peptide. The resultant solution was allowed to stir for 1 h at room temperature to generate the peptide Ac-T(3-NH_2-Tyr)PN-NH_2. The solution containing the crude peptide was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.9 Solid-phase synthesis of the peptide Ac-T(3-SH-Tyr)PN-NH_2

Following the previously developed copper-mediated cross-coupling reaction on solid phase,[169] the resin containing the protected peptide Ac-T(3-I-Tyr)PN-NH_2 (the tyrosine hydroxyl was protected by an acetyl group as described above, 20-30 mg, 7-10 μmol) was placed in a glass vial. Toluene (400 μL), DIPEA (66 μL, 0.4 mmol), copper(I) iodide (3.6 mg, 20 μmol), 1,10-phenanthroline (7.2 mg, 40 μmol), and thiolacetic acid (34.0 μL, 0.48 μmol) were added to the resin sequentially. The vial was sealed and the mixture was allowed to stir in an oil bath at 110 °C for 16-18 h. The reagents were removed from the resin by filtration. The resin was washed with DMF (4 mL × 4), CH_2Cl_2 (4 mL × 2), and MeOH (4 mL × 2) and dried with ether. The resin was subjected to cleavage and deprotection reactions under standard conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The

peptide was dissolved in a solution containing LiOH (10 mM) and DTT (25 mM) in phosphate buffer (pH 7.2, 50 mM) to generate the peptide Ac-T(3-SH-Tyr)PN-NH_2. The resultant solution was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.10 General solid-phase reaction to generate peptides containing 4*S*-fluoroproline (flp)[374]



The peptide Fmoc-HypN-NH-Resin was synthesized on Rink amide resin using standard Fmoc- solid-phase peptide synthesis. In order to protect the hydroxyproline, 10 equivalents (2.5 mmol) of trityl chloride dissolved in 2 mL of CH_2Cl_2 were added to the resin. To the resin was added 8% DIPEA in DMF (2 mL), and the resultant mixture was allowed to agitate on a rotary shaker for 60 minutes at room temperature. The reagents were removed from the resin by filtration, and the resin was washed with DMF (4 mL × 4). The peptide was synthesized using standard Fmoc- solid-phase peptide synthesis. Following the peptide synthesis and acetylation reaction, the trityl group was selectively removed from the protected peptide by subjecting the resin to reaction with a solution of 5% triethylsilane (TES) and 2% TFA in CH_2Cl_2 (4 mL × 1 min × 3), and the resin was then washed with CH_2Cl_2 (4 mL × 3). A solution of 0.5 M DAST in anhydrous CH_2Cl_2 was added to the resin, and the resultant mixture was allowed to agitate on a rotary shaker at room temperature for 4 h. The resin was washed with CH_2Cl_2 and MeOH, and dried with ether. The resin was

subjected to cleavage and deprotection reactions under standard conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.11 Solid-phase synthesis of the peptide Ac-T(4-SH-Phe)flpN-NH_2

The resin containing the protected peptide Ac-T(4-I-Phe)flpN-NH_2 (20-30 mg, 7-10 μmol, synthesized using the general procedure described above) was placed in a glass vial. Toluene (400 μL), DIPEA (66 μL, 0.4 mmol), copper(I) iodide (3.6 mg, 20 μmol), 1,10-phenanthroline (7.2 mg, 40 μmol), and thiolacetic acid (34.0 μL, 0.48 μmol) were added to the resin sequentially. The vial was sealed, and the resultant mixture was allowed to stir in an oil bath at 110 °C for 16-18 h. The reagents were removed from resin by filtration. The resin was washed with DMF (4 mL × 4), CH_2Cl_2 (4 mL × 2), and MeOH (4 mL × 2) and dried with ether. The resin was subjected to cleavage and deprotection reactions under standard conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The precipitated peptide was subjected to reaction with a solution of DTT (25 mM) in phosphate buffer (pH 7.2, 50 mM) to generate the peptide Ac-T(4-SH-Phe)flpN-NH_2. The solution containing the peptide was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

3.4.3.12 Modified solid-phase synthesis of the peptide Ac-TWflpF-NH_2

A protocol was developed for synthesis of the difficult peptide sequence, TWflpF. The peptide Fmoc-HypF-NH-Resin was synthesized via Fmoc- solid-phase peptide synthesis. A solution of 0.5 M DAST in anhydrous CH_2Cl_2 was added to the resin, and the resultant mixture was allowed to agitate on a rotary shaker at room temperature for 4 h. The reagents were removed from the resin by filtration. A solution of 0.5 M DAST in anhydrous CH_2Cl_2 was added to the resin, and the resultant mixture was allowed to agitate on a rotary shaker at room temperature for an additional 4 h. The reagents were removed from the resin by filtration. In order to protect the hydroxyproline, 10 equivalents (2.5 mmol) of trityl chloride dissolved in 2 mL of CH_2Cl_2 were added to the resin. To the resin was added 8% DIPEA in DMF (2 mL), and the resultant mixture was allowed to agitate on a rotary shaker for 60 minutes at room temperature. The reagents were removed from the resin by filtration, and the resin was washed with DMF (4 mL × 4). The peptide was then synthesized using Fmoc solid-phase peptide synthesis. The resin was washed with DMF (4 mL × 4), CH_2Cl_2 (4 mL × 2), and MeOH (4 mL × 2) and dried with ether. The resin was subjected to cleavage and deprotection reactions under standard conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried to generate the peptide Ac-TWflpF-NH_2. The peptide was dissolved in phosphate buffer (pH 4) and the solution containing the peptide was filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC using the conditions described in Table 3.16.

Attempts to subject the synthesized, protected peptide to the reaction with DAST were examined as described in Chapter 3.4.3.9, but the reaction proceeded with

low conversion, likely due to aromatic stacking of the Phe and Trp, which could sterically hinder the reaction site.

3.4.3.13 General solid-phase synthesis of peptides containing 2S,4*S*-(*p*-nitrobenzoyl)-hydroxyproline (4-PNB-hyp)

The peptides were synthesized to contain trityl-protected hydroxyproline as described above (Chapter 3.4.3.9). Following the peptide synthesis and acetylation reaction, the trityl group was selectively removed from the protected peptide by subjecting the resin to reaction with a solution of 2% TFA and 5% triethylsilane (TES) in CH_2Cl_2 (4 mL × 1 min × 3), and the resin was then washed with CH_2Cl_2 (4 mL × 3). The resin was then placed in CH_2Cl_2 (2 mL), and triphenylphosphine (262 mg, 1 mmol) and 4-nitrobenzoic acid (167 mg, 1 mmol) were added. The resultant mixture was allowed to chill on ice for 15 minutes. DIAD (197 μL, 1 mmol) was added dropwise to the mixture under nitrogen. The reaction vessel was sealed under nitrogen. The resultant mixture was allowed to incubate on ice for 30 minutes, and then the mixture was allowed to agitate on a rotary shaker at room temperature for 18-24 h. The resin was washed with CH_2Cl_2, MeOH, and dried with ether. The resin was subjected to cleavage and deprotection reactions under standard conditions. The TFA was removed by evaporation under nitrogen. Peptide was precipitated from the mixture with cold ether, and the precipitate was dried. The peptide was dissolved in water or phosphate buffer, and then filtered using a 0.45 μm syringe filter. The peptide was purified via HPLC.

3.4.4 NMR Spectroscopy

Purified, lyophilized peptides were dissolved in buffer containing 5 mM phosphate (pH 4.0 or as indicated), 25 mM NaCl, and 100 μM TSP in 90% H_2O/10%

D_2O. The pH was adjusted as needed using dilute HCl (100 mM) or NaOH (100 mM). 100 μM TCEP was added to peptides containing thiophenylalanine to prevent disulfide formation. 100 μM EDTA was added to peptides containing disubstituted aromatic amino acids to prevent potential metal binding. Peptide concentrations were 10 μM-200 μM. NMR spectra were collected at 298 K on a Brüker AVN 600 MHz NMR spectrometer equipped with a triple resonance cryoprobe or a TXI probe. Spectra were internally referenced with TSP. 1-D spectra were collected with a Brüker w5 watergate pulse sequence and a relaxation delay of 1.7-2.0 s. 2-D spectra were collected with a watergate TOCSY pulse sequence.

Well-resolved peaks in the NMR spectra were integrated after phasing and baseline correction. $K_{trans/cis}$ was calculated based on the average of integrated ratios of 2-3 pairs of peaks. In general, backbone amide peaks and Thr methyl peaks were used for integration, but other pairs of resonances were used when amide peaks were too broad or obscured by spectral overlap.

For NMR experiments in non-aqueous solvents, MeCN-d_3, 90% MeOH-d_3/10% MeOD-d_4 , or benzene-d_6, trace TMS was added as an internal reference. NMR spectra were integrated after baseline normalization, and the $K_{trans/cis}$ was calculated based on the average integrated ratios of 2-3 pairs of peaks. Amide protons were used preferentially because they were generally well resolved and exhibit rapid relaxation, but when necessary, Thr methyl protons, N-terminal acetyl methyl protons, Thr Hα or Hβ protons, or aromatic protons were used for integration. For van′t Hoff analyses, ^{1}H NMR spectra were collected after allowing the sample to equilibrate at each temperature for at least 20 minutes. The data represents the average of at least two independent trials unless otherwise indicated.

3.4.5 Fluorescence Spectroscopy

The purified, lyophilized peptides Ac-T(4-pyrrolidyl-Phe)PN-NH_2 or Ac-TYPN-NH_2 were dissolved in buffer containing 5 mM phosphate (pH 2.5 or 7.2). Samples contained 100 μM of the peptide. Fluorescence spectra were collected in a 1 cm quartz cell (Starna) on a Photon Technology International fluorescence spectrometer model QM-3/2003 with a CW source and a Hamamatsu R928 PMT. All slit widths were 3 nm. Excitation scans were obtained from 200 nm to 400 nm, with emission detection at 365 nm. Emission scans were obtained from 275 nm to 475 nm with excitation at 265 nm, or from 305 nm to 505 nm with excitation at 300 nm. Data were the average of at least three independent trials. Data were background-corrected but were not smoothed. Error bars are shown and indicate standard error.

3.4.6 UV-Vis Spectroscopy

The purified, lyophilized peptide Ac-T(4-pyrrolidyl-Phe)PN-NH_2 was dissolved in buffer containing 50 mM phosphate; samples contained 250-75 μM peptide, determined via NMR. The purified, lyophilized peptide Ac-T(3-SH-Tyr)PN-NH_2 was dissolved in buffer containing 50 mM phosphate and 1 mM TCEP; samples contained 45-65 μM peptide, determined via Ellman's test.[109] Absorbance spectra were collected on a Perkin-Elmer Lambda 25 UV-Vis spectrometer in a 1 cm cell. Absorbance spectra were obtained from 400 nm to 220 nm with a slit width of 1 nm. After each measurement, the pH of each sample was verified with a pH electrode (Mettler Toledo), and the absence of disulfide formation was confirmed via HPLC reinjection. Data were the average of at least three independent trials. Error bars are shown and indicate standard error.

3.4.7 Synthesis of dipeptides in solution phase

3.4.7.1 Synthesis of the dipeptide series Ac-AP-XMe and Ac-WP-XMe: X = O or NH

The dipeptides based on Ac-AP or Ac-WP (0.15 mmol) were synthesized manually on 2-chlorotrityl resin via standard Fmoc solid-phase peptide synthesis using HBTU as a coupling reagent. 60 minute coupling reactions were performed with 3 equivalents of Fmoc amino acid and HBTU. The peptides contained an acetyl group at the N-terminus, which was incorporated by subjecting the peptide on resin to a solution of DMF/2,6-lutidine/acetic anhydride (89/6/5 v/v/v) for 2 minutes (4 mL × 2). The resin was washed thoroughly with CH_2Cl_2. The resin containing the peptides were subjected to cleavage and deprotection reactions for 90 minutes using standard conditions (90% TFA/5% TIS/5% H_2O). The TFA was removed via evaporation under nitrogen. The crude peptide products were precipitated using ether. The mixture was centrifuged, and the ether was removed. The precipitate was dried, and then dissolved in water (10 mL). The mixture was filtered, and the solution containing the peptide was lypophilized.

Methyl ester modification at the C-terminus (to generate the peptides Ac-AP-OMe or Ac-WP-OMe) was conducted in solution phase on the crude, dried precipitate containing the peptide from resin. The precipitate containing the peptide Ac-AP or Ac-WP was dissolved in THF (500 μL), and the solution was placed under a nitrogen atmosphere and chilled on ice. Trimethylsilyldiazomethane (0.5 mL) was added dropwise to the chilled solution via syringe, and the solution was allowed to stir while chilled on ice for 1 hour. The reaction was quenched by slowly adding drops (10 μL) of acetic acid until the evolution of gases subsided. The resulting solution was diluted

with water (3 mL) and filtered. The crude peptides Ac-AP-OMe or Ac-WP-OMe were purified via HPLC.

Methyl amide modification at the C-terminus (to generate the peptides Ac-AP-NHMe or Ac-WP-NHMe) was conducted in solution phase on the crude, dried peptides from resin. The crude precipitate containing the peptide Ac-AP or Ac-WP was dissolved in 1,4-dioxane (350 μL). DIPEA (49 μL, 0.23 mmol) and PyBOP (86 mg, 0.17 mmol) were added to the solution containing the peptide. The peptide was allowed to stir at room temperature for 30 minutes, and then chilled on ice. Separately, methylamine hydrochloride ($MeNH_2 \cdot HCl$, 100 mg, 1.5 mmol) was dissolved in cold water (350 μL) and placed on ice. LiOH (72 mg, 3.0 mmol) was added to the mixture, and the solution was maintained on ice in a closed container. The solution containing methylamine added to the chilled solution containing the peptide, and the resultant mixture was allowed to stir in a sealed container. The solution was allowed to warm to room temperature over 6 hours while stirring. After 6 hours, an additional volume of the chilled solution containing methylamine (10 equivalents) and LiOH (20 equivalents) in chilled water (350 μL) was prepared as described. The additional solution of methylamine was added to the solution containing the peptide. The resultant solution was allowed to stir in a sealed container at room temperature overnight. The solution containing the peptide was filtered and diluted in water. The crude peptides Ac-AP-NHMe or Ac-WP-NHMe were purified via HPLC.

The peptides were purified using reverse phase HPLC on a Vydac C18 semi-preparative column (250 × 10 mm, 5-10 μm particle, 300 Å pore) or on a Varian Microsorb MV C18 analytical column (250 × 4.6 mm, 3-5 μm particle, 100 Å pore) using a linear gradient of buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A

(98% water, 2% MeCN, 0.06% TFA). Peptide purity was verified via reinjection on an analytical HPLC column. Peptides were characterized by ESI-MS (positive ion mode) on an LCQ Advantage (Finnigan) mass spectrometer.

Table 3.18. Purification and ESI-MS of dipeptides

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-AP-OMe	30 minutes, isocratic A for 10 minutes then 0-10% buffer B over 20 minutes	25.5	242.1	243.0 $(M+H)^+$
Ac-AP-NHMe	40 minutes, isocratic A for 20 minutes then 0-5% buffer B over 20 minutes (at 75% normal flow rate)	18.9	241.1	242.2 $(M+H)^+$
Ac-WP-OMe	40 minutes 15-65% buffer B	17.6	357.2	358.1 $(M+H)^+$
Ac-WP-NHMe	30 minutes 0-35% buffer B	23.3	356.2	357.2 $(M+H)^+$

3.4.7.2 Synthesis of Boc-2*S*,4*R*-hydroxyproline

2*S*,4*R*-Hydroxyproline (2.0 g, 15 mmol) and Boc_2O (3.3 mg, 15 mmol) were dissolved in 1,4-dioxane (30 mL). A solution of NaOH (1.0 M, 30 mL) in water was added to the solution containing 2*S*,4*R*-hydroxyproline, and the resultant mixture was allowed to stir at room temperature for 18 hours, or until the starting material was depleted as observed via thin-layer chromatography. Dilute HCl (1 M, 40 mL) was added to the solution (pH 2), and the product was extracted into ethyl acetate (20 mL × 3). The organic layers were combined, and the solvent was removed under reduced pressure, generating Boc-2*S*,4*R*-hydroxyproline (2.9 g, 13 mmol) in 84% yield. The

crude product was used in the next step without further purification. The ^{1}H NMR and ^{13}C NMR spectra of the resultant product corresponded with literature values.[393]

3.4.7.3 Synthesis of Boc-2*S*,4*R*-hydroxyproline methyl ester

Crude Boc-2*S*,4*R*-hydroxyproline (24.9 g, 108 mmol) was dissolved in acetone (1 L), and potassium carbonate was added (45.5 g, 329 mmol). Dimethyl sulfate (15.3 mL, 161 mmol) was added to the mixture while the mixture was allowed to stir. The mixture heated to reflux and allowed to stir for 3 hours. The solvent was removed under reduced pressure, the crude residue was redissolved in ethyl acetate (200 mL × 3). The organic layer was washed with water (200 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (0 to 4% methanol in CH_2Cl_2 (v/v)) to yield Boc-2*S*,4*R*-hydroxyproline methyl ester (20 g, 82 mmol) as a colorless oil in 76% yield. The ^{1}H NMR and ^{13}C NMR spectra of the resultant product corresponded with literature values.[393]

3.4.7.4 Synthesis of Boc-2*S*,4*S*-fluoroproline methyl ester

Using a previously established protocol,[375] purified Boc-2*S*,4*R*-hydroxyproline methyl ester (400 mg, 1.6 mmol) was dissolved in anhydrous CH_2Cl_2 (815 μL), and the solution was placed under a nitrogen atmosphere. The solution was placed on ice and allowed to stir for 10 minutes. DAST (0.37 mL, 2.8 mmol) was added to the solution dropwise, and the resultant solution was allowed to stir on ice for 2 hours. The solution was then allowed to warm to room temperature, and then allowed to stir for an additional 16 hours at room temperature. DAST was quenched by adding a solution of saturated sodium bicarbonate (2 mL). When the evolution of gases had subsided, water (25 mL) was added to the resultant mixture, and the crude product was

extracted into CH_2Cl_2 (20 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (0 to 30% ethyl acetate in hexanes (v/v)) to yield Boc-2*S*,*4S*-fluoroproline methyl ester (260 mg, 1.05 mmol) as a colorless oil in 64% yield. NMR spectra reflected a mixture of cis and trans proline rotamers (1:1 ratio, respectively). ^{1}H NMR (600 MHz, $CDCl_3$): δ 5.25 (d, J = 3.8 Hz), 5.17 (d, J = 4.0 Hz) (sum of 5.25 ppm and 5.17 ppm resonances = 1H), 4.53 (d, J = 9.7 Hz), 4.42 (d, J = 9.1 Hz) (sum of 4.53 ppm and 4.42 ppm resonances = 1H), 3.85-3.60 (m), 3.74 (s) (sum of 3.74 ppm and 3.73 ppm resonances = 5H), 2.50-2.30 (m, 2H), 1.48 (s), 1.43 (s) (sum of 1.48 ppm and 1.43 ppm resonances = 9H). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 172.17, 171.81, 153.93, 153.5 6, 92.74, 91.67, 91.58, 90.50, 80.25, 80.21, 57.55, 57.18, 53.20, 53.04, 52.87, 52.71, 37.40, 37.26, 36.56, 36.41, 28.27, 28.15. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ -172.94, -173.25. ESI-MS m/z: $[M]^+$ calcd for $C_{11}H_{18}FNO_4$ 247.1, found 270.0 ($M + Na^+$). These data corresponded with literature values.[375]

3.4.7.5 Synthesis of Boc-Trp-(2*S*,*4S*-fluoroproline) methyl ester

Boc-2*S*,*4S*-fluoroproline methyl ester (1.5 g, 6.0 mmol) was dissolved in a solution of 10% TFA in CH_2Cl_2 (24 mL), and the solution was allowed to stir at room temperature for 3 hours, or until disappearance of starting material was observed via thin-layer chromatography. The solvent was removed under reduced pressure, and the crude product was redissolved in acetonitrile (12 mL). DIPEA was added to the mixture until the solution was at pH 8 (approximately 3 mL). Separately, Boc-Trp-OH (2.7 g, 7.0 mmol) and PyBOP (4.7 g, 7.0 mmol) were dissolved in a solution of 10% DIPEA in acetonitrile (12 mL), and the resultant solution was allowed to stir at room temperature for 30 minutes. The solution of Boc-Trp-OH and PyBOP was added to the

mixture of Boc-2*S*,*4S*-fluoroproline methyl ester, and then the resultant mixture was allowed to stir at room temperature for 24 hours. DIPEA was added as necessary during the reaction to maintain basic conditions (pH 8). Distilled water was added to the resultant reaction mixture (20 mL), and the crude product was extracted into ethyl acetate (20 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (50 to 100% hexanes in ethyl acetate (v/v)) to yield Boc-Trp-(2*S*,*4S*-fluoroproline) methyl ester (780 mg, 1.8 mmol) as a white solid in 30% yield. NMR spectra reflected a mixture of cis and trans proline rotamers (1:1.2 ratio, respectively). ^{1}H NMR (600 MHz, $CDCl_3$): δ 8.46 (br s), 8.37 (br s) (sum of 8.46 ppm and 8.37 ppm resonances = 1H), 7.70 (d, J = 7.7 Hz), 7.50 (d, J = 7.9 Hz), 7.35 (d, J = 8.0 Hz), 7.32 (d, J = 8.1 Hz), 7.20-7.13 (m), 7.09 (br s), 7.04 (t, J = 7.4 Hz) (sum of 7.70 ppm, 7.50 ppm, 7.35 ppm, 7.32 ppm, 7.17 ppm, 7.09 ppm, and 7.04 ppm resonances = 5H), 5.64 (d, J = 8.3 Hz), 5.32 (d, J = 8.5 Hz) (sum of 5.64 ppm and 5.32 ppm resonances = 1H), 5.09 (d, J = 52.3 Hz), 4.80 (d, J = 9.8 Hz), 4.76-4.67 (m), 4.52-4.48 (m) (sum of 5.09 ppm, 4.72 ppm, and 4.50 ppm resonances = 2H), 3.80-3.61 (m), 3.72 (s), 3.63 (s) (sum of 3.72 ppm and 3.63 ppm resonances = 3H), 3.42 (d, J = 9.1 Hz), 3.38-3.35 (m), 3.33-3.26 (m), 3.19-3.10 (m) (sum of 4.80 ppm, 3.70 ppm, 3.42 ppm, 3.36 ppm, 3.30 ppm, and 3.15 ppm resonances = 5H), 2.47-2.41 (m), 2.27-2.15 (m), 2.07-2.02 (m) (sum of 2.44 ppm and 2.04 ppm resonances = 1H), 1.47 (s), 1.40 (s) (sum of 1.47 ppm and 1.40 ppm resonances = 9H), 0.56-0.44 (m) (sum of 2.22 ppm and 0.50 ppm resonances = 1H). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 171.83, 171.64, 170.86, 155.41, 154.89, 136.10, 136.95, 122.44, 121.94, 119.57, 119.53, 118.54, 111.27, 111.21, 111.16, 111.00, 109.58, 92.62, 91.44, 90.82, 89.66, 79.82, 79.62, 57.24, 57.16, 54.08, 53.03, 52.99,

52.87, 52.82, 52.50, 52.43, 36.63, 36.49, 35.59, 35.44, 30.96, 28.62, 28.40, 28.35. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ -173.17, -173.70. ESI-MS m/z: $[M]^+$ calcd for $C_{22}H_{28}FN_3O_5$ 433.2, found 456.2 ($M + Na^+$).

3.4.7.6 Synthesis of Boc-Trp-(2*S*,*4S*-fluoroproline)

Boc-Trp-(2*S*,*4S*-fluoroproline) methyl ester (200 mg, 0.46 mmol) was dissolved in 1,4-dioxane (0.92 mL). A solution of LiOH in water (0.75 mM, 0.92 mL) was added to the solution containing the dipeptide, and the resultant solution was allowed to stir at room temperature for 10 hours, or until disappearance of starting material was observed via thin-layer chromatography. Dilute HCl (1 M, 5 mL) was added to the mixture to reduce the pH to 2, and the crude product was extracted into ethyl acetate (10 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (0 to 4% methanol in CH_2Cl_2 (v/v)) to yield Boc-Trp-(2*S*,*4S*-fluoroproline)-OH (180 mg, 0.43 mmol) as a white solid in 93% yield. NMR spectra reflected a mixture of cis and trans proline rotamers (1:1.5 ratio, respectively). A polar cosolvent was required for NMR analysis due to the limited solubility of Boc-Trp-(2S,*4S*-fluoroproline) in $CDCl_3$. The cosolvent broadened the proton resonances, particularly on the exchangeable amide and carbamate protons. ^1H NMR (600 MHz, $CDCl_3$ + EtOAc): δ 8.77 (br s), 8.72 (br s) (sum of 8.77 ppm and 8.72 ppm resonances = 0.3H), 7.66 (d, J = 7.6 Hz), 7.49 (d, J = 7.6 Hz), 7.35-7.32 (m), 7.18-7.11 (m), 7.05 (t, J = 7.1 Hz) (sum of 7.66 ppm, 7.49 ppm, 7.34 ppm, 7.15 ppm, 7.05 ppm = 5H), 5.81 (d, J = 7.9 Hz), 5.32 (br s) (sum of 5.64 ppm and 5.32 ppm resonances = 0.3H), 5.10 (d, J = 52.1 Hz), 4.79 (d, J = 9.4 Hz), 4.68 (br s), 4.48-4.47 (m) (sum of 5.10 ppm, 4.68 ppm, and 4.48 ppm resonances = 2H), 3.81 (dd, J = 35.0, 11.5 Hz), 3.81 (dd, J = 26.6, 14.2 Hz), 3.42-3.11 (m) (sum

of 4.79 ppm, 3.81 ppm, 3.66 ppm and 3.27 ppm = 5H), 2.60-2.55 (m), 2.32-2.19 (m), 2.10-2.06 (m) (sum of 2.58 ppm and 2.07 ppm resonances = 1H), 1.45 (s), 1.41 (s) (sum of 1.45 ppm and 1.41 ppm resonances = 9H), 0.56-0.46 (m) (sum of 2.24 ppm and 0.51 ppm resonances = 1H). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 172.40, 172.09, 171.80, 155.52, 155.32, 127.52, 127.19, 124.40, 123.50, 122.34, 121.78, 119.42, 118.83, 118.20, 111.39, 111.20, 110.57, 108.51, 92.46, 91.27, 90.83, 89.67, 80.08, 80.04, 57.64, 57.06, 54.12, 53.48, 53.31, 52.94, 52.52, 36.48, 36.34, 35.02, 34.92, 30.76, 29.68, 28.32. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ -173.53, -173.84. ESI-MS m/z: $[M]^+$ calcd for $C_{21}H_{26}FN_3O_5$ 419.2, found 442.2 ($M + Na^+$).

3.4.7.7 Synthesis of Boc-Trp-(2*S*,*4S*-fluoroproline) methyl amide

Boc-Trp-(2*S*,*4S*-fluoroproline) (80 mg, 0.19 mmol) and PyBOP (110 mg, 0.21 mmol) were dissolved in a solution of 10% DIPEA in 1,4-dioxanes (450 μL). The resultant solution was allowed to stir for 30 minutes at room temperature. Separately, methylamine hydrochloride (130 mg, 1.9 mmol) was added to a solution of LiOH in water (8.4 M, 450 μL) and allowed to chill on ice. The neutralized solution of methylamine was added to the dipeptide in solution, and the resultant solution was allowed to stir in a sealed container at room temperature for 6 hours. An additional 10 equivalents of chilled, neutralized methylamine solution were added to the reaction mixture, and the mixture was allowed to stir in a sealed container at room temperature for an additional 18 hours. Distilled water was added to the mixture (3.0 mL), and the crude product was extracted into ethyl acetate (10 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (50 to 100% ethyl acetate in hexanes (v/v)) to yield Boc-Trp-(2*S*,*4S*-fluoroproline)-methyl ester (50 mg, 0.12 mmol) as a white solid in 61% yield. NMR spectra reflected

a mixture of cis and trans proline rotamers (1.8:1 ratio, respectively). ^{1}H NMR (600 MHz, $CDCl_3$): δ 8.98 (br s), 8.87 (br s) (sum of 8.98 ppm and 8.87 ppm resonances = 1H), 7.73 (d, *J* = 7.8 Hz), 7.65 (d, *J* = 4.1 Hz), 7.45 (d, *J* = 8.0 Hz), 7.39 (d, *J* = 8.0 Hz), 7.35 (d, *J* = 8.2 Hz), 7.26 (t, *J* = 7.1 Hz), 7.18-7.15 (m), 7.08 (d, *J* = 2.1 Hz), 7.05 (t, *J* = 7.4 Hz) (sum of 7.73 ppm, 7.45 ppm, 7.39 ppm, 7.35 ppm, 7.26 ppm, 7.17 ppm, 7.08 ppm, and 7.05 ppm resonances = 5H), 5.43 (d, *J* = 4.1 Hz) (sum of 7.65 ppm and 5.43 ppm resonances = 1H), 5.34 (d, *J* = 8.3 Hz), 5.30 (d, *J* = 6.1 Hz) (sum of 5.34 ppm and 5.30 ppm resonances = 1H), 5.04 (d, *J* = 52.1 Hz), 4.50 (dd, *J* = 15.5, 8.1 Hz), 4.68 (d, *J* = 52.2 Hz) (sum of 5.04 ppm and 4.68 ppm resonances = 1H), 4.68 (d, *J* = 9.9 Hz), 4.23 (m) (sum of 4.50 ppm and 4.23 ppm resonances = 1H), 3.92 (ddd, *J* = 36.2, 12.4, 3.6 Hz), 3.61-3.44 (m), 3.33-3.19 (m), 3.13 (dd, *J* = 13.6, 4.7 Hz) (sum of 4.68 ppm, 3.92 ppm, 3.53 ppm, 3.43 ppm and 3.13 ppm resonances = 5H), 2.75 (d, *J* = 4.7 Hz), 2.65 (t, *J* = 16.3 Hz), 2.45 (d, *J* = 4.7 Hz) (sum of 2.75 ppm and 2.45 ppm resonances = 3H), 2.19 (t, *J* = 15.5 Hz) (sum of 2.65 ppm and 2.19 ppm resonances = 1H), 2.10-1.98 (m), 1.45 (s, 9H) 0.22-0.10 (m) (sum of 2.04 ppm, 0.16 ppm resonances = 1H). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 173.08, 172.80, 170.65, 170.54, 155.86, 155.37, 136.26, 136.13, 126.93, 126.79, 123.71, 122.93, 122.70, 122.63, 120.17, 119.61, 118.85, 118.51, 111.79, 111.49, 109.72, 92.56, 91.38, 91.10, 89.94, 80.57, 80.34, 59.28, 58.74, 54.76, 54.23, 54.07, 53.84, 53.68, 52.41, 36.46, 36.32, 34.90, 34.76, 29.25, 28.77, 28.37, 28.30, 26.38, 26.28. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ -173.98, -174.38. ESI-MS m/z: $[M]^+$ calcd for $C_{22}H_{29}FN_4O_4$ 433.2, found 456.2 (M + Na^+).

3.4.7.8 Synthesis of Boc-TrpPro methyl ester

Proline methyl ester (HCl salt, 125 mg, 0.754 mmol) and LiOH (20 mg, 0.83 mmol) were dissolved in acetonitrile (500 μL). Boc-Trp-OH (300 mg, 0.986 mmol), PyBOP (513 mg, 0.987 mmol), and cesium carbonate (493 mg, 1.52 mmol) were dissolved in acetonitrile (1 mL). Both of the solutions containing each amino acid were stirred vigorously or sonicated until the reagents were fully dissolved. The solution containing proline methyl ester was added to the solution containing activated Boc-Trp-OH, and the resultant solution was allowed to stir for 24 hours at room temperature. Distilled water was added to the mixture (10 mL), and the crude product was extracted into ethyl acetate (10 mL × 3). The solvent was removed under reduced pressure. The crude product was purified via column chromatography (10 to 50% ethyl acetate in hexanes (v/v)) to yield Boc-TrpPro methyl ester (140 mg, 0.34 mmol) as a white solid in 45% yield. NMR spectra reflected a mixture of cis and trans proline rotamers (1:3.4 ratio, respectively). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.43 (br s), 8.39 (br s) (sum of 8.43 ppm and 8.39 ppm resonances = 1H), 7.69 (d, J = 7.7 Hz), 7.58 (d, J = 7.9 Hz) (sum of 7.69 ppm and 7.58 ppm resonances = 1H), 7.35 (d, J = 7.8 Hz), 7.32 (d, J = 7.6 Hz) (sum of 7.35 ppm and 7.32 ppm resonances = 1H), 7.19-7.06 (m, 3H), 5.55 (d, J = 8.7 Hz), 5.33 (d, J = 8.6 Hz) (sum of 5.55 ppm and 5.33 ppm resonances = 1H), 4.78 (dd, J = 14.7, 7.1 Hz, 1H), 4.50 (m, 1H), 3.71 (s), 3.64 (s) (sum of 3.71 ppm and 3.64 ppm resonances = 3H), 3.57-3.52 (m), 3.48-3.41 (m), 3.28-3.22 (m), 3.18-3.11 (m) (sum of 3.55 ppm, 3.45 ppm, 3.25 ppm and 3.15 ppm resonances = 4H), 2.18-2.09 (m), 1.93-1.81 (m), 1.66-1.50 (m), 1.45 (s), 1.40 (s) (sum of 1.45 ppm and 1.40 ppm resonances = 9H), 0.91-0.83 (m) (sum of 2.14 ppm, 1.87 ppm, 1.58 ppm and 0.87 ppm resonances = 4H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 172.69, 172.12, 171.49, 171.25, 155.39, 154.92, 136.07, 136.01, 127.91, 127.33,

123.88, 123.21, 122.29, 121.86, 119.62, 119.50, 119.04, 118.70, 111.20, 111.13, 110.87, 109.90, 79.65, 79.52, 58.89, 58.81, 53.78, 52.74, 52.41, 52.27, 46.92, 46.17, 30.79, 30.09, 29.08, 28.67, 28.43, 28.41, 24.94, 22.07. ESI-MS m/z: $[M]^+$ calcd for $C_{22}H_{29}N_3O_5$ 415.2, found 438.2 ($M + Na^+$).

3.4.7.9 Synthesis of Ac-Trp(2*S*,*4S*-fluoroproline) methyl ester

Purified Boc-Trp(2*S*,*4S*-fluoroproline) methyl ester (55 mg, 0.13 mmol) was dissolved in 1,4-dioxane (0.64 mL). A solution of HCl in water (4 M, 0.64 mL) was added to the dipeptide, and the resultant solution was stirred at room temperature for 2 hours, and then at 40 °C for 2 hours. Reaction completion was verified by observed depletion of Boc-Trp-(2*S*,*4S*-fluoroproline) methyl ester via thin-layer chromatography and NMR. The solvent was removed under reduced pressure, and the crude residue was redissolved in acetonitrile (1 mL). DIPEA was added dropwise until the solution was neutralized (pH 7-8). While stirring the solution containing Trp(2*S*,*4S*-fluoroproline) methyl ester, acetyl chloride (13.5 μL, 0.19 mmol) was added to the solution, and the resultant solution was allowed to stir at room temperature for 3 hours, or until the disappearance of starting material was observed via thin-layer chromatography. Dilute HCl (1 M, 5 mL) was added to the solution, and the crude product was extracted into ethyl acetate (10 mL × 3). The solvent was removed under reduced pressure, and the crude residue was redissolved in 50% acetonitrile in water to yield Ac-Trp(2*S*,*4S*-fluoroproline) methyl ester. The peptide was purified using reverse phase HPLC on a Varian Microsorb MV C18 analytical column (250 × 4.6 mm, 3-5 μm particle, 100 Å pore) using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes: t_R 16.5 min, exp. 375.2, obs. 398.4 $(M + Na)^+$.

NMR spectra reflected a mixture of cis and trans proline rotamers (1:1.2 ratio, respectively). ^{1}H NMR (600 MHz, $CDCl_3$): δ 8.16 (br s), 8.14 (br s) (sum of 8.16 ppm and 8.14 ppm resonances = 1H), 7.70 (d, *J* = 7.8 Hz), 7.51 (d, *J* = 7.9 Hz), 7.37 (d, *J* = 8.0 Hz), 7.35 (d, *J* = 8.1 Hz), 7.31 (d, *J* = 2.1 Hz), 7.21-7.14 (m), 7.07 (t, *J* = 7.8 Hz) (sum of 7.70 ppm, 7.51 ppm, 7.37 ppm, 7.35 ppm, 7.31 ppm, 7.18 ppm, and 7.07 ppm resonances = 5H), 6.64 (d, *J* = 7.6 Hz), 6.34 (d, *J* = 7.9 Hz) (sum of 6.64 ppm and 6.34 ppm resonances = 1H), 5.14 (dt, *J* = 52.3, 3.8 Hz), 4.99 (dt, *J* = 7.9, 5.3 Hz), 4.81 (d, *J* = 9.8 Hz), 4.78-4.67 (m) (sum of 5.14 ppm, 4.99 ppm, and 4.72 ppm resonances = 2H), 3.86-3.62 (m), 3.76 (s), 3.63 (s) (sum of 3.76 ppm and 3.63 ppm resonances = 3H), 3.42-3.31 (m), 3.22 (dd, *J* = 14.6, 5.2 Hz), 3.08 (dd, *J* = 13.7, 11.0 Hz) (sum of 4.81 ppm, 3.74 ppm, 3.37 ppm, 3.22 ppm, and 3.08 ppm resonances = 5H), 2.52-2.46 (m), 2.32-2.21 (m), 2.08-2.03 (m) (sum of 2.49 ppm and 2.05 ppm resonances = 1H), 2.07 (s), 1.94 (s) (sum of 2.07 ppm and 1.94 ppm resonances = 3H), 0.52-0.41 (m) (sum of 2.26 ppm and 0.47 ppm resonances = 1H). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 171.53, 171.26, 170.68, 170.58, 170.00, 169.27, 136.04, 127.75, 127.23, 124.14, 123.41, 122.62, 122.09, 119.74, 119.67, 118.99, 118.47, 111.28, 111.17, 111.02, 109.35, 92.54, 91.35, 90.71, 89.55, 57.18, 57.17, 53.31, 53.15, 53.04, 52.93, 52.89, 52.48, 51.30, 36.52, 36.38, 35.62, 35.48, 30.40, 28.23, 23.34, 23.12. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ -173.32, -173.68

Chapter 4

SYNTHESIS OF 2-THIOPHENYLALANINE AND ITS APPLICATION INTO NATIVE CHEMICAL LIGATION OF PEPTIDES AND PROTEINS

4.1 Introduction

Advancements in cloning and transformation techniques have transformed the landscape of molecular biology over the past 50 years. The ability to incorporate non-natural amino acids has broadened the genetic code, and the ability to incorporate amino acids with unique functionalities into proteins has revolutionized our ability to rationally engineer proteins and other biomolecules. However, there are limitations to the nature and size of the non-natural amino acids that can be expressed into proteins, depending on the specific technique utilized, thus limiting the scope of amino acids that can be studied and utilized in this manner.[394, 395] Thus, the ability to synthesize proteins in a chemical manner is a crucial part of the general "toolbox" in molecular science. Indeed, our understanding about the structural, chemical, and biological consequences of protein phosphorylation and glycosylation has relied on the ability to site-selectively modify proteins, but new methods and techniques are required to broaden the scope for synthetic proteins.

Peptides and small proteins can be chemically synthesized in a step-wise, sequence-specific manner from amino acids via Fmoc- or Boc- solid-phase peptide synthesis (SPPS). With this step-wise methodology, modified amino acids can be easily incorporated into a protein sequence, without the need for recombinant DNA or expression in cell cultures. Solid-phase peptide synthesis has provided tools for

studying residue effects on local protein conformations and the implications of post-translational modifications on protein structure, but SPPS becomes much more difficult with peptides longer than 50 amino acids. While protein expression has become routine practice for larger peptides and proteins, it is difficult to incorporate site-specific modifications (such as glycosylation).[394, 395] Chemical ligation, which combines protein expression and chemical synthetic methods, has been an active area of research over the past 50 years. A major advancement in the chemical ligation of proteins was pioneered by Dawson *et al.*[396] where a 72-amino acid protein was synthesized completely in a chemical manner via "native chemical ligation" (NCL). NCL and related methodologies have allowed for the chemical synthesis and semi-synthesis of proteins and peptides.[394, 395] NCL is a practical means for synthesizing native proteins and modified proteins that contain large or bulky non-natural amino acids, which cannot be otherwise expressed into proteins. A recent "tour-de-force" of NCL was reported by Wang *et al.*,[397, 398] which culminated in a ten-year effort to synthesize a single glycoform of human erythropoietin, a 166-amino acid protein containing 4 different oligosaccharide modifications.

NCL relies on the nucleophilicity of thiolates to undergo acyl transfer reactions at physiological pH, which occurs widely in natural biological processes (i.e. acetyl Coenzyme A). NCL was initially developed as a chemical synthetic strategy using cysteine, but the synthesis and applications of other thiol-containing amino acids and functional auxiliaries have been developed for use in NCL since the early 2000s.[399] Combined with desulfurization reactions (e.g. desulfurization reaction on cysteine to generate alanine[400]), nearly any amino acid can be used for ligation of peptides and proteins, including valine, alanine, lysine, and phenylalanine, among others.[399]

Aromatic amino acids are crucial for protein folding, intermolecular protein-protein interactions, and substrate recognition and binding.[18] β-Mercaptophenylalanine has been synthesized and demonstrated as a suitable means to ligate peptides and proteins using NCL with an aromatic amino acid, and can generate phenylalanine at the ligation site following a desulfurization reaction. However, β-mercaptophenylalanine was generated in 12% overall yield in 9 synthetic steps[401, 402] and has not been applied since it was first demonstrated for use in NCL in the mid-2000s,[401, 403] suggesting a need for an alternative strategy to synthesize a thiolated phenylalanine derivative. Given the practicality of the copper-mediated cross-coupling approach described in Chapter 1,[169, 404] we sought to synthesize and apply 2-thiophenylalanine for NCL of peptides and proteins. 2-Thiophenylalanine may also have unique advantages in NCL, given that aryl thiols are more acidic than alkyl thiols, and aryl thiolates are potentially more reactive than alkyl thiolates under mildly acidic conditions. The synthesis and application of 2-thiophenylalanine for NCL of model peptides and miniproteins is herein described.

4.1.1 Chemical synthesis of peptides

One of the crucial advancements towards the total chemical synthesis of proteins was the development of solid-phase peptide synthesis (SPPS) by Merrifield.[405, 406] By 1960, the synthesis of peptides and proteins was possible via solution phase, step-wise synthesis using carbodiimide coupling reagents and protected amino acids. By the time Merrifield introduced his solid-phase methodology, several peptides had been chemically synthesized using solution approaches, including vasopressin derivatives[407-410] and oxytocin.[411] However, the

numerous purification steps that were required for a solution-phase synthesis of a long peptide fragment undoubtedly limited broader applications of peptide synthesis.[412]

In 1963, Merrifield reported the first step-wise synthesis of a peptide bound to a solid, polystyrene support (Figure 4.1).[405] By immobilizing the peptide to a solid support, a simple washing procedure replaced the tedious purification and recrystallization steps that followed each amino acid coupling reaction in solution. The simplicity of SPPS allows for the automated peptide synthesis of peptides and proteins, and also allows for the practical incorporation of non-natural amino acids or the synthesis of small molecules and peptoids. After step-wise chemical synthesis, the full-length peptide can be released from the solid support via a cleavage reaction, and then the peptide is purified. In addition, the protected peptide can be subjected to additional chemistries for modifications, including C-terminal labelling or side-chain modifications. Many biologically active peptides and proteins have been synthesized using SPPS approaches derived from Merrifield's work, including the total synthesis of ribonuclease A,[413] both L- and D- variants of HIV-1 protease (99-amino acids),[414] the synthesis of the green fluorescent protein (238 residues),[415] and countless other examples.[394, 395, 412]

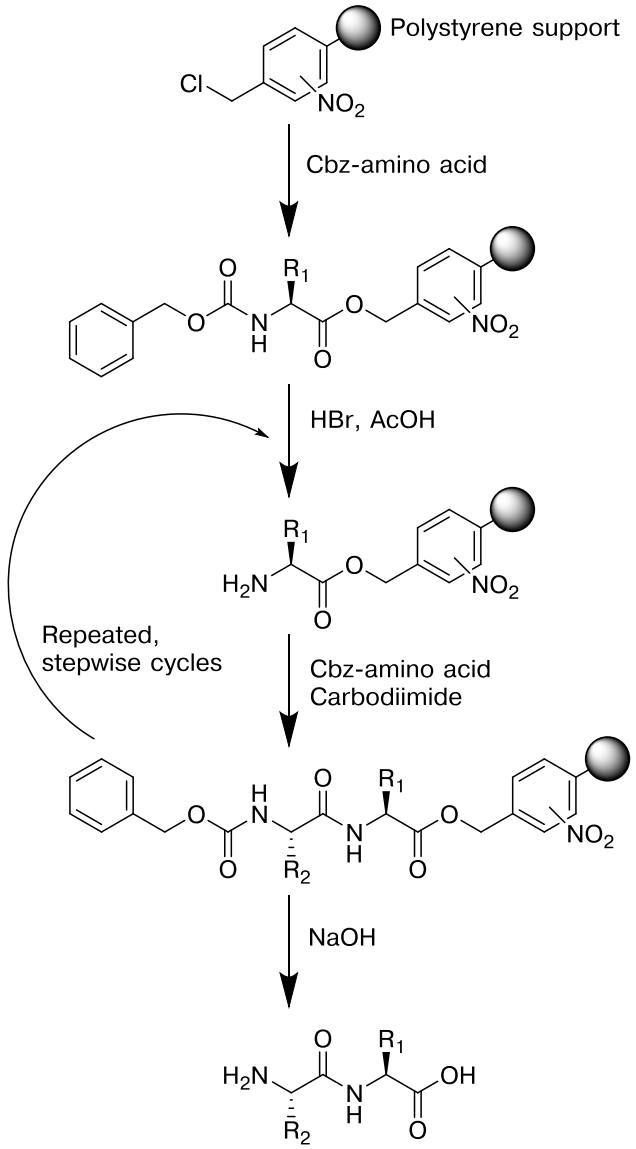


Figure 4.1 Synthetic strategy of stepwise, solid-phase peptide synthesis
Since it was initially proposed,[405] SPPS has been improved through broader availability of amine and side-chain protecting groups, improved efficiency in the coupling and deprotection reactions, and development of solid supports with improved stability.

With the widespread development and application of SPPS techniques, much has been discovered regarding protein structure, peptide-based inhibitors in medicinal chemistry, and modes of biological action of post-translational modifications, but there are some limitations associated with SPPS. As a synthetic peptide increases in length on resin, there is considerable steric crowding within the resin support, which can potentially result in decreased efficiency of the coupling reactions, and an increased likelihood of significant impurities in the final peptide product.[395] In peptides and proteins that are rich in aromatic amino acids or prolines, the problems associated with steric crowding on resin are more pronounced. With a step-wise approach on a long peptide, even the most efficient amino acid coupling reactions will ultimately result in a statistical accumulation of impurities that increases with peptide length (i.e. 98% coupling reaction efficiency over 45 amino acids = 40% overall yield).[416] Some peptide sequences are more prone to reaction inefficiency, such as short segments of sterically hindered amino acids or proline-rich segments, which further increases the likelihood of low overall yield and difficult purification. With these limitations in step-wise chemical synthesis, SPPS is generally limited to applications involving peptides approximately 50-60 amino acids in length or less, considerably smaller than the majority of proteins.[395, 416]

In order to chemically synthesize peptides and proteins that are longer than 50-60 amino acids, techniques were sought and developed for ligation reactions of peptide fragments, which can be synthesized via SPPS and then "stitched" together using ligation reactions following synthesis and purification.[417] Ideally, these peptide ligation reactions would be highly selective, would be orthogonal to biological functional groups, and would not introduce any large chemical linkers that could

potentially affect protein folding or dynamics. Numerous bioconjugation strategies have been introduced and reviewed using various synthetic methodologies.[51, 133, 417-419]

In one example, Schnölzer & Kent[420] reported a ligation methodology in the total synthesis of HIV-1 protease using a thioester in place of an amide bond. A thioester linkage is similar in length to a glycyl amide bond, and allows for bond rotation similar to amide bonds. In this work, two peptides of approximately 50 amino acids in length were synthesized via SPPS: HIV-$1_{1\text{-}50}$ containing a C-terminal thiocarboxylic acid, and HIV-$1_{53\text{-}99}$ containing an N-terminal bromoacetamide. The ligation reaction proceeded through nucleophilic attack of the thioacid on the methylene carbon to displace the bromide, uniting the two peptide fragments through a modified backbone (Figure 4.2).[420] The ligation reaction between these two peptide fragments was complete in 3 hours, and the resultant protein had similar enzymatic activity to the native protein.[420] It was noted that this "pseudo-peptide bond" could impact the hydrogen bonding capability of the full length protein at the ligation site, but it was not found to affect the activity or protein folding in this case.

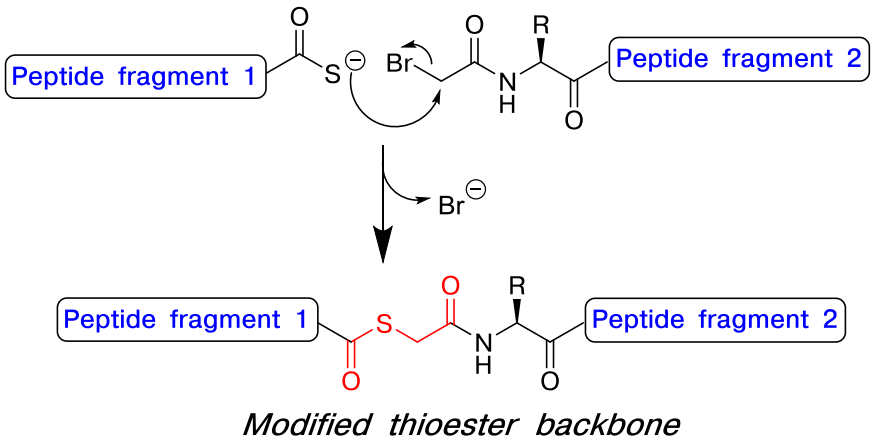


Figure 4.2 Chemical ligation of proteins with a modified thioester backbone[420] In Schnölzer & Kent's work, two peptide fragments were synthesized, one containing a C-terminal thioacarboxylic acid and the other containing an N-terminal bromoacetamide. Under neutral, aqueous conditions, the peptide fragments were ligated within 3 hours to form a full length, modified HIV-1 protease. At the ligation site, a glycine thioester is generated, where a sulfur atom replaces an amide N–H. The modified HIV-1 protease exhibited similar activity to the native protein.

In order to incorporate potential hydrogen bond donors and acceptors into modified amide bonds, Rose[421] reported a ligation strategy employing an oxime ligation reaction between peptide fragments to generate dendrimers nearly 20 kD in size. In this example, two peptide fragments containing either an N-terminal aminoxy moiety or an aldehyde were allowed to react under aqueous conditions to generate a peptide with an internal oxime in place of an amide bond (Figure 4.3). Although this reaction is slower than Kent's chemical ligation approach (16 h vs. 3 h), thiol chemistry is completely avoided with Rose's method,[421] which can potentially form unreactive disulfides during the ligation reaction, and oxime formation is highly selective over imine formation or other potential products.

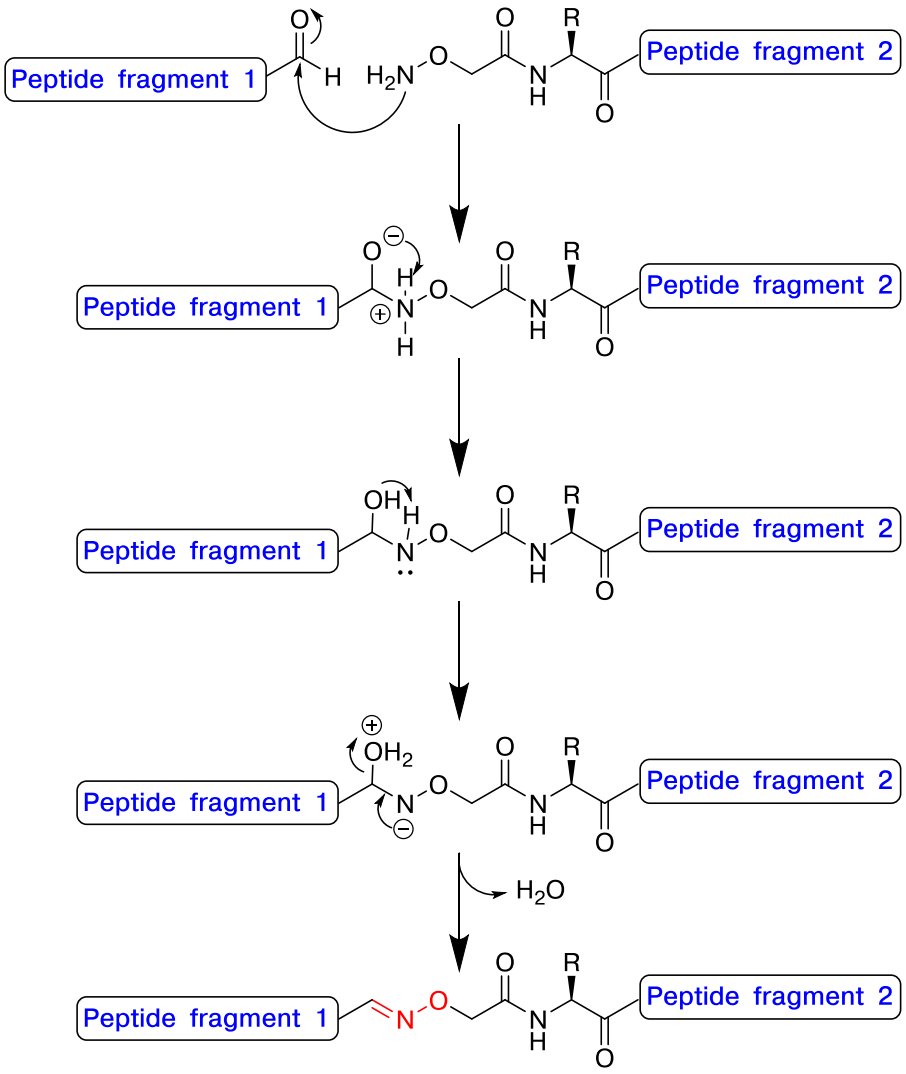


Figure 4.3 Synthesis of proteins with a via oxime ligation reaction[421]
In Rose's work, two peptide fragments were synthesized, one containing a C-terminal aldehyde and the other containing an N-terminal oxyamide. Under mildly acidic, aqueous conditions, the peptide fragments were ligated within 18 hours, with an oxime generated at the ligation site.

Development of chemical ligation reactions for synthetic peptides allowed for the total synthesis of large bioactive proteins with site-specific modifications. Many other chemical ligation and bioconjugation methods have been reported and

extensively utilized, including Staudinger ligation, Huisgen [3 + 2] cycloaddition, Diels-Alder cycloaddition, olefin cross-metathesis reaction, thiol-ene and thiol-yne reaction, among others.[51, 418, 422] These bioconjugation reactions can also be utilized for generating hybrid materials, such as electrode surfaces functionalized with proteins or protein-therapeutic conjugates.[362, 418, 423-426] Although some of the unnatural linkers generated from the ligation reactions were well-tolerated within the protein target, this is not always be the case, and "traceless" ligation methods that generate a native amide bond have been sought for the chemical synthesis of proteins.

4.1.2 Chemical synthesis of proteins via native chemical ligation

One of the earliest reported traceless ligation approaches in peptides was described by Kemp and coworkers, termed as "thiol capture".[427, 428] Kemp *et al.*[429] had reported amide bond formation via a template-assisted, intramolecular acyl transfer reaction (Figure 4.4), and highlighted the potential applications of this approach in peptide synthesis. It was noted that the reaction efficiency would be dictated by the "capture" of the peptide fragments, since entropic activation of the peptide fragments effectively replaced the need for carbonyl activation using carbodiimides.[427] Furthermore, cysteine presented a unique opportunity for extending "thiol capture" to ligation of peptides, where a cysteine can be "captured" as a disulfide or using an organomercury compound. However, these intramolecular acyl transfer reactions to form amide bonds were comparatively slow (requiring several hours to several days),[427, 428] and these methods were seldom applied in peptides and proteins, predominantly due to the subsequent report of native chemical ligation.

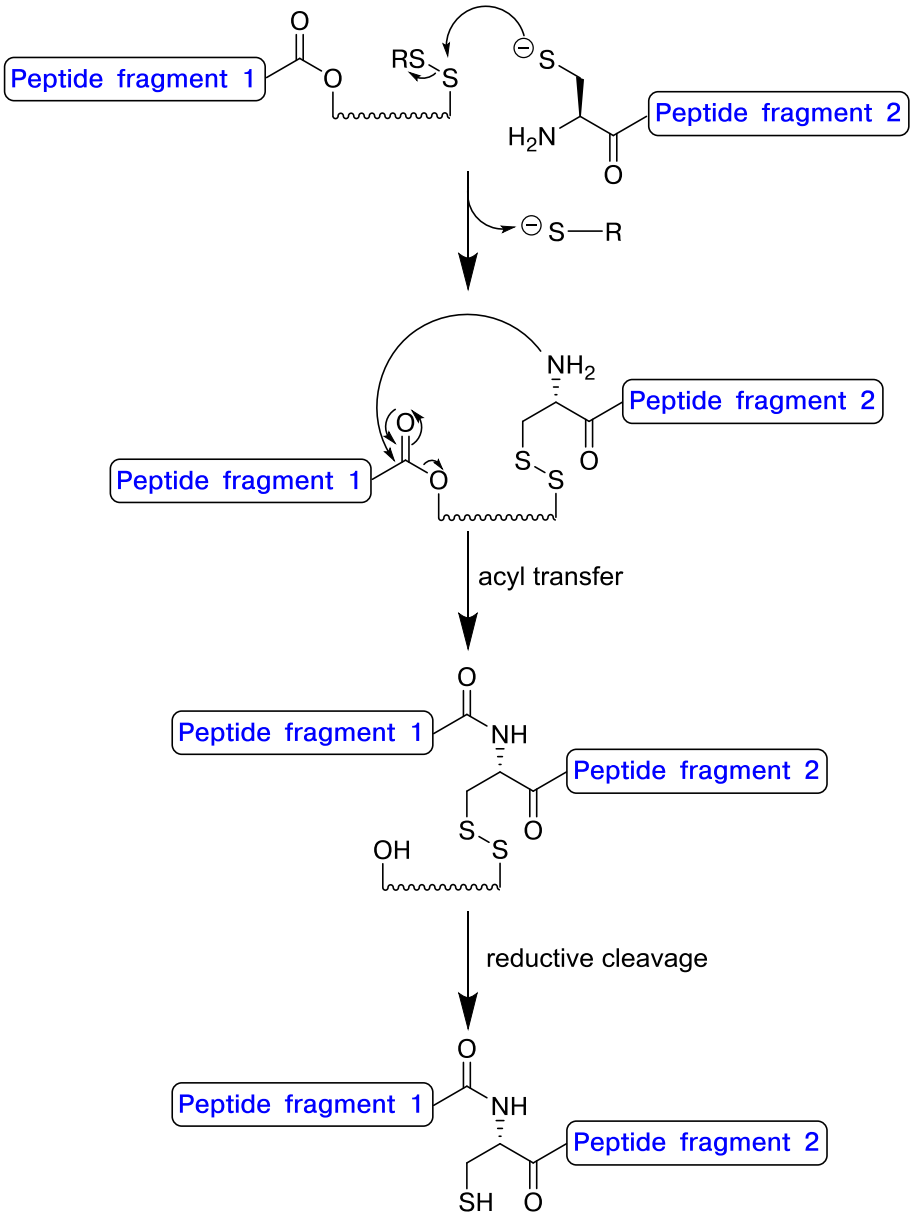


Figure 4.4 Amide bond formation by "thiol-capture" strategy[427-430]
Kemp and coworkers suggested the use of activated templates to assist amide bond formation.[427-430] The concept was to entropically activate the peptide fragments to accelerate acyl transfer. This strategy ultimately guided development of native chemical ligation.[396]

Following inspiration from Kemp's "thiol capture" strategy[430] and the successful chemical ligation of HIV-1 protease (Figure 4.2), Dawson *et al.*[396] described a rapid, traceless ligation technique that required only an N-terminal cysteine and a thioester. This elegant strategy, termed "native chemical ligation" (NCL, for the "native" amide bond that is formed from the ligation reaction) derives principles from Kemp's thiol-capture strategy, where cysteine is used to entropically activate an intramolecular acyl transfer reaction that forms an amide bond. Rather than "capture" the cysteine thiol with an intermediary template molecule, in NCL the cysteine thiolate reacts directly with a peptide containing a C-terminal thioester. Formation of the thioester intermediate entropically activates the amine to undergo *S*→*N* acyl transfer reaction, which releases the cysteine thiolate and forming the amide bond (Figure 4.5). The reaction conditions can be conducted at room temperature, and only a reductant is required to prevent disulfide formation between cysteine thiolates. In fact, it was found that including aryl thiol additives in the reaction significantly accelerated the ligation reaction rate.[431, 432] NCL is a mild, chemoselective, and rapid methodology, and yields an amide bond at the ligation site, and was immediately applied to proteins (Dawson *et al.* used human interleukin-8 as a model to demonstrate NCL).[396]

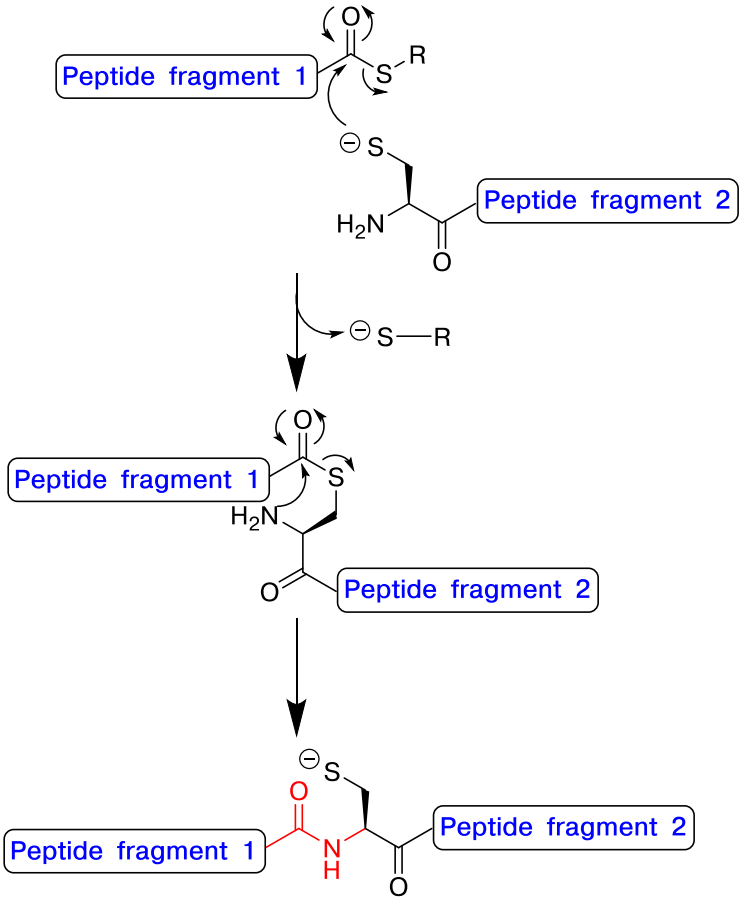


Figure 4.5 Amide bond formation via native chemical ligation[396]
This chemical ligation method relies on "thiol-capture" of the cysteine thiolate using a peptide containing a C-terminal thioester, which entropically activates the intermediate to undergo a 5-membered intramolecular acyl transfer reaction, releasing the cysteine thiolate and forming a native amide bond.

Dawson *et al.*[396] noted that the nature of the C-terminal thioester group, or more precisely the leaving group of the intermolecular acyl transfer reaction, impacted the rate of the ligation reaction. This aspect of NCL has been utilized for "kinetically controlled ligation," where a cysteine thiolate will react with an aryl thioester faster than with an alkyl thioester. This difference in reactivity introduced a means for site-selectivity in the presence of multiple thioesters, allowing for the convergent synthesis

of multiple peptide fragments (Figure 4.6). Using this approach, combined with masking cysteine residues as thiazolidines, Bang *et al.*[433] synthesized the 46-amino acid protein, crambin, using six peptide fragments.

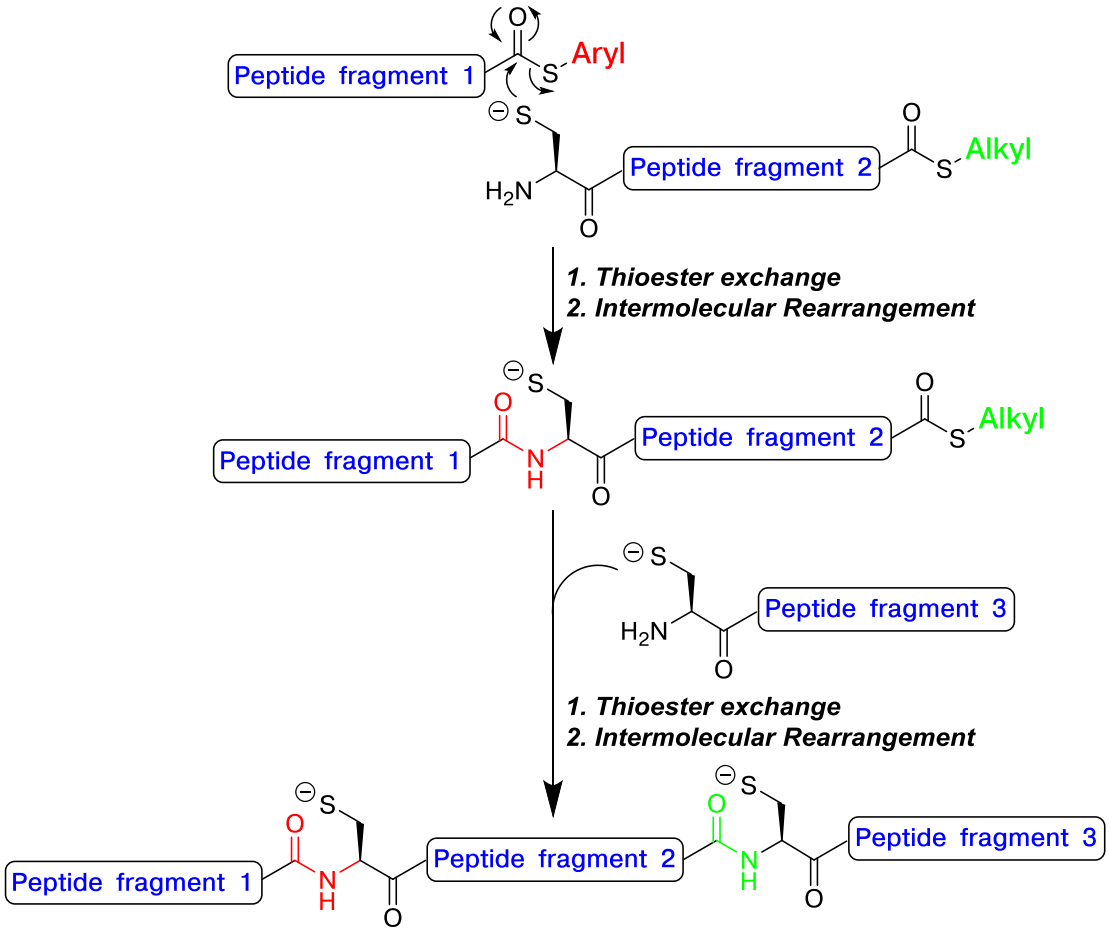


Figure 4.6 Convergent synthesis of peptides using kinetically controlled ligation[433]

Aryl thioesters have better leaving groups than alkyl thioesters, and are more reactive to *S*→*N* acyl transfer reaction with cysteine. In the presence of alkyl- and aryl-thioesters, a cysteine thiolate will react preferentially with the aryl thiolate.This "kinetic control" of ligation reaction rate with thioester groups can be used for the convergent synthesis of multiple peptide or protein fragments.

4.1.3 Expressed protein ligation and semi-synthetic methodologies

The simplicity of NCL resembles of the mechanism of protein splicing that has evolved in Nature. Protein splicing mechanisms are conserved throughout bacteria and eukaryotes, and function as a means of protein synthesis and regulation.[36] An expressed gene may contain multiple protein fragments within a peptide fragment, a precursor protein. This precursor protein must be processed to excise internal peptide fragments (inteins), and this process ligates the flanking protein sequences (exteins) together as a native amide bond.[38] In auto-processing precursor proteins, the intein region adopts a conformation that promotes their own excision from the precursor protein, and they require no external source for initiation (Figure 4.7).

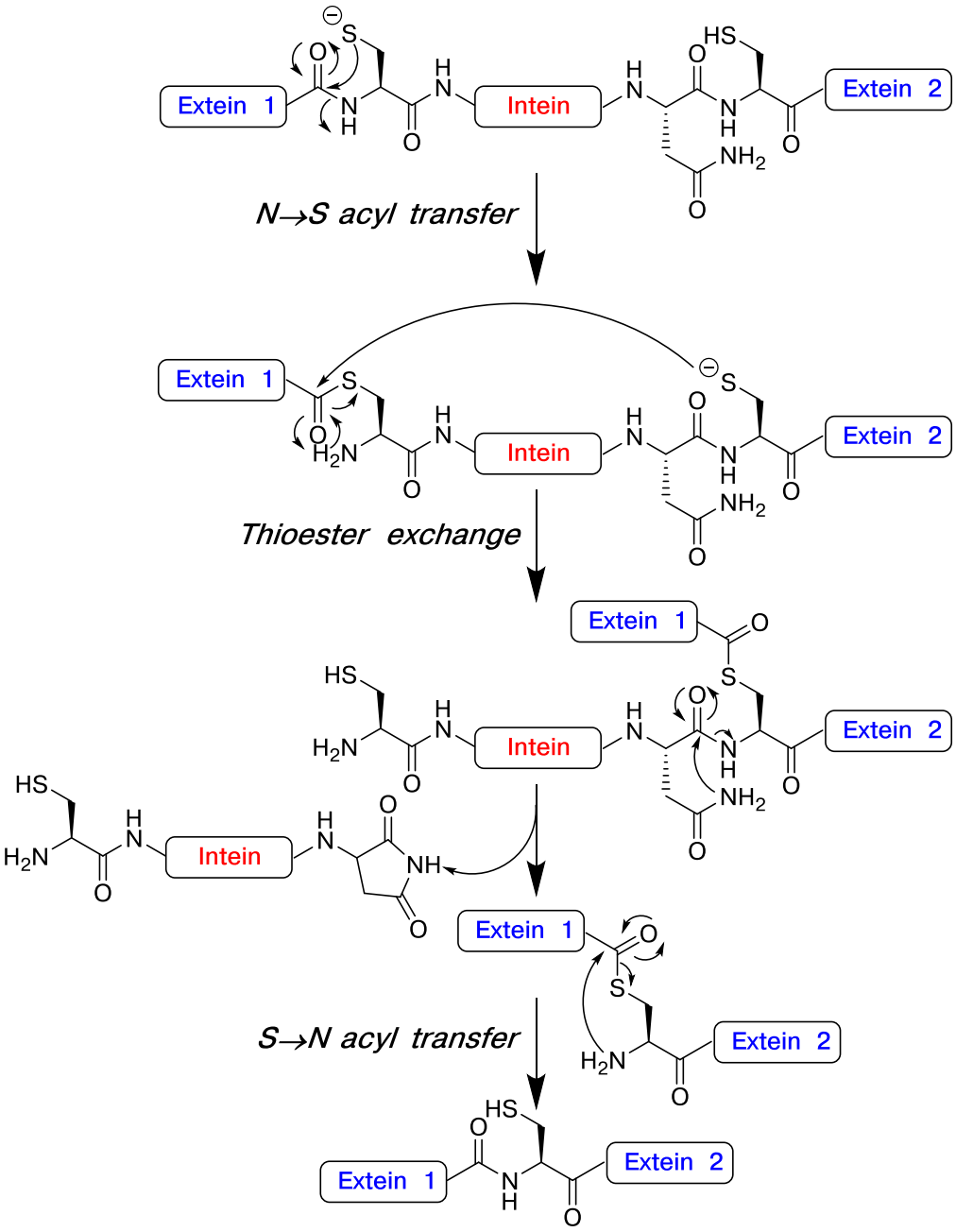


Figure 4.7 Intein-mediated protein splicing, NCL reactions evolved in Nature
Protein splicing occurs in bacteria and eukaryotes as a means of post-translational modification of proteins. The intein adopts a conformation that promotes excision and ligation of the exteins.[38]

In NCL, incorporation of an N-terminal cysteine in a peptide or protein via expression or SPPS is generally not difficult, but modifying a C-terminus of a peptide or protein with a thioester is much more challenging. In the early development of NCL, peptides containing C-terminal thioesters could be synthesized by chemical modification of the peptide, if the side chains were protected from the reaction, and no method had been presented for expression of proteins containing C-terminal thioesters. Muir and coworkers[36, 434, 435] recognized that intein-mediated protein splicing presented an unique opportunity for the chemical synthesis of proteins via expression of proteins containing a C-terminal thioester group. By expressing a fusion protein containing an engineered intein, addition of an excess thiol could initiate excision of the intein (Figure 4.8), and form a protein with a C-terminal thioester.

With this advancement, expressed proteins could be used in NCL on *either* side of the ligation site, paving the way for semi-synthesis of proteins and protein hybrid materials. Muir *et al.* initially demonstrated expressed protein ligation (EPL) for the semi-synthesis of a modified 50 kDa C-terminal Src kinase (Csk). Csk catalyzes phosphorylation of a tyrosine in Src kinase, which initiates a conformational change and an interdomain association. The structure of Csk resembles its Src kinase substrates, but does not contain the C-terminal region with the phosphorylation site.[434] Muir *et al.*[434] examined the conformational effects of adding the phosphorylated C-terminal peptide "tail" to Csk, in order to determine if this modified Csk could undergo the same conformational changes as phosphorylated Src kinases.[434] Protein expression of a modified Crk domain would not allow for site-selective phosphorylation, and automated peptide synthesis of Csk containing a C-terminal phosphotyrosine would be impossible, given the size of the kinase and the location of

the post-translational modification at the C-terminus. To overcome this synthetic challenge, the C-terminal peptide fragment was chemically synthesized containing an the phosphotyrosine and an N-terminal cysteine. Csk was expressed as a fusion protein with an engineered intein, and was modified to contain a C-terminal thioester as shown in Figure 4.8.[434] The Csk containing a C-terminal thioester was then used as a substrate for NCL with the synthesized phosphopeptide containing an N-terminal cysteine (Figure 4.8).

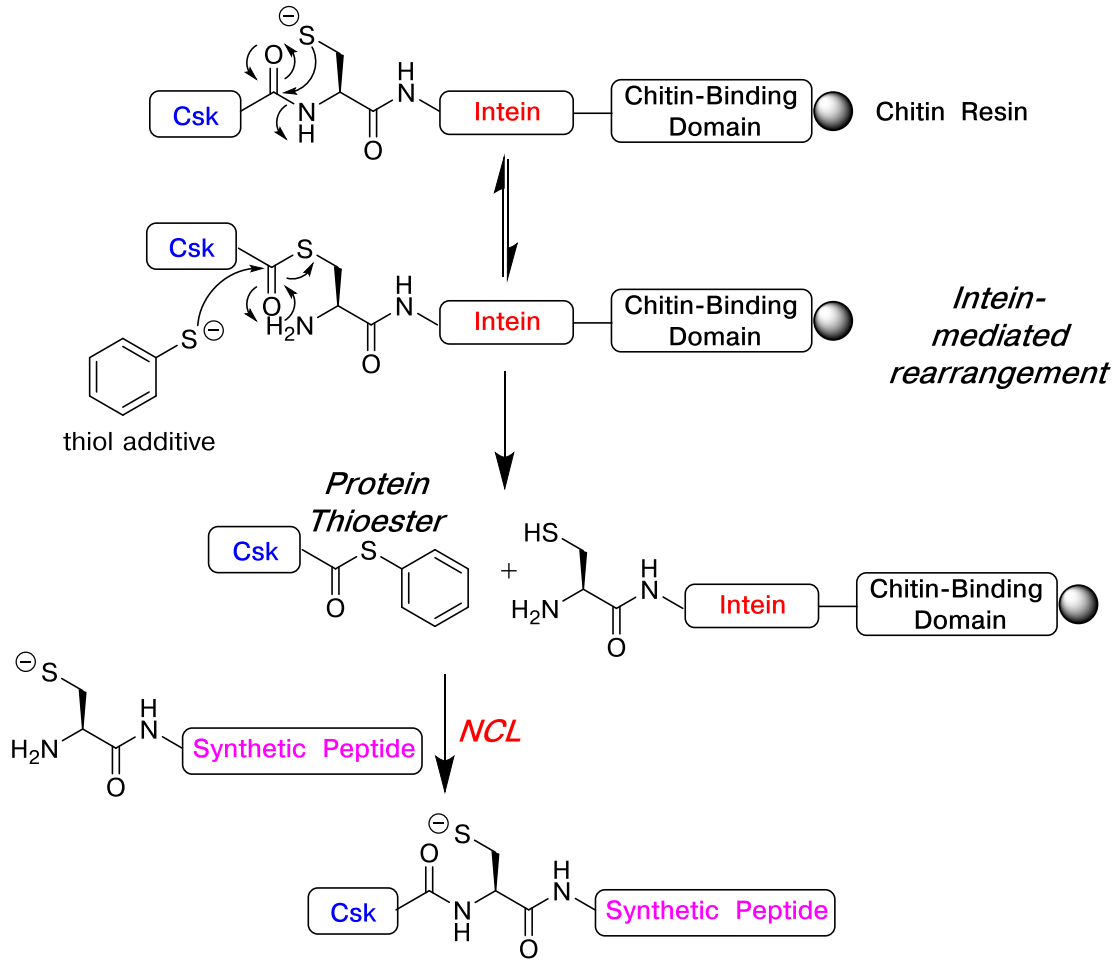


Figure 4.8 Semi-synthesis of proteins via expressed protein ligation
Deriving from intein-mediated protein splicing, Muir[434] developed a methodology where an expressed protein can be chemically modified to contain a C-terminal thioester, after treatment of an intein-fusion protein with a thiol additive. This method allowed for semi-synthesis of proteins using NCL, where expressed proteins could be used on either side of the ligation site.

EPL methodology has also been useful for synthesizing modified proteins where the modification is at the N-terminus, such as in histones. Histones are proteins that scaffold and organize DNA into chromatin and chromosomes, and the post-translational modifications of histones are thought to influence protein regulation by

mitigating access to DNA for transcription, replication, signaling transcription enzymes.[436-438] Histones are structured as a core group of globular proteins that coordinate DNA, and contain a long N-terminal "tail" is less structured.[438] Histones can be modified via phosphorylation, ubiquitination, and methylation, with more than 60 modification sites identified, suggesting their importance in regulating transcription pathways.[437, 438] With histones, a single post-translational modification can influence the modifications at other sites, which complicates mapping of individual modifications and understanding the cellular "histome".[436, 437] EPL has allowed for the practical semi-synthesis of histones with specific post-translational modifications near the C-terminus, such as residues that interface with DNA, in order to identify the structural and regulatory features of single post-translational modifications.[437]

Muir and coworkers have utilized EPL and NCL methodologies to unravel some of the intricate details in histone post-translational modification, and this work a current area of research.[36, 41, 121, 434, 439, 440] Since Muir reported this powerful methodology, harnessing protein splicing for chemical semi-synthesis of proteins, many researchers have utilized EPL and intein-based methods for synthesizing engineered proteins and peptides, including incorporation of synthetic β-turn motifs into RNase A;[441] generating synthetic mutants of sodium-potassium ion channels;[442] generating cyclic proteins,[443] and numerous additional examples.[41, 434, 440, 444]

4.1.4 Expansion of native chemical ligation to non-cysteine amino acids

NCL was revolutionary for the synthesis of engineered proteins, but there still remained the requirement for a cysteine residue at the ligation site. Many proteins do not contain any cysteine residues, and some proteins may not be able to structurally or functionally tolerate a cysteine mutation. In addition, given that NCL proceeds less

efficiently at thioesters of proline or stereically hindered amino acids, not all cysteines are amenable to NCL reactions.[445] Yan & Dawson[400] demonstrated that cysteine can be desulfurized following NCL reaction, which generates alanine at the ligation site (Figure 4.9a). In this work, the cyclic peptide Microcin J25 was synthesized using NCL, and the thiol functionality on cysteine was removed via Pd/Al_2O_3 or Raney nickel.[400] With desulfurization, the method of NCL was expanded beyond peptides and proteins containing cysteine.

Given the harsh reaction conditions that are required for removal of cysteine thiols, milder desulfurization approaches were developed. Wan & Danishefsky[446] demonstrated that cysteine can be desulfurized within a glycopeptide using a water-soluble radical initiator (VA-044) in the presence of TCEP. Without requiring Raney nickel or harsh hydrogenation reaction conditions, VA-044 desulfurization reactions are compatible with delicate peptide modifications.[446] Wan & Danishefsky[446] also demonstrated a variant of this strategy where selenocysteine was deselenized to form alanine using VA-044 and TCEP (Figure 4.9b). The development of mild desulfurization and deselenization methods established that NCL and EPL could be utilized in the absence of cysteine residues.

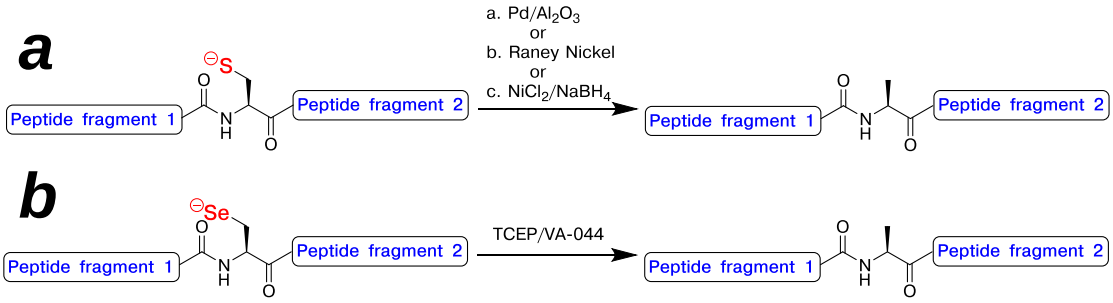


Figure 4.9 Desulfurization and deselenization methods on cysteine and selenocysteine to generate alanine in peptides
Pioneered by Yan & Dawson,[400] peptides containing cysteine can be converted to alanine using desulfurization reactions. Deselenization approaches where demonstrated by Wan & Danishefsky,[446] as a means to generate alanine at the ligation site without desulfurizing other cysteine residues. These methods have been applied to other thiol- and selenol-containing amino acids.

The initial requirement of cysteine in a peptide or protein imposed an inherent barrier to the broader use of NCL in the synthesis of engineered proteins. Tam & Yu[447] were some of the pioneers in extending NCL beyond the requirement for cysteine. Tam & Yu[447] proposed "methionine ligation," where the NCL reaction is mediated by an N-terminal homocysteine, which can be methylated after the ligation reaction to generate methionine at the ligation site. After the introduction of desulfurization and deselenization strategies,[399, 400, 448] NCL reactions could be mediated by other non-natural thiolated amino acids. That is, non-natural amino acids containing a thiol could be incorporated into synthetic peptides or expressed proteins to mediate the NCL reaction, and could be subsequently desulfurized to generate almost any amino acid. The growing repertoire of thiolate amino acids that have been synthesized and utilized for NCL includes β-mercaptophenylalanine,[401-403] penicillamine and γ-thiovaline,[449, 450] β-mercaptoleucine,[451] γ-mercaptothreonine,[452] α-thiolysine and ε-thiolysine,[453] and 4*R*-mercaptoproline,[454] (Figure 4.10).[399] Amino

acids containing selenols have also been synthesized for NCL reactions and subsequent deselenization with VA-044 and TCEP.[399] Alternative functional group have also been demonstrated for mediating NCL reactions, including side-chains containing phosphorylations,[455] *o*-hydroxythiophenols,[456] and amidyl auxiliaries.[457-460] Although some of these thiolated amino acids are commercially available (4-mercaptoproline; penicillamine, valine precursor), most of those shown in Figure 4.10 require several synthetic steps, and generate the thiolated amino acid with low overall yield.

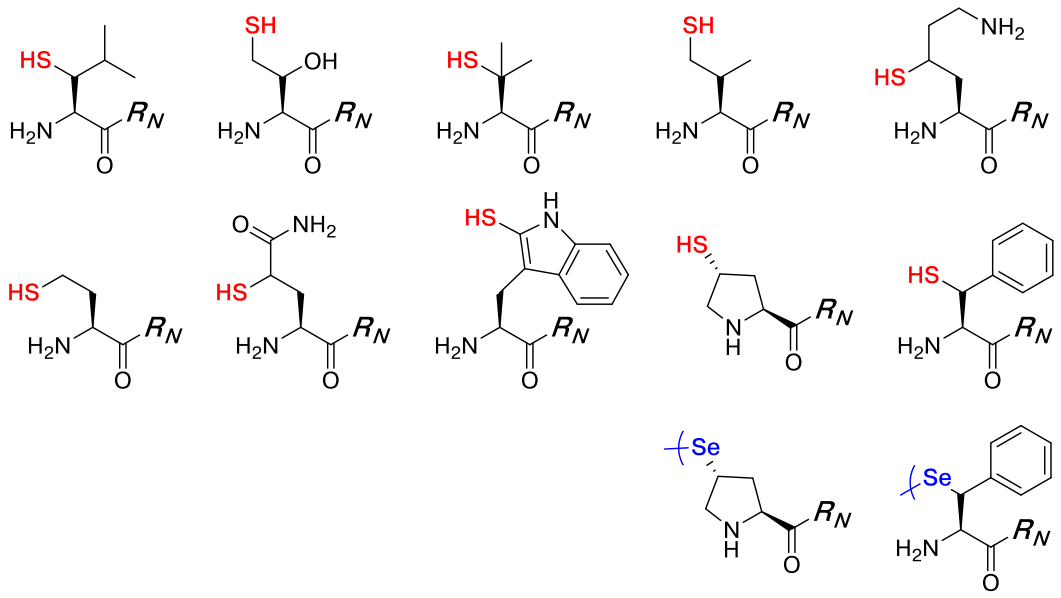


Figure 4.10 Thiol- and selenol-modified amino acids for mediating NCL
With the advent of desulfurization and deselenization methods, numerous thiolated or selenol-containing amino acids were synthesized and applied for NCL. This list of available amino acids is continually growing.[399]

Aromatic amino acids, including phenylalanine, are crucial residues for tailoring protein structure and molecular recognition via non-covalent interactions (see Chapters 2 and 3), and the development of NCL methodologies involving aromatic amino acids must be improved to enable further studies on these important amino acids.[161, 403] β-Mercaptophenylalanine and β-selenophenylalanine have been demonstrated for application in NCL with model peptides and proteins, but these amino acids require 6-9 synthetic steps with an overall yield that is less than 6-12%, depending on the synthetic routes used.[401-403, 461] Furthermore, NCL using β-mercaptophenylalanine is relatively slow (24 h), potentially due to the steric bulk and hydrophobicity of this aryl thiol.

Given the need for more practical approaches to NCL at aromatic amino acids, and the practicality of synthesizing peptides containing 4-thiophenylalanine using solid-phase modification approaches,[169, 404] we hypothesized that 2-thiophenylalanine at the N-terminus of a peptide could be utilized for NCL at phenylalanine. The copper-mediated cross-coupling approach that was used to synthesize 4-thiolphenylalanine (described in Chapter 1) could be utilized for the synthesis of peptides containing N-terminal 2-thiophenylalanine, using commercially available Boc-2-iodophenylalanine. This synthetic strategy on the protected peptide on solid-phase avoids the multiple, solution-phase synthetic steps required to generate the protected amino acid for incorporation via solid-phase peptide synthesis. With the increased acidity of 4-thiophenylalanine relative to cysteine,[169] we anticipated that 2-thiophenylalanine could potentially act as a nucleophile in NCL reactions under mildly acidic conditions. In addition to desulfurization of 2-thiophenylalanine to generate phenylalanine, derivatives of thiophenylalanine-mediated NCL could potentially be used to introduce

novel aromatic amino acids into peptides and proteins in a practical manner (i.e. 3-mercaptotyrosine). We describe synthesis of peptides containing N-terminal thiophenylalanine, and demonstrate NCL reactions mediated by 2-thiophenylalanine-mediated in model peptides, and applications into the chemical synthesis of proteins.

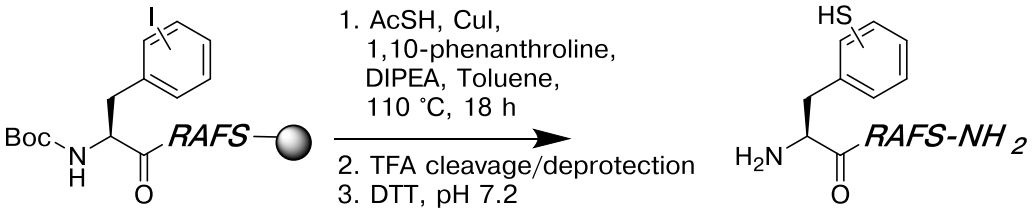
4.2 Results

Cysteine-mediated NCL reactions proceed through via a thioester exchange reaction between a peptide containing an N-terminal cysteine and a peptide containing a C-terminal thioester. The intermediate species entropically activates the N-terminus to undergo a rapid intramolecular *S→N* acyl transfer through a 5-membered reaction intermediate (Figure 4.5).[396] The intramolecular *S→N* acyl transfer reaction has been known to proceed slowly through larger intermediates (15- and 16-membered).[462] With these mechanistic aspects about NCL in consideration, we sought to determine if NCL reactions could be mediated using an N-terminal thiophenylalanine in place of cysteine. A practical synthesis of peptides containing 4-thiophenylalanine was previously demonstrated using a copper-mediated cross-coupling reaction on resin-bound peptides containing 4-iodophenylalanine,[169] and this approach was utilized for synthesis of peptides containing N-terminal thiophenylalanine for NCL reactions.

4.2.1 Initial synthesis of peptides containing N-terminal thiophenylalanine

For our preliminary studies on NCL mediated by thiophenylalanine, we synthesized a series of model peptides containing Boc-iodophenylalanine at the N-terminus. The position of the thiol group could have significant consequences on the ligation reaction rate, depending on the transition state for the intramolecular *S→N* acyl transfer to generate the amide bond. While cysteine-mediated ligation reactions

proceed through a 5-membered ring intermediate (Figure 4.5), larger intermediates have been demonstrated with some success.[463] In order to determine the optimal position of the thiol group on the N-terminal thiophenylalanine, the series of model peptides were synthesized to contain 2-, 3-, or 4-iodophenylalanine (2-I-Phe, 3-I-Phe, or 4-I-Phe, respectively) at the N-terminus. The model peptide, XRAFS (X = iodophenylalanine), was selected because this and related sequences have been utilized in previous model peptide studies of NCL reactions.[464] Boc-2-, 3-, and 4-iodophenylalanine are all commercially available amino acids, and were readily incorporated at the N-terminus of the peptides on solid phase. The peptides containing Boc-iodophenylalanine were subjected to solid-phase copper-mediated cross-coupling conditions that were previously established (described in Chapter 1).[169] All of these peptides containing N-terminal thiophenylalanine were successfully generated using the previously established cross-coupling reaction (Figure 4.11). However, the reaction on the peptide containing 2-iodophenylalanine resulted in an unexpected reaction product, which strongly suggested exceptional reactivity of this peptide for NCL reactions (Figure 4.12).



2-, 3-, or 4-iodophenylalanine

Figure 4.11 Synthesis of model peptides containing N-terminal thiophenylalanine via solid-phase cross-coupling reaction

The reaction scheme for the cross-coupling reaction on the resin-bound peptides containing N-terminal iodophenylalanine, and subsequent reduction and thiolysis reactions in solution phase to generate the peptides containing N-terminal thiophenylalanine. The resultant peptides contained either 2-, 3-, or 4-thiophenylalanine at the N-terminus.

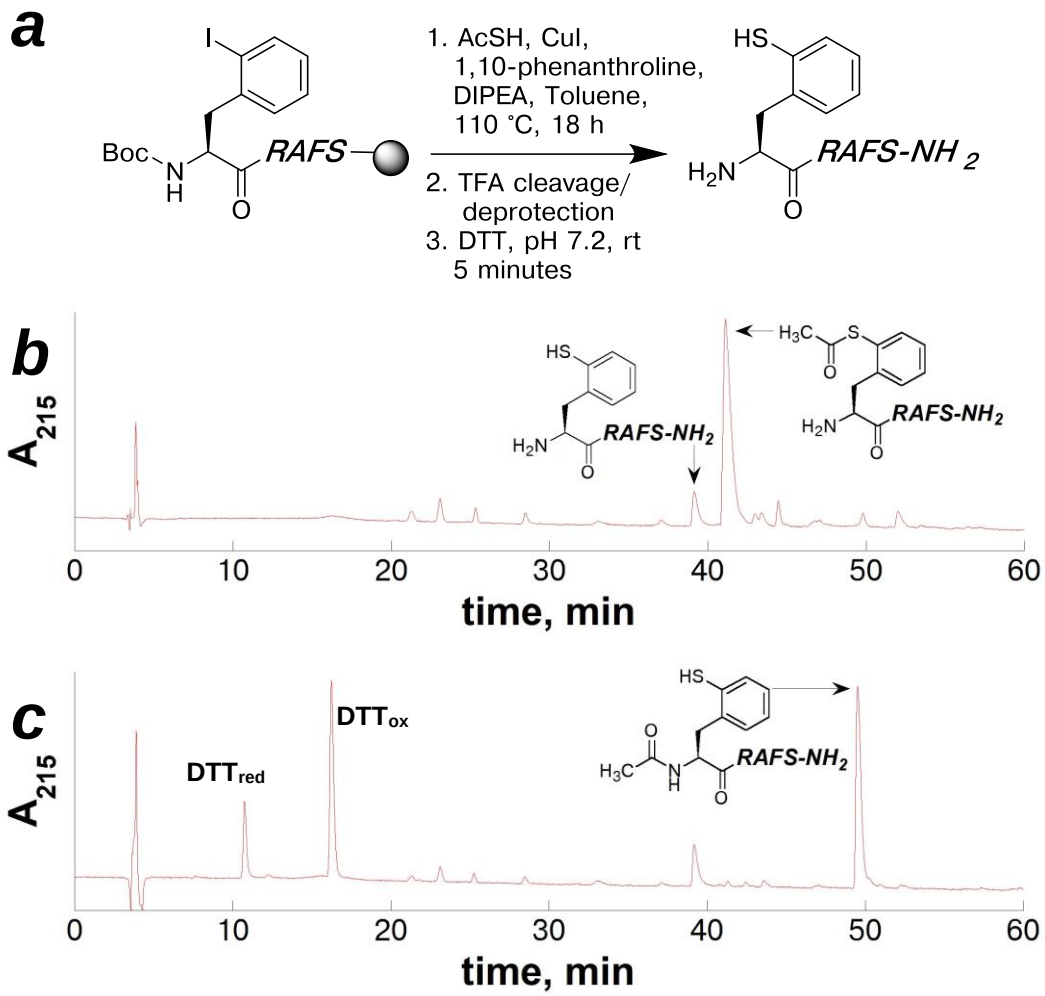


Figure 4.12 Initial synthesis of a model peptide containing N-terminal 2-thiophenylalanine via solid-phase cross-coupling reaction

(a) Reaction scheme for the cross-coupling reaction on the resin-bound peptide containing 2-iodophenylalanine and subsequent reduction and thiolysis reactions in solution to generate the peptide (2-SH-Phe)RAFS-NH_2; (b) HPLC chromatogram of the resultant peptide products from the cross-coupling reaction using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the resultant peptide products from the thiolysis and reduction reactions in solution using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

DTT_{red} and DTT_{ox} indicate reduced and oxidized dithiothreitol, respectively.

In the synthesis of a peptide containing N-terminal 2-thiophenylalanine (2-SH-Phe), the major product observed from the cross-coupling reaction was Ac-(2-SH-Phe)RAFS-NH_2, which was confirmed by ESI-MS, NMR, and a positive result from Ellman's test[109] (indicating a free sulfhydryl group). The NMR spectrum of this resultant product (Figure 4.13) indicated the presence of an acetyl group, and five amide protons (not the expected four), indicating the peptide contained an N-terminal acetyl group. This product potentially resulted from an intramolecular *S*→*N* acyl transfer reaction between the free amine and thioacetyl peptide derivative (Figure 4.14). In the context of the mechanism of NCL reaction, this result is completely expected: the acetyl modification on 2-thiophenylalanine bears the same electrophilic carbonyl group that is presented in NCL reactions with a good aryl thiolate leaving group, and the acyl transfer reaction proceeds through a 7-membered-ring reaction intermediate. 7-Membered-ring intramolecular *S*→*N* acyl transfer intermediates have been reported previously with NCL at homocysteine.[447]

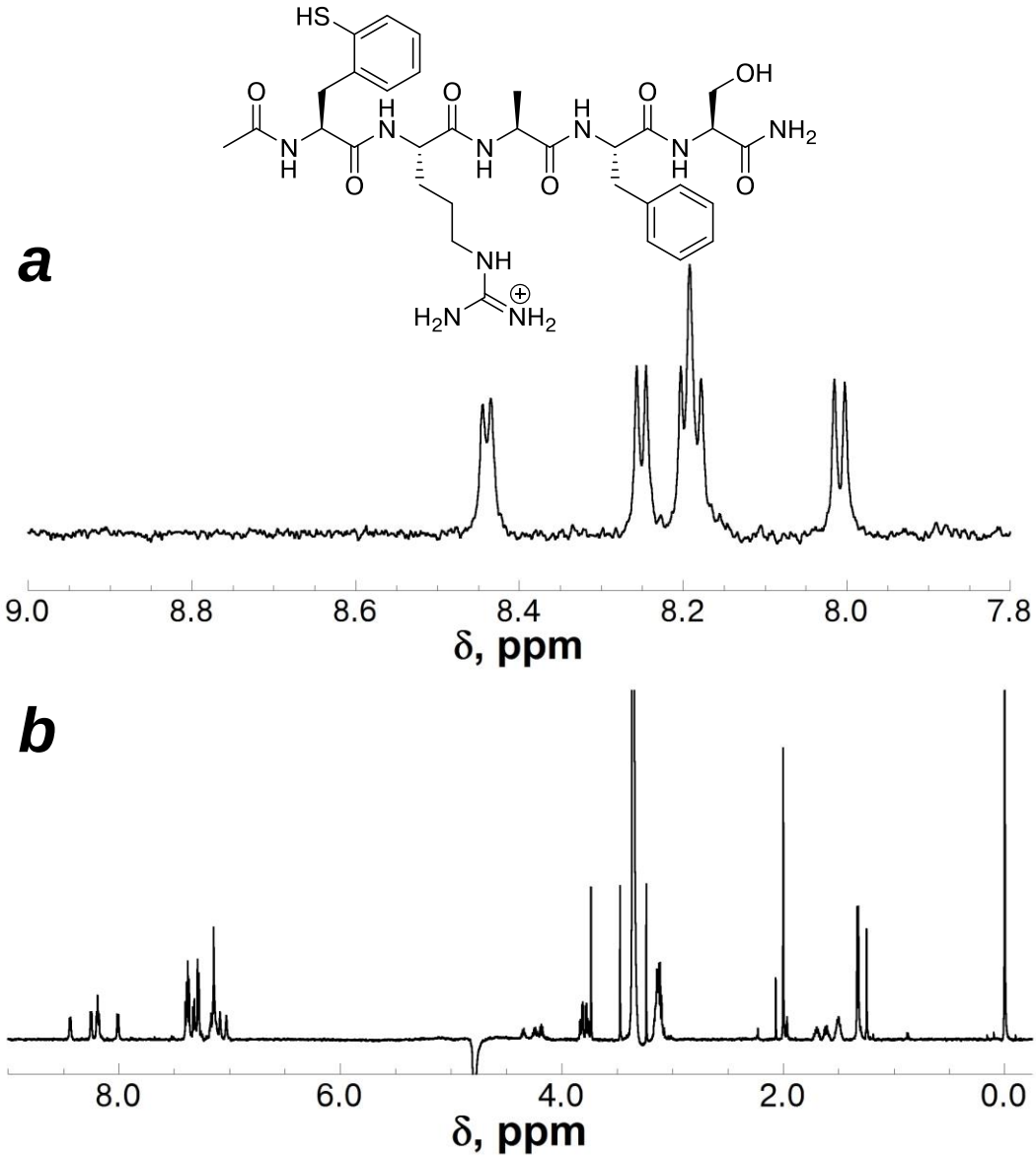


Figure 4.13 ^{1}H NMR spectrum of the resultant product from the solid-phase cross coupling reaction on the peptide Boc-(2-I-Phe)RAFS

^{1}H NMR spectrum of the peptide Ac-(2-SH-Phe)RAFS-NH_2 in 90% H_2O/10% D_2O with 5 mM phosphate buffer (pH 4), 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP. (a) amide region; (b) full spectrum. The singlet at 2.00 ppm (3H) is assigned to an acetyl CH_3 group, and 5 amide doublets are present (compared to 4 amide peaks in the peptide (2-SH-Phe)RAFS-NH_2). The large peak at 3.34 ppm is due to methanol.

Figure 4.14 Intramolecular acyl transfer mechanism of 2-thiophenylalanine
Since the peptide is subjected to deprotection reaction conditions before thiolysis reaction conditions, a free amine can potentially react with the thioacetyl carbonyl on 2-thioacetyl-phenylalanine, resulting in the peptide Ac-(2-SH-Phe)RAFS-NH_2. These data indicate that 2-thiophenylalanine can undergo rapid and efficient intramolecular acyl transfer in NCL reactions.

The observation of an intramolecular *S→N* acyl transfer reaction in the synthesis of a peptide containing 2-thiophenylalanine suggested that this amino acid can undergo rapid NCL reactions with thioester peptides. Formation of the *S→N* acyl transfer product peptide was complete in less than 5 minutes at room temperature. However, with the amine converted to an amide via an acetyl group, this peptide can not be used for NCL reactions. In order to synthesize the peptide containing 2-thiophenylalanine at the N-terminus, the peptide would need to be subjected to the thiolysis reaction on solid-phase, prior to the TFA deprotection reaction, which removes the Boc-protecting group on the terminal amine, in order to prevent the acyl transfer reaction that reacts at the N-terminus.

4.2.2 Optimized synthesis of a model peptide containing 2-thiophenylalanine via copper-mediated cross-coupling reaction on solid phase

Conditions for the thiolysis reaction on peptides containing 4-S(acetyl)-thiophenylalanine on solid-phase were described previously (Chapter 1.2.10).[169] These reaction conditions were utilized to generate the peptide containing 2-

thiophenylalanine at the N-terminus of the protected peptide on solid-phase. This synthetic strategy avoids the potential for an intramolecular *S*→*N* acyl transfer reaction between the N-terminus and the thioacetyl group (Figure 4.14). The HPLC chromatogram of the peptide products that resulted from this modified synthesis is shown in Figure 4.15. The peptide product containing 2-thiophenylalanine was confirmed via NMR (Figure 4.16, 4.17).

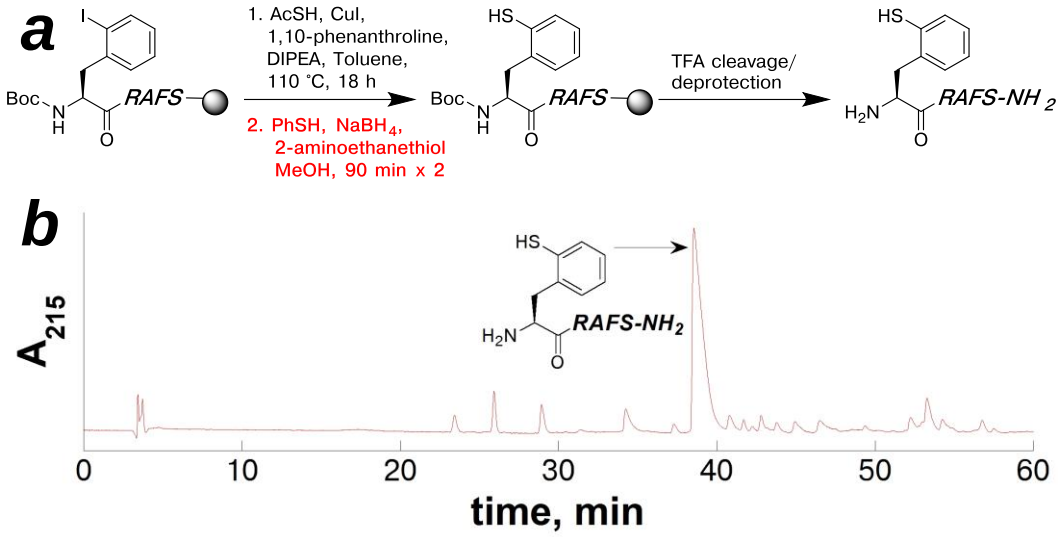


Figure 4.15 Synthesis of a model peptides containing N-terminal 2-thiophenylalanine via cross-coupling reaction and thiolysis on solid-phase
(a) The reaction scheme for the cross-coupling reaction on the resin-bound peptide and subsequent reduction and thiolysis on solid phase to generate the protected peptide (2-SH-Phe)RAFS-NH_2; (b) HPLC chromatogram of the peptide products that resulted from the cross-coupling reaction and subsequent thiolysis reaction on solid-phase using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).
No reductants were added to the crude peptide in solution, indicating that the acetyl group was removed from the thiol on solid phase (TFA cleavage conditions do not effect thiolysis).

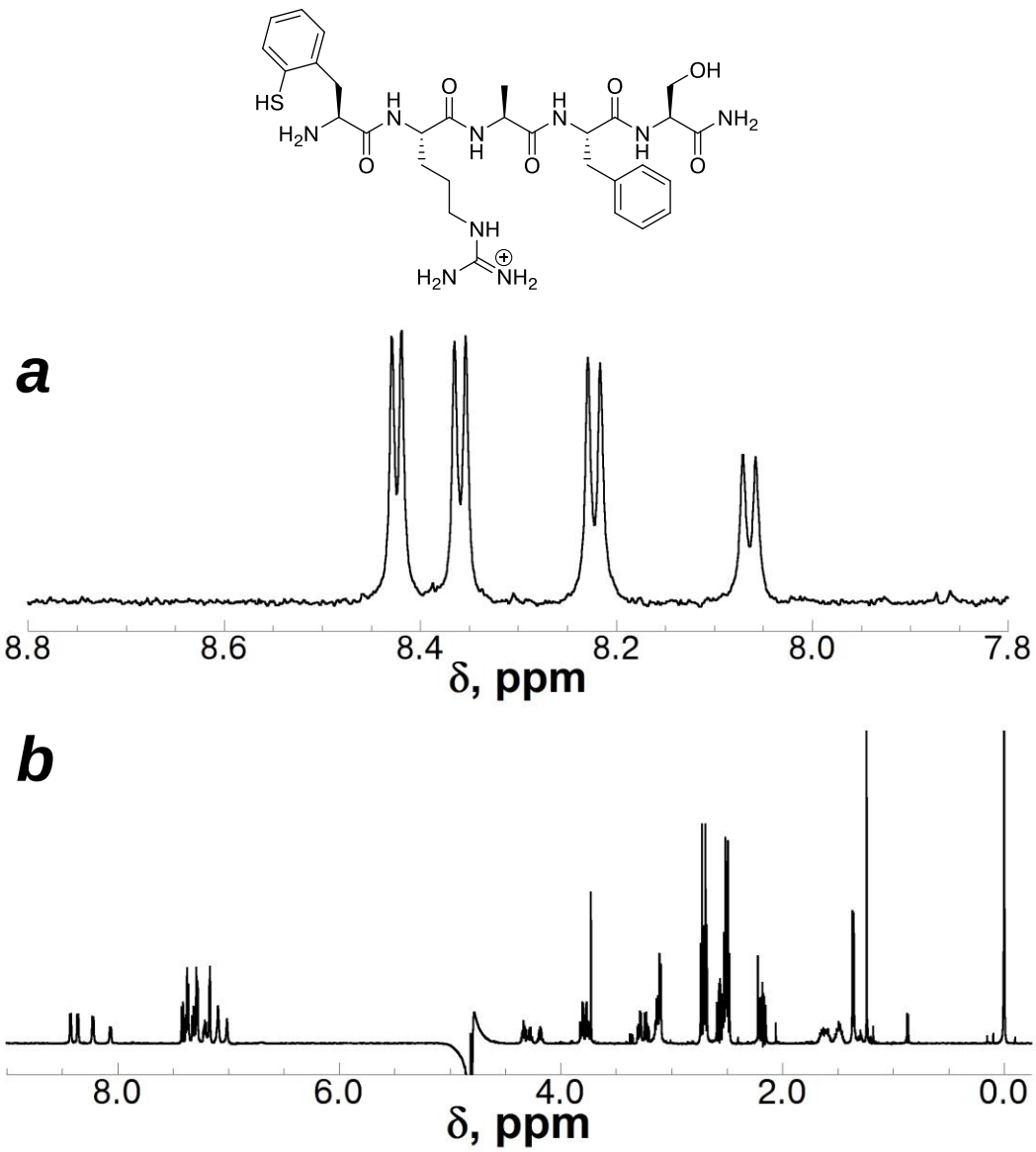


Figure 4.16 ^{1}H NMR spectrum of the peptide (2-SH-Phe)RAFS-NH$_2$
(a) ^{1}H NMR spectrum of the amide region for the peptide (2-SH-Phe)RAFS-NH$_2$; (b) ^{1}H NMR spectrum for the peptide (2-SH-Phe)RAFS-NH$_2$. The NMR sample contained in 10% D$_2$O/90% H$_2$O with 5 mM phosphate pH 4, 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP.

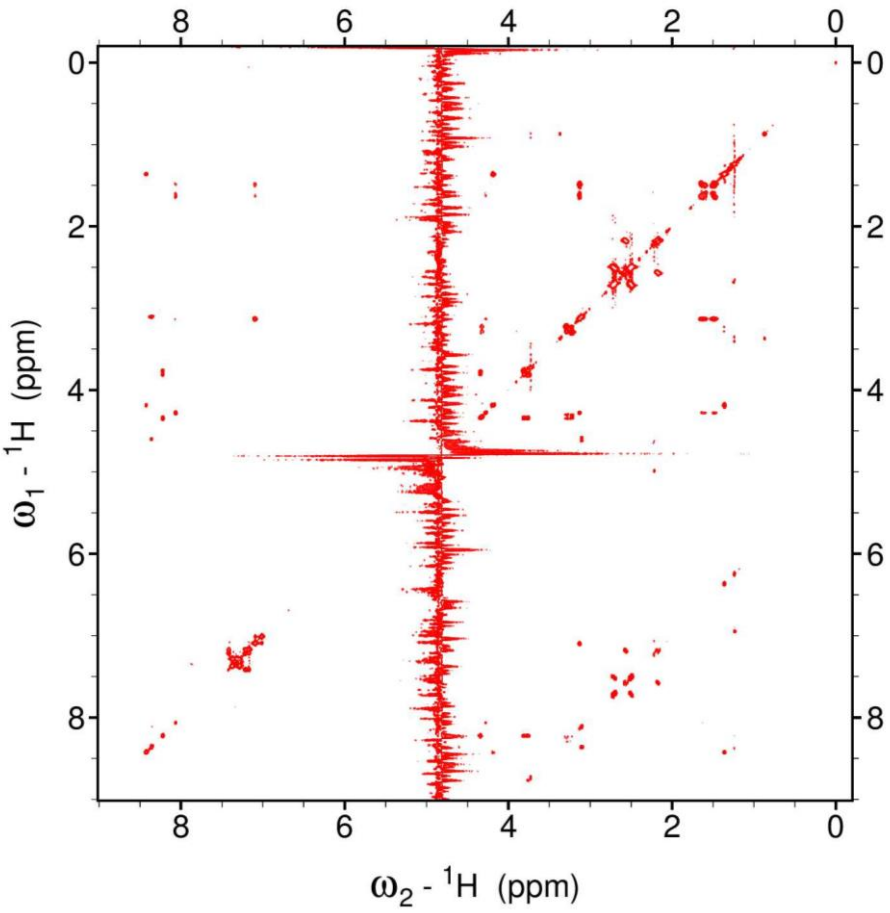


Figure 4.17 TOCSY spectrum of the peptide (2-SH-Phe)RAFS-NH_2
TOCSY spectrum for the peptide (2-SH-Phe)RAFS-NH_2. The NMR sample contained in 10% D_2O/90% H_2O with 5 mM phosphate pH 4, 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP.

By subjecting the peptide to the solid-phase thiolysis conditions after the cross-coupling reaction, the (2-SH-Phe)RAFS peptide was synthesized in 88% conversion in two steps from the peptide on solid phase. From completion of the peptide synthesis on resin, the peptide containing 2-thiophenylalanine can be generated within 24 hours with only one standard HPLC purification step. This solid-phase synthetic strategy significantly reduces the amount of time, reagents, and resources that would be needed for synthesizing the protected amino acid for incorporation via SPPS.

4.2.3 Retention of stereochemistry under cross-coupling conditions

The copper-mediated cross-coupling conditions subject the protected peptide to high reaction temperatures and a high concentration of DIPEA. Under these reaction conditions, α-epimerization on the peptide could potentially occur. Although epimerization was not observed in synthesis of 4-thiophenylalanine (Chapter 1), the possibility for epimerization was further evaluated for the model peptide containing 2-thiophenylalanine at the N-terminus. The model peptide XRAFS was synthesized containing Boc-2-iodo-D-phenylalanine at the N-terminus, and the resin-bound, protected peptide was subjected to the copper-mediated cross-coupling conditions with thiolacetic acid as previously described. The chromatogram of the crude peptide products that resulted from the cross-coupling reaction on XRAFS (X = 2-iodo-D-phenylalanine) is shown in Figure 4.18.

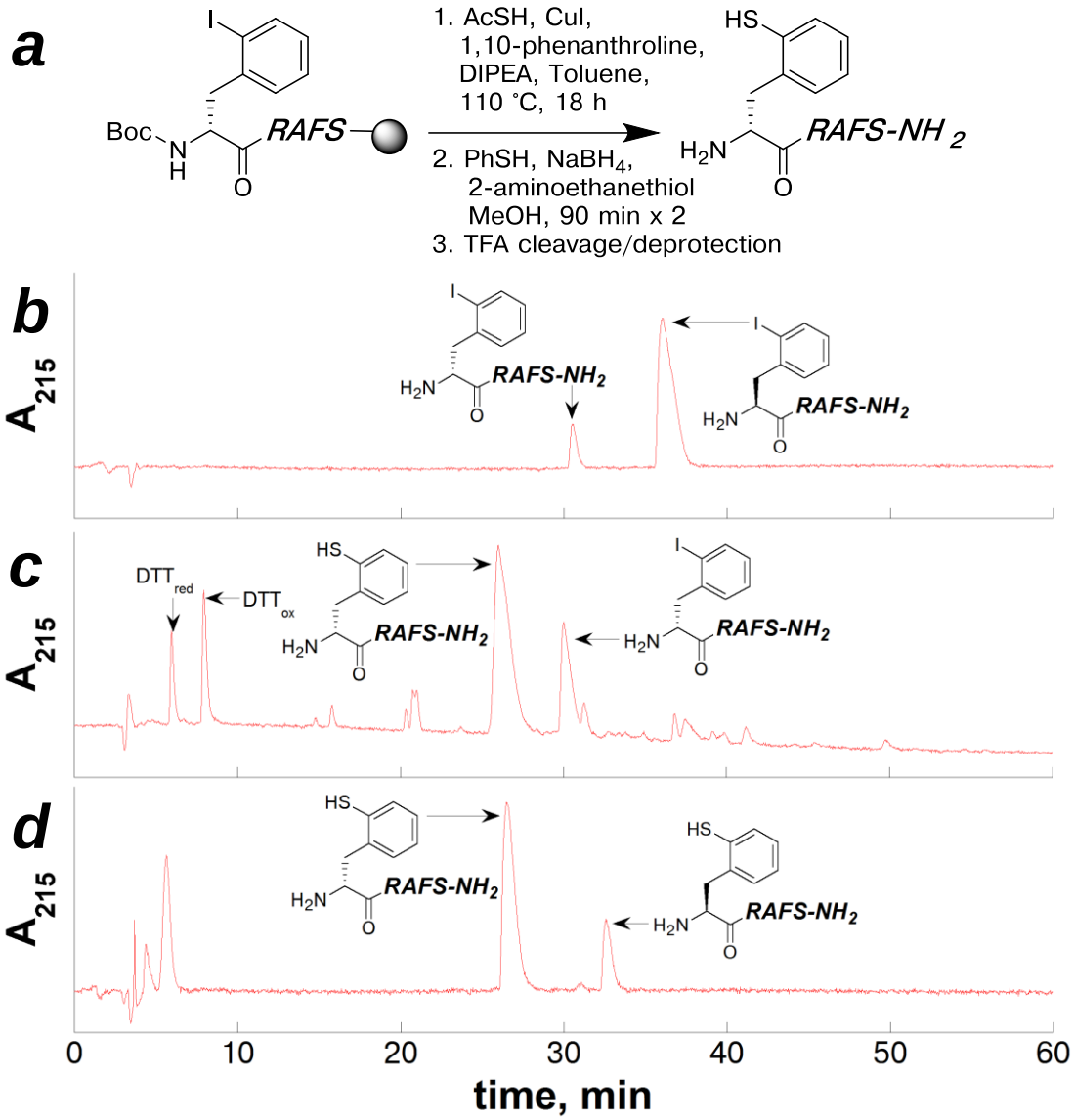


Figure 4.18 Synthesis of model peptides containing 2-thio-D-phenylalanine at the N-terminus

(a) Reaction scheme for the cross-coupling reaction on the resin-bound peptide and subsequent reduction and thiolysis in solution phase to produce the peptide (2-SH-D-Phe)RAFS-NH_2; (b) HPLC chromatogram of the coinjection of the purified peptides containing 2-iodophenylalanine at the N-terminus, both D- and L- variants; (c) HPLC chromatogram of the crude reaction products that resulted from the cross-coupling reaction on the peptide containing 2-iodo-D-phenylalanine on solid phase; (d) HPLC chromatogram of the coinjection of the purified peptide containing 2-thiophenylalanine, both D- and L- variants. All chromatograms were conducted using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

Based on the synthesis of the peptide (2-SH-D-Phe)RAFS-NH_2 and comparison via HPLC chromatograms of coinjection with material generated from the solid phase cross-coupling reaction and subsequent thiolysis reaction, there was no significant epimerization that resulted from the cross-coupling reaction on solid phase.

4.2.4 Alternative synthesis of a model peptide containing 2-thiophenylalanine via solid-phase peptide synthesis with Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine

The copper-mediated cross-coupling conditions subject the protected peptide to high temperatures (100-110 °C) in the presence of high concentrations of DIPEA and thiolacetic acid for an extended period of time. These reaction conditions have the potential to remove side chain protecting groups, which could lead to side reactions in peptides that have delicate or reactive side-chain modifications, such as glycosylated amino acids. In order to avoid any potential side-reactions on the peptide as a result of the cross-coupling reaction conditions, an alternative synthetic strategy was developed to generate peptides containing 2-thiophenylalanine at the N-terminus. Solution synthesis of a Boc-protected 2-thiophenylalanine and incorporation into peptides via solid-phase peptide synthesis would avoid any potential side reactions that could occur during the cross-coupling reaction on solid phase. With development of a synthesis for

the Boc-protected 2-thiophenylalanine for use in SPPS, this amino acid could be more broadly utilized in the semi-synthesis of proteins.

The synthetic strategy for synthesizing the protected Boc-2-thiophenylalanine is shown in Figure 4.19. The copper-catalyzed cross-coupling reaction was adapted for Boc-2-iodophenylalanine in solution phase. Ideally, the thiol functional group should have a protecting group that can be readily removed during or after the TFA cleavage and deprotection reaction conditions. Rather than using a sterically demanding S(trityl) group, the thiol can be protected as a disulfide for practical removal in solution with reductants such as DTT. Ragjapalan[23] has demonstrated that Boc-4-iodophenylalanine can be used directly as a substrate for palladium-catalyzed cross-coupling conditions, without protecting the carboxylic acid. Utilizing a similar strategy, Boc-2-iodophenylalanine was used directly in the copper-catalyzed cross-coupling reaction without protecting the carboxylic acid (Figure 4.19). However, the major products from this reaction were not the expected Boc-2-(S-acetyl)-thiophenylalanine and corresponding disulfide products (Figure 4.20).

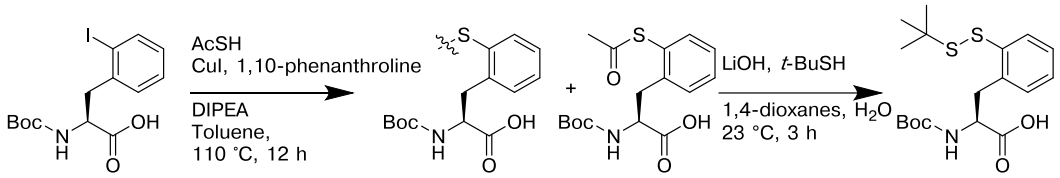


Figure 4.19 Initial synthetic strategy for Boc-2(S-*tert*-butyl disulfide)-thiol-L-phenylalanine

Utilizing commercially available Boc-2-iodophenylalanine directly, the copper-catalyzed cross-coupling reaction can generate a mixture of Boc-2(S-acetyl)-thiophenylalanine and the corresponding disulfides (observed in the model iodobenzene reaction, Chapter 1.2.1). Subsequent thiolysis reaction in the presence of excess *tert*-butyl mercaptan can generate the disulfide-protected Boc-2-thiophenylalanine. This amino acid can be coupled to the N-terminus of a peptide via solid-phase peptide synthesis, and deprotection of the side chain can be effected on the crude peptide in solution using DTT.

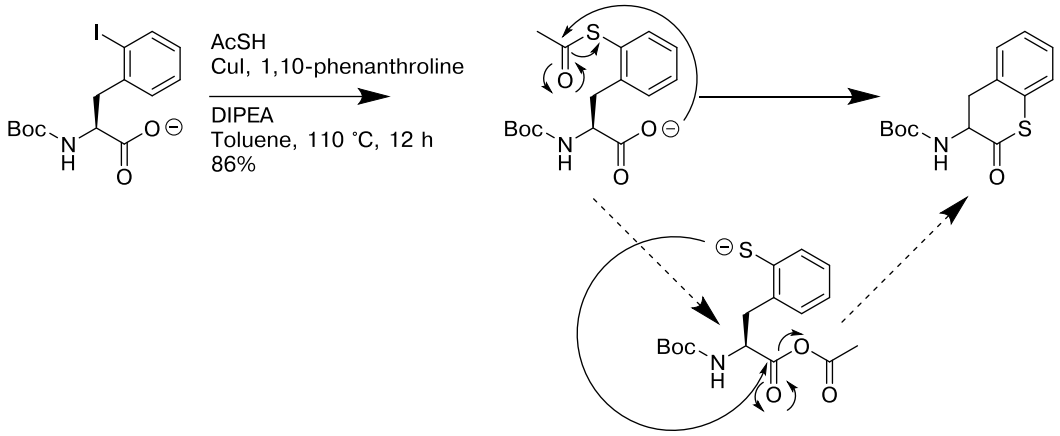


Figure 4.20 Reaction scheme for synthesis of the observed Boc-protected-3-amino-δ-thiochromanone

Boc-2-iodophenylalanine was subjected to the copper-catalyzed cross-coupling conditions in solution, and the reaction resulted in a racemic, thiochromanone product (86% isolated yield). The product potentially resulted from an acyl transfer between the carboxylate the thioacetyl product. The resultant anhydride would increase the acidity of the α-proton and increase likelihood of racemization at this position. The thiolate can undergo nucleophilic attack of the anhydride (inter- or intra-molecularly), releasing acetic acid and generating the Boc-protected-3-amino-δ-thiochromanone. Racemization could also occur via α-deprotection of the δ-thiochromanone.

The resultant Boc-protected 3-amino-δ-thiochromanone is a unique and interesting building block for synthesis. This product was generated in 86% yield in one step from commercially available Boc-2-iodophenylalanine. However, chiral HPLC analysis of the thiochromanone product indicated that the product was racemic (Figure 4.21). The Boc-protected 3-amino-δ-thiochromanone crystallized via slow evaporation from chloroform, and the crystal structure was obtained (Figure 4.22). The crystal structure confirmed the structure of this racemic product.

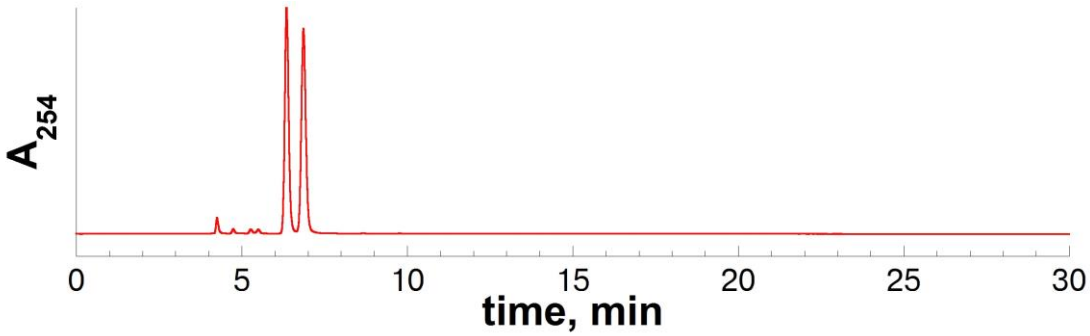


Figure 4.21 Chiral HPLC chromatogram of the purified Boc-protected-3-amino-δ-thiochromanone
Chiral HPLC analysis of showed complete racemization of the product Boc-protected-3-amino-δ-thiochromanone. Analytical chiral HPLC chromatogram (UV detection at 254 nm) of purified Boc-protected-3-amino-δ-thiochromanone using a mixture of isocratic 10% isopropanol in hexanes over 30 minutes on a Daicel ChiralPak 1A column (250 x 4.6 mm, 5 μm particle, 1.0 mL/min).

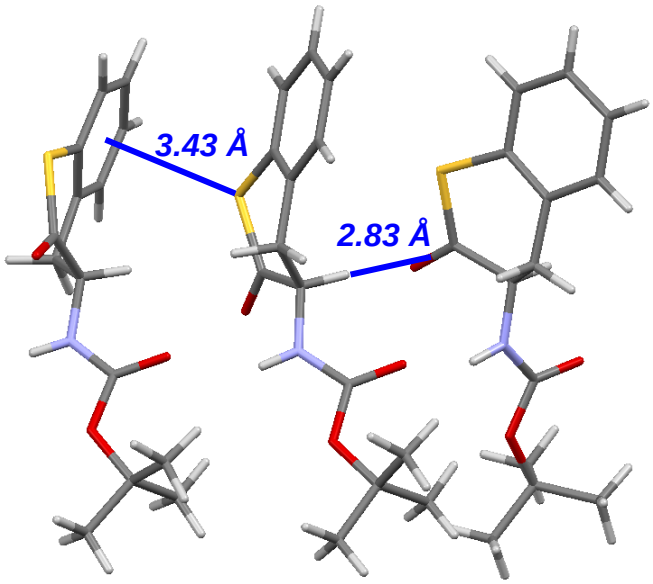


Figure 4.22 Crystal structure of the Boc-protected-3-amino-δ-thiochromanone Boc-protected 3-amino-δ-thiochromanone crystallized via slow evaporation from chloroform, and the crystal structure was obtained (0.75 Å resolution). This structure confirmed the identity of this synthesis product.

The crystal structure of the Boc-protected-3-amino-δ-thiochromanone product showed three molecules within the asymmetric unit. A short contact distance (2.83 Å) between the Hα proton and the carbonyl double bond was observed to stabilized the crystal packing, potentially as a result of a $\pi \rightarrow \sigma^*$ molecular orbital interaction. Furthermore, the crystal packing of the thiochromanone is also stabilized by an intermolecular sulfur-π interaction with the aryl thioether (S··centroid = 3.43 Å). The Boc-protected-3-amino-δ-thiochromanone adopted an extended conformation in the crystal structure ((ϕ, ψ) = (–115°, –157°), (–133°, –161°), (–94°, –168°), based on the *S* stereoisomer).

A synthetic strategy was developed to generate the Boc-protected amino acid with a disulfide-protected thiol using the racemic thiochromanone product. If the

cross-coupling reaction conditions were modified and optimized to generate enantiopure thiochromanone product, then this unique building block could be used in broader applications. The racemic 3-amino-δ-thiochromanone was hydrolyzed with lithium hydroxide in the presence of excess *tert*-butyl 2-pyridyl disulfide, which generated Boc-2-S(S-*tert*-Butyl)-thiol-D,L-phenylalanine in 49% yield (Figure 4.23). This product also crystallized from chloroform, and the crystal structure was solved to 0.68 Å resolution (Figure 4.24).

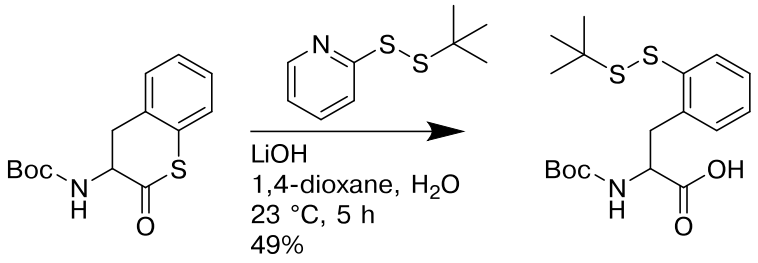


Figure 4.23 Synthesis of Boc-2-S(S-*tert*-butyl)-thiol-D,L-phenylalanine via 2-thiochromanone
Base-catalyzed hydrolysis conditions with the Boc-3-amino-δ-thiochromanone in the presence of excess *tert*-butyl 2-pyridyl disulfide cleanly generated Boc-2(2-S-*tert*-butyl disulfide)-thiophenylalanine, which can be used to incorporate 2-thiophenylalanine at the N-terminus of a peptide via solid-phase peptide-synthesis.

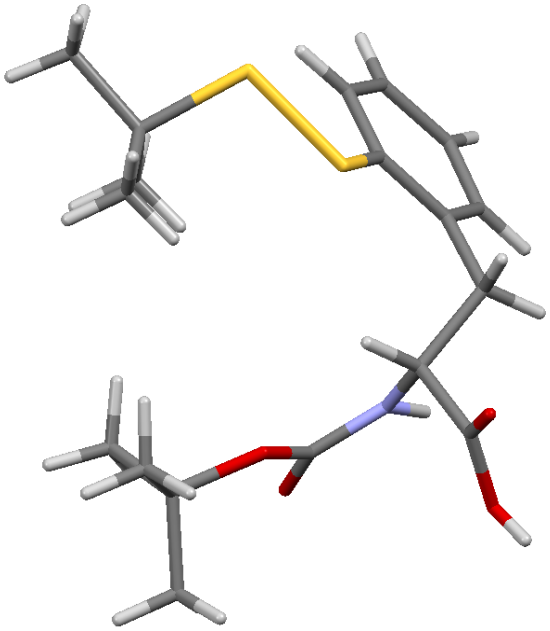


Figure 4.24 Crystal structure of Boc-2-S(S-*tert*-butyl)-thiol-D,L-phenylalanine. Boc-2-S(S-*tert*-Butyl)-thiol-D,L-phenylalanine, synthesized via hydrolysis of the racemic δ-thiochromanone product, was crystallized from chloroform and the crystal structure was obtained (0.68 Å). The crystal structure confirmed that the sample was racemic.

Optimization of the cross-coupling reaction was attempted to increase the enantiopurity of the thiochromanone product. Reaction conditions were screened to determine the role of solvents (toluene, NMP, DMF, 1-butanol, *tert*-amyl alcohol, isopropanol, ethanol, methanol, acetonitrile, 1,4-dioxane), copper ligands (1,10-phenanthroline, 1,2-cis-cyclohexanediol), the thiol source (thiolacetic acid, potassium thioacetate), the base (DIPEA, K_3PO_4, base-free), concentration of reagents, reaction temperatures, and reaction duration. Using conditions that included 1,10-phenanthroline with 1.2 equivalents of DIPEA and 1.2 equivalents of thiolacetic acid in isopropanol at 80 °C for 4 hours, the resultant δ-thiochromanone product had increased enantiopurity (80% ee), but with a reduced reaction yield (approx. 70%). In

this case, alternative cross-coupling conditions (i.e. palladium-catalyzed[23]), which can be conducted at lower reaction temperatures, could potentially avoid the side-reactions that cause racemization of the product. Although the δ-thiochromanone is an interesting synthetic precursor, enantiopurity is requisite for most peptide and protein applications, and the synthesis of enantiopure Boc-2-S(S-*tert*-Butyl)-thiol-L-phenylalanine was necessary for incorporation into peptides that can be used for NCL.

Without established conditions that generated an enantiopure δ-thiochromanone product, the synthetic strategy was revised. An alternative synthesis was developed in order to avoid formation of the 3-amino-δ-thiochromanone product during the reaction conditions, with the carboxylic acid protected as a methyl ester. A methyl ester can be hydrolyzed in the presence of hydroxide to produce a carboxylic acid,[139] which was similar to the final step in the δ-thiochromanone strategy. A methyl ester protection step was added into the synthetic sequence, which only added one additional reaction and column purification compared to the δ-thiochromanone synthetic strategy, while allowing product formation with high enantiopurity (Figure 4.25).

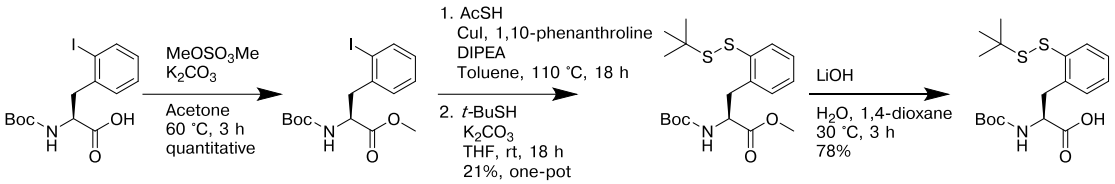


Figure 4.25 Optimized synthesis of Boc-2(2-S-*tert*-butyl disulfide)-thiophenylalanine
Enantiopurity is maintained during the cross-coupling reaction when the carboxylic acid is protected as a methyl ester, avoiding formation of the δ-thiochromanone product. Thiolysis and generation of the asymmetric disulfide were conducted as a one-pot reaction from the cross-coupling reaction. The asymmetric disulfide product was hydrolyzed with LiOH to produce Boc-2(2-S-*tert*-butyl disulfide)-thio-L-phenylalanine with retention of enantiopurity.

This reaction sequence generated enantiopure Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine in three steps from commercially available Boc-2-iodophenylalanine, requiring only two purification steps. Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine was used in SPPS with the peptide RAFS on resin using standard coupling reaction conditions (Figure 4.26). The model peptide was synthesized 2-thiophenylalanine at the N-terminus without subjecting the peptide to high temperatures that were required in the cross-coupling reaction conditions. The disulfide was readily reduced in solution using DTT. The resultant peptides (2-SH-Phe)RAFS-NH_2 synthesized via solid phase cross-coupling reaction and via SPPS with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine, were verified to be identical by HPLC coinjection (Figure 4.26e) and by NMR spectroscopy (Figure 4.27).

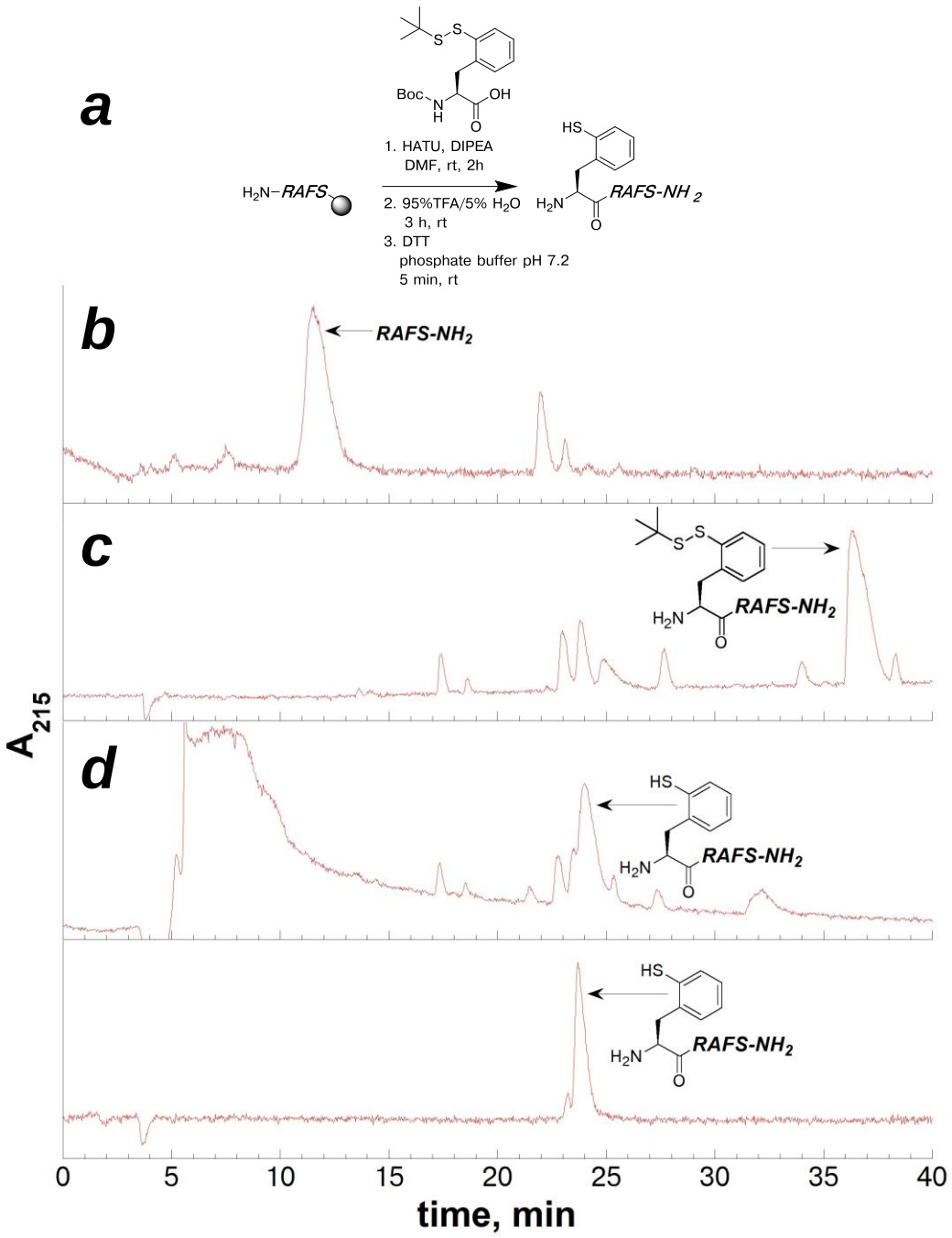


Figure 4.26 Synthesis of the peptide (2-SH-Phe)RAFS-NH$_2$ via SPPS with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine

(a) Reaction scheme to synthesize the peptide (2-SH-Phe)RAFS-NH_2 via SPPS using Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine; (b) HPLC chromatogram of the model peptide RAFS-NH_2 prior to any reactions; (c) the HPLC chromatogram of the peptide products that resulted from the coupling reaction with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine; (d) the HPLC chromatogram that resulted from the reaction of the peptide products with DTT; (e) the HPLC chromatogram of the coinjection of the peptides (2-SH-Phe)RAFS-NH_2 synthesized by two different strategies. All chromatographic steps were conducted using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS.

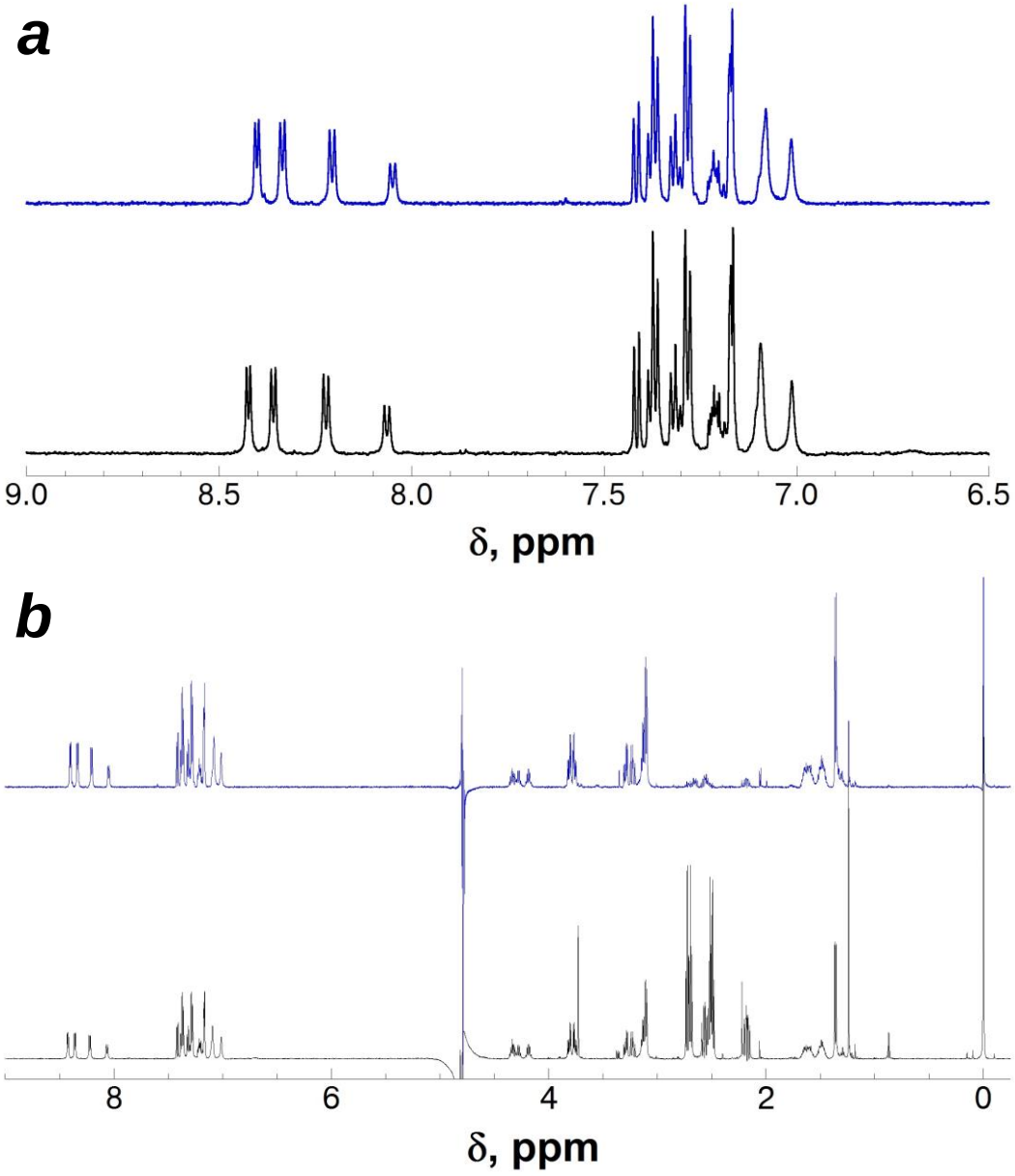


Figure 4.27 ^{1}H NMR spectra for the peptide (2-SH-Phe)RAFS-NH_2 synthesized by two strategies

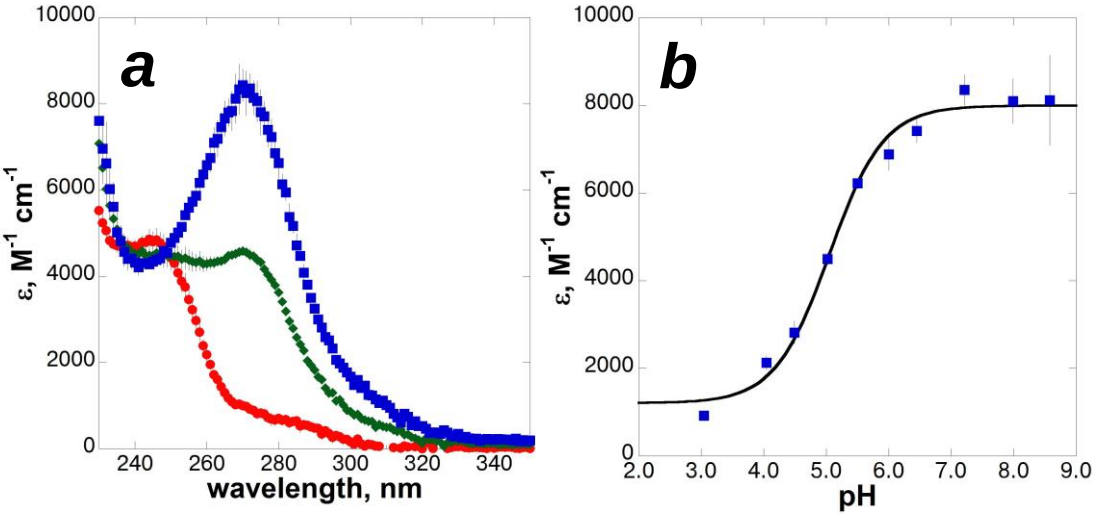
^{1}H NMR spectra of the peptide (2-SH-Phe)RAFS-NH_2 in 90% H_2O/10% D_2O with 5 mM phosphate pH 4, 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP. Blue: the peptide synthesized by SPPS using Boc-2-S(S-*tert*-butyl)-thiolphenylalanine; black: the peptide synthesized by the cross-coupling reaction on the peptide on solid phase (the peaks at 2.71 and 2.50 ppm are assigned to TCEP). (a) the amide and aromatic region; (b) and full spectra.

Using a 3-step synthesis, Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine was synthesized and incorporated into the model peptide, XRAFS. Following the cleavage and deprotection reaction from resin, and reduction in solution phase using DTT, the peptide (2-SH-Phe)RAFS-NH_2 was cleanly synthesized without any exposure to elevated reaction temperatures. The model peptide containing 2-thiophenylalanine at the N-terminus was synthesized in two ways: by the cross-coupling reaction on the peptide on solid-phase and by the amide coupling reaction with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine. Both of the peptides (2-SH-Phe)RAFS-NH_2 that were synthesized using two different methodologies were established to be identical via ^{1}H NMR and coinjection on HPLC. The copper-mediated cross-coupling reaction methodology on solid phase generated the peptide with no silica column chromatography, but could potentially be associated with side reactions with amino acid chains within the protected peptide on solid phase. The methodology using SPPS with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine does not have the same potential for side reactions as the cross-coupling reaction conditions, but requires a 3-step synthesis to generate the protected amino acid in 16% overall yield. Having established two different approaches to synthesizing peptides containing 2-thiophenylalanine at the N-terminus, the reactivity of 2-thiophenylalanine for NCL reactions were examined in model peptides and proteins.

4.2.5 UV-Vis characterization and pK_a of peptides containing N-terminal 2-thiophenylalanine

As an aryl thiol, 4-thiophenylalanine is more acidic (pK_a = 6.4) compared to cysteine (pK_a ~ 8.5).[169, 465] NCL reactions rely on the thiolate form of cysteine for the thioester exchange reaction to proceed.[396] We hypothesized that 2-thiophenylalanine

would be more acidic than cysteine, which would allow it to react with thioesters under mildly acidic conditions. The pK_a of the (2-SH-Phe)RAFS-NH_2 peptide was measured via UV-Vis spectroscopy. A red shift was observed in the UV absorbance with increasing pH (Figure 4.28a). From these data, the pK_a of 2-thiophenylalanine at the N-terminus of a peptide was measured by fitting the pH dependence of the molar extinction coefficient to Equation 6 (Figure 4.28).



$$\varepsilon_{272} = \frac{\varepsilon_{272,thiolate} 10^{(pH-pKa)} + \varepsilon_{272,thiol}}{1+10^{(pH-pKa)}}$$

$pK_a = 5.05 \pm 0.08$ (6)

Figure 4.28 pK_a Determination and UV Spectra of the peptide (2-SH-Phe)RAFS-NH_2
UV Spectra were measured in aqueous buffered solutions (50 mM phosphate) in the presence of TCEP (1 mM) to prevent disulfide formation. Concentrations of stock solutions of the peptide (2-SH-Phe)RAFS-NH_2 were measured via Ellman's test.[109] Error bars indicate standard error. (a) UV spectra of the peptide (2-SH-Phe)RAFS-NH_2 at pH 3.0 (red, λ_{max} 244 nm), pH 5.0 (green, λ_{max} 270 nm), and 7.2 (blue, λ_{max} 270 nm); (b) pH dependence of the molar extinction coefficient of (2-SH-Phe)RAFS-NH_2 at 272 nm (blue squares). The pK_a was calculated based on a curve fit of pH-dependence on molar absorbance at 272 nm, following Equation 6.

Interestingly, the pK_a of 2-thiophenylalanine at the N-terminus of a peptide was considerably lower (pK_a = 5.1) than the pK_a for 4-thiophenylalanine within a peptide (pK_a = 6.4). Potentially, the thiolate anion is stabilized by an intramolecular hydrogen bond or by a salt-bridge interaction with the N-terminal amine (Figure 4.29).

Figure 4.29 Potential stabilizing interaction between anionic 2-thiophenylalanine and the N-terminal amine
The measured pK_a of 2-thiophenylalanine at the N-terminus of a peptide was lower than the pK_a of 4-thiophenylalanine within a peptide.[169] The amine at the N-terminus can potentially interact with the 2-thiophenylalanine thiolate, decreasing the thiolacid dissociation constant.

At physiological pH, 2-thiophenylalanine at the N-terminus of a peptide is mostly anionic, and can act as a nucleophile in NCL reactions. The pK_a of 2-thiophenylalanine at the N-terminus of a peptide is slightly lower than selenocysteine (Sec pK_a = 5.2). [466] Selenocysteine has been utilized for NCL reactions and in expressed protein ligation with many proteins and peptides.[448, 467-469] Selenocysteine can act as a nucleophile in NCL reactions under mildly acidic conditions (e.g. pH 5).[468] The increased acidity of 2-thiophenylalanine can potentially allow for site-selective NCL reactions in the presence of cysteines under mildly acidic conditions.

4.2.6 Initial 2-thiophenylalanine-mediated native chemical ligation of a glycine-containing thioester model peptide

2-Thiophenylalanine was synthesized in a practical manner at the N-terminus of a peptide using either a solid-phase cross-coupling reaction approach, or by synthesizing the protected amino acid for incorporation via SPPS. In the initial synthesis of the peptide (2-SH-Phe)RAFS-NH_2, a rapid intramolecular *S*→*N* acyl transfer reaction was observed between the amine at the N-terminus and the 2-thioacetyl-phenylalanine (Chapter 4.2.1). This *S*→*N* acyl transfer reaction was

complete in less than 5 minutes, and the pH of the solution under these conditions was potentially slightly acidic, due to residual TFA in the peptide precipitate from the cleavage and deprotection reactions. The rapid reaction strongly suggested that the peptide (2-SH-Phe)RAFS-NH_2 could rapidly react with peptides containing thioesters.

We subjected the purified peptide (2-SH-Phe)RAFS-NH_2 to ligation reaction conditions with a model peptide containing a thioester. Ac-LYRAX thioester peptides have been well studied as model peptides to demonstrate NCL reactions.[401, 470, 471] In general, the peptides contain a C-terminal thioester that is air stable and provides a good leaving group for the thioester exchange reaction.[396] Peptides containing S-benzyl (-SBn) thioesters are reasonably air-stable, and can react with an aryl thiol additive in the ligation reaction conditions (such as thiophenol) to generate a more reactive thioester *in situ*.[432, 470] Initially, the (2-SH-Phe)RAFS-NH_2 peptide was subjected to ligation reaction conditions with the peptide Ac-LYRAG-SBn in the presence of excess thiophenol (Figure 4.30), which maintained reductive reaction conditions and generated the S-phenyl thioester (Ac-LYRAG-SPh) *in situ*, which contains a better leaving group for NCL reaction than Ac-LYRAG-SBn.

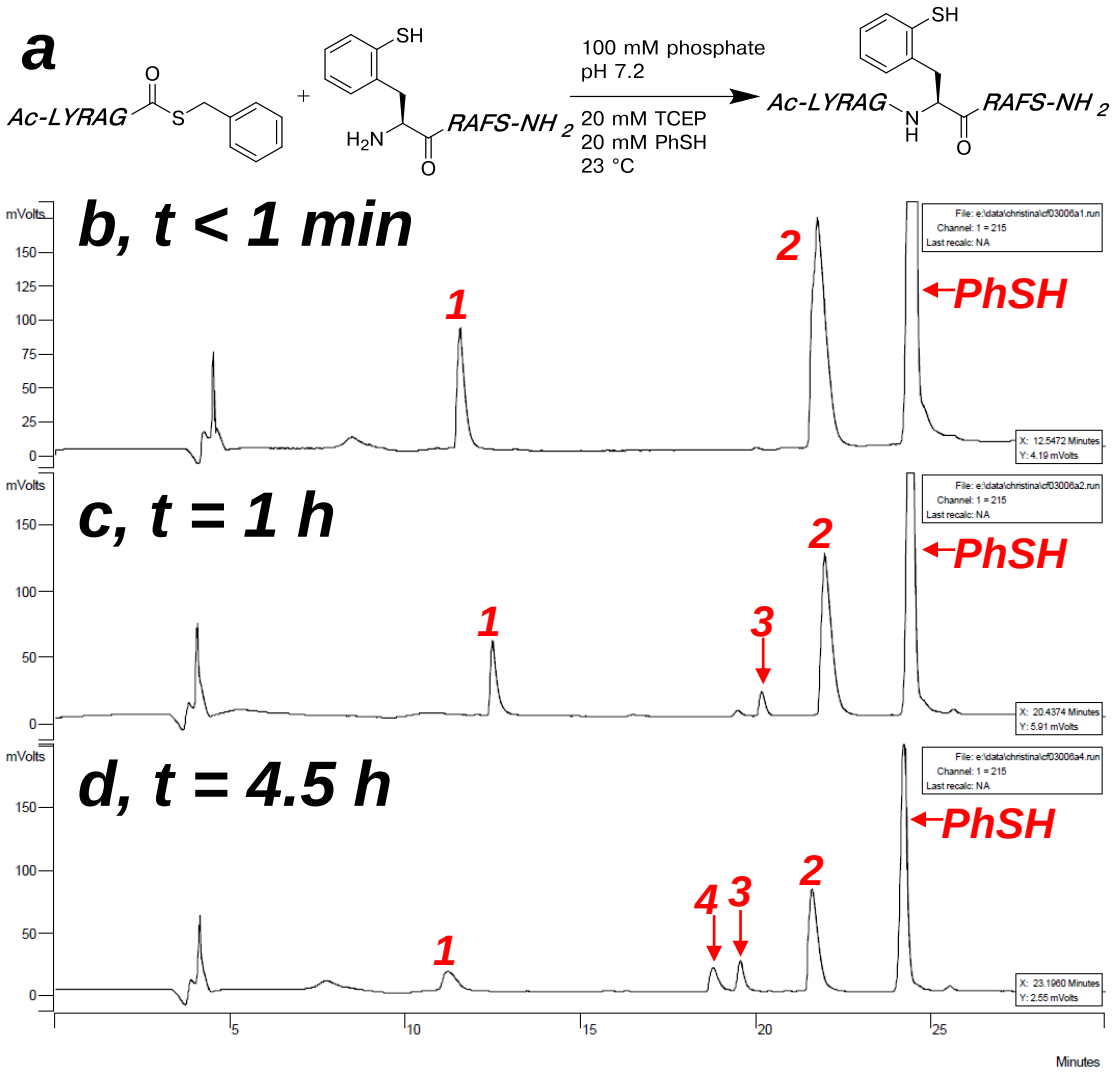


Figure 4.30 Ligation reactions between (2-SH-Phe)RAFS-NH$_2$ and Ac-LYRAG-SBn in the presence of thiophenol

(a) Reaction scheme for the ligation reaction between the peptides (2-SH-Phe)RAFS-NH$_2$ and Ac-LYRAG-SBn; (b, c, d) HPLC chromatograms of the crude reaction products that resulted from the NCL reaction at indicated time points using a linear gradient of 20-80% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 30 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS.

1 indicates the (2-SH-Phe)RAFS-NH$_2$ peptide

2 indicates the Ac-LYRAG-SBn peptide

3 indicates the Ac-LYRAG-SPh peptide (generated *in situ*)

4 indicates the Ac-LYRAG(2-SH-Phe)RAFS-NH$_2$ ligation reaction product

Thiophenol was included as an additive in the NCL reaction to react with the peptide Ac-LYRAG-SBn to generate Ac-LYRAG-SPh *in situ*, as thiophenolate is a better leaving group than benzyl thiolate, and to maintain reductive reaction conditions and prevent disulfide formation of the peptide (2-SH-Phe)RAFS-NH_2.[432] After 1 hour of the ligation reaction with the peptides Ac-LYRAG-SBn and (2-SH-Phe)RAFS-NH_2 (Figure 4.30c), there was little formation of the ligation reaction product (Ac-LYRAG(2-SH-Phe)RAFS-NH_2, **4** in Figure 4.30) and an observable quantity of the peptide Ac-LYRAG-SPh (**3** in Figure 4.30). After 4.5 hours of the ligation reaction (Figure 4.30d), more of the ligated product peptide was generated, while the amount of the peptide Ac-LYRAG-SPh was nearly the same as in 1 hour of ligation reaction (**3** in Figure 4.30c and 4.30d). These results suggest that the rate of the ligation reaction was limited by the rate of formation of the peptide Ac-LYRAG-SPh, and was not limited by the reaction with the peptide containing 2-thiophenylalanine. In order to more accurately determine the relative reaction rates of peptides containing 2-thiophenylalanine, the peptide containing the S-phenyl thioester was used directly in the ligation reaction. Using the peptide Ac-LYRAG-SPh directly in the ligation reaction avoids the rate-determining step for generating this substrate *in situ*. The (2-SH-Phe)RAFS-NH_2 peptide was subjected to ligation reaction conditions with the peptide Ac-LYRAG-SPh in the presence of excess thiophenol (Figure 4.31).

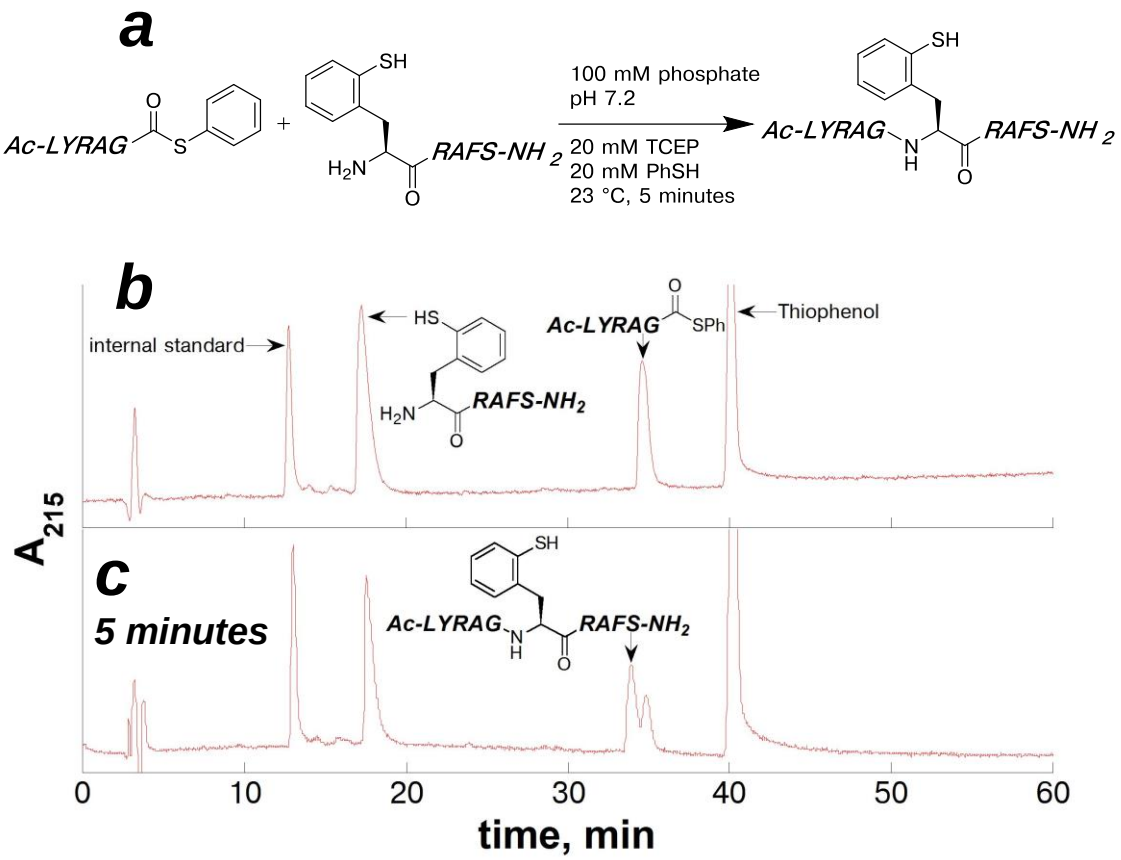


Figure 4.31 Ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh in the presence of thiophenol
(a) Reaction scheme for the ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh; (b) HPLC chromatogram of a coinjection representing the initial NCL reaction with an internal standard (2 mM phenol). The substrates were maintained under acidic conditions to prevent reaction; (c) HPLC chromatogram of the crude peptide products resulting from the NCL reaction after 5 minutes at 23 °C using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

The NCL reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh was significantly faster than the NCL reaction using the peptide Ac-LYRAG-

SBn. The NCL reaction was 62% complete within 5 minutes at 23 °C. These reaction conditions used lower concentrations of the reactant peptides (1.0-2.0 mM peptide) than are typically reported in the literature for comparable ligation reactions (~10 mM peptide containing β-mercaptophenylalanine[401]).

Thiol additives have been known to accelerate ligation reaction rates in studies with other model peptides.[431] Peptides containing C-terminal thioesters are susceptible to hydrolysis, and thioesters that have better leaving groups for NCL reactions are also more sensitive to hydrolysis. Therefore, aryl thiol additives are primarily used to generate a more reactive peptide containing a thioester *in situ*. However, we did not observe significant hydrolysis of the peptide containing an S-phenyl thioester, provided the peptides were stored at -20 °C with minimal exposure to air. Given that the peptide containing the S-phenyl thioester was used directly, there was no need to include thiophenol as an additive. The TCEP included in the ligation reaction buffer should be sufficient to reduce disulfides. Therefore, in order to examine the role of thiophenolate in the ligation reaction mediated by 2-thiophenylalanine, the peptide (2-SH-Phe)RAFS-NH_2 was subjected to ligation reaction conditions with the peptide Ac-LYRAG-SPh in the absence of thiol additives (Figure 4.32).

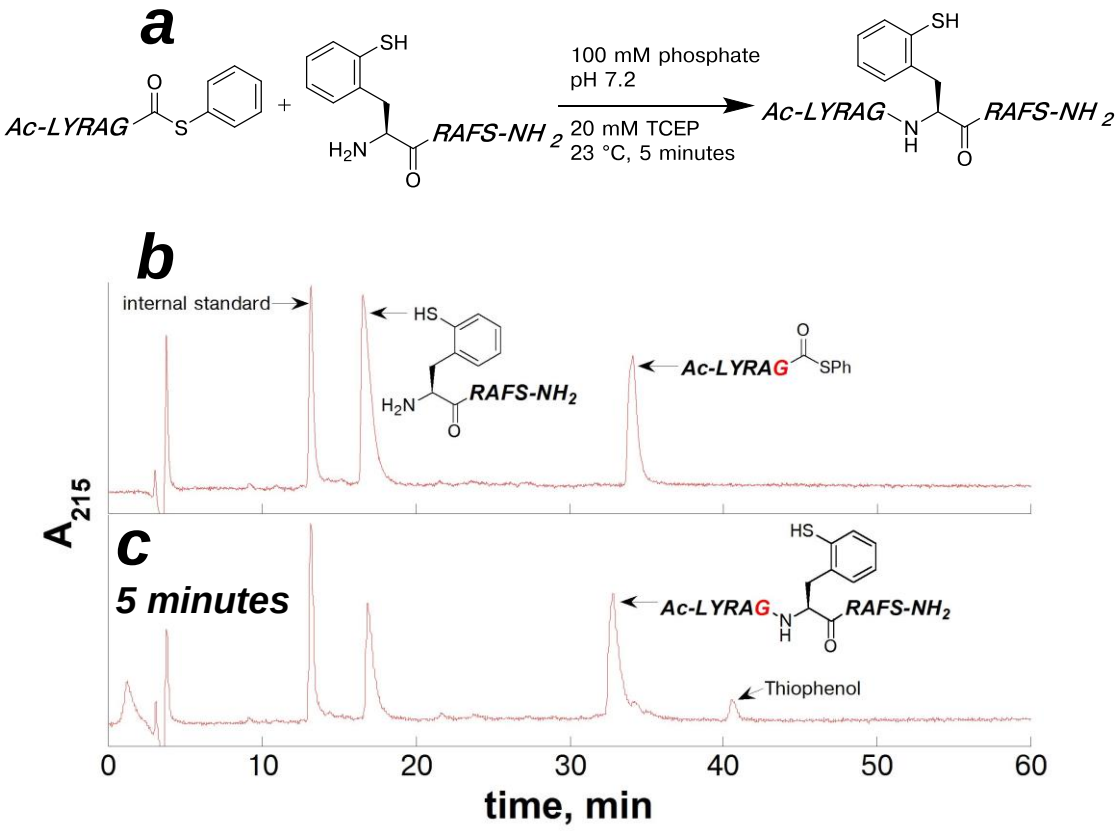


Figure 4.32 Ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh in the absence of thiol additives

(a) Reaction scheme for the ligation reaction between the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh in the absence of thiophenol; (b) HPLC chromatogram of a coinjection representing the initial NCL reaction with an internal standard (2 mM phenol). The substrates were maintained under acidic conditions to prevent reaction; (c) HPLC chromatogram of the crude peptide products resulting from the NCL reaction after 5 minutes at 23 °C without thiophenol using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

Surprisingly, the NCL reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh was nearly complete in 5 minutes *without* thiophenol as an additive. The ligation reaction mediated by 2-thiophenylalanine was faster in 5 minutes in the absence of thiol additives (>95% complete) than with thiophenol (62% complete). While thiol additives such as thiophenol or 4-mercaptophenylacetic acid are used to accelerate NCL reaction rates,[431] NCL reactions that are mediated by 2-thiophenylalanine are fundamentally different from NCL reactions that are mediated by alkyl thiols. In comparison to cysteine or related alkyl thiols, 2-thiophenylalanine has increased acidity, increased steric bulk, and proceeds through a 7-membered ring intermediate to generate the amide bond. These properties of 2-thiophenylalanine effects the reactivity of this amino acid, and affects the role of the thiol additives during the ligation reaction.

The NCL reaction is initiated as 2-thiophenylalanine or cysteine undergoes nucleophilic attack on the peptide containing the S-phenyl thioester, and a thioester intermediate peptide is formed (Figure 4.33, k_1). This reaction to form the thioester intermediate is reversible, and so the intermediate species can either: undergo an intramolecular *S*→*N* acyl transfer reaction to irreversibly form the ligation product (Figure 4.33, k_3), or react with thiophenolate and regenerate the reactant peptides (Figure 4.33, k_2). In NCL reactions that are mediated by cysteine, the irreversible intramolecular rearrangement to form the ligation product proceeds through a 5-membered ring transition state, while NCL reactions mediated by 2-thiophenylalanine proceed through a 7-membered ring transition state. A 5-membered ring transition state is more favored than a 7-membered ring transition state, and so k_3 is presumably faster for cysteine than for 2-thiophenylalanine (Figure 4.33). Furthermore, the leaving

group for k_2 is different between NCL reactions mediated by cysteine or 2-thiophenylalanine. In NCL reactions mediated by cysteine, an alkyl thiolate is the leaving group for k_2, while an aryl thiolate is the leaving group for k_2 in 2-thiophenylalanine-mediated NCL reactions. Therefore, k_2 is slower for cysteine-mediated NCL reactions than for 2-thiophenylalanine-mediated NCL reactions. The reaction that proceeds through k_2 is dependent on the concentration of thiophenolate, either as an additive or generated from the initial nucleophilic attack on the peptide containing the S-phenyl thioester (k_1). In the presence of added thiophenolate, the overall NCL reaction is slowed due promotion of the reaction proceeding through k_2 (Le Chatelier's principle). In essence, for 2-thiophenylalanine-mediated NCL the reversible step k_2 is more rapid in comparison to cysteine-mediated NCL, and the irreversible step k_3 is slower compared to cysteine. However, with the increased acidity of 2-thiophenylalanine compared to cysteine, under acidic pH conditions a larger mole fraction of 2-thiophenylalanine is anionic and available to react with an S-phenyl thioester. Therefore, 2-thiophenylalanine could potentially react selectively over cysteine under acidic conditions, and this possibility is discussed further in section 4.2.8.

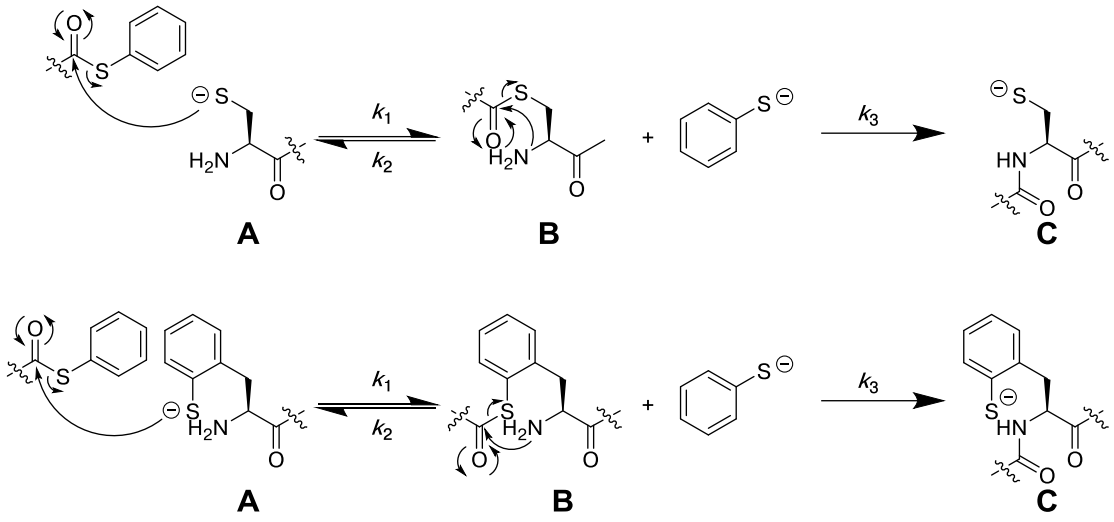


Figure 4.33 Kinetics of the ligation reaction between cysteine or 2-thiolphenylalanine and a thioester
For NCL reactions mediated by cysteine (a), $k_3 > k_2[RS^-]$, where cysteine proceeds through a more favored 5-membered ring transition state to form the ligated product. In contrast, for NCL reactions mediated by 2-thiophenylalanine (b), the rate of k_3 and $k_2[RS^-]$ are nearly the same, as 2-thiophenylalanine proceeds through a less favored 7-membered ring transition state. In addition, $k_2[RS^-]$ is faster for 2-thiophenylalanine than for cysteine, since the aryl thiolate is a better leaving group than the alkyl thiolate. Therefore, the intermediate species B proceeds to C more rapidly for NCL reactions mediated by cysteine than for reactions mediated by 2-thiophenylalanine.

At the time of this work, NCL reactions with aryl thiolated amino acids had not been accomplished, and so these mechanistic details were unique to 2-thiophenylalanine. Practical access to 2-thiophenylalanine, either via solid-phase cross-coupling reaction on solid phase or via SPPS with the protected amino acid, has provided additional insights into the mechanism of NCL reactions. After initial submission but prior to publication of this work, Malins *et al.*[472] has since reported similar mechanistic details with an aryl thiolated tryptophan, corroborating our observations.

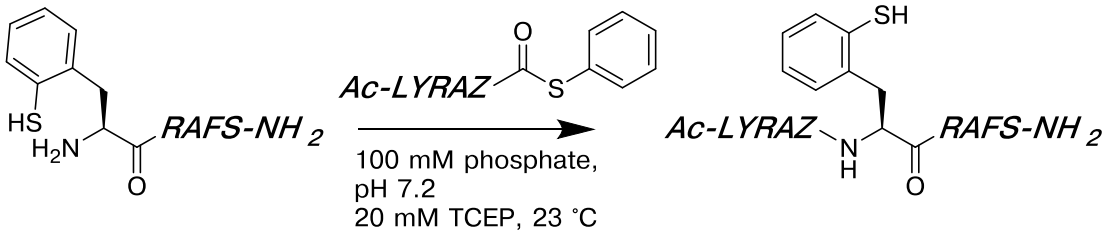
4.2.7 Substrate scope of 2-thiophenylalanine-mediated native chemical ligation with thioester peptides

Having determined that thiol additives slow the overall reaction rate in the context of NCL reactions mediated by amino acids with aryl thiolates, we sought to examine the scope of the ligation reactions for peptides containing S-phenyl thioesters. The reaction efficiency can vary depending on the nature of the amino acid at the C-terminus of the peptide containing the thioester.[470] For example, the conformational flexibility and low steric hindrance of glycine at the C-terminus allows for enhanced reactivity of peptides containing glycine S-phenyl thioesters in NCL reactions. In contrast, peptides containing prolyl S-phenyl thioesters are known to be among the least efficient for NCL reactions.[445] β-Branched amino acids at the C-terminus of the peptide containing the thioester are generally significantly slower than glycine for NCL reactions.[470] Side-chain groups that can potentially interact favorably with the thiolate can enhance the efficiency of NCL reactions.[470] Intriguingly, aromatic amino acids at the C-terminus of the peptides containing thioesters have increased reactivity for NCL reactions over other hydrophobic amino acids,[470] potentially via an S··π aromatic interaction between the cysteine thiolate and the aromatic amino acid (Chapter 2).

In order to evaluate the efficiency and broader applicability of ligation reactions mediated by 2-thiophenylalanine, the scope of the peptides containing S-phenyl thioesters was screened for non-glycine amino acids at the C-terminus. Hydrophobic, β-branched, and sterically hindered amino acids were examined for ligation reaction efficiency using peptides Ac-LYRAZ-SPh (where where Z = Gly, Ala, Val, Leu, Phe, or Pro). The ligation reaction efficiency of these peptides Ac-LYRAZ-SPh with the peptide (2-SH-Phe)RAFS-NH_2 are shown in Table 4.1.

Table 4.1. Substrate scope of peptides Ac-LYRAZ-SPh in ligation reactions with (2-SH-Phe)RAFS-NH_2.

Ligation reaction conversion and efficiency were evaluated via HPLC. Product identity was established via ESI-MS. Product conversions were calculated based the depletion of the peptide containing S-phenyl thioester compared to an initial time-point of the reaction or a representative coinjection. Peak integrations were normalized against an internal standard (2 mM phenol), and are averaged from at least two independent reactions.
*Indicates an estimated conversion, since the amount of remaining thioester at the given time-point was too small to accurately calculate an integration.



Entry	Z =	Ligation Reaction Duration	Conversion to Ligated Peptide
1	Gly	5 minutes	88%
2	Ala	1 h	95%
3	Leu	6 h	90%
4	Val	1 h	>95%*
5	Phe	1 h	90%
6	Pro	12 h	84%

For all peptides examined, the ligation reactions employing 2-thiophenylalanine resulted in higher yields in significantly less reaction time than similar ligation reactions reported in the literature (including thiolated tryptophan-mediated NCL with peptides containing S-phenyl thioesters at the C-terminus).[470, 472] All of the ligation reactions using the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAZ-SPh proceeded rapidly at 23 °C in the absence of any thiol additives. Ligation

reactions with peptides containing branched amino acids at the C-terminus preceeded to near completion in 1-6 hours using (2-SH-Phe)RAFS-NH_2 (Table 4.1, entries 3 and 4). In the work by Malins *et al.*[472] ligation reactions were examined with peptides containing aryl-thiolated tryptophan and peptides containing S-phenyl thioesters. Malins *et al.*[472] observed 65% ligation reaction completion in 24 hours with the peptide containing phenylalanine-S-phenyl thioester, using 4 mM ligation reaction concentrations. In contrast, a similar reaction using 2-thiophenylalanine was 90% complete in only 1 hour using 1-2 mM peptide concentrations (Table 4.1, entry 5). For 2-thiophenylalanine-mediated NCL reactions, the least efficient reaction was with the prolyl thioester peptide, which was 85% complete in 12 hours (Table 4.1, entry 6). Malins *et al.*[472] reported 58% ligation reaction yield in 30 hours with a peptide containing a prolyl-S-phenyl thioester using a peptide containing an aryl-thiolated tryptophan. The NCL reactions with the peptide (2-SH-Phe)RAFS-NH_2 proceeded efficiently and rapidly with a broad scope of peptides Ac-LYRAZ-SPh.

The rapid ligation reaction rates with peptides containing 2-thiophenylalanine can potentially be attributed to the increased acidity of this amino acid. At pH 7, there is a larger mole fraction of 2-thiophenylalanine that is anionic than there is for cysteine, due to the higher pK_a for cysteine (pK_a = 8.5[465]) than for 2-thiophenylalanine pK_a = 5.1, Chapter 4.2.5). The thiolate form is required for initiating the NCL reaction and formation of the thioester intermediate (Figure 4.33). With a larger mole fraction of anionic 2-thiophenylalanine at pH 7.2 (compared to cysteine), more of the aryl thiolated amino acid can react with peptides containing C-terminal thioesters. However, NCL reactions reported in the literature often have different thioester groups

(e.g. S-benzyl), higher reactant peptide concentrations, thiol additives, and different peptide sequences in comparison to those examined in the present study.[431, 432, 445, 473]

4.2.8 Competitive native chemical ligation with other thiolated amino acids

Prior examples of NCL reactions in the literature have different reactant peptide concentrations and different thioester functional groups, and these differences in reaction conditions and functional groups can have significant consequences on the ligation reaction efficiency (e.g. S-phenyl thioester vs. S-benzyl thioester, Chapter 4.2.6). The ligation reactions that are mediated by 2-thiophenylalanine cannot be directly compared to the literature examples of cysteine-mediated NCL reactions, since few examples have been conducted in the absence of thiol additives. In order to compare the ligation reactions mediated by 2-thiophenylalanine with other thiolated amino acids, additional peptides XRAFS-NH_2 were synthesized containing either cysteine or 2*S*,4*R*-mercaptoproline (Mpt) at the N-terminus (Figure 4.34a). Mpt is a thiolated amino acid that was expected to react with lower efficiency in NCL reactions in comparison to either cysteine or 2-thiophenylalanine, due to the conformational restriction on the secondary amine. These peptides XRAFS-NH_2 were also compared for ligation reaction efficiency under mildly acidic conditions. It was expected that the increased acidity of 2-thiophenylalanine would allow for superior reactivity for NCL reactions compared to cysteine or Mpt under mildly acidic conditions. Competitive ligation reactions were conducted with the peptide Ac-LYRAA-SPh using peptides XRAFS-NH_2, where X = 2-thiophenylalanine, cysteine, or 2*S*,4*R*-mercaptoproline (Figure 4.34). The peptides were subjected to ligation reaction conditions at pH 7.2, pH 5.0, or pH 4.0. The peptide containing Mpt was synthesized via modification of 2*S*,4*R*-hydroxyproline on peptides on solid phase,[373] discussed in Chapter 4.2.11.

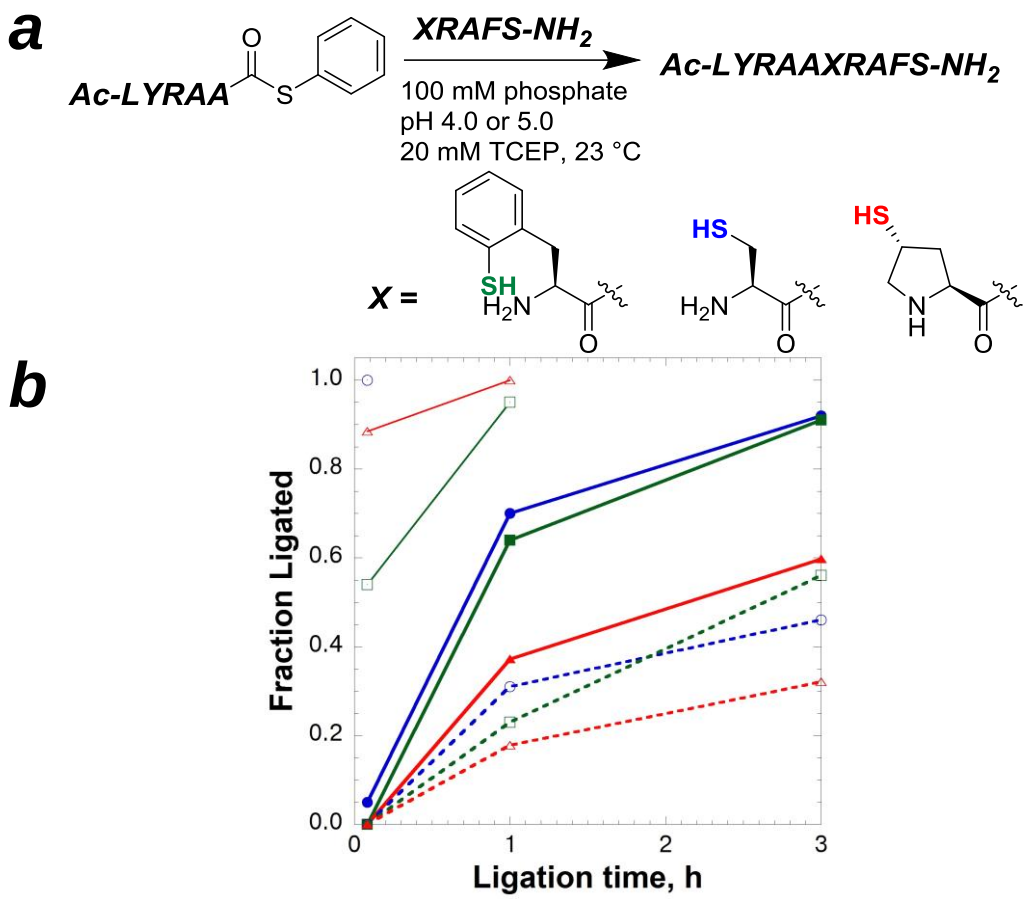


Figure 4.34 Comparison of ligation reactions of the peptide Ac-LYRAA-SPh with the peptides XRAFS-NH_2 (X = 2-SH-Phe, Cys, or Mpt)
Competitive ligation reactions using peptides Ac-LYRAA-SPh and XRAFS-NH_2, where is X either 2-thiophenylalanine (green squares), cysteine (blue circles), or 2*S*,4*R*-mercaptoproline (red triangles). The peptides were subjected to ligation reactions conditions at different pH: pH 7.2 (open figures, thin solid line), pH 5.0 (closed figures, thick solid line), or pH 4.0 (open figures, dashed line). (a) Reaction scheme for the competitive ligation reactions; (b) reaction conversion, calculated based on depletion of the peptide Ac-LYRAA-SPh evaluated by HPLC, compared to a representative coinjection. Peak integrations from the HPLC trance were normalized against an internal standard of 2 mM phenol). For the ligation reaction with CRAFS-NH_2 and Ac-LYRAA-SPh at pH 7.2, no Ac-LYRAA-SPh was observed after 5 minutes. The lines are interpolations employed to guide the eye and do not represent a mathematical fit.
The un-rearranged thioester intermediate (not containing an amide bond) was observed for ligation reactions with 2*S*,4*R*-mercaptoproline, and was included in the reaction conversion, although this species could potentially regenerate the reactant peptides.

Proline is the only cyclic natural amino acid containing a secondary amine, and it is widely known to be a difficult substrate for ligation reactions using peptides containing C-terminal prolyl-thioesters.[445, 473] For the synthesis of peptides via NCL that are proline-rich,[474] this decreased ligation reaction efficiency can be detrimental. As a means to conduct NCL reactions in peptides involving proline, Shang *et al.*[454] used peptides containing commercially available mercaptoproline at the N-terminus for NCL reactions, which generates proline after desulfurization. Shang *et al.*[454] found that peptides containing 2*S*,4*R*-mercaptoproline reacted with aryl-ester peptides (85% ligation reaction in 2.5 h), while 2*S*,4*S*-mercaptoproline was relatively unreactive under the same condition (<5% ligation reaction in 28 h).[454] In the work herein, peptides containing 2*S*,4*R*-mercaptoproline at the N-terminus were utilized as a control for ligation reactions compared to peptides containing cysteine and 2-thiophenylalanine.

At pH 7.2, the NCL reaction mediated by 2-thiophenylalanine proceeded efficiently at room temperature in the absence of thiol additives, but was the slowest compared to NCL reactions mediated by cysteine or Mpt. Within 5 minutes at room temperature at pH 7.2, no thioester peptide was observed in the NCL reaction with the peptide CRAFS-NH_2. The NCL reaction efficiency of the peptides XRAFS-NH_2 at pH 7.2 was Cys > Mpt > 2-thiophenylalanine, with all reactions nearly complete within 1 hour. While a larger mole fraction of 2-thiophenylalanine is deprotonated at pH 7.2 compared to Mpt or cysteine, the relative reactivity of this amino acid at the N-terminus of a peptide was *not* enhanced in NCL reactions. These results suggest that the 7-membered ring intermediate of 2-thiophenylalanine may hinder the overall NCL reaction rate, even though this amino acid is more acidic than cysteine or Mpt.

For the ligation reactions under acidic conditions using the peptide Ac-LYRAA-SPh, the reactions using the peptide CRAFS-NH_2 or (2-SH-Phe)RAFS-NH_2 proceeded with comparable overall rates, while the reaction using MptRAFS-NH_2 was considerably slower. The ligation reactions mediated by cysteine or 2-thiophenylalanine were nearing completion after 3 hours at pH 5.0 (Figure 4.34b). The reaction between the peptides Ac-LYRAA-SPh and MptRAFS-NH_2 was more rapid than expected at pH 5.0, with 78% of the peptide Ac-LYRAA-SPh having reacted after 3 hours. However, this 78% reaction conversion was a result of the reaction between Ac-LYRAA-SPh and MptRAFS-NH_2 to generate the non-rearranged thioester intermediate and not for formation of the final ligated peptide product. As can be seen in Figure 4.35, showing the HPLC chromatograms for the ligation reaction with Ac-LYRAA-SPh and MptRAFS-NH_2 at pH 5, the product is mostly the thioester intermediate, with only a small fraction of the final, amide-bond-containing ligated product formed. These data indicate that the *S*→*N* acyl transfer reaction with 2*S*,4*R*-mercaptoproline is considerably slower[454] compared to cysteine- or 2-thiophenylalanine, potentially as a result of a more hindered secondary amine.

At pH 4.0, the differences in reactivity between the three different thiolated-amino acids became more pronounced. As in the ligation reaction at pH 5.0, the reaction between MptRAFS-NH_2 and LYRAA-SPh at pH 4.0 was slower, only 24% complete in 3 hours, with no observed formation of the final ligated product. In comparing the ligation reactions at pH 4.0 using CRAFS-NH_2 or (2-SH-Phe)RAFS-NH_2, the cysteine-mediated ligation reaction was more complete in 1 hour, but the reaction with the peptide containing 2-thiophenylalanine was more complete after 3 hours (56%) compared to cysteine (46%).

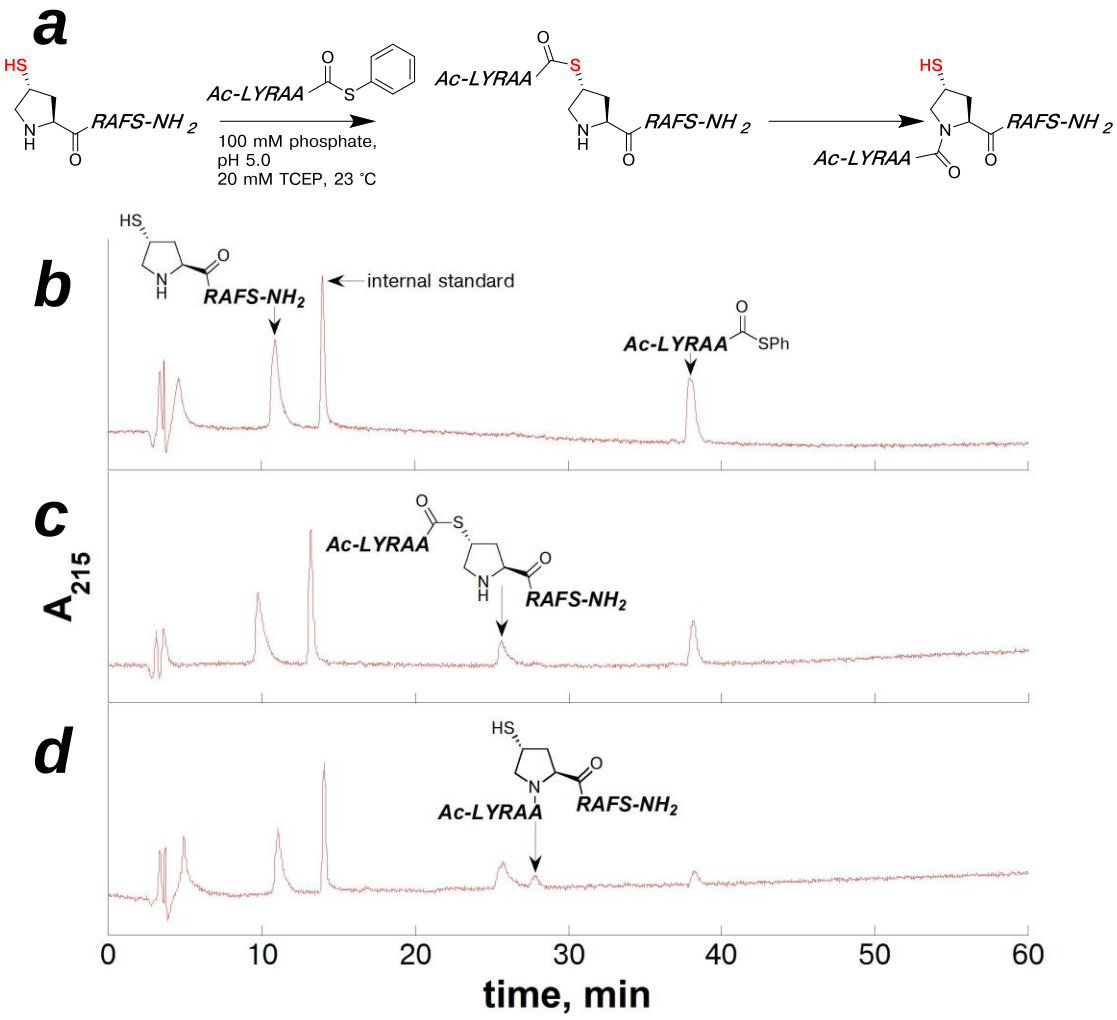


Figure 4.35 Ligation reaction using the peptides Ac-LYRAA-SPh and MptRAFS-NH$_2$ at pH 5.0

(a) Reaction scheme for the ligation reaction between the peptides MptRAFS-NH$_2$ and Ac-LYRAA-SPh at pH 5.0; HPLC chromatograms of the ligation reaction after 5 minutes (b), 1 hour (c), and 3 hours (d) between the peptides (4*R*-mercaptoproline)RAFS-NH$_2$ (19, 2.0 mM) and Ac-LYRAA-SPh (5, 0.5 mM) in ligation buffer (100 mM phosphate, pH 5.0, 20 mM TCEP) at 23 °C with an internal standard (phenol, 2 mM). The identity of the unrearranged thioester intermediate was verified via mass spectrometry, and via reinjection in water or ligation buffer, where this product had converted to the product of the ligation reaction.

In Figure 4.34, the "fraction ligated" is not representative of the ligation reaction conversion to the final product containing a native amide bond, but demonstrates that the reaction efficiency between Ac-LYRAA-SPh and MptRAFS-NH_2 was the slowest, compared to cysteine- or 2-thiophenylalanine-mediated NCL reactions. 2*S*,4*R*-Mercaptoproline is a conformationally restricted alkyl thiol, which proceeds through a 6-membered transition state with a secondary amine in the *S*→*N* acyl transfer reaction in NCL. Consistent with our observations, Shang *et al.*[454] reported that the thioester exchange reaction between the Mpt peptide and an O-ester N-peptide proceeded to 90% conversion within 10 minutes, and conversion to ligated product was complete in 85% yield in 2.5 hours. The steric and conformational restriction in 2*S*,4*R*-mercaptoproline potentially hindered the ability of its amine to undergo intramolecular nucleophilic attack on the thioester carbonyl. With the less hindered thiolates, it was hypothesized that the ligation reaction with the peptide containing 2-thiophenylalanine would be more efficient than cysteine-mediated reactions under acidic conditions, given that a larger mole fraction of 2-thiophenylalanine is anionic (and reactive) under these conditions. However, with the peptide Ac-LYRAA-SPh at pH 5, both of the ligation reactions with CRAFS-NH_2 and (2-SH-Phe)RAFS-NH_2 were more that 90% complete in 3 hours. At pH 4, the ligation reaction with the peptide (2-SH-Phe)RAFS-NH_2 was comparable to the reaction with the peptide with CRAFS-NH_2, 56% and 46% complete, respectively. For the cysteine- and 2-thiophenylalanine-mediated ligation reactions, the final irreversible peptide product containing a native amide bond was observed as the major product, with only trace observed thioester intermediate.

Combined, these data indicate that the increased acidity of 2-thiophenylalanine only marginally enhanced reactivity for NCL reactions compared to cysteine at pH 4. However, cysteine and 2-thiophenylalanine were competitive at pH 4 and 5, whereas cysteine was substantially more rapid at pH 7. The ring size of the transition state for the *S*→*N* acyl transfer reaction between cysteine and 2-thiophenylalanine did not have significant consequences on overall reactivity at lower pH, since ligation reactions with both of these thiolated amino acids comparably formed the ligated product containing an amide bond. The most significant impact on reactivity for NCL reactions was due to the steric and conformational restriction in the thiolated amino acid. 2*S*,4*R*-Mercaptoproline exhibited the least reactivity for NCL reactions, considering both the *S*→*N* acyl transfer reaction and the initial reaction with the peptide containing the thioester. Potentially, any enhancement in reactivity for NCL reactions with the 2-thiophenylalanine due to increased acidity were compensated by the steric hindrance of the aryl thiol. The steric bulk of the aryl thiolate in 2-thiophenylalanine could have potentially "clashed" with the S-phenyl thioester in the initial reaction with the peptide Ac-LYRAA-SPh, effectively reducing its overall reactivity compared to cysteine-mediated NCL reactions.

4.2.9 Synthesis of peptides containing 2*S*,4*R*-mercaptoproline via "proline editing" on solid phase

In proteins, proline is involved in turn formation, activation and inactivation of proteins, protein folding, and substrate recognition.[282-284, 286, 474, 475] The ability to conduct ligation reactions near proline residues in the semisynthesis of proteins is an unsolved problem in NCL, as peptides containing prolyl thioesters are widely known for their low reactivity in NCL reactions.[445, 470] In order to avoid the low reactivity of

peptides containing prolyl thioesters, Shang *et al.*[454] demonstrated the utility and application of commercially available 4-mercaptoproline to mediate NCL reactions. Shang *et al.*[454] observed a substantial difference in reactivity for NCL reactions between peptides containing 2*S*,4*R*-mercaptoproline and peptides containing 2*S*,4*S*-mercaptoproline, which was hypothesized due to their different ring puckers. Using 2*S*,4*R*-mercaptoproline for NCL reaction with a peptide containing a thioester, the reaction was 85% complete in 2.5 hours. In contrast, 2*S*,4*S*-mercaptoproline exhibited little reactivity with a peptide containing a thioester, with less than 5% conversion to ligated peptide in 28 hours.[454] While commercially available, the high cost of 2*S*,4*R*-mercaptoproline can potentially inhibit its broader applications in NCL of peptides and proteins (Fmoc-2*S*,4*R*-mercapto(S-trityl)-proline, €330/1 g or $226/mmol, as of May 2016, PolyPeptide Group, San Diego, CA). To overcome the cost and to increase accessibility of 2*S*,4*R*-mercaptoproline, we developed a convenient synthesis from inexpensive starting materials, utilizing the previously described proline-editing[373, 374] approach.

2*S*,4*R*-Hydroxyproline has been previously modified on peptides on solid phase to generate peptides with a variety of proline derivatives with defined stereoselectivity and structural preferences, and is inexpensive (Fmoc-2*S*,4*R*-hydroxyproline $1.27/mmol, as of May 2016, ChemImpex, Wood Dale, IL).[373, 374] An approach to synthesize 4-mercaptoproline derivatives was described previously via Mitsunobu reaction with thiolacetic acid on peptides on solid-phase containing hydroxyproline.[373] We developed an alternative strategy to synthesizing 2*S*,4*R*-mercaptoproline in peptides, via an S_N2 reaction with sodium thiolate and a peptide on solid-phase containing 2*S*,4*S*-bromoproline, which was synthesized on peptides on

solid phase using the proline editing methodology[373, 374] (Figure 4.36). This synthetic approach was applied to incorporate 2*S*,4*R*-mercaptoproline at the N-terminus of the peptide XRAFS-NH_2 for use in NCL reactions (Chapter 4.2.8).

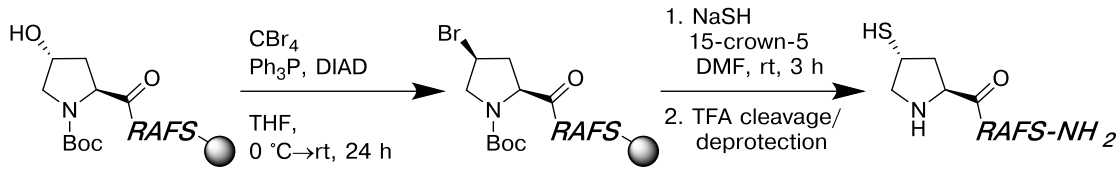


Figure 4.36 Solid-phase synthesis of peptides containing N-terminal 2*S*,4*R*-mercaptoproline via proline editing
Using proline editing, peptides containing 2*S*,4*R*-hydroxyproline can be modified using reactions on peptides on solid-phase. The protected peptide on solid phase containing 2*S*,4*R*-hydroxyproline was subjected to Mitsunobu reaction conditions with carbon tetrabromide to generate the peptide containing 2*S*,4*S*-bromoproline on solid phase. Subsequently, the peptide on solid phase containing 2*S*,4*S*-bromoproline was allowed to react with sodium thiolate, which generated the protected peptide on solid-phase containing 2*S*,4*R*-mercaptoproline.

The first reaction in this approach utilizes Mitsunobu reaction conditions with carbon tetrabromide on the resin-bound, protected peptide containing 2*S*,4*R*-hydroxyproline. Boc-2*S*,4*R*-hydroxyproline was incorporated at the N-terminus of the XRAFS peptide via SPPS. This Mitsunobu reaction incorporates a bromide leaving group in place of the secondary alcohol, and inverts the stereocenter in 2*S*,4*R*-hydroxyproline. In one step following solid-phase peptide synthesis, 2*S*,4*R*-hydroxyproline was readily converted to 2*S*,4*S*-bromoproline within a peptide (Figure 4.37).

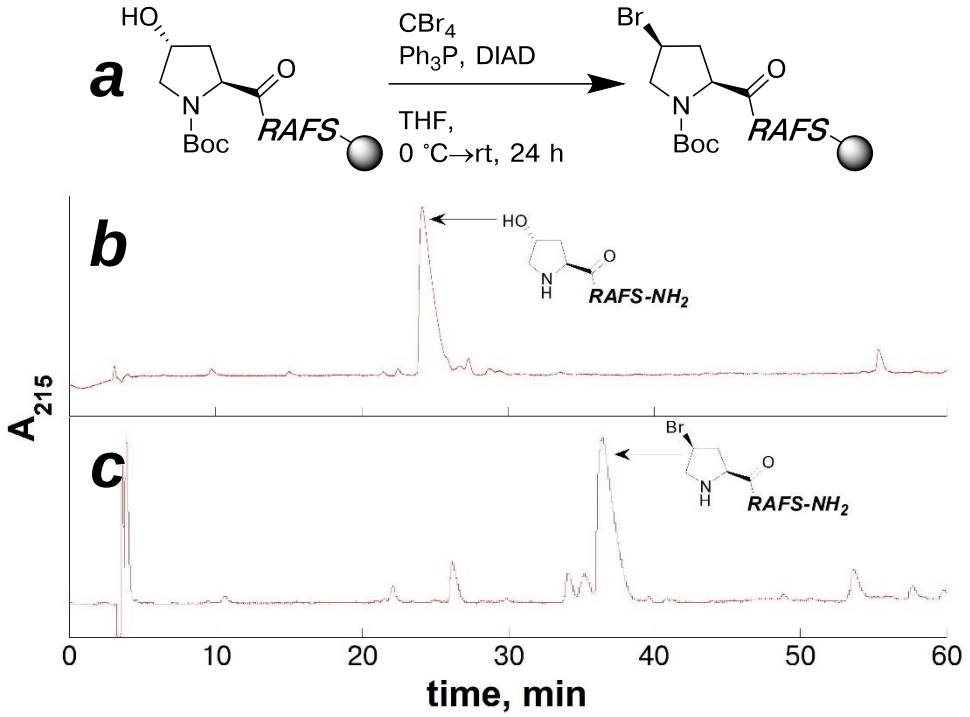


Figure 4.37 Synthesis of peptides containing 2*S*,4*S*-bromoproline via solid-phase Mitsunobu reaction
(a) Reaction scheme for the Mitsunobu reaction to generate the protected peptide containing 2*S*,4*S*-bromoproline on solid phase; (b) HPLC chromatogram of the peptide HypRAFS-NH_2 using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the crude reaction products that result from the solid phase Mitsunobu reaction with carbon tetrabromide and the protected peptide containing 2*S*,4*R*-hydroxyproline on solid phase using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

Halogens are excellent leaving groups in S_N2 alkylation reactions, and these reactions proceed through inversion of stereochemistry. Addition of sodium thiolate to the peptide containing 2*S*,4*S*-bromoproline on solid phase substitutes a thiol group for the bromine atom and inverts the stereochemistry (Figure 4.38). A similar synthetic approach was utilized by Townsend *et al.* in synthesizing 4-selenoproline

derivatives.[476] The synthetic strategy developed here generated the peptide containing 2*S*,4*R*-mercaptoproline in 59% conversion over two steps following solid phase peptide synthesis.

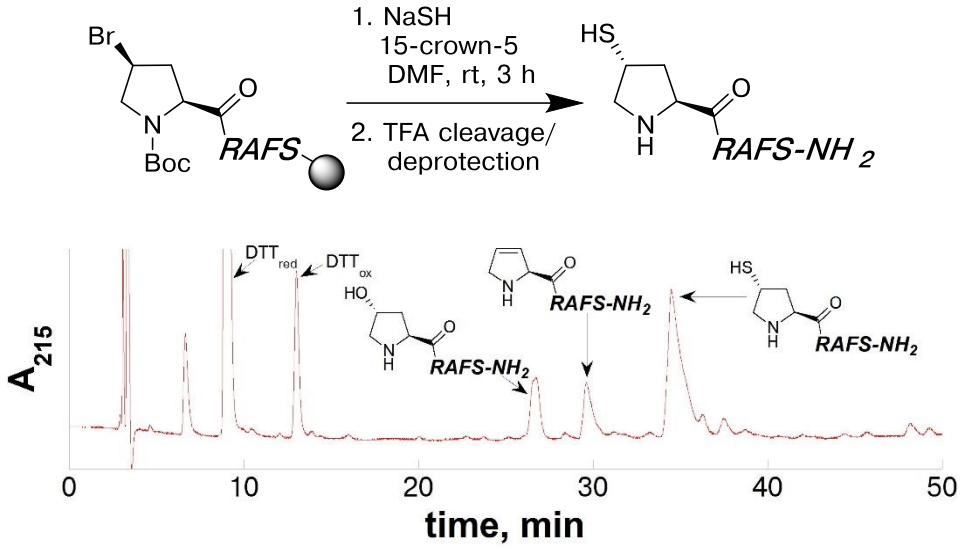


Figure 4.38 Synthesis of peptides containing N-terminal 2*S*,4*R*-mercaptoproline on solid phase
HPLC chromatogram of the crude peptide products that result from the S_N2 reaction with sodium thiolate on the protected peptide containing 2*S*,4*S*-bromoproline on solid phase using a linear gradient of 0-20% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). The crude reaction products were treated with DTT in solution phase to prevent disulfide formation.

By utilizing the proline editing technique[373, 374] on protected peptides on solid phase, peptides containing 2*S*,4*R*-mercaptoproline were readily synthesized in a practical, inexpensive manner. In two steps following solid-phase peptide synthesis, 2*S*,4*R*-mercaptoproline was generated within a peptide in less than two days total reaction time at a considerably lower cost than purchasing the protected amino acid and considerably less time than the solution phase synthesis of this amino acid. 2*S*,4*R*-

Hydroxyproline is an inexpensive, readily available amino acid, and through simple and efficient transformations on protected peptides on solid phase, can be elaborated to generate a variety of different proline derivatives.[373]

4.2.10 Validation of non-racemization under the ligation reaction conditions

During the NCL reactions, racemization of the amino acids at the ligation site has been suggested as a potential epimerization side product, given that amino acid coupling reagents can sometimes cause racemization via formation of an oxazolone.[133, 470, 477, 478] Valine and phenylalanine are the amino acids most prone to racemization during coupling reactions.[133, 464] In order to evaluate the possibility of epimerization occurring during the 2-thiophenylalanine-mediated ligation reactions, the D-valine and D-phenylalanine variants of the ligated peptide products were independently synthesized and characterized. These D-amino acid variants can be compared via HPLC coinjection and NMR spectroscopy against the products that form as a result of the ligation reactions between the peptides (2-SH-Phe)RAFS-NH_2 and the corresponding peptide containing a C-terminal thioester. The peptides Ac-LYRAZ(2-I-Phe)RAFS, where Z = D-Val or D-Phe, were synthesized via SPPS. The peptides containing 2-iodophenylalanine were subjected to the solid-phase cross-coupling reaction to generate the peptides Ac-LYRAZ(2-SH-Phe)RAFS-NH_2, where Z was D-Val or D-Phe. The purified products of the cross-coupling reaction were compared to the crude products that resulted from the corresponding ligation reactions. The comparison between the peptides Ac-LYRAZ(2-SH-Phe)RAFS-NH_2, where Z = Val or D-Val, is shown in Figures 4.39 and 4.40. The comparison between the peptides Ac-LYRAZ(2-SH-Phe)RAFS-NH_2, where Z = Phe or D-Phe, is shown in Figures 4.41 and 4.42.

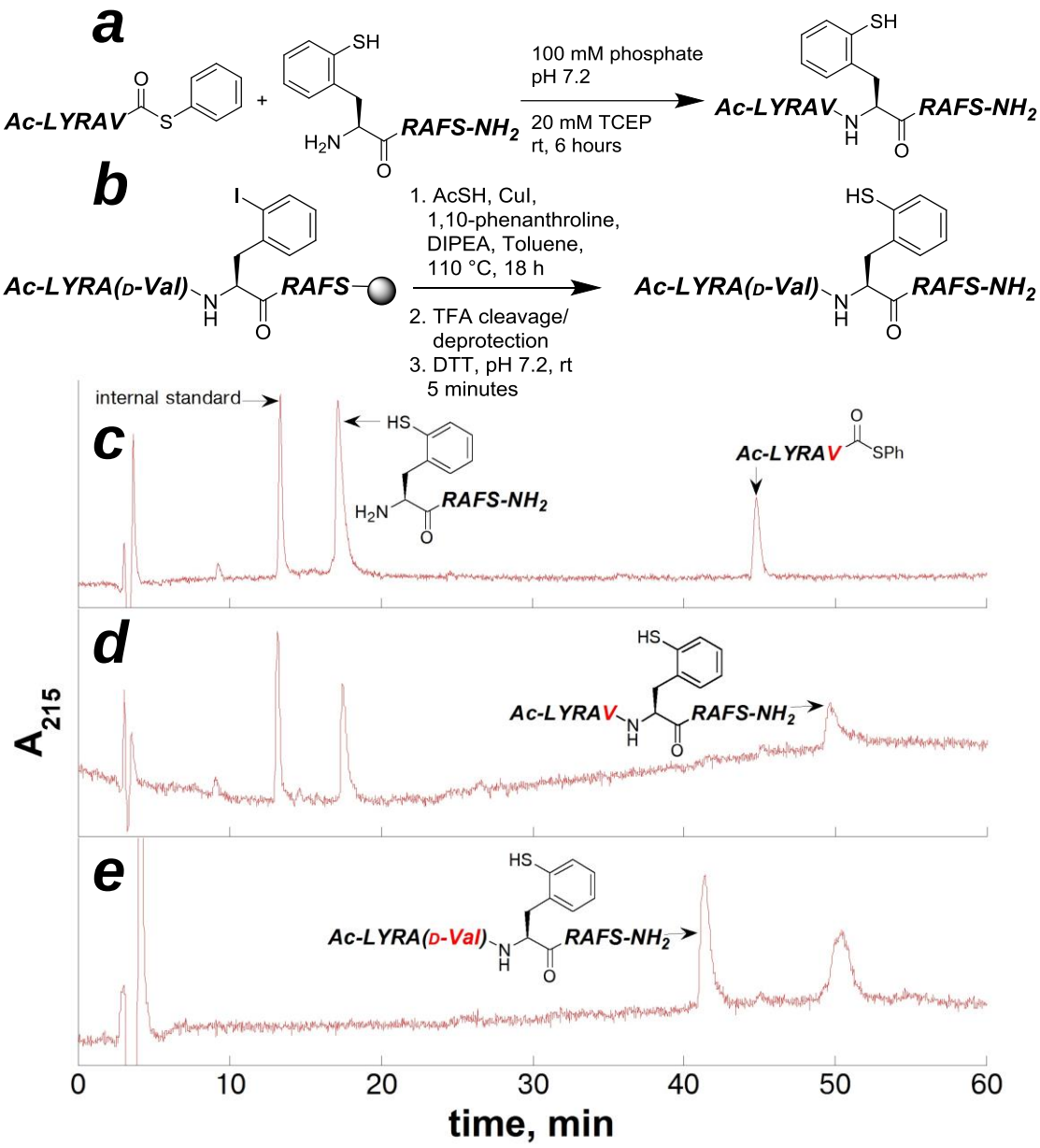


Figure 4.39 Comparison of the crude products from the ligation reaction with Ac-LYRAV-SPh and (2-SH-Phe)RAFS-NH_2 to the D-valine variant of the peptide Ac-LYRAV(2-SH-Phe)RAFS-NH_2

HPLC analysis was conducted using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). (a) Reaction scheme for the ligation reaction between the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAV-SPh; (b) Reaction scheme for the independent synthesis of the peptide Ac-LYRA(D-Val)(2-SH-Phe)RAFS-NH_2; (c) HPLC chromatogram of the crude peptide products that resulted from the ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAV-SPh after 5 minutes; (d) HPLC chromatogram of the crude peptide products that resulted from the ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAV-SPh after 6 hours; (e) HPLC chromatogram of the coinjection of the purified peptide product resulting from the ligation reaction with the independently synthesized D-amino acid variant.

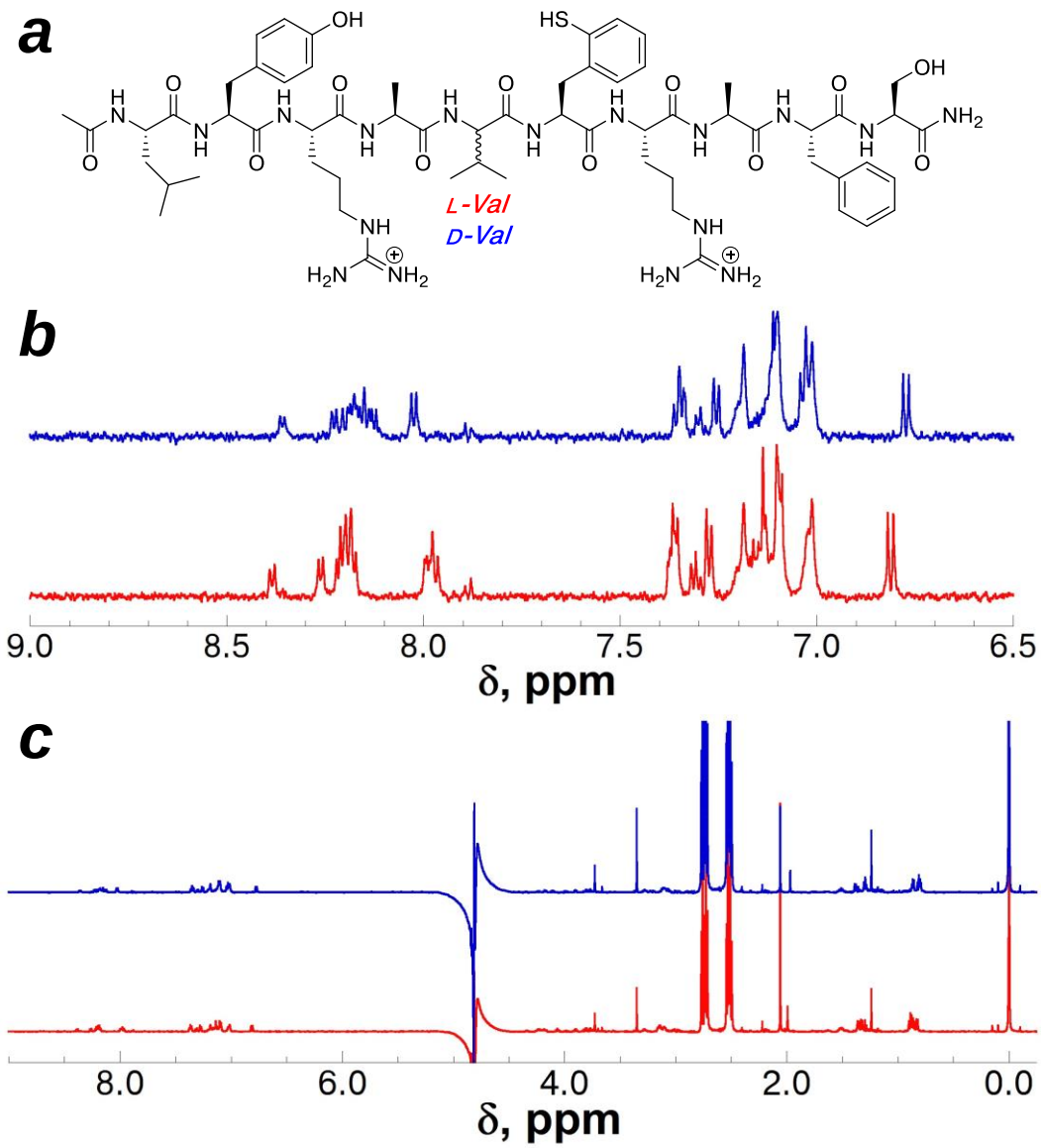


Figure 4.40 Superposition of the ^{1}H NMR spectra of the ligation reaction product peptide Ac-LYRAV(2-SH-Phe)RAFS-NH_2 and the independently synthesized peptide Ac-LYRA(D-Val)(2-SH-Phe)RAFS-NH_2

Samples were obtained in 10% D_2O/90% H_2O with 5 mM phosphate pH 4, 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP. The ligation reaction product peptide Ac-LYRAV(2-SH-Phe)RAFS-NH_2 is shown in red; the independently synthesized peptide Ac-LYRA(D-Val)(2-SH-Phe)RAFS-NH_2 is shown in blue. (a) Structure of the peptide product; (b) ^{1}H NMR spectra of the amide region of the indicated peptides; (c) full ^{1}H NMR spectra for the indicated peptides.

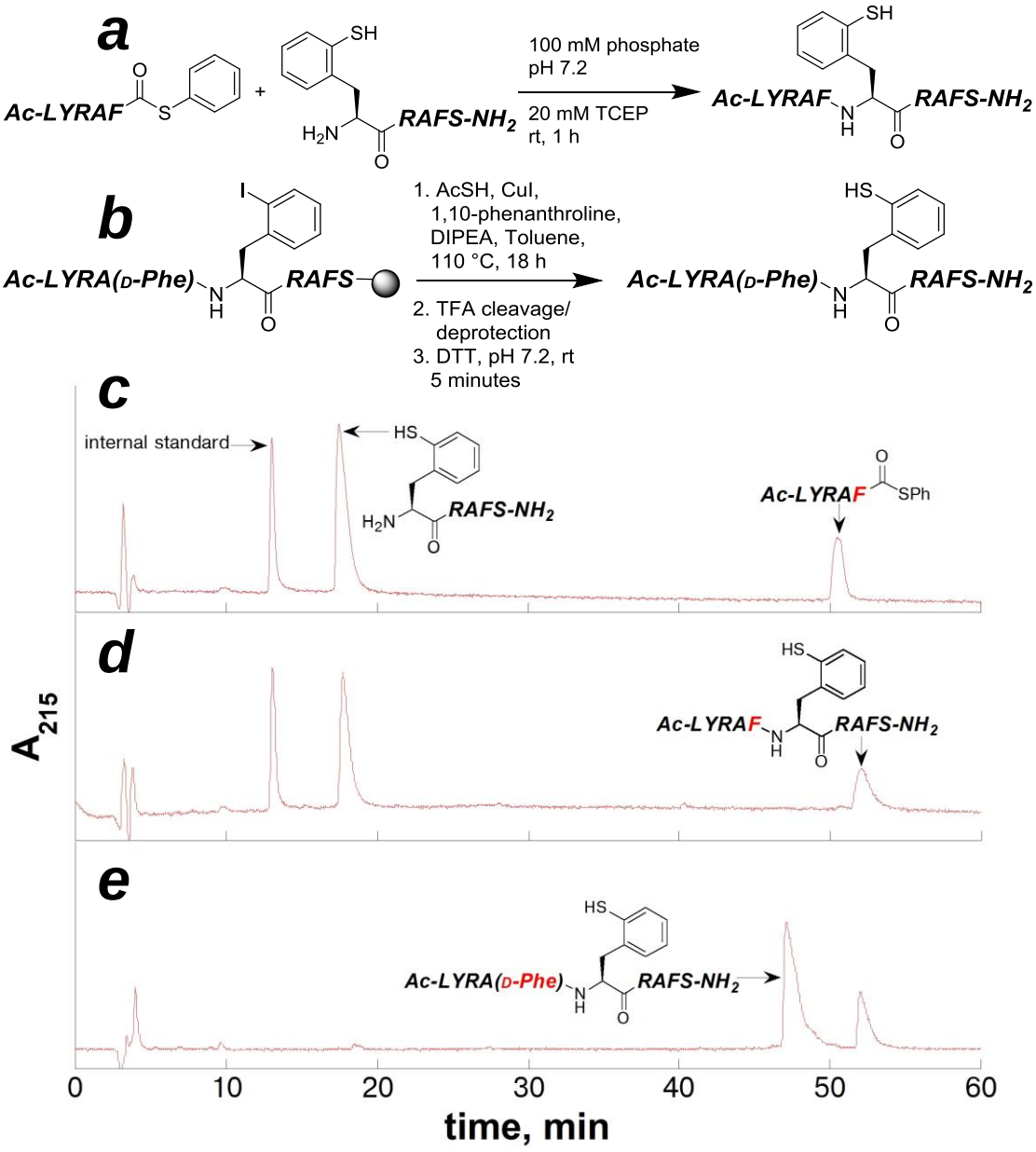


Figure 4.41 Comparison of the crude products from the ligation reaction with Ac-LYRAF-SPh and (2-SH-Phe)RAFS-NH$_2$ to the D-phenylalanine variant of the peptide Ac-LYRAV(2-SH-Phe)RAFS-NH$_2$

HPLC analysis was conducted using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). (a) Reaction scheme for the ligation reaction between the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAF-SPh; (b) Reaction scheme for the independent synthesis of the peptide Ac-LYRA(D-Phe)(2-SH-Phe)RAFS-NH_2; (c) HPLC chromatogram of the crude peptide products that resulted from the ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAF-SPh after 5 minutes; (d) HPLC chromatogram of the crude peptide products that resulted from the ligation reaction with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAF-SPh after 1 hour; (e) HPLC chromatogram of the coinjection of the purified peptide product resulting from the ligation reaction with the independently synthesized D-amino acid variant.

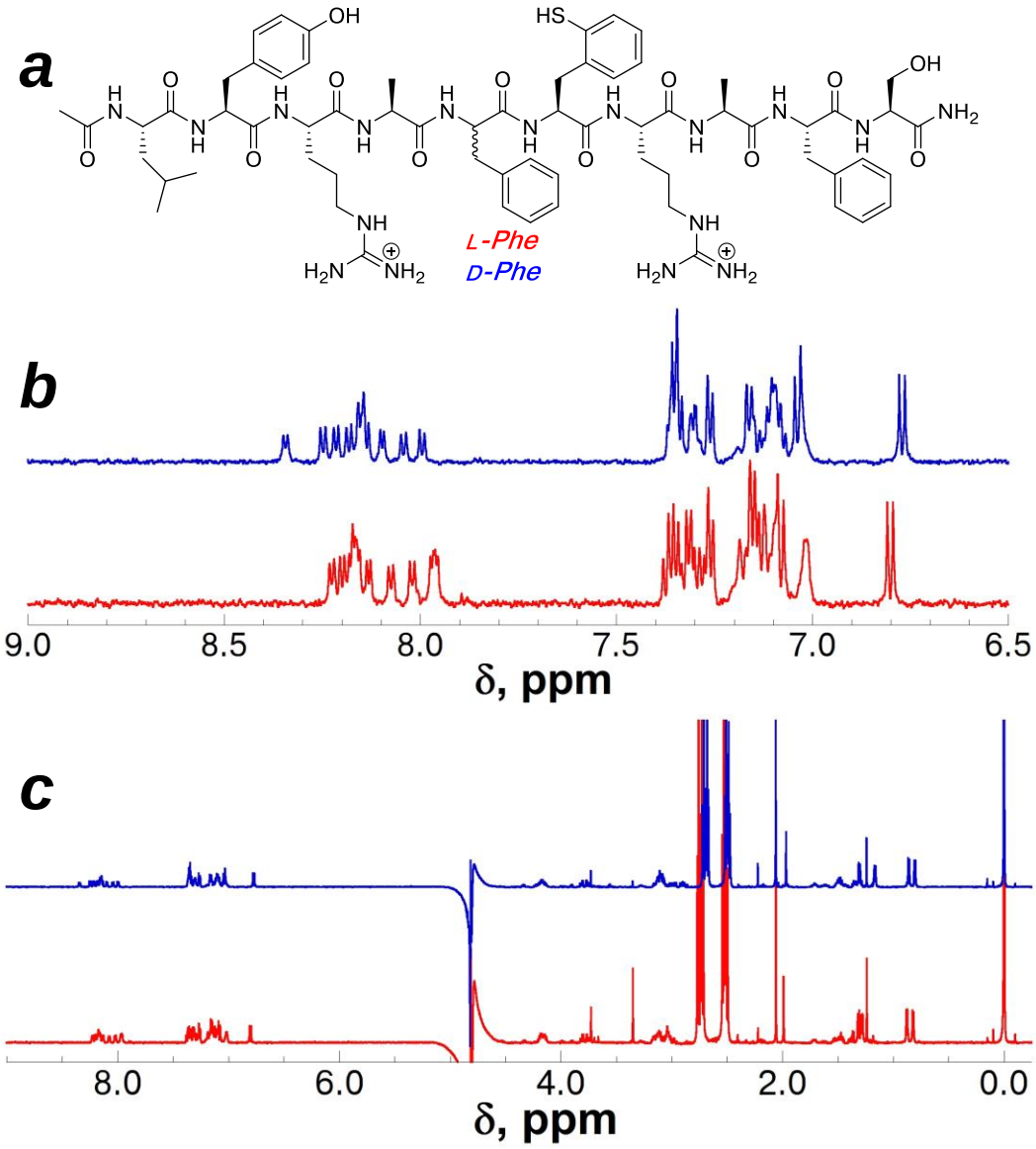


Figure 4.42 Superposition of ^{1}H NMR spectra of the ligation reaction product peptide Ac-LYRAF(2-SH-Phe)RAFS-NH$_2$ and the independently synthesized peptide Ac-LYRA(D-Phe)(2-SH-Phe)RAFS-NH$_2$

Samples were obtained in 10% D_2O/90% H_2O with 5 mM phosphate pH 4, 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP. The ligation reaction product peptide Ac-LYRAF(2-SH-Phe)RAFS-NH$_2$ is shown in red; the independently synthesized peptide Ac-LYRA(D-Phe)(2-SH-Phe)RAFS-NH$_2$ is shown in blue. (a) Structure of the peptide product; (b) ^{1}H NMR spectra of the amide region of the indicated peptides; (c) full ^{1}H NMR spectra for the indicated peptides.

Based on the HPLC coinjections and comparison of the ^{1}H NMR spectra of the purified peptide products resulting from the ligation reaction and the independently synthesized D-amino acid variants, no significant epimerization was observed under the ligation reaction conditions. The HPLC chromatogram of the coinjected peptides shows a significant difference in the retention time between the native and D-amino acid variants (Figures 4.39e and 4.41e). Comparing the HPLC chromatograms of the purified peptides against the HPLC chromatogram of the crude ligation reactions shows no observable product that contained the D-amino acid variant (Figures 4.39d and 4.41d). In comparing the NMR spectra of the products that resulted from the ligation reactions to the independently synthesized D-amino acid variants, there does not appear to be any significant D-amino acid variant impurity in the products that result from the native chemical ligation reaction.

4.2.11 Desulfurization of model peptides containing 2-thiophenylalanine to produce phenylalanine

Following the ligation reaction, the thiol functional group must be removed via a desulfurization reaction to generate a canonical amino acid at the ligation site. Raney nickel and nickel boride are two common methods for desulfurization of thiol functionalities.[399, 400] Metal-based approaches were initially used in the development of peptide desulfurization following NCL reactions (i.e. nickel boride, Raney nickel, Pd/Al_2O_3), but adsorption of the peptides on the metal surfaces can significantly impact the isolated yield, and so radical initiated methods have become more common.[479] Danishefsky and coworkers have described a milder approach to desulfurization using a radical initiator VA-044 in the presence of TCEP, and this approach is applicable to thiols and selenols.[399, 446]

In order to generate phenylalanine at the ligation site, the peptides containing 2-thiophenylalanine were subjected to desulfurization reaction conditions. Initial attempts to effect desulfurization on the peptides containing 2-thiophenylalanine using VA-044 were unsuccessful, and resulted in a mixture of products. Malins *et al.*[472] also reported difficulty with desulfurization reactions using VA-044 on peptides containing 2-thiol-tryptophan (with a thiol at the 2-position in the indole ring), attributed to a greater C–S bond strength in thiolated indoles. Employing the desulfurization reaction conditions that were utilized by Crich & Banerjee[401] in desulfurization of peptides containing β-mercaptophenylalanine, the ligated model peptides containing 2-thiophenylalanine were subjected to desulfurization reaction using nickel boride (Figure 4.43). Desulfurization reaction conditions were demonstrated on the model peptides Ac-LYRAG(2-SH-Phe)RAFS-NH_2 (Figure 4.44), Ac-LYRAA(2-SH-Phe)RAFS-NH_2 (Figure 4.45), and Ac-LYRAL(2-SH-Phe)RAFS-NH_2 (Figure 4.46).

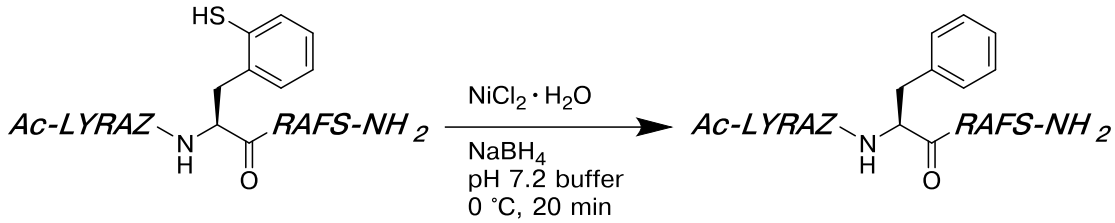


Figure 4.43 General scheme for desulfurization reaction of peptides containing 2-thiophenylalanine

The model peptides that resulted from the ligation reactions containing 2-thiophenylalanine were subjected to desulfurization using nickel borohydride following standard protocols.[401]

Z = Gly, Ala, Leu

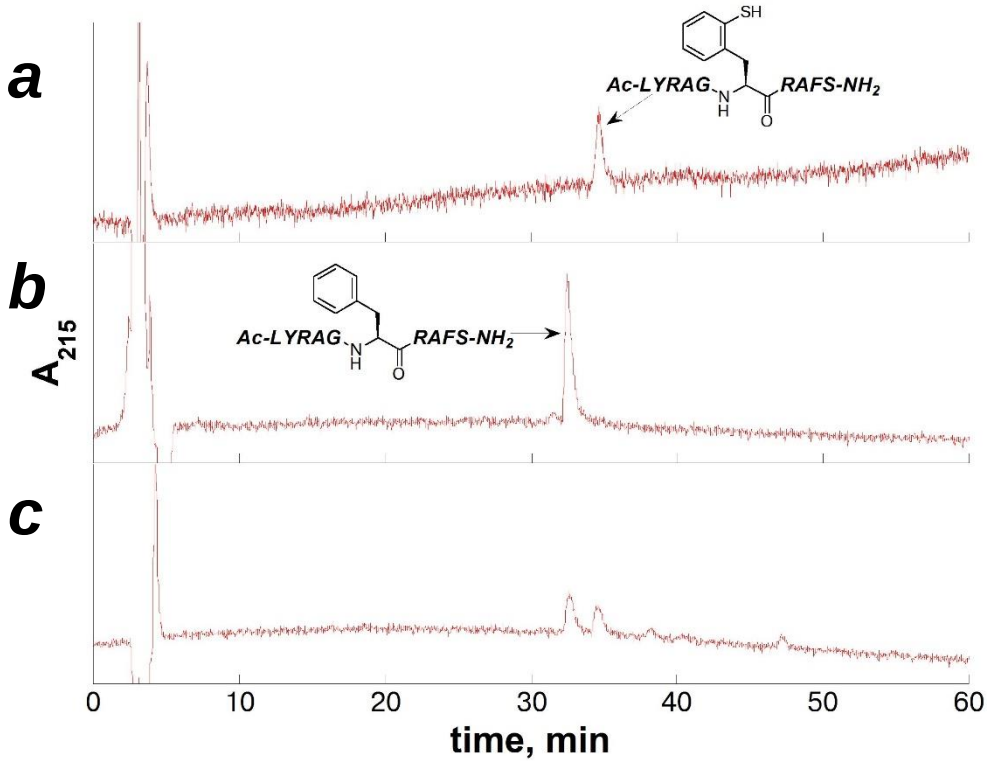


Figure 4.44 Desulfurization reaction on the peptide Ac-LYRAG(2-SH-Phe)RAFS-NH_2

(a) HPLC chromatogram of the purified peptide Ac-LYRAG(2-SH-Phe)RAFS-NH_2 that resulted from the ligation reaction; (b) HPLC chromatogram of the crude desulfurization reaction to generate the peptide Ac-LYRAGFRAFS-NH_2; (c) HPLC chromatogram of the coinjection of the purified peptides Ac-LYRAG(2-SH-Phe)RAFS-NH_2 and Ac-LYRAGFRAFS-NH_2, which was synthesized via ligation and desulfurization reactions. HPLC analysis was conducted using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS.

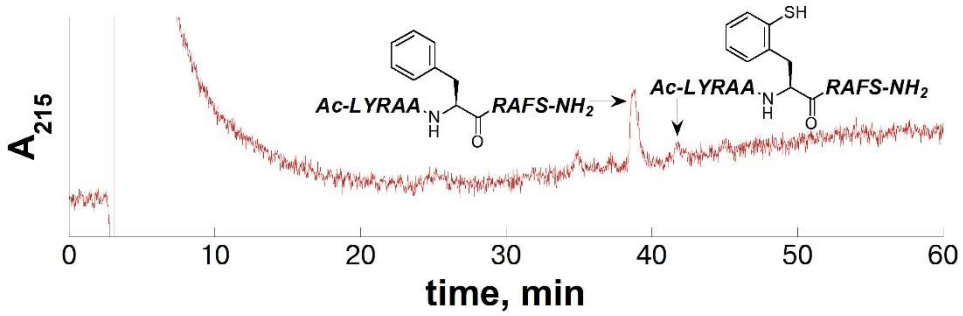


Figure 4.45 Desulfurization reaction on the peptide Ac-LYRAA(2-SH-Phe)RAFS-NH_2

HPLC chromatogram of the crude desulfurization reaction to generate the peptide Ac-LYRAAFRAFS-NH_2. HPLC analysis was conducted using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS.

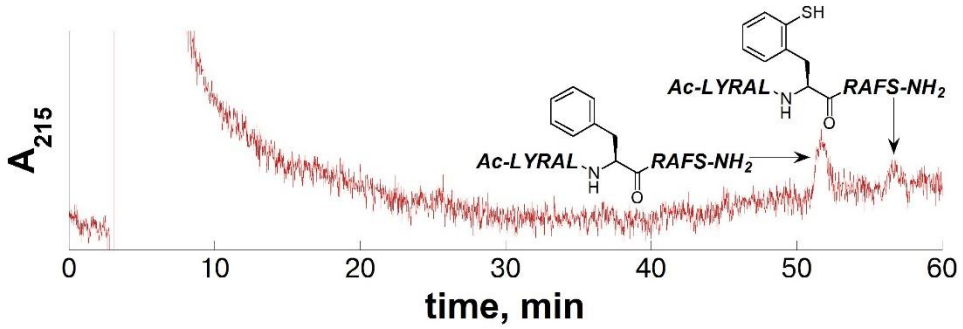


Figure 4.46 Desulfurization reaction on the peptide Ac-LYRAL(2-SH-Phe)RAFS-NH_2

HPLC chromatogram of the crude desulfurization reaction to generate the peptide Ac-LYRALFRAFS-NH_2. HPLC analysis was conducted using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). Products were identified via ESI-MS.

Desulfurization reactions on the selected peptides containing 2-thiophenylalanine, which were generated from ligation reactions, were accomplished via nickel borohydride, generated *in situ* from nickel(II) chloride and sodium borohydride.[401] Although the desulfurization reactions were fast (20 minutes) and

cleanly resulted in peptides containing phenylalanine in place of 2-thiophenylalanine, there was a significant loss of peptide product from the nickel boride reaction for all peptides. The nickel boride forms as a solid catalyst, and after reaction with the peptide containing 2-thiophenylalanine, the solids are subjected to high concentrations of acid and the solids are filtered and removed. The desulfurization reaction relies on adsorption of the thiol to the nickel surface, and subsequent cleavage of the C–S bond.[480, 481] It has previously been noted that metal-based desulfurization approaches in peptides result in low isolated yields, due to peptide aggregation and adsorption to metal surfaces.[479] The aromatic nature of 2-thiophenylalanine potentially enhanced adsorption of the peptide to the nickel boride solids, through π-backbonding, which could have resulted in low recovery of the peptide after desulfurization reaction. Malins *et al.*[472] reported excellent isolated yields from the desulfurization reaction utilizing Pd/Al_2O_3 on peptides containing 2-thio-tryptophan (66% isolated yield). Potentially, desulfurization reaction on the peptides containing 2-thiophenylalanine utilizing Pd/Al_2O_3, as opposed to nickel boride, would result in improved isolated yields.

4.2.12 Synthesis of a modified trp cage miniprotein containing 2-thiophenylalanine for native chemical ligation reactions

Having shown in model peptides that 2-thiophenylalanine can be used effectively for NCL reactions with peptides containing thioesters, and can be subsequently desulfurized to generate phenylalanine, we sought to synthesize a larger peptide containing 2-thiophenylalanine, to demonstrate its utility for NCL. The trp cage miniprotein and related variants are well characterized via NMR, and we have successfully utilized this protein in previous model studies for reactivity in the solid-

phase cross-coupling reaction conditions (Chapter 1.2.4).[94, 95, 169, 404] The trp cage miniprotein is 20-residue, highly stabilized motif that is comprised of an α-helix and a polyproline helix (Figure 1.27). The trp cage miniprotein is a truncation from the peptide exendin-4, isolated from Gila monster saliva.[94, 102] The 39-residue exendin-4 is not well structured in water, but the truncated trp cage is 95% folded in water, and has been widely utilized for protein folding studies in solution and by computational chemistry.[94, 95] This stabilized peptide framework makes the trp cage miniprotein an ideal scaffold for studying the structural effects of post-translational modifications,[482] or as a scaffold for designing stabilized peptide-based therapeutics.[102, 483] With the potential applications of trp cage variants for protein folding studies and for synthetic fusion proteins, we sought to demonstrate the utility of 2-thiophenylalanine-mediated NCL using the trp cage miniprotein.

The trp cage miniprotein has a tyrosine at position 3, near the N-terminus.[94] We previously synthesized a variant of the trp cage miniprotein where Tyr3 was replaced with 4-thiophenylalanine, via solid-phase cross-coupling reaction with the peptide containing 4-iodophenylalanine (Chapter 1.2.4). Having previously demonstrated that the cross-coupling reaction did not affect the folding and stability of the trp cage miniprotein, and that a modified aromatic amino acid is well tolerated at position 3, Tyr3 was selected as the ligation site for incorporating 2-thiophenylalanine. Two peptides were synthesized via SPPS: Ac-NL-SPh (trp-cage$_{1\text{-}2}$–SPh), and Boc-(2-I-Phe)IQWLKDGGPSSGRPPPS (trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine at the N-terminus) on Rink amide resin. The protected peptide trp-cage$_{3\text{-}20}$ containing Boc-2-iodophenylalanine at the N-terminus was subjected to the previously established cross-coupling conditions on solid phase[169] (Figure 4.47).

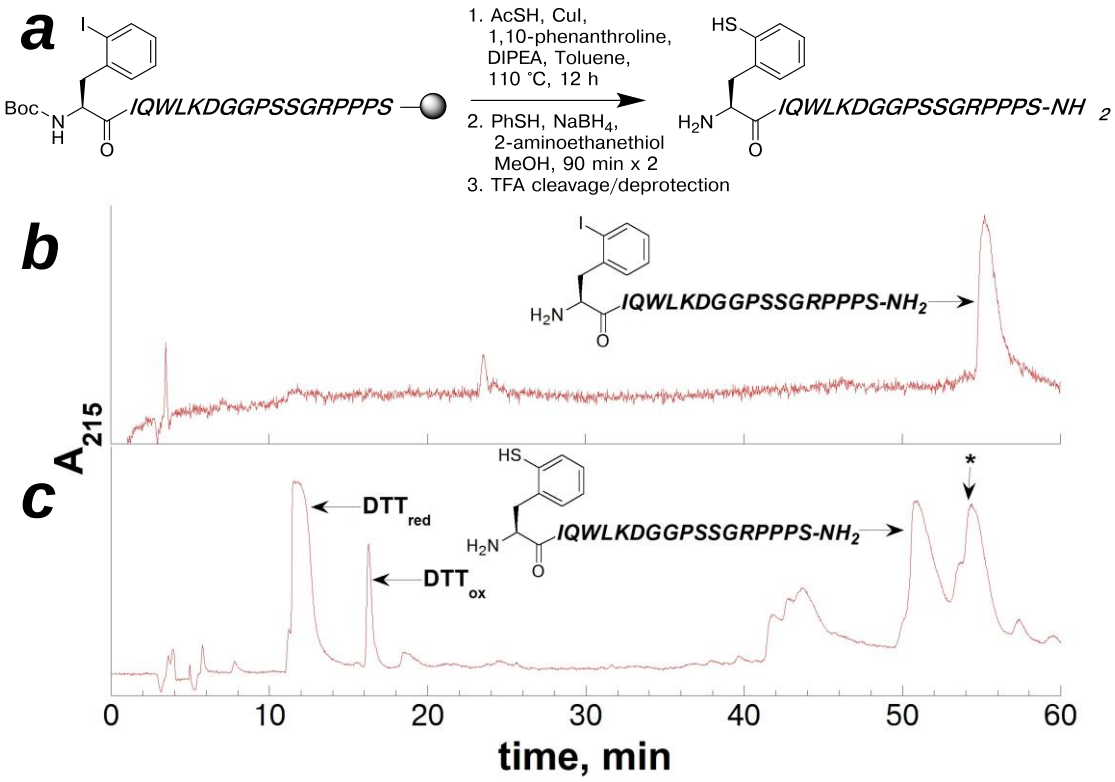


Figure 4.47 Synthesis of the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine via cross-coupling reaction

(a) Reaction scheme for cross-coupling reaction on the protected peptide and subsequent reduction and thiolysis on solid phase to generate the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine; (b) HPLC chromatogram of the peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the crude peptide products that result from the solid-phase cross-coupling reaction on the peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine and subsequent reaction with DTT in solution using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). The identity of the products were identified via ESI-MS.

*A significant side-product was observed, with a molecular weight that was consistent with a peptide Ac-trp-cage$_{4\text{-}20}$ (described below).

In addition to the trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine, a substantial side-product was observed in the products that resulted from the copper-mediated cross-coupling reaction on the protected peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine (denoted with an *, Figure 4.47c, t_R = 53.4 min). The molecular weight of this side-product was less than the expected product mass (1818.9 vs. 1956.0, respectively), suggesting that the peptide trp-cage$_{3\text{-}20}$ had been truncated during the cross-coupling or thiolysis reaction on solid phase. The molecular weight of this side product was consistent with the peptide the peptide Ac-trp-cage$_{4\text{-}20}$, where the phenylalanine residue at the N-terminus had been hydrolytically cleaved and an additional acetyl-group was attached to the peptide (Figure 4.48a). A product of this nature was not previously observed as a result of the cross-coupling reaction on solid phase to generate the peptide Y3(4-SH-Phe) trp cage (Chapter 1.2.4),[169] or to generate the model peptides (2-SH-Phe)RAFS-NH_2 (Chapter 4.2.2). The peptide Ac-trp-cage$_{4\text{-}20}$ was not observed prior to the cross-coupling reaction (Figure 4.47b), and can only result from the modification reactions on solid phase. NMR characterization of this product confirmed that the N-terminal phenylalanine was not present (only the tryptophan aromatic resonances were observed, Figure 4.48c, black box), and an additional acetyl group was present (Figure 4.48b, black arrow).

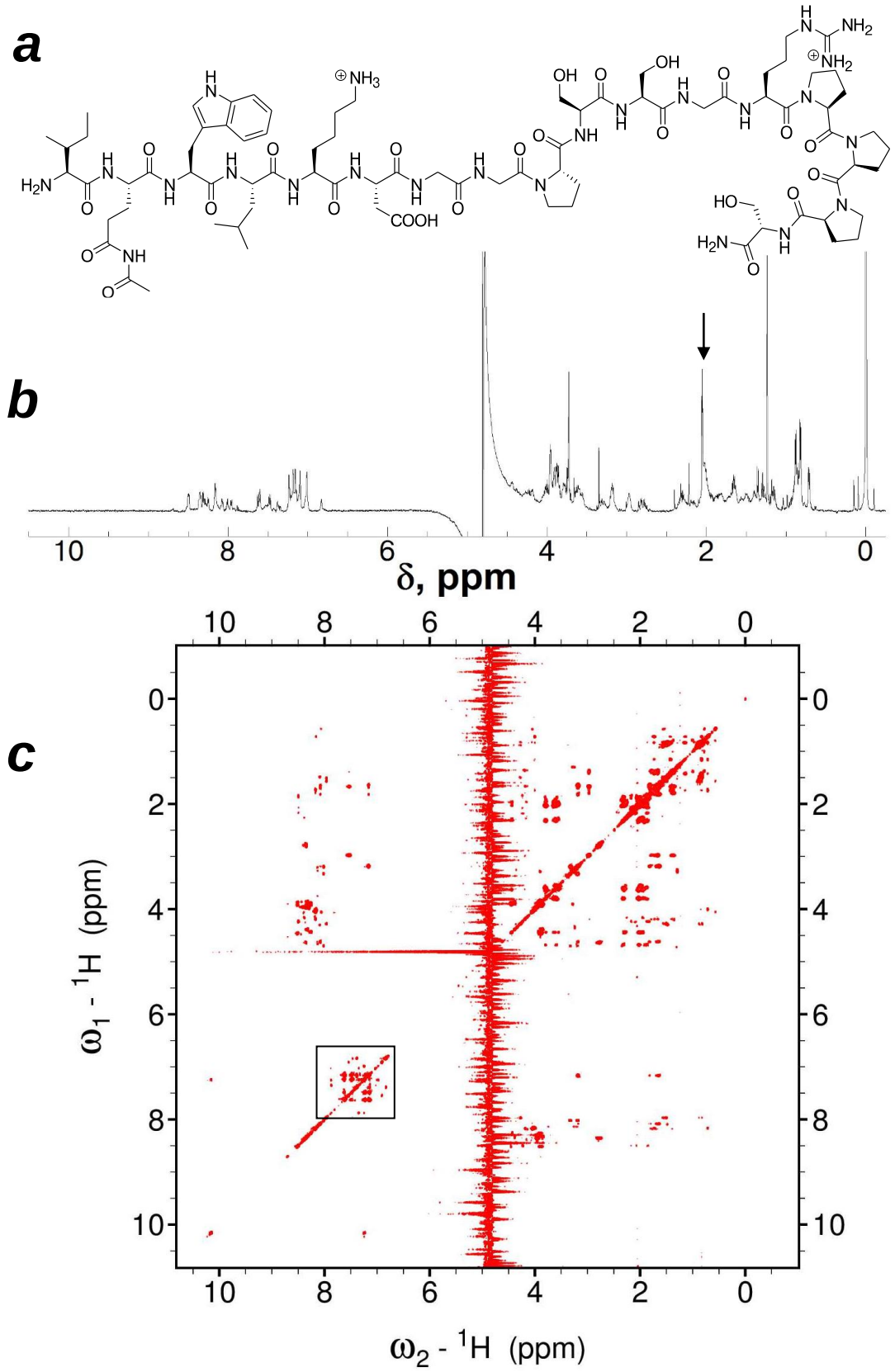

a
b
c
δ, ppm
ω1 - 1H (ppm)
ω2 - 1H (ppm)

Figure 4.48 TOCSY NMR Spectra of the major side product that resulted from the solid phase cross-coupling reaction of the trp cage miniprotein
(a) the proposed structure of the major side-product that resulted from the solid-phase cross-coupling reaction on the protected peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine; (b) ^{1}H NMR spectra of the major side product, proposed to be Ac-trp-cage$_{4\text{-}20}$. The resonance assigned to an acetyl CH_3 is indicated at δ 2.05 ppm (black arrow); (c) TOCSY spectra of the major side product, proposed to be Ac-trp-cage$_{4\text{-}20}$. In the aromatic region, only the tryptophan aromatic peaks are seen (black box), indicating that the N-terminal phenylalanine is not present (either as 2-iodo or 2-thio-phenylalanine).The NMR sample was in 90% H_2O/10% D_2O with 5 mM phosphate pH 4, 25 mM NaCl, and 0.1 mM TSP.
In the synthesized peptide, prior to any solid-phase modification reactions, a truncated product was not observed by HPLC (Figure 4.47b), so this product could potentially result from a side-reaction during the cross-coupling reaction with thiolacetic acid on the peptide on solid phase.

Potentially, deprotection of one of the side-chain groups during the cross-coupling reaction on the peptide at high temperature exposed a functional group that could proceed through an intramolecular *S→N* acyl transfer with the 2-thio(acetyl)-phenylalanine at the N-terminus. Following an acyl transfer reaction from the 2-thio(acetyl)-phenylalanine to a deprotected side chain, the resultant aryl thiolate could potentially proceed through an intramolecular *S→N* acyl transfer reaction with its own amide carbonyl, forming a Boc-protected-3-amino-δ-thiochromanone product and a truncated, acylated peptide, Ac-trp-cage$_{4\text{-}20}$ (Figure 4.48a). A similar proteolytic cleavage of an amide bond and formation of a thiolactone was previously observed with homocysteine-mediated NCL reactions,[447] and was attributed to prolonged exposure to basic reaction conditions.

However, the location of the acetyl group, and the side-chain that could have been deprotected during the cross-coupling reaction, was not definitively identified from the NMR spectrum. The reactive side chains that could potentially undergo intramolecular *S→N* acyl transfer include: the Lys8 amine, the Arg16 guanidine, and

the Gln5 carboxamide. These side-chains were protected as a Boc-carbamate (Lys8), as a guandino-Pbf (Arg16), and as a trityl amide (Gln5). Potentially, the Pbf group on Arg16 was cleaved during the cross-coupling reaction conditions and the guanidine could participate in an acyl transfer reaction,[59] but this modification was not observed in the synthesis of the model peptide, XRAFS-NH_2, which contains an arginine residue that is much closer to the 2-thiophenylalanine at the N-terminus. The arginine guanidino-group was consistent between the peptide Ac-trp-$cage_{4\text{-}20}$ and an independently synthesized peptide Y3F trp cage, indicating that the arginine side-chain was not irreversibly acetylated during the reaction. In the TOCSY spectrum of the peptide Ac-trp-$cage_{4\text{-}20}$, a resonance at 7.52 ppm was assigned to the Lys8 NHε, which was consistent with the chemical shift for lysine NHε in the peptide Y3F trp cage (7.56 ppm), independently synthesized via SPPS. The multiplicity of the resonance assigned to Lys8 NHε is broad from the ^{1}H NMR spectrum of the peptide Ac-trp-$cage_{4\text{-}20}$, which was consistent with a free amino group under acidic conditions (an NHε-acetyl amide would appear as a triplet under these conditions). In addition, there was no evidence of deprotection of the lysine side-chain during the cross-coupling reaction on the peptide Ac-KKHMC(4-I-Phe)-NH_2 (Chapter 1.2.4). Combined, these data suggest that the acetyl group is not located on the Lys8 amine or the Arg16 guanidine.

However, the trityl-protecting group previously exhibited sensitivity to the cross-coupling reaction conditions (deprotection of trityl-protected cysteine, Chapter 1.2.4), and the Gln5 side-chain may have been deprotected during the reaction to generate the peptide containing 2-thiophenylalanine. If the Gln5 side-chain was deprotected during the cross-coupling reaction conditions, then the Gln5 side chain

could participate in an intramolecular *S*→*N* acyl transfer reaction with 2-thio(acetyl)-phenylalanine, generating the *N*ε-acetyl imide product (Figure 4.48). The thiolate on 2-thiophenylalanine could act as a nucleophile on its own amide bond, releasing the Boc-3-amino-δ-thiochromanone product into solution and generating the peptide Ac-trp-cage$_{4\text{-}20}$ (Figure 4.49). Although the singlet proton from the acyl-imide on glutamine was not clearly identified in the ^{1}H NMR or TOCSY spectra, this mechanism is the most plausible based on comparison to the NMR spectra of the peptides (2-SH-Phe)RAFS-NH_2 and Y3F trp cage. To confirm the location of the acyl group and to confirm this proposed reaction mechanism, the potential side-product peptides would need to be independently synthesized and compared to the data shown in Figure 4.48.

Described further in Chapter 4.2.14, a different protected peptide containing Boc-2-iodophenylalanine at the N-terminus was subjected to the solid-phase cross-coupling reaction, and a similar corresponding product was also observed (with a molecular weight consistent with an additional acetyl group and truncation of the N-terminal phenylalanine). This peptide contained an Asn at the *i*,*i*+2 position to the N-terminal residue, which contained a trityl-protecting group. Combined, the side products observed in these two peptides, with the N-terminal motif Boc-(2-I-Phe)XQ/N, suggests that a trityl-protected asparagine or glutamine is prone to deprotection during the reaction conditions, which can potentially react with 2-S(acetyl)-phenylalanine at the N-terminus.

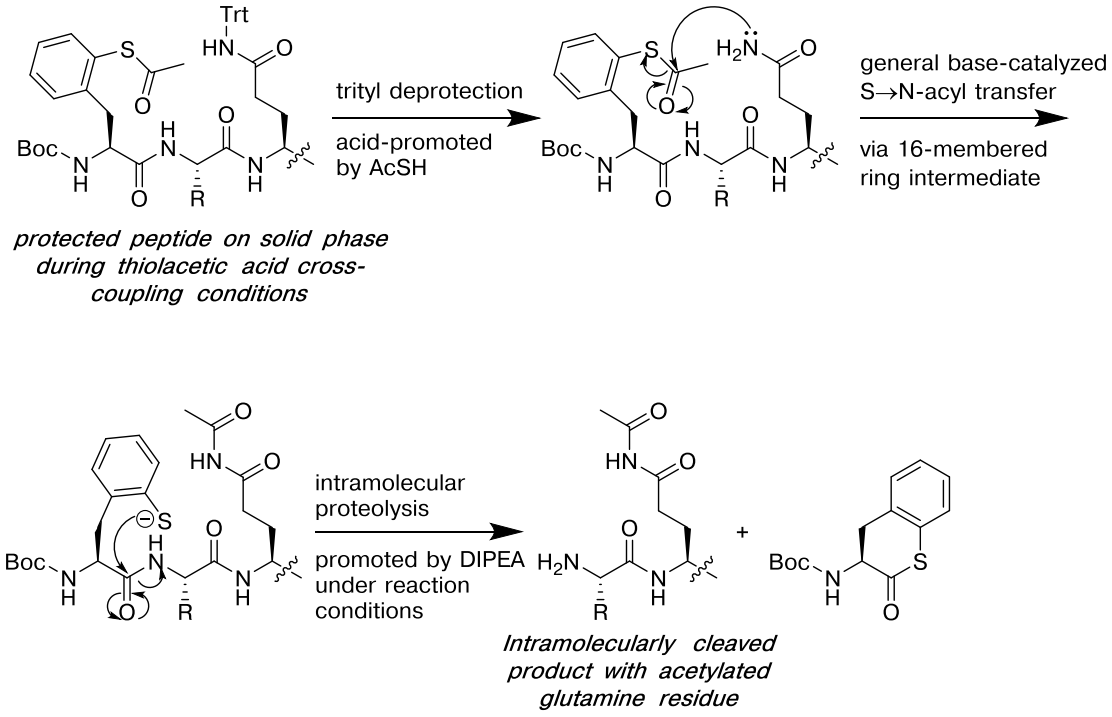


Figure 4.49 Proposed mechanism for formation of the peptide Ac-trp-cage$_{4\text{-}20}$ during the cross-coupling reaction on the protected peptide

Under the cross-coupling reaction conditions on the protected peptide trp-cage$_{3\text{-}20}$ containing Boc-2-iodophenylalanine at the N-terminus, the trityl-protecting group on the Gln5 side-chain was removed (potentially by the presence of thiolacetic acid and high reaction temperatures). The resultant primary amide can undergo intramolecular *S→N* acyl transfer with the 2-S(acetyl)-thiophenylalanine at the N-terminus via a 15-membered ring intermediate. Subsequently, the exposed thiolate on 2-thiophenylalanine can undergo a 5-membered proteolytic cleavage reaction with its amide bond, similar to side reactions observed in homocysteine-mediated NCL reactions.[447] This proteolytic cleavage reaction would produce the Boc-3-amino-δ-thiochromanone product and a truncated, acetylated variant of the trp cage C-peptide (Ac-trp-cage$_{4\text{-}20}$), consistent with the major side-product observed in Figure 4.47c. Products of this nature were not observed as a result of the cross-coupling reaction on the model peptide (2-I-Phe)RAFS, but were observed as a result of the cross-coupling reaction on a similar peptide containing an *i*,*i*+2 asparagine (described in Chapter 4.2.14). Furthermore, the Ac-trp-cage$_{4\text{-}20}$ product was not observed under the ligation reaction conditions, suggesting that the proteolytic cleavage reaction only occurs at high temperatures.

In order to suppress formation of the Ac-trp-cage$_{4\text{-}20}$ product, the cross-coupling reaction conditions were modified. In order to avoid formation of the Ac-trp-cage$_{4\text{-}20}$ product, the reaction conditions would need to prevent trityl deprotection of the Gln5 side chain, and to prevent the subsequent proteolytic cleavage reaction. Reaction conditions were briefly screened to examine the role of the base (DIPEA), concentration of acidic reagent (thiolacetic acid), reaction time and temperature, and role of the solvent. The potential for acid-mediated trityl-deprotection suggested that the concentration of thiolacetic acid should be reduced. The previous observation of proteolysis under prolonged exposure to base in homocysteine-mediated NCL reactions[447] suggested that the reaction conditions should have reduced reaction duration and reduced concentration of the base. In observing that the proteolytic cleavage did not occur at room temperature, it was proposed that the high reaction temperatures could play a role in the proteolytic cleavage reaction. The products resulting from the optimized conditions for the cross-coupling reaction on the peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine is shown in Figure 4.50c.

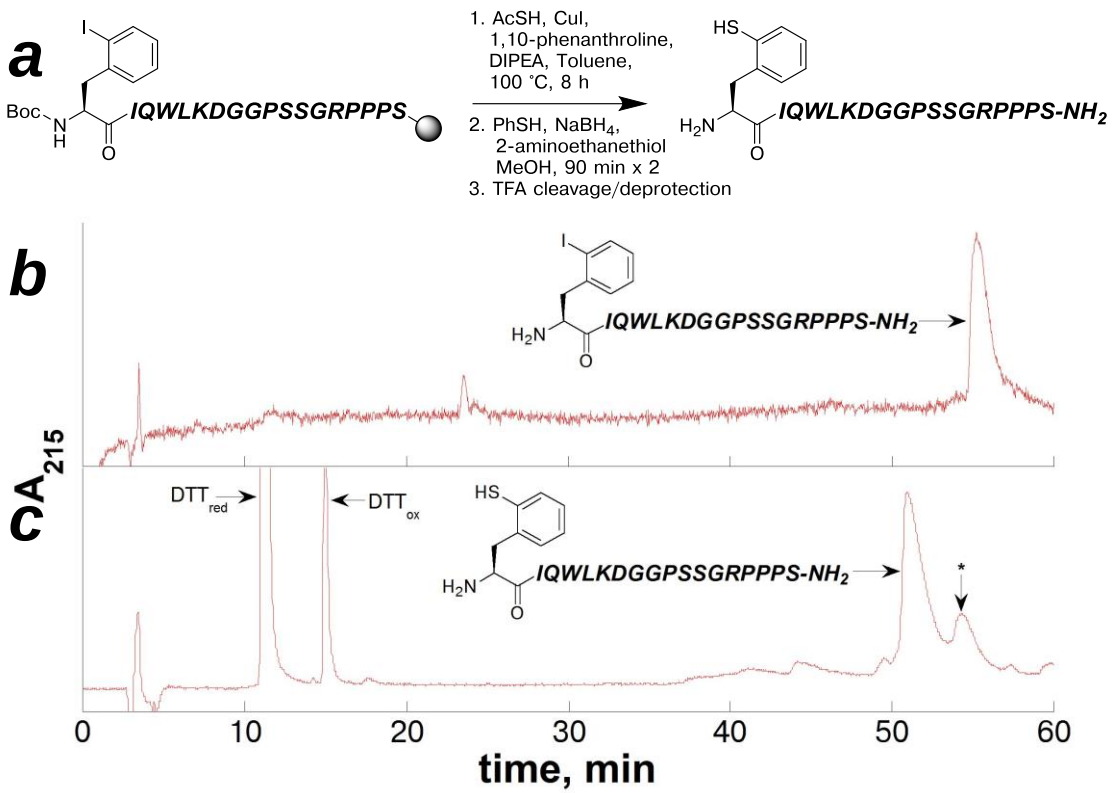


Figure 4.50 Optimized reaction conditions to produce the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine at the N-terminus

(a) Reaction scheme for the cross-coupling reaction on the protected peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine on solid phase; (b) HPLC chromatogram of the peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine and subsequent reaction with DTT in solution using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore); (c) HPLC chromatogram of the crude peptide products that result from the optimized solid-phase cross-coupling reaction on the peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenylalanine using a linear gradient of 0-45% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

*The Ac-trp-cage$_{4\text{-}20}$ side-product was observed in lower quantities under the optimized reaction conditions. Formation of Ac-trp-cage$_{4\text{-}20}$ was completely avoided if the reaction temperature was at 90 °C, but with very low conversion to the product peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine.

In the initial syntheses of the peptides (2-SH-Phe)RAFS-NH_2 and trp-cage$_{3-20}$ containing 2-thiophenylalanine using the solid-phase cross-coupling reaction conditions, the 2-thiophenylalanine at the N-terminus reacted readily as a nucleophile in intramolecular acyl transfer reactions. In both the model peptide XRAFS-NH_2 and the peptide trp-cage$_{3-20}$, consideration of the potential side-reactions and modification of the synthetic strategy improved the resultant yield of the peptide containing 2-thiophenylalanine. In cases where the cross-coupling reaction on solid-phase is impractical or potentially causes detrimental side-reactions, 2-thiophenylalanine could alternatively be incorporated via SPPS with the protected Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine (Chapter 4.2.4). However, the synthetic strategy for peptides containing 2-thiophenylalanine via cross-coupling reaction on solid phase generates the reactant peptides for NCL reactions with less time and smaller quantities of reagents. Depending on the target peptide or protein, both synthetic strategies for generating peptides with 2-thiophenylalanine are viable options.

4.2.13 2-Thiophenylalanine-mediated native chemical ligation to synthesize a Y3F variant of the trp cage miniprotein

Having synthesized the peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine at the N-terminus, the utility of this amino acid for NCL reactions was established in the context of a larger protein. The purified peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine was subjected to ligation reaction conditions with the peptide trp-cage$_{1-2}$–SPh using previously established NCL reaction conditions from the model peptides (Chapter 4.2.7), with 6 M guanidinium added to fully solubilize the reactant peptides. Two reactions with different ratios of the two reactant peptides, trp-cage$_{3-20}$ containing 2-thiophenylalanine and trp-cage$_{1-2}$–SPh, were allowed to react under

ligation reaction conditions. One reaction contained an excess of the peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine relative to trp-cage$_{1-2}$–SPh, and the other reaction contained both peptides as a 1:1 molar ratio. The resultant HPLC chromatograms of the reaction to generate the Y3(2-SH-Phe) variant of the trp cage miniprotein are shown in Figures 4.51 and 4.52.

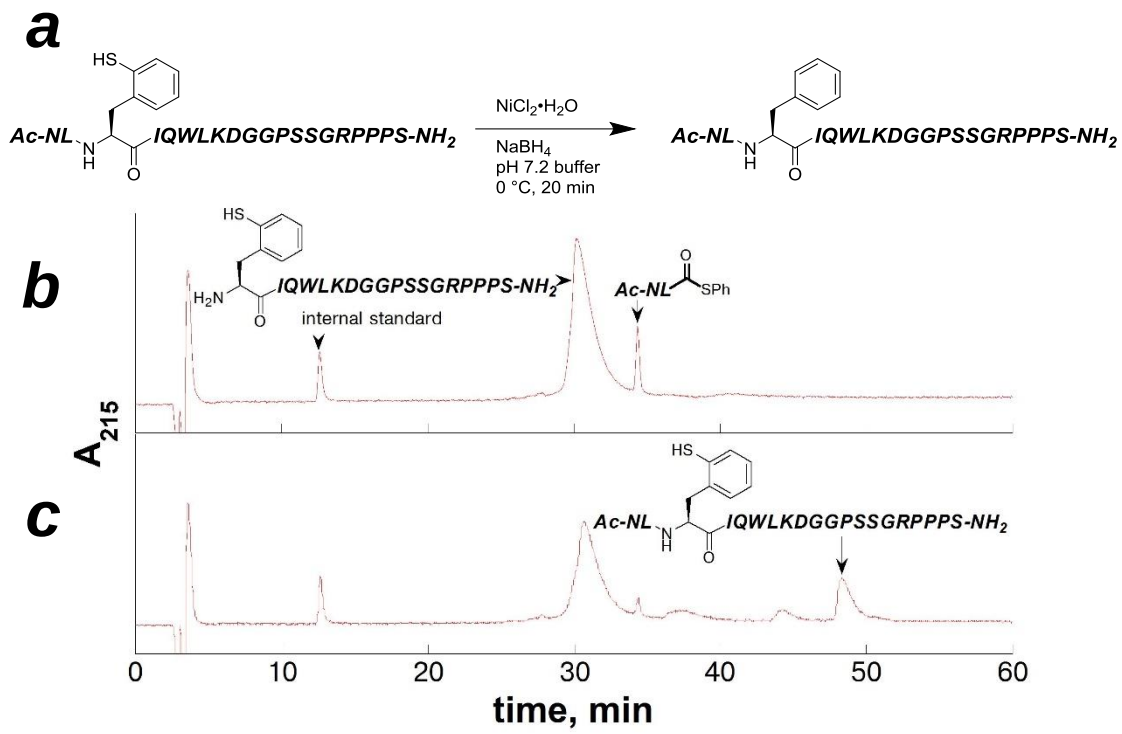


Figure 4.51 Native chemical ligation reaction to generate the Y3(2-SH-Phe)trp cage miniprotein using an excess of the peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine

(a) Reaction scheme for ligation reaction using the peptides trp-cage$_{1-2}$–SPh and trp-cage$_{3-20}$ containing 2-thiophenylalanine; (b) HPLC chromatogram of the coinjection of the reactant peptides with an internal standard (2 mM internal PhOH); (c) HPLC chromatogram of the peptide products that resulted from the ligation reaction using the peptides trp-cage$_{1-2}$–SPh and trp-cage$_{3-20}$ containing 2-thiophenylalanine after 3 hours using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

The reaction concentration of the peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine was 2.0 mM, and that of the peptide trp-cage$_{1-2}$–SPh was 0.8 mM.

Concentrations were determined by NMR (trp-cage$_{1-2}$–SPh) and Ellman's test (trp-cage$_{3-20}$ containing 2-thiophenylalanine), and reaction conversion was determined by integration of the HPLC chromatogram (normalized to the internal standard) by depletion of the peptide trp-cage$_{1-2}$–SPh. Reaction conversion was calculated to be 77%.

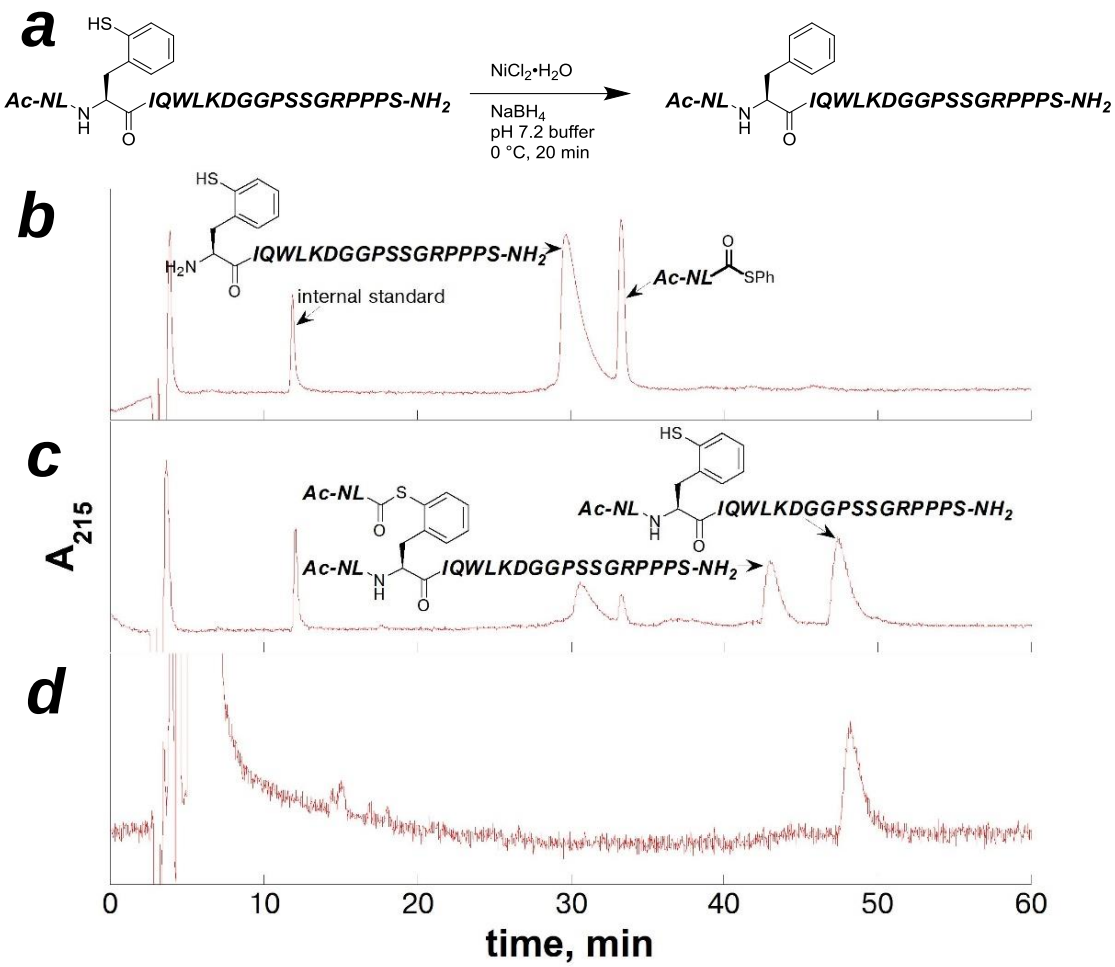


Figure 4.52 Native chemical ligation reaction to generate the Y3(2-SH-Phe)trp cage miniprotein using an equimolar ratio of reactant peptides

(a) Reaction scheme for ligation reaction using the peptides trp-cage$_{1-2}$–SPh and trp-cage$_{3-20}$ containing 2-thiophenylalanine; (b) HPLC chromatogram of the coinjection of the reactant peptides with an internal standard (2 mM internal PhOH); (c) HPLC chromatogram of the peptide products that resulted from the ligation reaction using the peptides trp-cage$_{1-2}$–SPh and trp-cage$_{3-20}$ containing 2-thiophenylalanine after 3 hours; (d) HPLC chromatogram of the isolated, diacylated reaction product (indicated in c) that was allowed to react with DTT in solution, generating the Y3(2-SH-Phe) trp cage variant. HPLC chromatograms shown used a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). The reaction concentration of the peptide trp-cage$_{3-20}$ containing 2-thiophenylalanine was 2.0 mM, and that of the peptide trp-cage$_{1-2}$–SPh was 1.95 mM. Concentrations were determined by NMR (trp-cage$_{1-2}$–SPh) and Ellman's test (trp-cage$_{3-20}$ containing 2-thiophenylalanine), and reaction conversion was determined by integration of the HPLC chromatogram (normalized to the internal standard) by depletion of the peptide trp-cage$_{1-2}$–SPh. Reaction conversion was calculated to be 74%, including the diacylated reaction product.

Using the ligation reaction conditions, 2-thiophenylalanine was successfully utilized in NCL reactions for synthesis of a Y3(2-SH-Phe) trp cage variant, representing a larger protein with broader scope of amino acid functional groups, particularly in the cross-coupling reaction. The ligation reaction was 74-77% complete in 3 hours at 23 °C, depending on the relative molar concentrations of the reactant peptides. This conversion compares well to the model ligation reaction with the peptides (2-SH-Phe)XRAFS-NH_2 and Ac-LYRAL-SPh, which was 90% complete in 6 hours at 23 °C. For the ligation reaction using a 1:1 molar ratio of the reactant peptides, an additional side-product was observed. The side-product was consistent with the product peptide, Y3(2-SH-Phe) trp cage miniprotein, with an additional trp-cage$_{1-2}$ moiety. This side product is consistent with the product peptide containing 2-thiophenylalanine proceeding through a thioester exchange reaction with unreacted peptide trp-cage$_{1-2}$–SPh, generating a diacylated product (Figure 4.52c). In NCL reactions that were mediated by homocysteine, a similar product was observed, and

this product was avoided if a slight excess of the peptide containing the thioester group was used.[447] The diacylated product was isolated and allowed to react in solution with DTT to generate the Y3(2-SH-Phe) trp cage variant product (Figure 4.52d). These reactions demonstrated that 2-thiophenylalanine can be effectively utilized to mediate ligation reactions in the context of larger peptides.

In order to generate phenylalanine at the ligation site, the product peptide Y3(2-SH-Phe) trp cage variant was subjected to desulfurization reaction conditions. Nickel boride desulfurization of peptides containing 2-thiophenylalanine had been established as a working methodology in model peptides, although the isolated yield was low (Chapter 4.2.11). The purified Y3(2-SH-Phe) trp cage variant from the ligation reaction was desulfurized using nickel boride to generate phenylalanine at the ligation site (Figure 4.53).

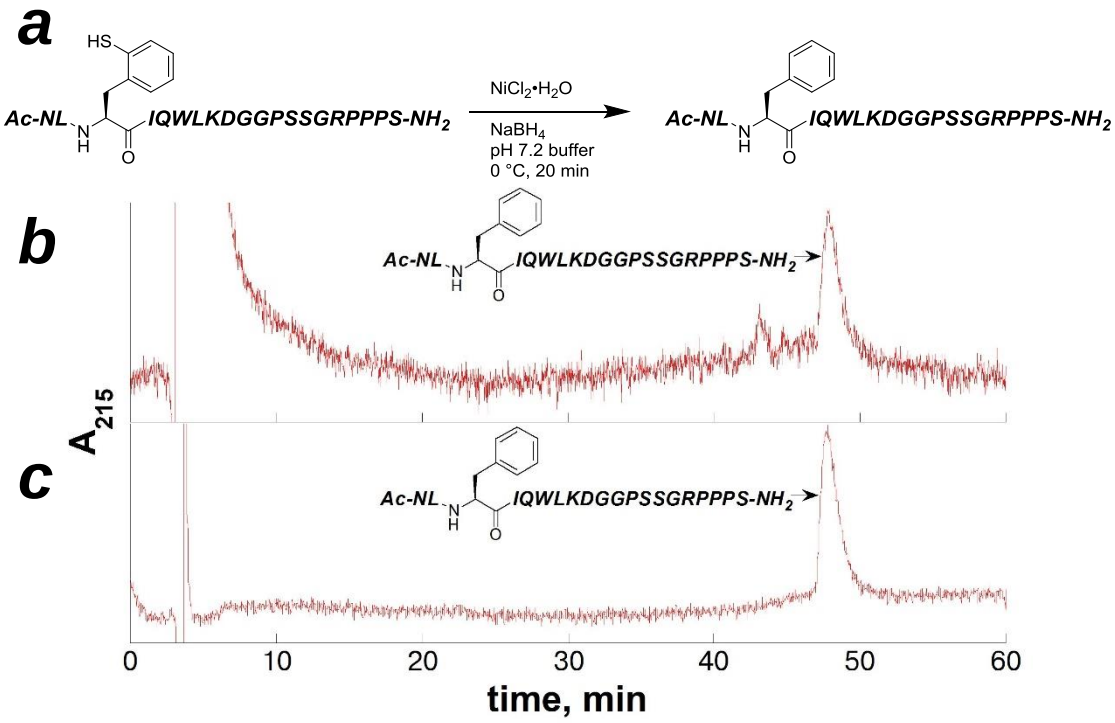


Figure 4.53 Desulfurization reaction on the peptide containing 2-thiophenylalanine to produce the Y3F trp cage variant
(a) Reaction scheme for the desulfurization reaction on the Y3(2-SH-Phe) trp cage variant from the ligation reaction; (b) HPLC chromatogram of the crude peptide products that resulted from the desulfurization reaction on the Y3(2-SH-Phe) trp cage variant from the ligation reaction; (c) HPLC chromatogram of the purified Y3F trp cage miniprotein that was independently synthesized via Fmoc-SPPS. HPLC conditions were using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).
The product was confirmed via ESI-MS.

The Y3(2-SH-Phe) trp cage variant, generated from the ligation reaction mediated by 2-thiophenylalanine, was cleanly desulfurized using nickel boride to generate the Y3F trp cage variant. The desulfurization reaction product was confirmed via ESI-MS, and the product had the same t_R via HPLC as an independently synthesized Y3F trp cage variant (Figure 4.53c). Similar to the desulfurization reactions on the model peptides, there was low isolated yield of the Y3F trp cage variant that was generated from the ligation and desulfurization reactions (Chapter 4.2.11), with observable aggregation of the peptide on nickel solids. Although the extent of aggregation and loss of the product peptide Y3F trp cage variant was not quantified, alternative approaches to desulfurization might be more appropriate, such as Pd/Al_2O_3.[472]

The utility of 2-thiophenylalanine for NCL and subsequent desulfurization reaction to generate the Y3F trp cage variant was demonstrated. The ligation reaction proceeded readily, with 74-77% reaction completion in 3 hours at 23 °C, and demonstrated utility of this unique amino acid in the context of larger proteins with broader scope of side-chain functional groups. The desulfurization reaction cleanly generated phenylalanine in place of 2-thiophenylalanine after the ligation reaction, but the recovery of the peptide was low, and it is suggested that alternative approaches to desulfurization (such as Pd/Al_2O_3 or Raney nickel)[399, 400, 448] are utilized.

4.2.14 Synthesis of a modified chicken villin headpiece ($cVHP_{17-35}$) containing 2-thiophenylalanine for native chemical ligation reactions

Having shown that 2-thiophenylalanine can be used as effectively for NCL reactions in the trp cage miniprotein, we sought to demonstrate 2-thiophenylalanine-mediated NCL reactions with a longer peptide containing a thioester group. The trp

cage miniprotein was selected as a model protein because it has been well characterized,[94-96, 103] and was previously utilized in substrate scope studies in optimizing the cross-coupling reaction on solid phase to generate peptides containing thiophenylalanine[169] (Chapter 1.2.4). In the trp cage, Tyr3 was selected as the site for the ligation reaction to result in a modest change in the structure, replacing tyrosine with phenylalanine after ligation reaction using 2-thiophenylalanine and subsequent desulfurization reaction. However, with the ligation site at position 3, the peptide containing the thioester group contained only 2 amino acids. Using the trp cage miniprotein as a target for NCL demonstrated that 2-thiophenylalanine could be synthesized at the N-terminus of a protected protein on solid phase, but the use of a dipeptide containing a thioester was not representative of an NCL reaction in the context of a larger protein.

We selected a well-studied, thermostable protein domain to synthesize via 2-thiophenylalanine-mediated NCL reaction, the chicken villin headpiece.[484] Villin is an actin-binding protein, which consists of two subdomains, an 84 kDa "core" and an 8 kDa "headpiece" domain.[484] The headpiece domain, cVHP-67 (Figure 4.54), contains the actin-binding domain, and due to its stability in solution, has been well characterized via NMR.[484] Within the villin headpiece is a 35-residue fragment that contains a hydrophobic "core" of three conserved phenylalanine residues (orange residues, Figure 4.54).[484-489] The thermostability and folding of the headpiece subdomain has been attributed to this 35-residue fragment (cVHP-35, cyan in Figure 4.54), although it exhibits little affinity for actin.[484, 489] We selected one of the three, highly conserved phenylalanine residues in the cVHP-35 domain (Phe17) for the ligation site to incorporate 2-thiophenylalanine. Positioning the ligation site at Phe17

can be advantageous for studying the role of the N-terminal α-helix, for stability studies with and without the actin binding component, or for understanding the role of phosphorylation headpiece domains related to cVHP.[487, 488, 490, 491] NCL at position 17 could also be useful for examining the kinetics of protein folding involving aromatic amino acids, which has been well examined with cVHP-35.[489] Phe17 is a conserved residue involved in folding and stability of the cVHP-35 fragment, and therefore the ligation reaction at this position must be mediated by a thiolated aromatic amino acid, such as 2-thiophenylalanine.

PTKLETFPLDVLVNTAAEDLPRGVDPSRKENHLSDEDFKAVFGMTRSA
FANLPLWKQQNLKKEKGLF-NH_2

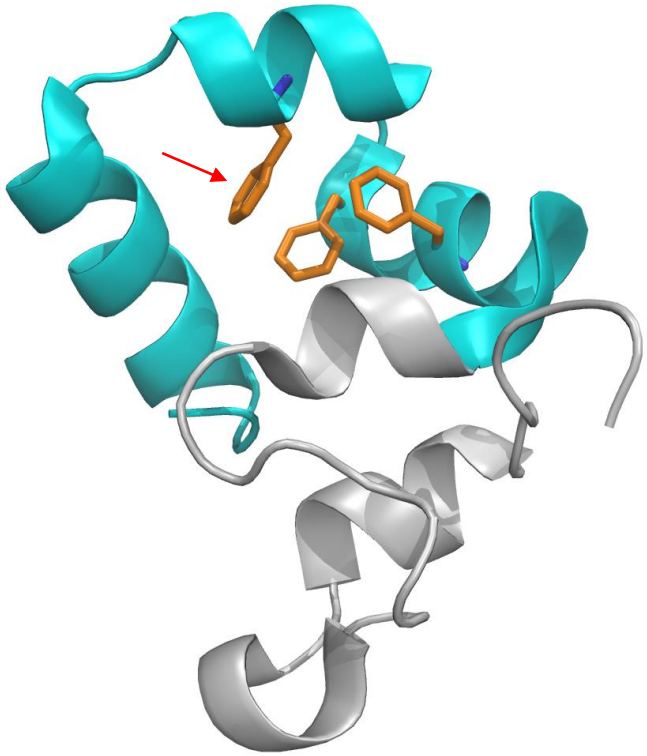


Figure 4.54 The chicken villin headpiece domain (cVHP-67)
Villin is an actin binding protein, and contains a stable 8 kDa subunit, termed the "villin headpiece." Within the villin headpiece domain, there is a 35-residue three-helix bundle (cyan, sequence shown) that is stabilized by 3 phenylalanines (orange). The 35-residue fragment (cVHP-35) has been attributed to the stability of the headpiece, is highly conserved, and has been the subject of numerous stability and peptide folding studies.[484-489] The ligation site for incorporating 2-thiophenylalanine is indicated by the underline in the sequence, and the arrow in the figure.

In order to demonstrate the 2-thiophenylalanine-mediated NCL reaction using the cVHP-35 domain, the peptide $cVHP_{17-35}$ containing Boc-2-iodophenylalanine at the N-terminus ((2-I-Phe)ANLPLWKQQNLKKEKGLF) was synthesized via SPPS on Rink amide resin. The protected peptide $cVHP_{17-35}$ containing Boc-2-iodophenylalanine was initially subjected to the cross-coupling reaction conditions

that were utilized for the model peptides XRAFS-NH_2 (Chapter 4.2.2). However, a major product was observed that corresponded to the peptide Ac-cVHP$_{18\text{-}35}$ (truncation of the phenylalanine residue at the N-terminus, with an additional acetyl group). A corresponding side-product was observed in the initial synthesis of the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine, which contained a trityl-protected glutamine residue at the *i*,*i*+2 position relative to the N-terminus (Chapter 4.2.12). Based on the common moiety between these two peptides, (2-SH-Phe)XQ/N where Q and N were trityl-protected during the cross-coupling reaction, it was thought that the peptide cVHP$_{17\text{-}35}$ was also subject to side reaction via the mechanism shown in Figure 4.49.

Given the potential for reactivity in the side-chain protecting groups, the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine was synthesized on solid phase via cross-coupling reaction conditions, but the reaction conditions were optimized for the peptide cVHP$_{17\text{-}35}$ containing 2-iodophenylalanine. Similar to optimization of the reaction conditions for the peptide trp-cage$_{3\text{-}20}$, described in Chapter 4.2.12, reaction conditions were screened for reaction solvent, reaction temperature, reacton duration, and concentrations of DIPEA and thiolacetic acid. The aim for reaction condition optimization for the peptide cVHP$_{17\text{-}35}$ containing 2-iodophenylalanine was to prevent trityl deprotection reaction on the asparagine reside (located 2 residues from the 2-iodophenylalanine), and to prevent reaction of thiolates with the carbamate bond at 2-iodophenylalanine. The optimized solid-phase cross-coupling reaction conditions for synthesis of the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine are shown in Figure 4.55.

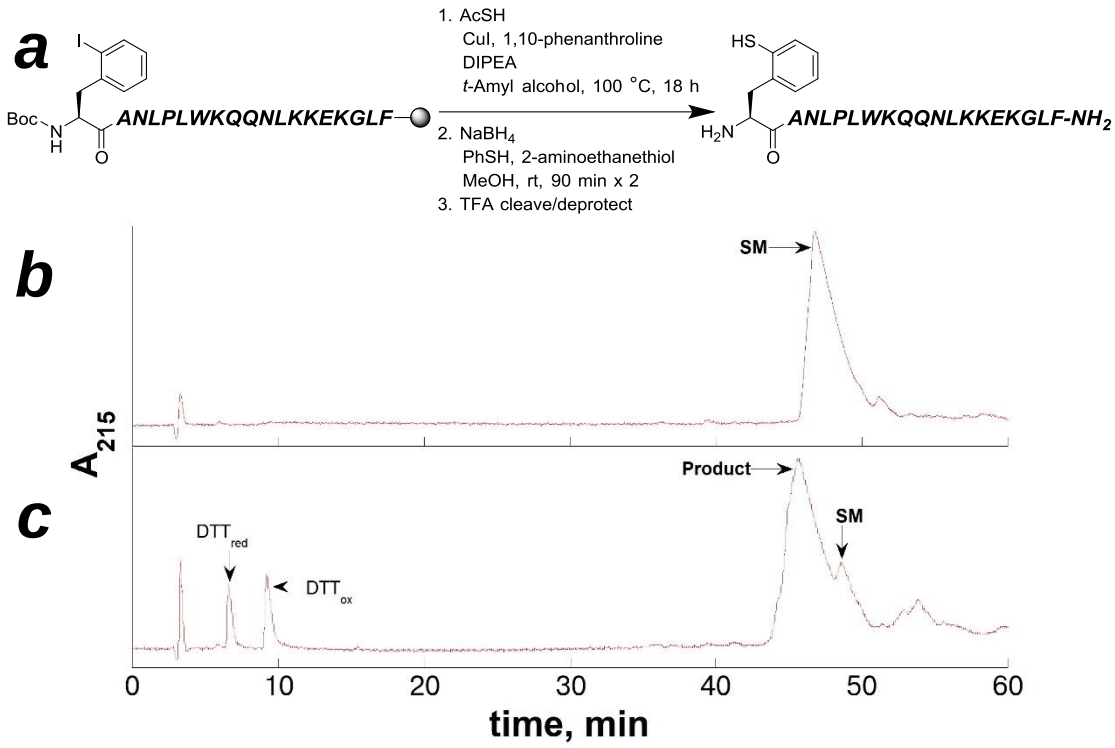


Figure 4.55 Synthesis of the peptide cVHP$_{17-35}$ containing 2-thiophenylalanine via cross-coupling reaction

(a) Reaction scheme for the cross-coupling reaction on the protected peptide cVHP$_{17-35}$ containing 2-iodophenylalanine on solid phase; (b) HPLC chromatogram of the peptide cVHP$_{17-35}$ starting material containing 2-iodophenylalanine, indicated as "SM"; (c) HPLC chromatogram of the crude peptide products that result from the optimized solid-phase cross-coupling reaction on the peptide trp-cage$_{3-20}$ containing 2-iodophenylalanine and subsequent reduction reaction in solution with DTT. HPLC chromatograms shown used a linear gradient of 0-60% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). The products were identified via ESI-MS.

DTT$_{red}$ and DTT$_{ox}$ indicate reduced and oxidized DTT, respectively.

"SM" indicates the peptide cVHP$_{17-35}$ containing 2-iodophenylalanine at the N-terminus.

"Product" indicates the peptide cVHP$_{17-35}$ containing 2-thiophenylalanine at the N-terminus.

The optimized cross-coupling reaction on the peptide $cVHP_{17\text{-}35}$ containing 2-iodophenylalanine on solid phase successfully generated the peptide containing 2-thiophenylalanine at the N-terminus. Similar to the synthesis of the peptide trp-$cage_{3\text{-}20}$ containing 2-thiophenylalanine, the truncated, acylated side-product (Ac-$cVHP_{18\text{-}35}$) was avoided if the optimized reaction conditions were employed, using a lower concentration of DIPEA, and lower reaction temperature. Furthermore, a lower concentration of thiolacetic acid was utilized in these reaction conditions, and a polar protic solvent was utilized to stabilize any thiolates that may have formed during the reaction. Reaction conversion to form the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine was improved by increasing the reaction time to 18 hours (from the 8 hours utilized in the optimized cross-coupling reaction conditions for the trp cage miniprotein). However, the overall cross-coupling reaction yield for the peptide $cVHP_{17\text{-}35}$ was considerably lower in comparison to the model peptide XRAFS-NH_2. The purification of the peptide product resulting from the cross-coupling reaction was difficult, due to the severe "tailing" of the peptide product in the HPLC purification conditions and co-elution with the reactant peptide $cVHP_{17\text{-}35}$ containing 2-iodophenylalanine (Figure 4.55).

The cross-coupling reaction on solid phase generated the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine in a practical manner. However, with the potential for side-chain reactivity involving asparagine, the modestly lower reaction yield resulting from the optimized cross-coupling reaction conditions, and coelution of the product peptide and the reactant peptide in the HPLC purification, an alternative synthetic approach to $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine was utilized. A protocol had been developed for synthesis of the Boc-protected 2-thiophenylalanine for

incorporation into peptides via SPPS, which was utilized as an alternative approach to synthesis of the model peptide (2-SH-Phe)RAFS-NH_2 (Chapter 4.2.4). Using the Boc-protected 2-thiophenylalanine, which was synthesized in three steps from the commercially available Boc-2-iodophenylalanine, the peptide $cVHP_{17\text{-}35}$ containing Boc-2-thiophenylalanine at the N-terminus was synthesized with via SPPS (Figure 4.56). The resultant peptides $cVHP_{17\text{-}35}$ containing an N-terminal 2-thiophenylalanine, synthesized via cross-coupling reaction on solid phase or via SPPS with Boc-protected 2-thiophenylalanine, were confirmed to be identical by NMR (Figure 4.57, Figure 4.58).

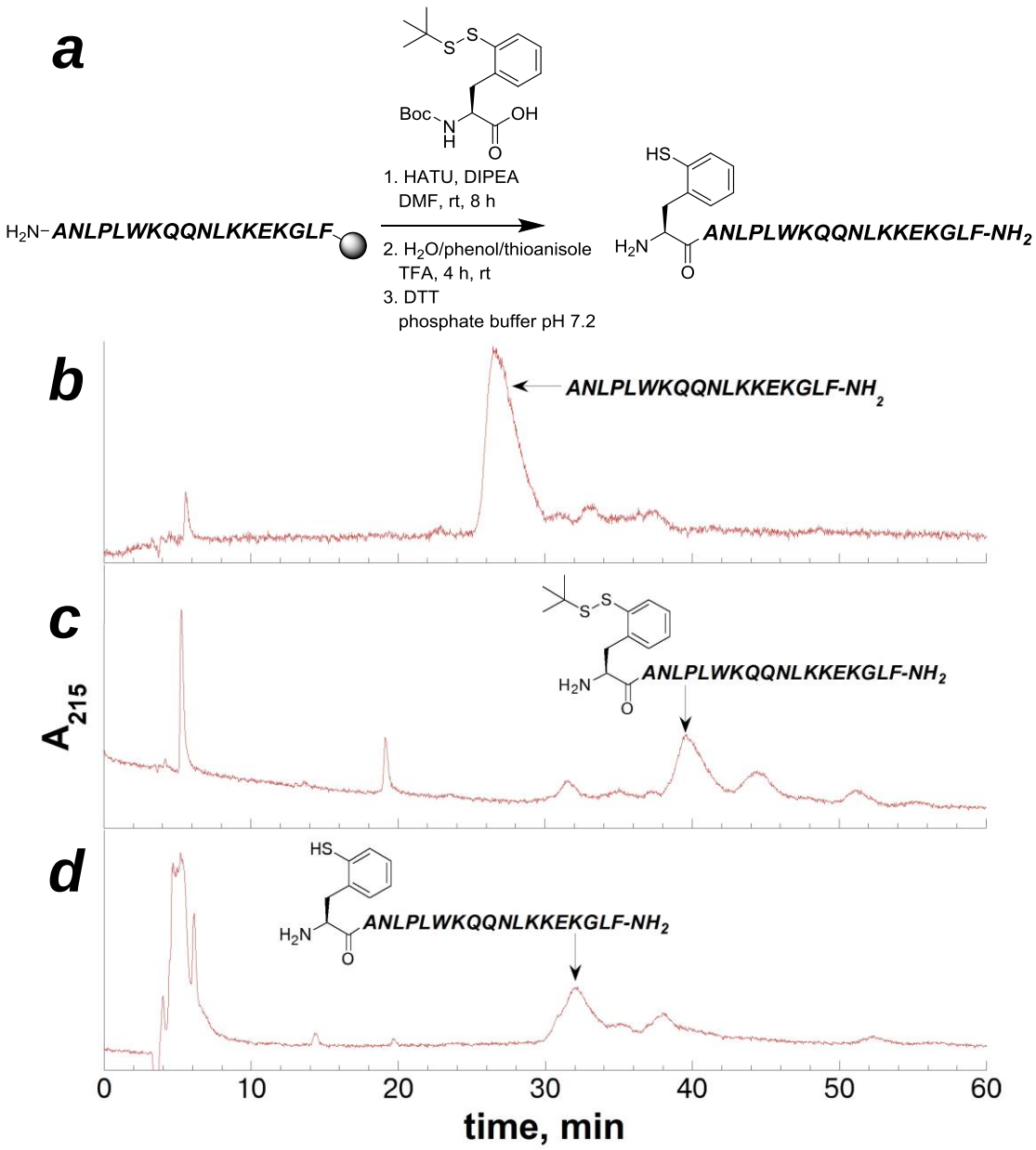


Figure 4.56 Synthesis of the peptide cVHP$_{17-35}$ via SPPS with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine

(a) Reaction scheme to synthesize the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine via SPPS using Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine; (b) the HPLC chromatogram of the peptide $cVHP_{18\text{-}35}$ prior to any reactions; (c) the HPLC chromatogram of the peptide products that resulted from the coupling reaction with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine; (d) the HPLC chromatogram that resulted from the reaction of the peptide products with DTT. HPLC chromatograms shown used a linear gradient of 15-75% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore). The products were identified via ESI-MS.

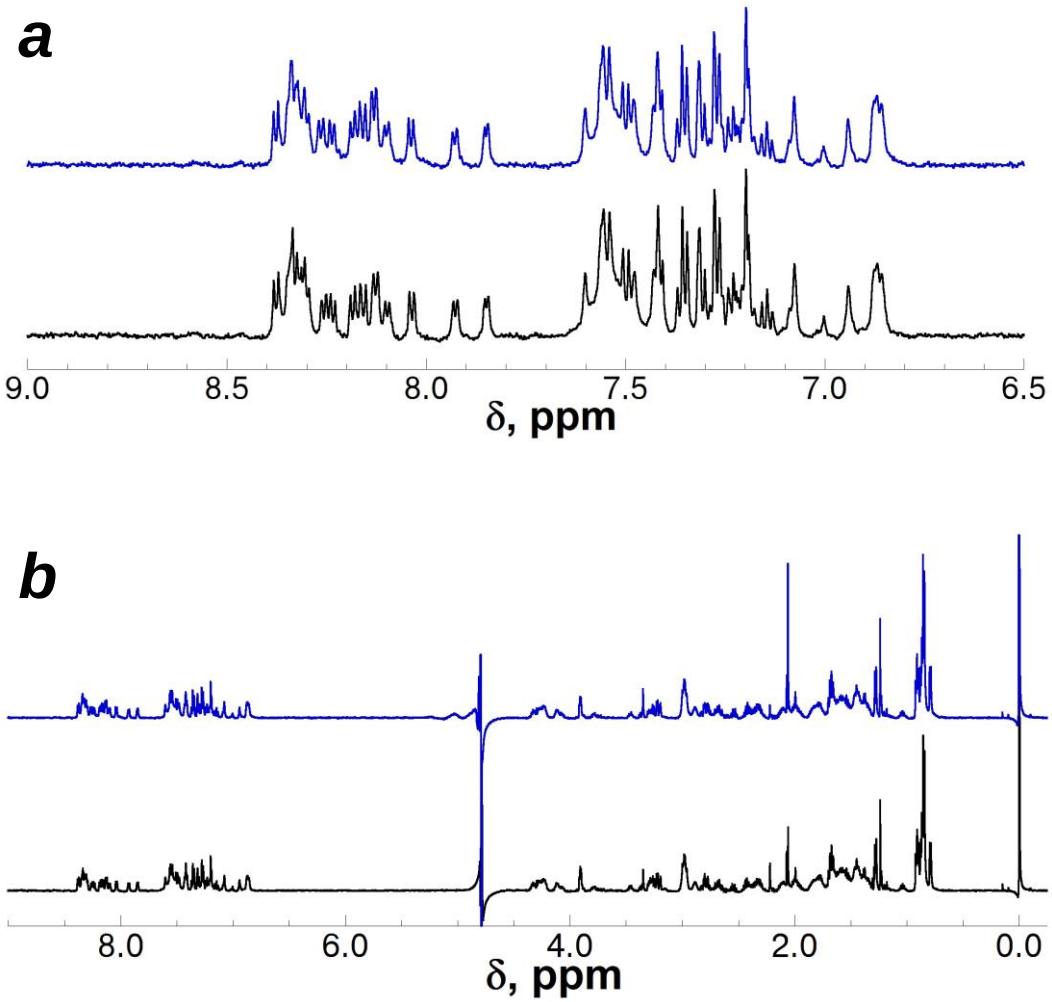


Figure 4.57 ^{1}H NMR spectra for the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus synthesized by two strategies
Superposition of the ^{1}H NMR spectra of the peptide cVHP$_{17\text{-}35}$ containing N-terminal 2-thiolphenylalanine synthesized by solid phase cross-coupling reaction (blue) or synthesized by solid-phase peptide synthesis using Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine (black). (a) Amide and aromatic regions of the ^{1}H NMR spectra; (b) full ^{1}H NMR spectra. The peptides were dissolved in 90% H_2O/10% D_2O with 5 mM phosphate buffer (pH 4), 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP.

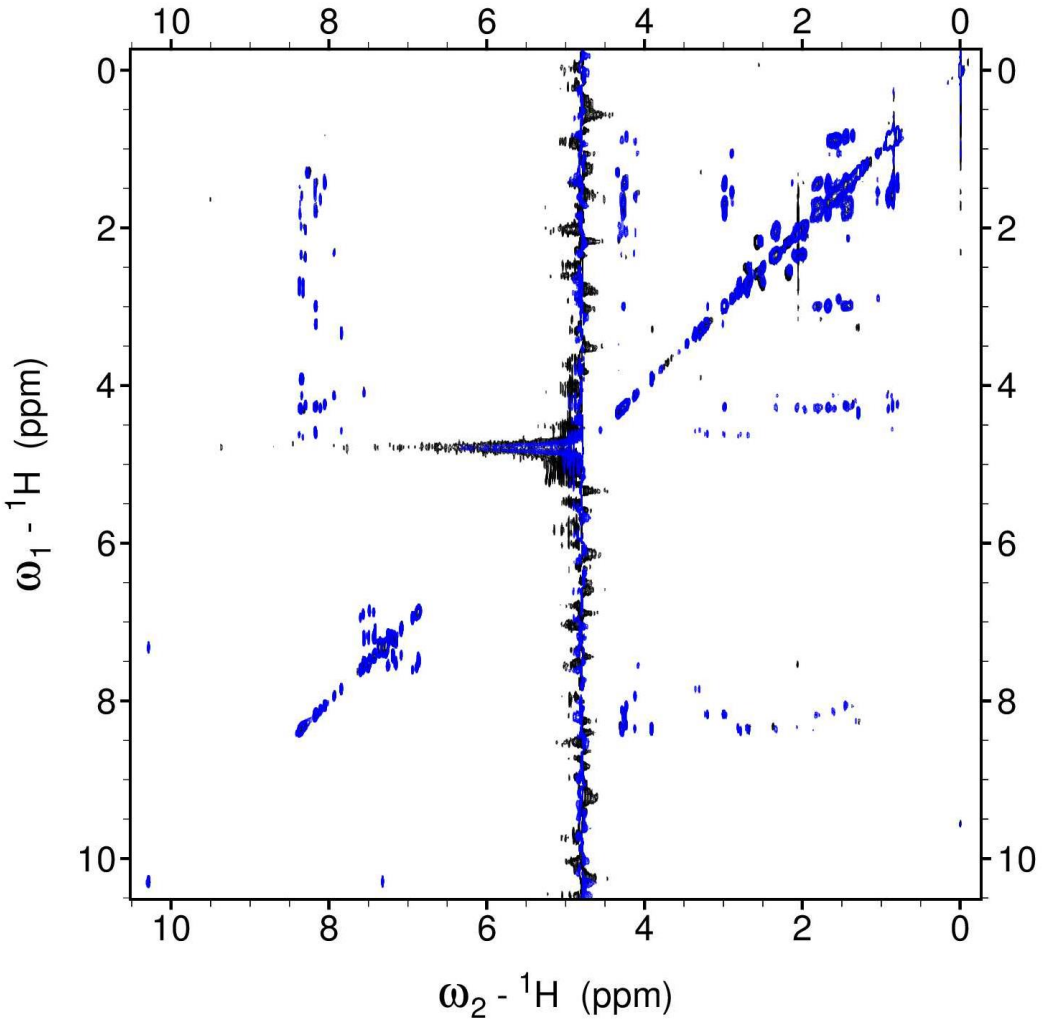


Figure 4.58 Superposition of the TOCSY spectra for the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus synthesized by two strategies
The peptide cVHP$_{17\text{-}35}$ containing N-terminal 2-thiolphenylalanine synthesized by solid phase cross-coupling reaction (blue) or synthesized by solid-phase peptide synthesis using Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine (black). The peptides were dissolved in 90% H_2O/10% D_2O with 5 mM phosphate buffer (pH 4), 25 mM NaCl, 0.1 mM TCEP, and 0.1 mM TSP.

2-Thiophenylalanine can be incorporated at the N-terminus of the peptide cVHP$_{17\text{-}35}$ using two different synthetic strategies. With the cross-coupling reaction on the peptide cVHP$_{17\text{-}35}$ containing Boc-2-iodophenylalanine on solid phase, the peptide

containing 2-thiophenylalanine can be rapidly generated with no solution synthesis of purification steps. However, with the cross-coupling reaction, there is a potential risk of side-chain deprotection and cross-reactivity with the 2-thiophenylalanine side chain of its thioester. The cross-coupling reaction yield on solid phase was modified to suppress potential side-chain reactions, but this change resulted in modestly lower product yield. Alternatively, the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine was synthesized by coupling the protected Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine via standard solid-phase peptide synthesis. Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine was synthesized from commercially available Boc-2-iodophenylalanine in 3 steps with 16% overall yield. This synthetic strategy has a lower risk of side reactions. Notably, in the 9 step synthesis of β-mercaptophenylalanine reported by Crich & Banerjee,[401, 402] an overall 11.5% yield was reported. Malins *et al*.[472] reported a synthetic route to β-selenol-phenylalanine with an overall 13% yield in 7 steps from commercially available Garner's aldehyde ($216/1g, $50/mmol, Sigma-Aldrich as of May, 2016). Utilizing either of these synthetic strategies developed herein, peptides containing 2-thiophenylalanine at the N-terminus can be generated in a practical manner, requiring less time and resources than other methods used to generate thiolated phenylalanine residues.

4.2.15 2-Thiophenylalanine-mediated native chemical ligation to synthesize the protein cVHP-35

Having generated the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus using two different synthetic methods, the utility of this amino acid for NCL reactions was established in the context of a larger protein with a 16-residue peptide containing a thioester group at the C-terminus. The purified peptide $cVHP_{17\text{-}35}$

containing 2-thiophenylalanine was subjected to ligation reaction conditions with the peptide cVHP$_{1\text{-}16}$–SPh using previously the established NCL reaction conditions from the model peptide (Chapter 4.2.7), with 6 M guanidinium added to fully solubilize the reactant peptides. The resultant HPLC chromatogram of the reaction that generated the F17(2-SH-Phe) variant of the cVHP-35 protein is shown in Figure 4.59.

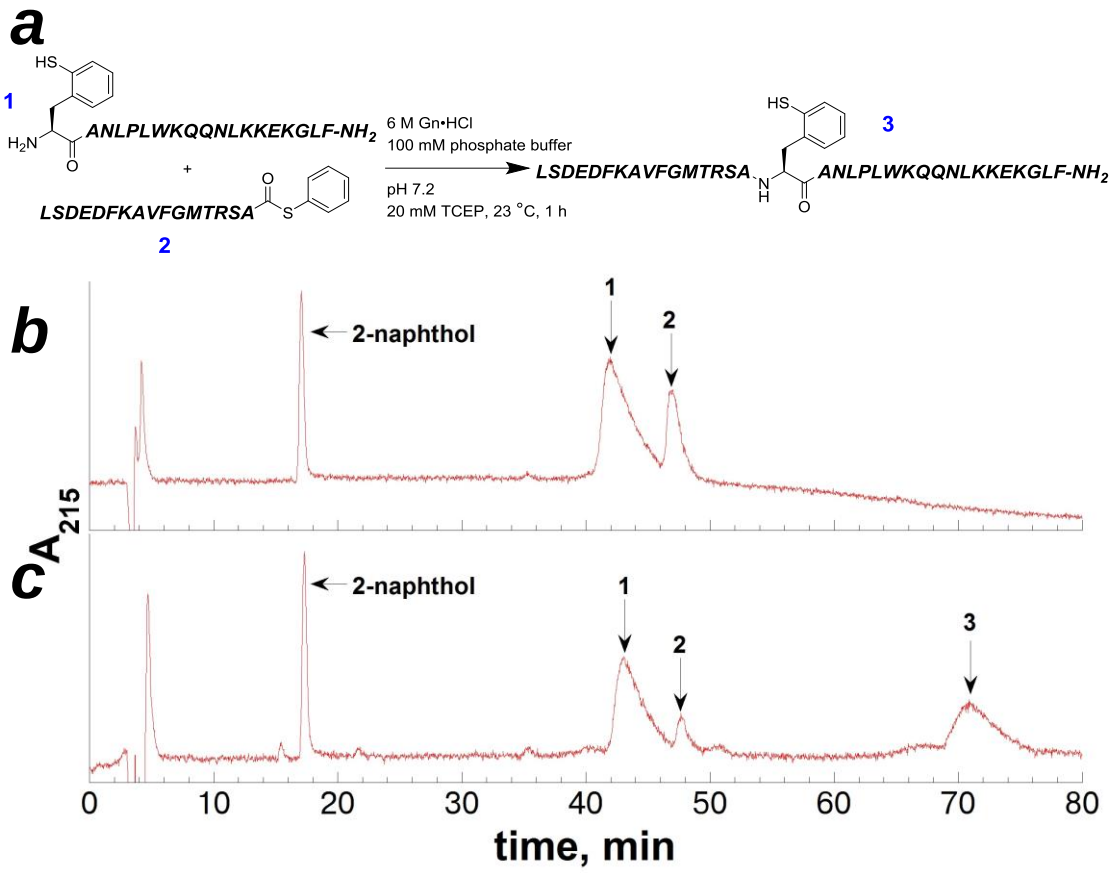


Figure 4.59 Native chemical ligation reaction to generate the F17(2-SH-Phe) cVHP-35 variant using 2-thiophenylalanine

(a) Reaction scheme for ligation reaction using the peptides cVHP1-16–SPh and cVHP17-35 containing 2-thiophenylalanine; (b) HPLC chromatogram of the coinjection of the peptides with 1 mM internal 2-naphthol internal standard; (c) HPLC chromatogram of the peptide products that resulted from the ligation reaction using the peptides cVHP1-16–SPh and cVHP17-35 containing 2-thiophenylalanine after 1 hours using a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 80 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).
1 is the peptide cVHP17-35 containing 2-thiophenylalanine
2 is the peptide cVHP1-16–SPh
3 is the F17(2-SH-Phe) variant of cVHP-35
The reaction concentration of the peptide cVHP17-35 containing 2-thiophenylalanine was 3.0 mM, and that of the peptide cVHP1-16–SPh was 1.0 mM.
Concentrations were determined by NMR (cVHP1-16–SPh) and Ellman's test (cVHP17-35 containing 2-thiophenylalanine), and reaction conversion was determined by integration of the HPLC chromatogram (normalized to the internal standard) by depletion of the peptide cVHP1-16–SPh. Reaction conversion was calculated to be 67% after 1 hour.

The ligation reaction to produce the F17(2-SH-Phe) cVHP-35 variant was 67% complete in only 1 hour at 23 °C. While this reaction conversion is somewhat less than the ligation reaction in model peptides for (2-SH-Phe)RAFS-NH_2 and Ac-LYRAA-SPh (95% in 1 hour), this reaction is still considerably faster than similar reported reactions using *S*-phenyl thioesters.[472] By comparison, ligation reactions using the peptides LYRAA and CRANK reported to be less than 50% complete in 4 hours at 37 °C, although the thioester exchange reaction to form the S-phenyl thioester (from the S-benzyl thioester) was potentially a rate-determining step.[470] With ligation reactions using peptides containing 2-thiol-tryptophan and Ac-LYRANA-SPh, the reaction was incomplete after 15 hours, requiring additional Ac-LYRANA-SPh for an isolated yield of 81% over 24 hours; peptides containing leucine thioesters were not examined in that study.[472]

In order to generate the cVHP-35 protein, the peptide F17(2-SH-Phe) cVHP-35 that resulted from the ligation reaction was subjected to desulfurization reaction conditions. The conditions for desulfurization used nickel boride, following the protocol that was used for desulfurization reaction on the model peptides and the Y3(2-SH-Phe) trp cage variant. The resultant chromatogram to generate the the cVHP-35 protein is shown in Figure 4.60.

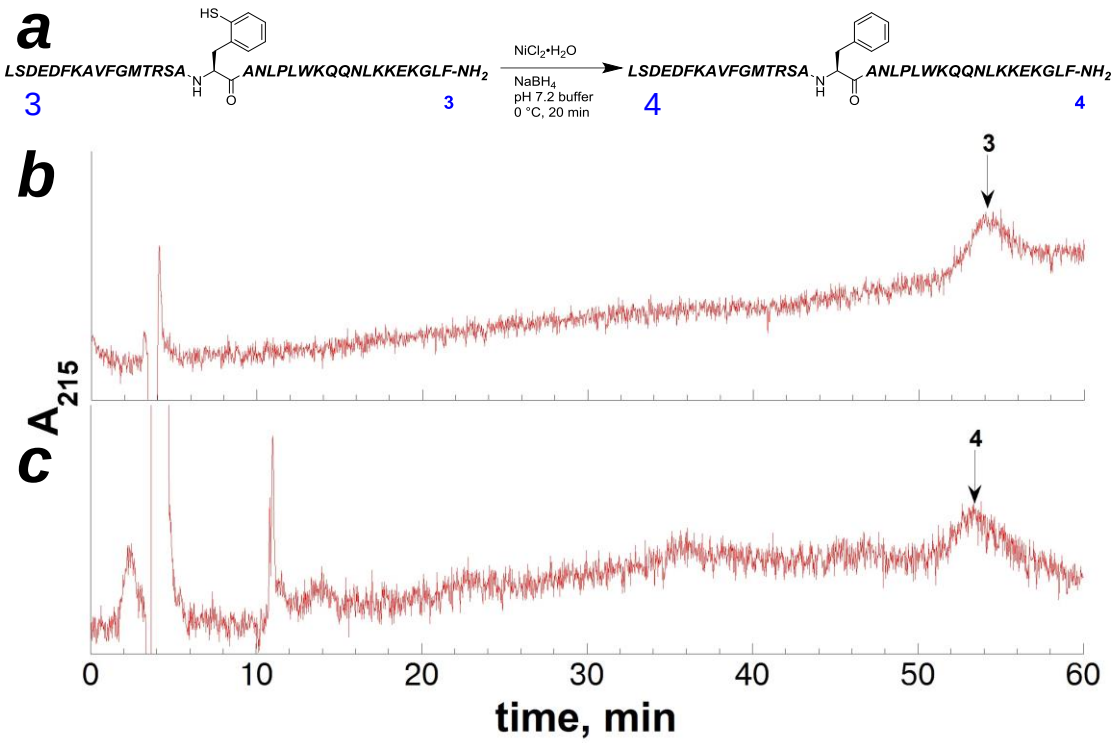


Figure 4.60 Desulfurization reaction on the peptide containing 2-thiophenylalanine to generate the peptide cVHP-35

(a) Reaction scheme for the desulfurization reaction on the F17(2-SH-Phe) variant of cVHP-35 variant from the ligation reaction; (b) HPLC chromatogram of the purified peptide F17(2-SH-Phe) variant of cVHP-35 that resulted from the ligation reaction; (c) HPLC chromatogram of the crude products that resulted from the desulfurization reaction. HPLC chromatograms shown used a linear gradient of 15-65% buffer B (20% water, 80% MeCN, 0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA) over 60 minutes on a Microsorb MV C18 column (4.6 mm × 250 mm, 100 Å pore).

3 is the peptide F17(2-SH-Phe) variant of cVHP-35 variant from the ligation reaction

4 is the product resulting from the desulfurization reaction, expected to be authentic cVHP-35. The UV detection via diode array suggested successful desulfurization reaction, but this product was not confirmed via mass spectroscopy due to low sample recovery from the reaction.

The desulfurization reaction was conducted on the peptide F17(2-SH-Phe) cVHP-35 in order to generate the peptide cVHP-35. The product resulting from the

desulfurization reaction was consistent with desulfurized product via UV-Vis (detected by diode array on the HPLC), but could not be confirmed via mass spectroscopy to be the desulfurized product due to low sample recovery. Similar to the desulfurization reaction on the Y3(2-SH-Phe)trp cage variant, the isolated yield of the cVHP-35 was very low. Potentially, the nickel boride solid catalysts aggregate or bind to the peptides (potentially via an aromatic interaction with the nickel surface), and so much of the product is lost upon filtration using this desulfurization strategy. An alternative desulfurization method would be required to improve the isolated yield of the reaction, such as Pd/Al_2O_3 reported by Malins *et al.*[472] in desulfurization of 2-thiol-tryptophan.

Herein, we demonstrated the utility of 2-thiophenylalanine-mediated NCL reaction with the cVHP-35 protein, at highly stabilized domain with an aromatic "core" of three phenylalanine residues. Two different synthetic strategies to obtain the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine, via solid-phase cross-coupling reaction and via SPPS with the Boc-protected amino acid.

4.3 Discussion

Prior studies on 4-thiophenylalanine demonstrated that the sulfur analogue of tyrosine had unique reactivity (Chapter 1) and effects on structure via non-covalent interactions (Chapter 2). The aryl thiol in 4-thiophenylalanine ($pK_a = 6.4$)[169] is more acidic than alkyl thiols, such as cysteine ($pK_a = 8.5$),[109] and is mostly anionic and nucleophilic under physiological conditions. Conventional native chemical ligation (NCL) reactions with thioesters rely on thiolates that are proximal to amine functional groups, where the thiolate displaces the thioester and forms a thioester-linked intermediate, and the proximal amine undergoes *S*→*N* acyl transfer to release the

thiolate and form an amide bond (Figure 4.5).[396] Much of the research on NCL since it was first proposed[396] has explored the utility of thiolated amino acids as surrogates for NCL reactions. Auxiliary-mediated NCL reactions have also been explored, where the N-terminus is functionalized with a thiolated mediator, which entropically activates the secondary amine to proceed through *S→N* acyl transfer to release the thiolate and form an amide bond (Figure 4.61).[492] The reactivity of aryl thiolates has been demonstrated in auxiliary-mediated NCL reactions,[457, 458, 460, 473, 492] but the mechanism and kinetics of auxiliary-mediated NCL reactions are distinct from thiolated amino acids, with a secondary amine initiating intramolecular *S→N* acyl transfer rather than a primary amine. Only one recent example, 2-thiol-tryptophan, [472] has demonstrated an aryl thiolated amino acid for NCL, and the results from this work were suggestive of somewhat different reaction kinetics compared to alkyl thiolated amino acids. We sought to further explore the mechanisms and applications for aryl thiolated amino acids in NCL reactions, a previously unexplored area of native chemical ligation chemistry.

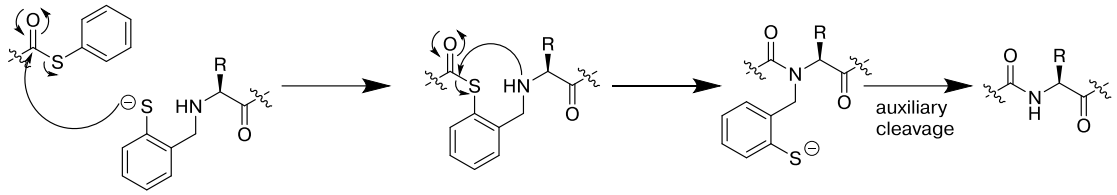


Figure 4.61 NCL reactions mediated by an aryl thiolated auxiliary
Alkyl thiolated amino acids have been widely utilized for mediating NCL reactions, with only one example of an aryl thiolated amino acid used for NCL.[399, 472, 479] Aryl thiolated auxiliaries have been utilized for NCL reactions, where a functional group containing a thiol is attached to the N-terminus, which mediates the *S*→*N* acyl transfer via a secondary amine.[492]

In the semi-synthesis of a protein via NCL, an amino acid with a thiol functional group is necessary to initiate the ligation reaction. In some cases a mutation of an amino acid to a cysteine residue is well tolerated in the structure and function of the protein.[431] However, in proteins where aromatic amino acids are conserved or critically involved in an active site mechanism or folding of a domain,[161, 489] mutation to cysteine residues is not well tolerated, and a thiolated aromatic amino acid must be used to mediate the NCL reaction. For phenylalanine, syntheses for β-mercaptophenylalanine and β-selenol-phenylalanine have been developed and successfully utilized for NCL reactions in peptides, requiring 7 or 9 step syntheses in solution and with overall product yields of 11% or 13%, respectively.[401, 493] These sources have been widely cited in literature as examples of NCL reactions with aromatic amino acids, but β-mercaptophenylalanine and β-selenol-phenylalanine have not been applied beyond their initial reports for the chemical synthesis of proteins, suggesting that the lengthy syntheses are strongly discouraging further application. By utilizing the practical reaction methodologies that we had developed for synthesizing

thiophenylalanine within peptides and in solution (Chapters 1 and 2), we sought to develop thiophenylalanine-mediated NCL reactions for the synthesis of proteins.

4.3.1 Peptides containing thiophenylalanine at the N-terminus for NCL reaction

We initially sought to examine the plausibility of NCL reactions using 4-, 3-, or 2-thiophenylalanine at the N-terminus of a model peptide. Model peptides that were previously well studied for model NCL reactions were selected for this study:[464, 470] Ac-LYRAZ-SR (Z = canonical amino acid, SR = a thioester group) and XRAFS-NH_2 (where X = thiophenylalanine). The previously optimized cross-coupling reaction conditions were utilized on peptides containing iodophenylalanine on solid-phase to generate thiophenylalanine.[169] In the initial synthesis of 2-thiophenylalanine (2-SH-Phe) at the N-terminus of the model peptide, the major product was Ac-(2-SH-Phe)RAFS-NH_2, which resulted from the *S*→*N* acyl transfer reaction between the N-terminal amine and 2-(S-acetyl)thiophenylalanine, with complete formation of the N-acetylated peptide product in less than 5 minutes at pH 7 at room temperature (Figures 4.12 and 4.13). This rapid *S*→*N* acyl transfer reaction perfectly mimicked the same intermediate steps that would be necessary for native chemical ligation reactions with peptides containing thioesters, which strongly suggested the unique reactivity of 2-thiophenylalanine for NCL reactions. 7-Membered ring intermediates in *S*→*N* acyl transfer reaction intermediates are uncommon in prior examples of NCL with thiolated amino acids; the closest example was reported by Tam & Yu[447] with homocysteine-mediated NCL, proceeding via a 6-membered ring intermediate. In light of the exceptional reactivity of 2-thiophenylalanine, and in comparison to prior ligation reactions,[401, 447, 492] the scope and application of the 2-thiophenylalanine-meditated NCL reactions were examined more closely.

In focusing these studies on 2-thiophenylalanine-mediated NCL reactions, a practical synthesis was required to obtain peptides containing 2-thiophenylalanine at the N-terminus. The modification reaction of peptides on solid phase to generate 2-thiophenylalanine was changed so that the protected peptide was subjected to the thiolysis reaction on solid phase immediately following the cross-coupling reaction.[169] The *S*→*N* acyl transfer reaction between the N-terminal amine and 2-(S-acetyl)thiophenylalanine was avoided by subjecting the peptide to thiolysis before the N-terminal amine was deprotected (Figure 4.15). In order to determine if epimerization occurred during the cross-coupling reaction on solid-phase, the peptide (2-SH-D-Phe)RAFS-NH_2 was independently synthesized via cross-coupling reaction on the peptide containing 2-iodo-D-phenylalanine on solid phase. By comparison to the product that resulted from the cross-coupling reaction, via HPLC and NMR, it was established that no significant epimerization occurred during the cross-coupling reaction to generate peptides containing 2-thiophenylalanine on solid-phase.

In addition, to avoid any potential challenges that resulted from the cross-coupling reaction on solid phase, an alternative methodology was developed for synthesizing peptides containing 2-thiophenylalanine at the N-terminus. While the cross-coupling approach can be generally utilized on solid-phase peptides containing iodophenylalanine, it may not be ideal to expose some sensitive side-chain protecting groups or delicate, modified amino acids (such as glycosylated residues) to high reaction temperatures in the presence of bases or thioacids. In order to avoid any potential for side reactions with sensitive peptide side chains, an alternative synthetic strategy was developed, where the Boc-protected 2-thiophenylalanine was incorporated via SPPS. Rajagopalan *et al.*[23] has previously shown that Boc-

iodophenylalanine with a free carboxylic acid can react with palladium and *tert*-butyl mercaptan to form thioethers, so Boc-iodophenylalanine was used directly, without protecting the carboxylic acid, in our initial synthesis.

Surprisingly, the copper-catalyzed cross-coupling reaction on Boc-2-iodophenylalanine generated Boc-protected-3-amino-δ-thiochromanone in 86% isolated yield. The thiochromanone product was subjected to hydrolysis with lithium hydroxide in the presence of (S-*tert*-butyl)-2-thiopyridyl disulfide to generate Boc-2-S(S-*tert*-butyl)-thiophenylalanine in 42% yield over 2 steps. Boc-2-S(S-*tert*-butyl)-thiophenylalanine can be incorporated into peptides via solid-phase peptide synthesis, and the 2-thiophenylalanine can be deprotected in solution using DTT or any reductant. Boc-2-S(S-*tert*-butyl)-thiophenylalanine could also potentially be incorporated into proteins using a modified tRNA synthetase, as the enzyme will resolve the enantiomers, using only the L-amino acid.[494] However, chiral HPLC analysis and x-ray crystallography showed that both the thiochromanone and Boc-2-S(S-*tert*-butyl)-thiophenylalanine products were completely racemic mixtures. The racemization potentially occurred during the cross-coupling reaction conditions, where formation of the δ-thiochromanone or the mixed anhydride rendered the Hα especially prone to abstraction, resulting in enolization at the Cα (Figure 4.62).[495] Attempts to modify the reaction conditions in order to increase the enantiopurity (decreasing the reaction temperature and reaction duration, using different solvents) were only partially successful. With increasing enantiopurity, the isolated reaction yield was significantly decreased.

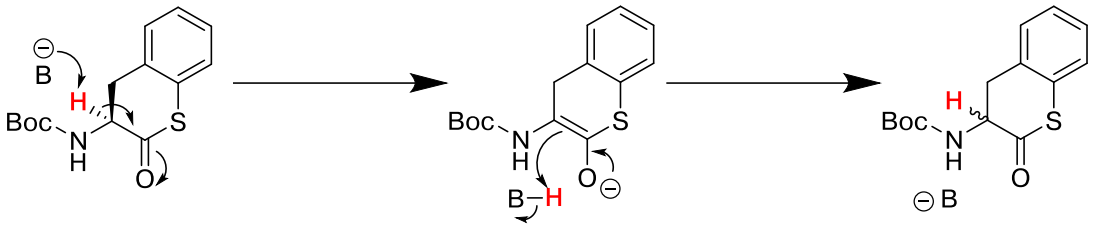


Figure 4.62 Potential racemization mechanism of the Boc-protected-3-amino-δ-thiochromanone
One potential mechanism of enolization at the Cα resulting in racemization of the thiochromanone product. The Hα can be abstracted from the thiochromanone, or a mixed anhydride intermediate (Figure 4.20).[495] Alternatively, racemization can proceed via oxazolone formation.[496] This enolate is stabilized via extended conjugation with the aromatic ring.

In order to avoid racemization, the carboxylic acid of Boc-2-iodophenylalanine was protected as a methyl ester, to prevent formation of the δ-thiochromanone or the mixed anhydride species. Boc-2-iodophenylalanine-methyl ester was subjected to the copper-catalyzed cross-coupling conditions, and followed by thiolysis and oxidation conditions in the presence of *tert*-butyl mercaptan to generate a one-pot synthesis of Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine-methyl ester. The Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine-methyl ester product was subjected to hydrolysis with lithium hydroxide to cleanly generate Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine in 3 steps in 16% overall yield with retention of enantiopurity. The synthetic sequence to generate the protected amino acid was considerably shorter than the syntheses to produce β-mercaptophenylalanine or β-selenol-phenylalanine (7-9 steps).[401, 402, 493] The protected amino acid Boc-2-thiophenylalanine was incorporated into the model peptide using standard solid-phase peptide synthesis. The model peptide (2-SH-Phe)RAFS-NH_2, synthesized using these two different methodologies, were shown to be identical by NMR.

Two different approaches to synthesize peptides containing N-terminal 2-thiophenylalanine were established, both utilizing copper-mediated cross-coupling reactions on 2-iodophenylalanine. With practical syntheses in hand, the reactive properties of peptides containing 2-thiophenylalanine were characterized. The pK_a of 2-thiophenylalanine at the N-terminus of the model peptide was measured to be 5.1, substantially more acidic than 4-thiophenylalanine within a peptide (pK_a = 6.4).[169] The increased acidity of N-terminal 2-thiophenylalanine was potentially due to stabilization via an intramolecular hydrogen bond between the terminal amine and the thiolate (Figure 4.29). The increased acidity of 2-thiophenylalanine suggested that NCL reactions can be conducted under mildly acidic conditions, potentially allowing for selective NCL reactions in the presence of cysteine.

While this current work closely examined the mechanism and applications of NCL reactions mediated by 2-thiophenylalanine, 3-thiophenylalanine could also potentially mediate NCL reactions. 4-Thiophenylalanine and 3-thiophenylalanine were successfully incorporated into the N-terminus of peptides via cross-coupling reaction on solid-phase using the described solid-phase modification approach. In preliminary studies, peptides that contained 3-thiophenylalanine at the N-terminus reacted successfully with the peptide Ac-LYRAG-SPh at pH 7.2 at room temperature, although NMR and HPLC reinjection of the ligation reaction product (observed after 1 hour of reaction time) suggested that the species was the thioester intermediate instead the final ligated product. Peptides containing N-terminal 4-thiophenylalanine were unable to undergo NCL reaction with thioester peptides under comparable conditions. NCL reactions mediated by 3-thiophenylalanine were slower than reactions mediated by 2-thiophenylalanine, and the thioester intermediate species was prone to hydrolysis

rather than the *S*→*N* acyl transfer reaction to form the ligated product peptide. NCL ligation reactions using 3-thiophenylalanine proceed through a sterically unfavorable 8-membered ring intermediate, which potentially slows the overall ligation reaction. Notably, if 3-thiophenylalanine is the second residue from the N-terminus of a peptide, then 3-thiophenylalanine could potentially mediate NCL reaction via an 11-membered ring intermediate (Figure 4.63). Furthermore, if 3-thiophenylalanine can under ligation reactions with peptides containing C-terminal thioesters, then 3-mercaptotyrosine (Chapter 3.4.3.9) could also potentially be utilized for mediating NCL reactions (Figure 4.64). In this manner, 3-mercaptotyrosine can potentially mediate NCL reactions in peptides and proteins, and also act as metal-coordinating ligand or as a structural switch that is sensitive to changes in pH or changes in oxidation state.

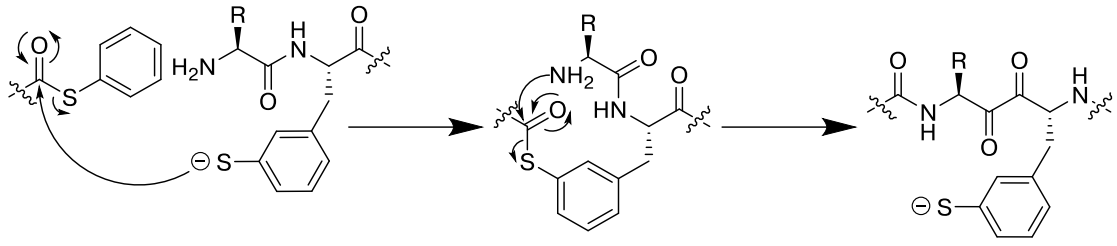


Figure 4.63 NCL reactions mediated by 3-thiophenylalanine via *i,i*–2 acyl transfer
Peptides containing 3-thiophenylalanine at the N-terminus were successfully synthesized and demonstrated to mediate NCL reactions with peptides containing a thioester at the C-terminus. However, 3-thiophenylalanine or 3-mercaptotyrosine could also potentially mediate NCL reactions when positioned two residues from the N-terminus via an 11-membered intermediate.

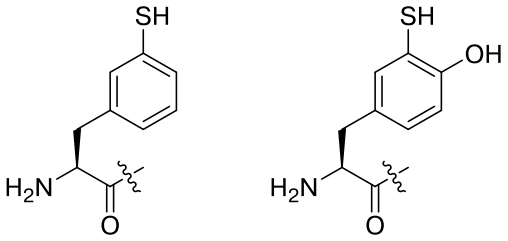


Figure 4.64 Aryl-thiolated amino acids that could potentially be used to mediate NCL reactions
3-Thiophenylalanine was successfully incorporated at the N-terminus of model peptides using the solid-phase cross-coupling reaction approach. 3-thiophenylalanine or 3-mercaptotyrosine could also be utilized for mediating NCL reactions.

4.3.2 Native chemical ligation reactions using 2-thiophenylalanine in model peptides

With practical access to peptides containing 2-thiophenylalanine at the N-terminus, and with the encouraging observation of *S*→*N* acyl transfer reaction between the N-terminal amine and 2-(S-acetyl)thiophenylalanine, the utility of 2-thiophenylalanine was explored for NCL reactions. The model peptide (2-SH-

Phe)RAFS-NH_2 was subjected to ligation reaction conditions with the peptide Ac-LYRAG-SBn (-SBn = S-benzyl thioester) in the presence of thiophenol as an additive, similar to ligation reaction conditions that have been utilized previously.[431, 464, 470] Peptides containing S-benzyl thioesters are more stable against hydrolysis, and are easier to handle and store, but are also less reactive substrates for NCL reactions. Therefore, it is common practice to generate an activated S-phenyl thioester from a more stable thioester group *in situ* using an additive such as thiophenol.[431, 432]

In the ligation reaction between (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SBn, the ligation product only formed after a substantial amount of Ac-LYRAG-SPh had accumulated, suggesting that the rate-determining step was formation of the activated thioester species (Figure 4.30). In order to more directly examine the rate of 2-thiophenylalanine-mediated NCL reactions, the peptide containing the S-phenyl thioester was used directly instead of generating this peptide *in situ*. Using the preactivated peptide Ac-LYRAG-SPh, the ligation reaction with (2-SH-Phe)RAFS-NH_2 was 62% complete in less than 5 minutes at 23 °C with thiophenol as an additive. In contrast, NCL reactions with peptides containing β-mercaptophenylalanine reported by Crich & Banerjee[401] required 16 hours to obtain 72-74% yield, while aryl thiolated auxiliaries[492, 497] required 0.5-12 hours to complete with peptides with S-phenyl thioester peptides.

Using the peptide containing an S-phenyl thioester directly, the thiophenol additive was no longer necessary in the NCL reaction. In the absence of thiophenol, the ligation reaction was >95% complete in 5 minutes at 23 °C between the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh. The increase in overall reaction rate for 2-thiophenylalanine-mediated NCL in the absence of thiophenol was, in retrospect,

not surprising: the reaction between (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SPh reversibly generates the thioester intermediate and thiophenol; excess thiophenol would shift the equilibrium to favor formation of the reactant peptides (Le Chatelier's principle). In essence, the bimolecular reaction between thiophenol and the thioester intermediate competes with the irreversible unimolecular intramolecular *S→N* acyl transfer to form the ligated peptide product (Figure 4.33). Combined, NCL reactions mediated by 2-thiophenylalanine are rapid, proceed under mild conditions, and do not require thiol additives provided that the S-phenyl thioester is used directly as a reaction substrate.

The reactivity of 2-thiophenylalanine in NCL was explored in the context of other peptides containing S-phenyl thioesters, since the residue at the C-terminus of the peptide containing the thioester group can significantly affect the reaction efficiency.[431, 445, 470] For example, proline thioester peptides are known to be among the slowest ligation reactions, due to an $n\rightarrow\pi^*$ interaction in trans-proline that blocks the thioester carbonyl from reaction with thiolates.[445] β-branched and sterically bulky amino acids at the C-terminus of peptides containing thioesters are also known to slow ligation reactions.[432, 473] The scope of peptides Ac-LYRAZ-SPh was examined for ligation reactions with the peptide (2-SH-Phe)RAFS-NH_2 (Table 4.1). It was found that all ligation reactions proceeded to excellent conversion in less than 12 hours. Ligation reactions with peptides containing valine, leucine, and phenylalanine at the C-terminus were all nearly complete in one hour at 23 °C. The slowest ligation reaction was between the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAP-SPh, which was still 84% complete in 12 hours. Considering the ligation reaction conditions, with only 1.0-2.0 mM peptide concentrations, no thiol additives, and the steric bulk of 2-

thiophenylalanine, these ligation reactions proceeded extremely efficiently using mild reaction conditions.

It is difficult to directly compare the ligation reactions using the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAZ-SPh to literature descriptions, because many ligation reaction studies much higher peptide concentrations (1-11 mM), and many use different thioester leaving groups (few examples use S-phenyl thioesters directly).[401, 431, 470, 472] A close comparison to the ligation reaction conditions utilized here involves the peptides Ac-LYRANX-SPh with (2-thiol-tryptophan)SPGYS-NH_2, using ligation reaction conditions at pH 6.6 at 37 °C (4 mM thiol-tryptophan peptide concentration).[472] The most efficient ligation reaction in this work utilized the peptide Ac-LYRANA-SPh with 85% isolated yield after 24 hours, while the slowest ligation reaction was with the peptide Ac-LYRANP-SPh with 58% isolated yield in 30 hours.[472] By comparison, 2-thiophenylalanine-mediated ligation reactions were remarkably fast, using lower reaction temperature and lower peptide concentrations.

Without a direct comparison to ligation reaction conditions via literature examples, it was difficult to determine if ligation reactions mediated by 2-thiophenylalanine were comparable to those at cysteine. Cysteine is not sterically hindered, and proceeds through a 5-membered *S*→*N* acyl transfer intermediate, but has a higher pK_a than 2-thiophenylalanine. However, given the increased acidity of 2-thiophenylalanine, a larger mole fraction of the aryl thiolated amino acid is anionic at pH 7.2 (and nucleophlic), allowing for potentially more efficient reaction under neutral (or even acidic) reaction conditions. It was hypothesized that cysteine and 2-thiophenylalanine-mediated NCL reactions would be comparable: 2-thiophenylalanine-mediated ligation reactions could be slowed by the less favorable 7-

membered ring intermediate,[432] but a larger mole-fraction of 2-thiophenylalanine is anionic (and reactive) at pH 7.2 compared to cysteine. In contrast, 2*S*,4*R*-mercaptoproline-mediated NCL reactions were anticipated to be much slower than either cysteine or 2-thiophenylalanine, due to the conformational restriction on this cyclic secondary amine. Shang *et al.*[454] demonstrated that 2*S*,4*R*-mercaptoproline can mediate NCL reactions, with a ligation reaction with an glycinyl-O-phenyl ester proceeding to 85% completion in 2.5 hours. We sought to compare the ligation reactions mediated by 2-thiophenylalanine against peptides containing cysteine and 2*S*,4*R*-mercaptoproline, and the key differences between these ligation reactions are shown in Figure 4.65.

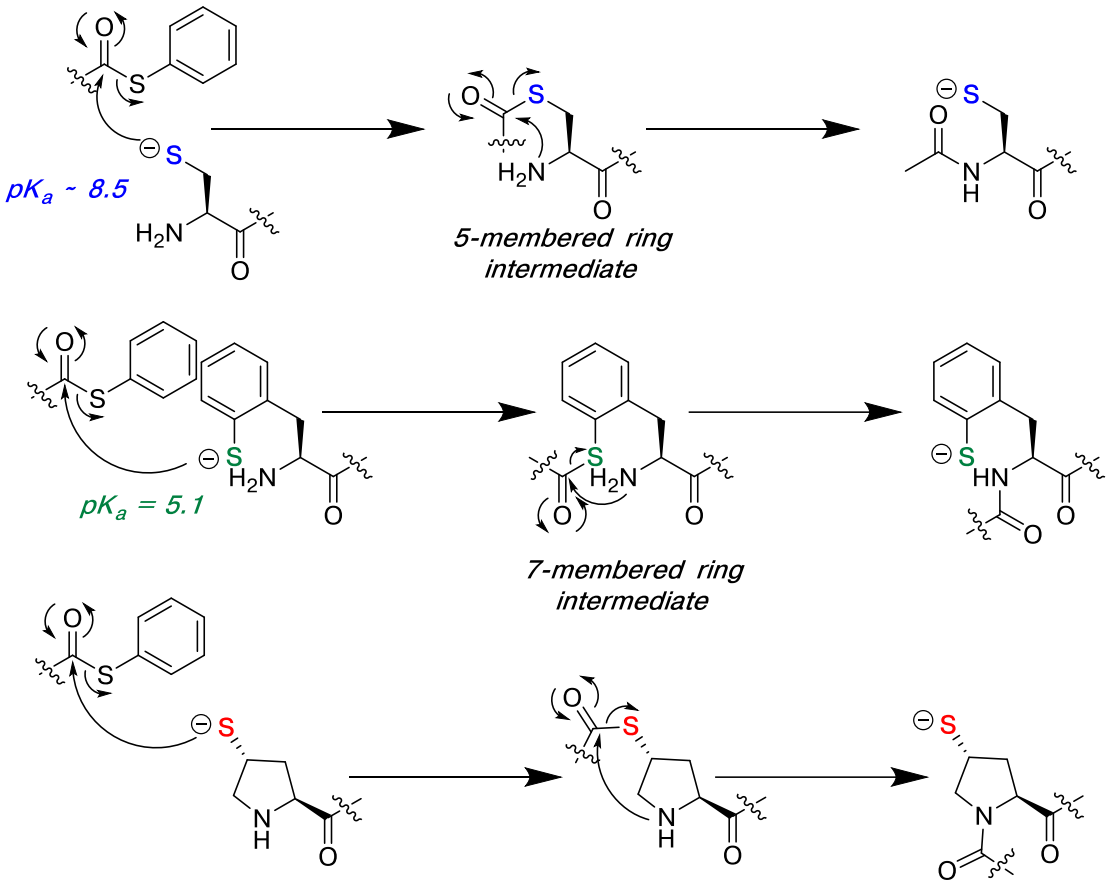


Figure 4.65 Comparison of NCL reactions using different thiolated amino acids Reactivity for ligation reaction with the peptide Ac-LYRAA-SPh were compared in peptides containing either 2-thiophenylalanine, cysteine, or 2*S*,4*R*-mercaptoproline. In consideration of differences in acidity, steric hindrance, and ring size of the intermediate to form the amide bond, it was anticipated that 2-thiophenylalanine and cysteine would exhibit greater reactivity than 2*S*,4*R*-mercaptoproline for ligation reaction, and 2-thiophenylalanine would be more reactive than cysteine at low pH.

At pH 7.2, it was found that the ligation reactions with the peptides Ac-LYRAA-SPh and XRAFS-NH_2 proceeded efficiently in the absence of thiol additives at room temperature, with relative reactivity as cysteine > 2*S*,4*R*-mercaptoproline > 2-thiolphenylalanine. All of these reactions with XRAFS-NH_2 at pH 7.2 were nearly complete in 1 hour, with the reaction with the peptide CRAFS-NH_2 complete in 5

minutes. In order to identify the key differences in reactivity between the different thiolated amino acids, the ligation reactions were modified to more acidic conditions, at pH 4 and 5. Using mildly acidic ligation reaction conditions, the efficiency of ligation with peptides containing cysteine and 2-thiophenylalanine reactions were comparable (Figure 4.34). As anticipated, the peptide containing 2*S*,4*R*-mercaptoproline exhibited lower reactivity compared to cysteine or 2-thiophenylalanine, via reactions with the peptide Ac-LYRAA-SPh at pH 4 and 5 compared. However, in ligation reactions with the peptide containing 2*S*,4*R*-mercaptoproline, it was observed that the peptide did not completely rearrange to generate the ligated product with an amide bond, and there was a substantial amount of thioester intermediate present after 3 hours (at both pH 4 and 5). These observations with 2*S*,4*R*-mercaptoproline suggests that the steric hindrance of a secondary amine may slow the amide bond formation (relevant to auxiliary-mediated NCL[492]), and the steric hindrance may slow the initial reaction with the peptide containing an S-phenyl thioester. It was anticipated that 2-thiophenylalanine ligation reactions would proceed more efficiently than cysteine at low pH, given a greater mole fraction of the thiolate pH 5.0 and 4.0 with the aryl thiolated amino acid. However, peptides containing 2-thiophenylalanine and cysteine exhibited similar reactivity with the peptide Ac-LYRAA-SPh under acidic conditions. Potentially the greater reactivity of 2-thiophenylalanine at lower pH is offset by the less-favorable 7-membered ring transition state to form the amide bond (Figure 4.65). Indeed, trace amounts of unrearranged thioester intermediate at pH 4.0 were observed in ligation reactions mediated with 2-thiophenylalanine, but not with cysteine.

During ligation reactions that require a longer reaction duration, peptides containing thioester groups can be prone to epimerization.[472] In order to establish that epimerization did not occur during the ligation reactions, the D-amino acid variants of the peptide products of the ligation reaction were synthesized. Two peptides were synthesized Ac-LYRAZ(2-SH-Phe)RAFS-NH_2 (X = D-Val or D-Phe) by subjecting the peptides containing 2-iodophenylalanine to the cross-coupling reaction conditions. By comparison of the D-amino acid variants to the products resulting from the ligation reaction, via HPLC coinjection and by NMR, it was established that no significant epimerization occurred during the ligation reactions (Val ligation reaction, 6 hours; Phe ligation reaction, 1 hour). Notably, successful cross-coupling reactions in these cases generated internal 2-thiophenylalanine residues, and indicates that the amide cleavage reaction observed in the trp cage and cVHP peptides was not inherent to internal 2-thiophenylalanine.

2-Thiophenylalanine was demonstrated to result in rapid, efficient native chemical ligation reactions with a variety of model peptides containing S-phenyl thioesters. In initial ligation reactions with the peptides (2-SH-Phe)RAFS-NH_2 and Ac-LYRAG-SBn, the rate-determining step in the ligation reaction was forming the activated S-phenyl thioester. Thiophenol as an additive in the NCL reaction *hindered* the overall rate of ligation, rather than accelerated ligation efficiency.[431] NCL reactions with 2-thiophenylalanine were comparable to cysteine-mediated ligation reactions, even under mildly acidic conditions. Although the thioester intermediate of 2-thiophenylalanine proceeds via a less-favorable 7-membered ring intermediate to form the amide bond, the increased acidity (and greater mole fraction of reactive thiolate) may compensate for a somewhat slower *S*→*N* acyl transfer reaction (Figure

4.65). 2-Thiophenylalanine-mediated ligation reactions were rapid with a variety of peptides containing S-phenyl thioesters, where even peptides with sterically hindered or hydrophobic amino acids at the C-terminus were mostly complete in 6-12 hours. These insights into the mechanisms of NCL reactions were obtained through practical access to peptides containing 2-thiophenylalanine, via cross-coupling reaction on solid-phase or via SPPS with Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine.

4.3.3 Desulfurization of peptides containing 2-thiophenylalanine to generate phenylalanine at the ligation site

The ability to desulfurize and deselenize amino acids was a major expansion on NCL reactions, as cysteine was no longer a requirement at the ligation site.[400, 448] Thiolated amino acids could be used to mediate the NCL reaction, and the canonical amino acid would be generated at the ligation site following desulfurization.[400] Several protocols have been developed and utilized for desulfurization reactions on peptides, including Raney nickel, Pd/Al_2O_3, nickel boride, or radical initiated approaches (VA-044).[399, 400, 446, 448] In order to generate phenylalanine at the site of the ligation reaction, the thiol functional "handle" on 2-thiophenylalanine must be removed via desulfurization reaction. For selected model peptides, the ligated peptide products Ac-LYRAZ(2-SH-Phe)RAFS-NH_2 (Z = Gly, Ala, Leu) were subjected to desulfurization using nickel boride, generated from $NiCl_2$ and $NaBH_4$.[400] The desulfurization reactions cleanly converted 2-thiophenylalanine to phenylalanine, although the isolated yields were low. Thiols adsorb to the solid metal catalysts used for desulfurization, and the adsorption can cause peptide or protein aggregation and potential loss of products.[479] Adsorption to the metal surface may be enhanced due to the aromatic nature of 2-thiophenylalanine, where the aromatic ring may interact favorably with the nickel

surface, causing peptide aggregation on the catalyst. Attempts to desulfurize 2-thiophenylalanine using radical-initiated methods were unsuccessful, resulting in a mixture of products. Potentially, the radical initiated approach to desulfurization is not compatible peptides containing aryl thiolates, although this approach was not fully explored. Malins *et al.*[472] also reported difficulty with desulfurization of an aryl thiolated tryptophan derivative, and Pd/Al_2O_3 was used for their desulfurization reactions. Potentially, a palladium-mediated approach to desulfurization would improve on the isolated yield of the final peptide products.

4.3.4 2-Thiophenylalanine-mediated ligation for the chemical synthesis of proteins

Having established that peptides containing 2-thiophenylalanine can efficiently react with several model thioester peptides, and that it could be subsequently desulfurized with nickel boride to generate phenylalanine at the ligation site, we sought to demonstrate the utility of 2-thiophenylalanine for NCL reactions within larger peptides. With consideration of protein folding, solubility, and the hydrophobic effect, proteins can be more challenging for ligation reactions than model peptides.

The trp cage miniprotein was selected as a model protein for 2-thiophenylalanine-mediated ligation reactions, because it has been previously well characterized and studied.[94, 95] The trp cage miniprotein, comprised of an α-helix folded against a polyproline helix, is a highly stabilized 20-residue fragment of exendin-4, a therapeutic peptide for treatment of type II diabetes isolated from Gila Monster saliva.[94, 95] This miniprotein has been the subject of numerous protein folding studies,[95, 96, 103, 104, 482, 498] In addition, the trp cage had been previously utilized to

establish the copper-mediated cross-coupling conditions for peptides containing 4-thiophenylalanine (Chapter 1.2.4).[169]

The peptide trp-cage$_{3\text{-}20}$ was synthesized containing N-terminal 2-iodophenylalanine, and this peptide was subjected to solid phase copper-mediated cross-coupling reaction conditions with thiolacetic acid, followed by thiolysis reaction conditions on solid-phase.[169] This approach generated an alternative species as the major product, which was consistent with the peptide Ac-trp-cage$_{4\text{-}20}$ (identified via ESI-MS and NMR), which was not observed in the reactant peptide trp-cage$_{3\text{-}20}$ containing 2-iodophenyalanine (Figure 4.47). The product Ac-trp-cage$_{4\text{-}20}$ potentially resulted from intramolecular *S*→*N* acyl transfer between 2-S(acetyl)thiophenylalanine and a deprotected side chain in the trp cage miniprotein. The exposed thiolate in 2-thiophenylalanine potentially reacted with its own amide carbonyl at the high reaction temperatures of the cross-coupling reaction, releasing a δ-thiochromanone into solution. This product Ac-trp-cage$_{4\text{-}20}$ was avoided by re-optimization of the cross-coupling reaction conditions to prevent side-chain deprotection and subsequent thiolate side reaction. Having established optimized reaction conditions to generate the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine at the N-terminus, this resultant peptide was subjected to native chemical ligation reaction conditions with the peptide trp-cage$_{1\text{-}2}$–SPh. The ligation reaction mediated by 2-thiophenyalanine efficiently formed the Y3(2-SH-Phe) trp cage in 77% conversion in 3 hours. Similar to the ligation reactions in the model peptides, the ligation reaction to generate the trp cage variant contained no thiol additives, although 6 M guanidinium was necessary to solubilize the reactant peptides. The ligated trp cage miniprotein was subsequently desulfurized using nickel boride to generate the Y3F trp cage miniprotein variant, as

confirmed via ESI-MS, in low isolated yield due to problems in desulfurization (consistent with the observations in the model peptide).

Although 2-thiophenylalanine was successfully utilized to synthesize a Y3F variant of the trp cage miniprotein, the ligation reaction utilized a peptide containing a thioester group that was only two residues in length, which is not representative of a protein target for ligation reactions. However, in synthesizing the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine at the N-terminus, we had discovered potential side-chain reactivity, and had modified the conditions for the solid-phase cross-coupling reaction. We synthesized another well-characterized and studied protein via 2-thiophenylalanine-mediated NCL, the chicken villin headpiece fragment (cVHP-35).[484] The villin headpiece is a 76-residue actin-binding domain, and the 35-residue subdomain at the C-terminus is a highly stabilized 3-helix miniprotein that is stabilized by three phenylalanines (Figure 4.54).[484, 488, 489] Similar to the trp cage, cVHP-35 has been the subject of many protein folding studies, given its exceptional stability for its short sequence.[484-487, 490] The three aromatic residues are well conserved, and are crucial in the stability of the villin headpiece.[489] Therefore, it is critical that these residues, and nearby hydrophobic residues, are not replaced with other thiolated amino acids, and that NCL reactions are mediated by an aryl thiolated amino acid that is synthesized by practical means.

The ligation site selected for the cVHP-35 was Phe17, which essentially divided the cVHP-35 in two parts: the peptide cVHP$_{1\text{-}16}$–SPh and the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus. The peptide cVHP$_{1\text{-}16}$–SPh contained two of the three crucial phenylalanine residues that stabilize the fold of the cVHP-35 subdomain, and an extension of this peptide can be expressed (via inteins) to

include the actin-binding domain (cVHP-67).[488] The peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine contains the third phenylalanine that is crucial for villin headpiece stability at the N-terminus; the C-terminal portion of this peptide contains an α-helix that may also play a role in folding the villin subdomain.[490] By using Phe17 for the ligation site, modifications on the C-terminal α-helix could be readily incorporated via SPPS, and then the modified peptide could be subjected to ligation reaction with a peptide containing a C-terminal thioester, which could include the full-length actin binding domain. For NCL reactions using 2-thiophenylalanine, any of the three phenylalanine residues in cVHP-35 could be utilized for the ligation site.

The cross-coupling reaction conditions that had been optimized for the peptide $trp\text{-}cage_{3\text{-}20}$ were utilized for synthesizing the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine on solid-phase. While conversion to generate the peptide containing 2-thiophenylalanine was excellent, purification via HPLC was difficult due to co-elution with the starting material peptide containing 2-iodophenylalanine (Figure 4.55). In addition to the cross-coupling reaction on solid-phase, an alternative synthetic strategy was utilized for generating the $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine, via SPPS with the synthesized peptide Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine (Chapter 4.2.4). Both of the peptides $cVHP_{17\text{-}35}$, containing 2-thiophenylalanine synthesized by two different methods (cross-coupling reaction on solid-phase or SPPS with the protected amino acid), were compared via NMR, and were established to be identical peptides (Figures 4.57, 4.58). The purified peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine was subjected to native chemical ligation reaction conditions with the peptide $cVHP_{1\text{-}16}$–SPh, and the ligation reaction proceeded smoothly to 95% conversion after 3 hours at 23 °C. With two different

synthetic approaches to generate the villin headpiece protein containing 2-thiophenylalanine, this aryl thiolated amino acid with unique reactivity should be broadly applied for NCL reactions in protein and peptide targets.

Herein, we have demonstrated the broad and efficient applicability of 2-thiophenylalanine as means to ligate model peptides and proteins. The peptides containing 2-thiophenylalanine were synthesized using two different strategies from commercially available 2-iodophenylalanine: via cross-coupling reaction on peptides on solid-phase, or via SPPS with the protected amino acid (3 steps from Boc-2-iodophenylalanine). Practical access to this aryl thiolated amino acid provided unique insights into the mechanism of NCL reactions, particularly in the context of more conventional alkyl thiolated amino acids. The fact that thiophenol as an additive *slows* the overall ligation reaction rate is against the dominant paradigm regarding NCL reactions,[395, 431, 432, 479] but an aryl thiolated amino acid such as 2-thiophenylalanine is not a conventional NCL reaction. Furthermore, an aryl thiolated amino acid such as 2-thiophenylalanine could potentially participate in unique interactions that enhance NCL reactions. In prior studies, aromatic amino acids and methionine at the C-terminus of peptides containing S-benzyl thioesters exhibited moderate reactivity for NCL reactions with cysteine,[470] potentially as a result of favorable interactions between the cysteine thiolate and the C-terminal amino acid.

Aromatic amino acids play important roles in biological recognition,[18, 161] and thiophenylalanine-mediated ligation approaches can be used to integrate this unique sulfur analogue of tyrosine into peptides and proteins. The scope of NCL reactions with 2-thiophenylalanine and different peptides Ac-LYRAZ-SPh were used as a proof-of-principle, but thiophenylalanine was demonstrated in the previous chapters to

have unique ability to participate in noncovalent aromatic interactions. These noncovalent aromatic interactions can potentially be used to accelerate ligation reaction rates in the context of 2-thiophenylalanine-mediated NCL reactions using designed peptides. For example, template-assisted ligation reactions rely on favorable interactions between a peptide "template" and the two peptide substrates for the ligation reaction.[499, 500] 2-Thiophenylalanine represents a unique amino acid that can both participate in noncovalent aromatic interactions and act as a mediator for NCL reactions. Furthermore, 3-thiophenylalanine or 3-mercaptotyrosine can be synthesized using synthetic methods described in Chapter 3, and these aryl thiolated amino acids can potentially be used for NCL reactions as well. The thiol group in 3-thiophenylalanine or 3-mercaptotyrosine is appropriately positioned to proceed through an 8-membered ring transition state in *i*,*i*–1 NCL reactions, or through an 11-membered ring transition state in *i*,*i*–2 NCL reactions. In this manner, aryl thiolated aromatic amino acids can be incorporated into peptides or proteins using NCL reactions, allowing for practical synthesis of biomolecules with unique reactive or spectroscopic properties, with the ability to coordinate metals and catalyze redox reactions, or with the ability to participate in noncovalent aromatic interactions.

4.4 Experimental

4.4.1 Materials

Fmoc-L-amino acids, Boc-2-L-iodophenylalanine, Boc-2-D-iodophenylalanine, and Boc-L-*trans*-hydroxyproline were purchased from Novabiochem (San Diego, CA), Bachem (San Carlos, CA), or Chem-Impex (Wood Dale, IL). Sulfamylbutyryl AM resin was purchased from Novabiochem. Rink amide MBHA resin, benzotriazol-1-yl-

oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP), O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU), and diisopropylethylamine (DIPEA) were purchased from Chem-Impex. Acetic anhydride (Ac_2O), diisopropylcarbodiimide (DIC), nickel(II) chloride hexahydrate, iodoacetonitrile, thiophenol, thiolacetic acid, 2-aminoethanethiol, trifluoroacetic acid (TFA), phenol, thioanisole, copper(I) iodide, 1,10-phenanthroline, lithium hydroxide, *tert*-butyl mercaptan, 2,2′-dipyridyldisulfide, dimethyl sulfate, and triisopropylsilane (TIS) were purchased from Acros. Ethanedithiol (EDT) was purchased from Pfaltz & Bauer (Waterbury, CT). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Hampton Research (Aliso Viejo, CA). O-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) was purchased from Senn Chemicals (San Diego, CA). Acetonitrile (MeCN), dimethylformamide (DMF), methylene chloride (CH_2Cl_2), methanol (MeOH), tetrahydrofuran (THF), ether, toluene, sodium borohydride, potassium carbonate, 1,4-dioxane, and pyridine were purchased from Fisher. Deionized water was purified by a Millipore Synergy 185 water purification system with a Simpak2 cartridge. Solid-phase post-synthetic modification reactions were performed in capped disposable fritted columns (Image Molding) or in glass vials (2 mL). All materials were used as purchased with no additional purification.

4.4.2 Peptide Synthesis and Characterization

Boc-(2-iodophenylalanine)RAFS, Boc-(2-iodo-D-phenylalanine)RAFS, Boc-CRAFS, and Boc-(4*R*-hydroxyproline)RAFS peptides (0.1 or 0.25 mmol) peptides were synthesized manually on Rink amide resin (0.9 mmol/g loading capacity) via standard Fmoc solid-phase peptide synthesis using HBTU as a coupling reagent. The

peptides trp-cage$_{3\text{-}20}$ and cVHP$_{17\text{-}35}$ containing N-terminal Boc-(2-iodophenylalanine) were synthesized on a Rainin PS3 peptide synthesizer (Protein Technologies, Tucson, AZ) on low-loading Rink amide resin (0.33 mmol/g loading capacity) via standard Fmoc solid-phase peptide synthesis with HBTU as a coupling reagent. The amino acid coupling reactions used 4 equivalents of Fmoc amino acid and HBTU for 60 minutes at room temperature. For coupling reactions using Boc-2-iodophenylalanine, 3 equivalents of amino acid were used. Peptides contained C-terminal amides. The peptides were subjected to cleavage from the resin and deprotection reaction for 2-4 hours under standard conditions (90% TFA/5% TIS/5% thioanisole). TFA was removed by evaporation. The peptides were precipitated with cold ether, the solvent was removed, and the crude precipitates were dried.

The peptides Ac-LYRAZ-SPh (0.25 mmol or 0.15 mmol, where Z = Gly, Ala, Val, Leu, Phe, or Pro) and the trp-cage$_{1\text{-}2}$–SPh peptide were synthesized manually on 4-sulfamylbutyryl Rink Amide AM resin (1.1 mmol/g maximum loading capacity, Novabiochem) via Fmoc solid-phase peptide synthesis. For the first amino acid, the resin was stirred with Fmoc-Z-OH (3 eq.) and DIPEA (9 eq.) in DMF (3 mL) at –20 °C for 30 minutes. PyBOP (4 eq.) was added to the mixture and the mixture was allowed to stir for 24 h while warming to room temperature.[501] The reagents were removed from the resin, and the coupling reaction with the first amino acid was repeated to ensure sufficient loading of the first amino acid on the resin. For the remaining amino acids, the Fmoc-Xaa-OH or Boc-Xaa-OH (5 eq.), HOBt (10 eq.), and DIC (10 eq.) in DMF (3 mL) were preavtivated at room temperature for 30 minutes.[501] The solution containing the activated amino acid was then added to the resin, and the mixture was stirred for 30 minutes at room temperature.[1] Following peptide synthesis,

the peptides on resin were subjected to a solution containing 5% acetic anhydride and 6% 2,6-lutidine in DMF (2 mL) for 2 minutes at room temperature, in order to incorporate an N-terminal acetyl group.[502] The reagents were removed from resin, and the resin was rinsed with CH_2Cl_2 and then dried with ether. C-terminal thioesters were incorporated by cleaving the peptides from 4-sulfamylbutyryl resin,[501] by activating the resin linker with iodoacetonitrile (136 μL, 1.9 mmol, freshly filtered over basic alumina) and DIPEA (136 μL, 0.82 mmol) in NMP (3 mL) for 24 h at room temperature while protected from light on a rotary shaker. The resin containing the activated peptide was then rinsed with NMP, CH_2Cl_2, and THF. To liberate the protected peptide from the resin with the corresponding thioester, the resin containing the activated peptide was mixed with thiophenol (400 μL, 4 mmol) in THF (3 mL) at room temperature for 24 h; benzyl mercaptan was used in place of thiophenol to generate the peptide Ac-LYRAG-SBn for initial experiments. The resin was filtered, and the solution was collected. The resin was washed twice with CH_2Cl_2, and the combined filtrates were dried under reduced pressure. The protected peptides containing C-terminal thioesters were subjected to deprotection reaction for 3 h under standard conditions (90% TFA/5% TIS/5% thioanisole). TFA was removed by evaporation and the peptides were precipitated using ether. The precipitate was centrifuged, solvent was removed, and the crude peptide was dried.

The peptide $cVHP_{17\text{-}35}$ containing a C-terminal thioester was synthesized on a Rainin PS3 peptide synthesizer (Protein Technologies, Tucson, AZ) on 2-chlorotrityl resin (1.51 mmol/g loading capacity) via standard Fmoc solid-phase peptide synthesis with HBTU as a coupling reagent. For the first amino acid, the resin was stirred with Fmoc-Ala-OH (1.5 eq.) and DIPEA (9 eq.) in CH_2Cl_2 (3 mL) at room temperature for

1 hour. The reagents were removed from resin, and the resin was subjected to the coupling reaction with the first amino acid again, in order to ensure sufficient loading of the first amino acid on the resin. The resin was washed with CH_2Cl_2 (3 mL) and then transferred to an automated synthesizer for the remainder of the sequence. The resin containing the protected peptide was subjected to cleavage reaction conditions using a solution of CH_2Cl_2:AcOH:TFE (3 mL, 8:1:1 v/v/v).[503] The mixture was agitated on a rotary shaker at room temperature for 20 minutes, and the filtrate was collected. The resin was subjected to cleavage reaction conditions and filtrates were collected, for a total of three cleavage reactions on the resin. The combined filtrates were neutralized with a solution of saturated $NaHCO_3$ (10 mL), and the protected peptide was extracted into CH_2Cl_2 (10 mL × 3). The combined organic layers were dried thoroughly under reduced pressure to remove any residual water. The crude, protected peptide was dissolved in 10% DIPEA in CH_2Cl_2 (1 mL). PyBOP (4 eq.) and thiophenol (400 μL) were added, and the mixture was stirred at room temperature for 18 h. The solvent was removed under reduced pressure, and the crude peptide was subjected to deprotection reaction conditions for 4 h under standard conditions (90% TFA/5% TIS/5% H_2O). TFA was removed by evaporation and the peptide was precipitated using cold ether. The precipitated peptide was centrifuged, the solvent was removed, and the precipitate was dried.

The crude peptides were dissolved in aqueous phosphate buffer and then filtered using a 0.45 μm syringe filter. The peptides were purified using reverse phase HPLC on a Vydac C18 semi-preparative column (250 × 10 mm, 5-10 μm particle, 300 Å pore) or on a Varian Microsorb MV C18 analytical column (250 × 4.6 mm, 3-5 μm particle, 100 Å pore) using a linear gradient of buffer B (20% water, 80% MeCN,

0.05% TFA) in buffer A (98% water, 2% MeCN, 0.06% TFA), unless otherwise indicated. Peptide purity was verified via reinjection on an analytical HPLC column. Peptides were characterized by ESI-MS (positive ion mode) or an LCQ Advantage (Finnigan) mass spectrometer. Purification conditions for all peptides are shown in Table 4.-4.4.

Thioester peptides Ac-LYRAZ-SPh were quantified via UV after establishing a molar extinction coefficient for Ac-LYRAG-SPh (*vide infra*, ε_{280} = 1,355 M^{-1} cm^{-1}, Figure 4.66). Peptides containing 2-thiophenylalanine-, cysteine-, and 4*R*-mercaptoproline were quantified via NMR using 4-hydroxybenzaldehyde (0.1 mM) as an internal standard. Errors in peptide concentration are estimated to be ±5%.

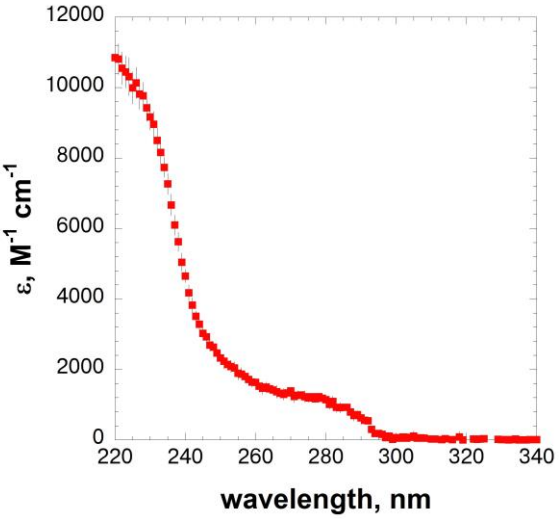


Figure 4.66 UV Absorbance Spectrum of Ac-LYRAG-SPh
Peptide concentration was established via ^{1}H NMR using 4-hydroxybenzaldehyde as an internal standard (0.1 mM). Data represents averages of at least 3 independent trials, and error bars indicate standard error of the mean.

Table 4.2. Purification and ESI-MS of canonical peptides and canonical peptides containing thioesters.

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-LYRAG-SPh	60 minutes 15-65% buffer B	35.4	712.3	713.5 $(M+H)^+$
Ac-LYRAA-SPh	60 minutes 15-65% buffer B	35.9	726.4	727.4 $(M+H)^+$
Ac-LYRAV-SPh	60 minutes 15-65% buffer B	43.1	754.4	755.4 $(M+H)^+$
Ac-LYRAL-SPh	60 minutes 15-65% buffer B	48.1	768.4	769.4 $(M+H)^+$
Ac-LYRAF-SPh	60 minutes 15-65% buffer B	50.6	802.4	803.4 $(M+H)^+$
Ac-LYRAP-SPh	60 minutes 15-65% buffer B	42.3	752.4	753.4 $(M+H)^+$
CRAFS-NH_2	60 minutes 0-45% buffer B	21.4	581.3	582.3 $(M+H)^+$
Ac-NL-SPh trp-$cage_{1-2}$–SPh	60 minutes 15-65% buffer B	33.7	379.2	402.1 $(M+Na)^+$
LSDEDFKAVFGMTRSA-SPh $cVHP_{1-16}$–SPh	60 minutes 15-45% buffer B	52.9	1866.0	933.7 $(M+2H)^{2+}$

Table 4.3. Purification and ESI-MS of peptides containing 2-thiophenylalanine

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
(2-SH-Phe)RAFS-NH_2	60 minutes 0-45% buffer B	38.6	657.3	658.3 $(M+H)^+$
(2-SH-Phe)IQWLKDGGPSSGRPPPS-NH_2 trp-$cage_{3-20}$	60 minutes 0-45% buffer B	53.4	1956.0	979.3 $(M+2H)^{2+}$
(2-SH-Phe)ANLPLWKQQNLKKEKGLF-NH_2 $cVHP_{17-35}$ via cross-coupling reaction	60 minutes 0-60% buffer B	46.0	2333.7	1167.4 $(M+2H)^{2+}$
(2-SH-Phe)ANLPLWKQQNLKKEKGLF-NH_2 $cVHP_{17-35}$ via coupling Boc-2-S(S-*tert*-butyl)-thiophenylalanine	60 minutes 15-75% buffer B	32.3	2333.7	584.5 $(M+4H)^{4+}$

Table 4.4. Purification and ESI-MS of peptides containing modified proline

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
(2*S*,4*S*-Bromoproline)RAFS-NH_2	60 minutes 0-20% buffer B	36.4	653.2	654.1 $(M+H)^+$
(2*S*,4*R*-Mercaptoproline)RAFS-NH_2	60 minutes 0-20% buffer B	34.5	607.3	608.4 $(M+H)^+$

4.4.3 Synthesis of Boc-protected-3-amino-δ-thiochromanone

Boc-2-iodophenylalanine (490 mg, 1.25 mmol), copper(I) iodide (20 mg, 0.11 mmol), and 1,10-phenanthroline (48 mg, 0.27 mmol) were placed in an oven-dried round bottom flask (50 mL) with a stirbar. Toluene (12.4 mL), DIPEA (445 μL, 2.56 mmol) and thiolacetic acid (109 μL, 1.53 mmol) were added sequentially at room temperature, and the mixture was heated to reflux at 110 °C for 12 hours. The reaction was cooled to room temperature, and the solvent was removed under reduced pressure. The crude residue was redissolved in CH_2Cl_2 and purified via column chromatography (0-0.2% methanol in CH_2Cl_2 (v/v)). The product Boc-protected-3-amino-δ-thiochromanone (300 mg, 1.08 mmol) was purified as a white solid in 86% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.30-7.18 (m, 4H), 5.53 (d, J = 4.6 Hz, 1H), 4.44 (m, 1H), 3.34 (dd, J = 14.9, 3.4 Hz, 1H), 3.09 (t, J = 14.1 Hz, 1H), 1.46 (s, 9H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 197.84, 155.05, 133.14, 129.78, 129.39, 127.85, 127.25, 126.41, 80.15, 56.93, 35.97, 28.19. LIFDI-HRMS m/z: $[M]^+$ calcd for $C_{14}H_{17}NO_3S$ 279.0929, found 279.0940.

4.4.4 Synthesis of 2-thiopyridyl-S(S-*tert*-butyl) disulfide

2-Thiopyridyl disulfide (500 mg, 2.27 mmol) was dissolved in THF (9.1 mL). K_2CO_3 (157 mg, 1.14 mmol) and *tert*-butyl mercaptan (640 μL, 5.68 mmol) were added, and the resultant solution was stirred at room temperature for 3 hours. Dilute

HCl (1 M, 20 mL) was added to the reaction mixture, and the product was extracted into ethyl acetate (20 mL × 3). The solution was filtered and the crude residue was diluted in CH_2Cl_2, and purified via column chromatography (CH_2Cl_2). The product 2-thiopyridyl-S(S-*tert*-butyl) disulfide (290 mg, 1.46 mmol, 1.134 g/mL) was purified as a slightly yellow liquid in 32% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.41 (ddd, *J* = 3.2, 1.1, 0.5 Hz, 1H), 7.79 (d, *J* = 8.1 Hz, 1H), 7.61 (dt, *J* = 5.4, 1.2 Hz, 1H), 7.04 (ddd, *J* = 7.4, 4.8, 1.0 Hz, 1H), 1.34 (s, 9H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 161.60, 149.13, 136.77, 120.35, 119.57, 49.21, 29.76. GC-EI-MS m/z: $[M]^+$ calcd for $C_9H_{13}NS_2$ 199.05, found 198, 142, 109, 78, 57. The ^{1}H and ^{13}C NMR spectra corresponded to literature data.[504]

4.4.5 Synthesis of Boc-2-S(S-*tert*-butyl)-thio-D,L-phenylalanine

Boc-protected-3-amino-δ-thiochromanone (300 mg, 1.08 mmol) and 2-thiopyridyl-S(S-*tert*-butyl) disulfide (327 μL, 2.16 mmol) were dissolved in 1,4-dioxane (5.4 mL). Lithium hydroxide (31 mg, 1.3 mmol) was dissolved in water (5.4 mL). The solution containing lithium hydroxide was added to the solution containing the thiochromanone, and the resultant solution was stirred at room temperature for 5 hours. Dilute HCl (1 M, 20 mL) was added to the reaction mixture, and the product was extracted into ethyl acetate (20 mL × 3). The organic layers were combined and the solvent was removed under reduced pressure. The crude residue was redissolved in CH_2Cl_2, and purified via column chromatography (0.5-3% methanol in CH_2Cl_2 (v/v)). The product Boc-2-S(S-*tert*-butyl)-thio-D,L-phenylalanine (205 mg, 0.531 mmol) was purified as a white solid in 49% yield. The NMR spectra reflected a mixture of rotamers about the carbamate bond (1:1 cis:trans). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.84 (d, *J* = 7.9 Hz, 1H), 7.29-7.13 (m, 3H), 6.95 (br d, *J* = 5.4 Hz), 5.22 (d, *J* = 8.1

Hz) (sum of 6.95 ppm and 5.22 ppm resonances, 1H), 4.65 (m, 1H), 3.61 (dd, *J* = 13.8, 3.8 Hz), 3.48 (dd, *J* = 14.3, 4.8 Hz) (sum of 3.61 ppm and 3.48 ppm resonances, 1H), 3.25 (dd, *J* = 14.3, 9.1 Hz), 2.99 (dd, *J* = 13.6, 10.6 Hz) (sum of 3.25 ppm and 2.99 ppm resonances, 1H), 1.38 (s), 1.31 (s), 1.30 (s), 1.15 (s) (sum of 1.38 ppm, 1.31 ppm, 1.30 ppm, and 1.15 ppm resonances, 18H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 176.49, 175.54, 156.51, 155.47, 138.08, 137.92, 134.83, 134.50, 131.09, 130.24, 128.30, 127.58, 127.52, 126.76, 126.29, 81.29, 80.26, 54.41, 54.23, 49.42, 49.24, 37.96, 35.02, 29.87, 28.23, 27.78. ESI-MS m/z: $[M]^+$ calcd for $C_{18}H_{27}NO_4S_2$ 385.14, found 408.1 ($M + Na^+$).

4.4.6 Synthesis of Boc-2-iodo-L-phenylalanine methyl ester

Boc-2-iodo-L-phenylalanine (500 mg, 1.28 mmol) and K_2CO_3 (529 mg, 3.84 mmol) were dissolved in acetone (12.8 mL). Dimethyl sulfate (182 μL, 1.92 mmol) was added dropwise, and the solution was stirred at reflux (60 °C) for 3 hours. After reaction completion was verified via TLC, the mixture was allowed to cool to room temperature. Dilute HCl (0.5 M, 100 mL) was added to the reaction mixture until the pH was below 5, and the product was extracted into CH_2Cl_2 (100 mL × 3). The organic layers were combined and the solvent was removed under reduced pressure. The crude product Boc-2-iodo-L-phenylalanine methyl ester was used without further purification (530 mg, quantitative yield). The NMR spectrum reflected a mixture of rotamers about the carbamate bond (3.6:1 trans:cis). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.82 (d, *J* = 7.8 Hz, 1H), 7.27 (t, *J* = 8.0 Hz, 1H), 7.20 (dd, *J* = 7.6, 1.5 Hz, 1H), 6.92 (t, *J* = 7.3 Hz, 1H), 5.19 (d, *J* = 8.6 Hz), 5.05 (br d, *J* = 6.5 Hz) (sum of 5.19 ppm and 5.05 ppm resonances, 1H), 4.64 (dd, *J* = 14.6, 8.4 Hz), 4.56 (br s) (sum of 4.64 ppm and 4.56 ppm resonances, 1H), 3.74 (s), 3.71 (s) (sum of 3.74 ppm and 3.71 ppm

resonances, 3H), 3.28 (dd, J = 13.9, 5.9 Hz, 1H), 3.09 (dd, J = 13.8, 8.6 Hz), 2.95 (br t, J = 10.8 Hz) (sum of 3.09 ppm and 2.95 ppm resonances, 1H), 1.37 (s), 1.26 (s) (sum of 1.37 ppm and 1.26 ppm resonances, 9H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 172.38, 154.96, 139.63, 139.43, 130.40, 128.71, 128.31, 101.19, 79.84, 53.64, 52.46, 43.10, 28.29. LIFDI-HRMS m/z: $[M]^+$ calcd for $C_{15}H_{20}NO_2I$ 405.0437, found 405.0438.

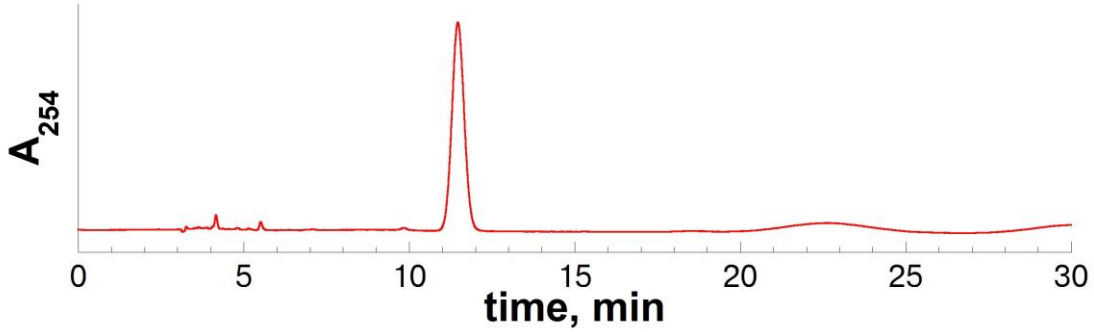


Figure 4.67 Chiral HPLC chromatogram of purified Boc-2-iodo-L-phenylalanine-methyl ester
Analytical chiral HPLC chromatogram (UV detection at 254 nm) of the purified Boc-2-iodo-L-phenylalanine-methyl ester using isocratic 10% isopropanol in hexanes over 30 minutes on a Daicel ChiralPak 1A column (250 x 4.6 mm, 5 µm particle, 1.0 mL/min).

4.4.7 Synthesis of Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine methyl ester

Boc-2-iodo-L-phenylalanine methyl ester (530 mg, 1.31 mmol), copper(I) iodide (63 mg, 0.33 mmol), and 1,10-phenanthroline (117 mg, 0.65 mmol) were placed in an oven-dried glass vial with a stir bar. Toluene (5.23 mL), DIPEA (456 µL, 2.62 mmol), and thiolacetic acid (149 µL, 2.09 mmol) were added sequentially at room temperature. The vial was sealed and placed in an oil bath set to 110 °C, and the mixture was stirred for 18 hours. The mixture was cooled to room temperature, and

the solvent was removed under reduced pressure. The crude residue was redissolved in THF (2.6 mL), and K_2CO_3 (540 mg, 3.93 mmol) and *tert*-butyl mercaptan (442 μL, 3.93 mmol) were added. The vial was sealed and the mixture was stirred at room temperature for 18 hours. Brine (50 mL) was added to the mixture, and the product was extracted using ethyl acetate (50 mL × 3). The organic layers were combined and the solvent was removed under reduced pressure. The crude residue was redissolved in CH_2Cl_2 and purified via column chromatography (0-12% ethyl acetate in hexanes (v/v)). The product Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine methyl ester (110 mg, 0.28 mmol) was generated in 21% yield over two steps (from Boc-2-iodo-L-phenylalanine). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.83 (d, *J* = 7.9 Hz, 1H), 7.25-7.22 (m), 7.16-7.10 (m) (sum of 7.24 ppm and 7.13 ppm resonances, 3H), 5.16 (d, *J* = 8.2 Hz, 1H), 4.68 (dd, *J* = 13.9, 8.1Hz, 1H), 3.74 (s, 3H), 3.37 (dd, *J* = 14.1, 5.5 Hz), 3.20 (dd, *J* = 14.1, 8.3 Hz) (sum of 3.37 ppm and 3.20 ppm resonances, 2H), 1.38 (s), 1.30 (s) (sum of 1.38 ppm and 1.30 ppm resonances, 18H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 172.57, 155.09, 139.67, 138.05, 134.56, 130.23, 128.37, 127.59, 126.63, 79.92, 54.09, 52.41, 49.38, 35.96, 29.91, 28.28. LIFDI-HRMS m/z: $[M]^+$ calcd for $C_{19}H_{29}NO_4S_2$ 399.1538, found 399.1537.

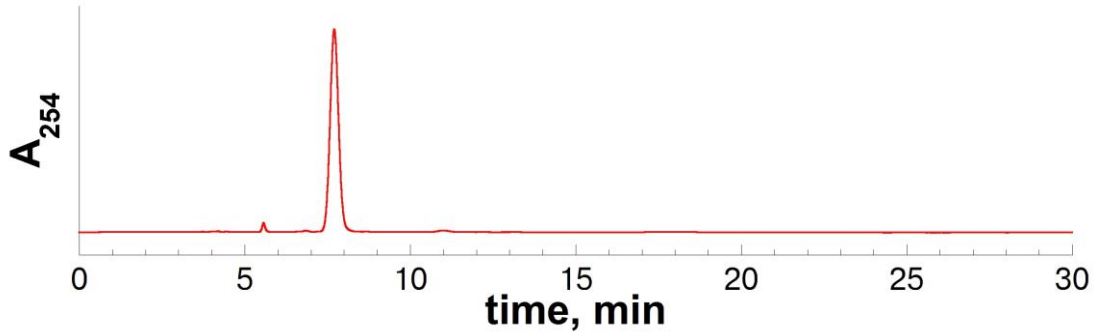


Figure 4.68 Chiral HPLC chromatogram of purified Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine-methyl ester
Analytical chiral HPLC chromatogram (UV detection at 254 nm) of the purified Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine-methyl ester using isocratic 10% isopropanol in hexanes over 30 minutes on a Daicel ChiralPak 1A column (250 x 4.6 mm, 5 µm particle, 1.0 mL/min).

4.4.8 Synthesis of Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine

Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine methyl ester (200 mg, 0.50 mmol) was dissolved in 1,4-dioxane (2.5 mL). Lithium hydroxide (18 mg, 0.75 mmol) was dissolved in water (2.5 mL). The LiOH solution was added to the Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine methyl ester, and the mixture was stirred at 30 °C for 3 hours. After reaction completion, as observed by disappearance of the starting material via TLC, dilute HCl (0.5 M, 50 mL) was added to the reaction mixture until the pH was below 3. The product was extracted into ethyl acetate (50 mL × 3). The organic layers were combined and the solvent was removed under reduced pressure. The crude product was redissolved in CH_2Cl_2 and purified via column chromatography (0-4% methanol in CH_2Cl_2 (v/v)). The product Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine (150 mg, 0.39 mmol) was generated in 78% yield. The NMR spectra reflected a mixture of rotamers about the carbamate bond (1:1.5 trans:cis). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.84 (d, J = 7.9 Hz, 1H), 7.25-7.15 (m, 3H), 7.04 (d, J = 7.7 Hz), 5.20 (d, J = 8.1 Hz) (sum of 7.04 ppm and 5.20 ppm resonances, 1H), 4.70-4.61 (m, 1H), 3.62

(dd, J = 13.9, 3.8 Hz), 3.47 (dd, J = 14.1, 4.5 Hz) (sum of 3.62 ppm and 3.47 ppm resonances, 1H), 3.25 (dd, J = 14.1, 9.0 Hz), 3.01 (dd, J = 13.3, 10.7 Hz) (sum of 3.25 ppm and 3.01 ppm resonances, 1H), 1.38 (s), 1.31 (s), 1.30 (s), 1.14 (s) (sum of 1.37 ppm, 1.31 ppm, 1.30 ppm, and 1.25 ppm resonances, 18H). ^{13}C NMR (100.6 MHz, $CDCl_3$) δ 176.27, 175.49, 156.65, 155.50, 138.20, 138.02, 135.01, 134.61, 131.16, 130.29, 128.44, 127.75, 127.63, 127.55, 126.82, 126.65, 126.36, 81.38, 80.30, 54.52, 54.21, 49.43, 49.26, 38.13, 35.14, 29.94, 28.28, 27.83. LIFDI-HRMS m/z: $[M]^+$ calcd for $C_{18}H_{27}NO_4S_2$ 385.1382, found 385.1400.

4.4.9 Optimized copper-mediated cross-coupling reaction on peptides on solid phase to synthesize peptides containing 2-thiophenylalanine

Resin containing the peptide Boc-(2-iodophenylalanine)RAFS (20-30 mg, 7-10 μmol) was placed in a glass vial with a small stir bar. Using prior conditions,[169] copper(I) iodide (3.8 mg, 20 μmol, 41 mM final concentration), 1,10-phenanthroline (7.2 mg, 40 μmol, 83 mM final concentration), toluene (400 μL), DIPEA (66 μL, 0.4 mmol, 0.83 M final concentration), and thiolacetic acid (17.0 μL, 0.24 mmol, 0.49 M final concentration) were added sequentially. The vial was sealed and the mixture was stirred in an oil bath set to 110 °C for 18 h. The resin was then washed with DMF (4 mL × 4), CH_2Cl_2 (4 mL × 2), and MeOH (4 mL × 2) and dried with ether to produce the peptide containing the protected peptide Boc-(2-S(acetyl)-Phe)RAFS on solid phase.

The resin was then subjected to thiolysis reaction conditions on solid phase in order to avoid the acyl-transfer side products. The resin containing the peptide Boc-(2-S(acetyl)-Phe)RAFS was placed into a fritted disposable tube containing MeOH (3 mL). 2-Aminoethanethiol (34 mg, 0.3 mmol), thiophenol (30.5 μL, 0.3 mmol), and

sodium borohydride (11 mg, 0.3 mmol) were added to the resin, and the mixture was stirred at room temperature for 90 min. The solution was removed, and the reagents were replenished, and the resin was subjected to thiolysis reaction conditions for another 90 minutes at room temperature. The resin was then washed with MeOH and dried with ether. The resin was subjected to cleavage and deprotection reaction under standard TFA conditions. The resultant peptide was precipitated using cold ether. The mixture was centrifuged, the ether was removed, and the precipitate was dried. The precipitate was dissolved in phosphate buffer (pH 7.2, 50 mM) containing DTT (25 mM, for reduction of disulfides) to generate the peptide (2-SH-Phe)RAFS-NH_2.

The peptides containing 2-thio-D-phenylalanine were synthesized in the same manner, using peptides containing 2-iodo-D-phenylalanine in peptides on phase as the reaction substrate.

4.4.10 Synthesis of peptides containing 2-thiophenylalanine via amide coupling reaction with Boc-2-S(S-*tert*-Butyl)-thiol-L-phenylalanine

In addition to synthesizing the peptide (2-SH-Phe)RAFS-NH_2 via modification of the protected peptide on solid-phase, an alternative synthesis was demonstrated via solid-phase peptide synthesis using Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine. The protected peptide Fmoc-RAFS on Rink amide resin (56 mg, 43 μmol) was placed in a disposable, fritted reaction vessel, and was swelled in DMF at room temperature for 20 minutes. The N-terminal amine was subjected to Fmoc-deprotection reaction using a solution of 20% piperidine in DMF (3 mL × 5 min × 3). The reaction solution was removed from the resin, and the resin was washed with DMF (3 mL × 3). The resin was transferred to a 2 mL Eppendorf tube. To a solution of DIPEA in DMF (8% v/v, 1.5 mL), Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine (25 mg, 65 μmol) and HATU

(33 mg, 87 μmol) were dissolved and then incubated at room temperature for 15 minutes in order to pre-activate the Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine. The solution containing the activated of Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine was added to the resin, and the resultant mixture was agitated on a rotary shaker for 2 hours at room temperature. The reaction solution was removed from the resin, and the resin was washed with DMF and CH_2Cl_2, and then dried with ether. The resin was subjected to cleavage and deprotection reaction conditions using 95% TFA/5% H_2O (1.5 mL) at room temperature for 3 hours. TFA was removed via evaporation under nitrogen, and the peptide was precipitated using cold ether. The precipitate was centrifuged, the ether was removed, and then the precipitated was dried. The crude precipitate was dissolved in a solution of DTT (25 mM final concentration) in phosphate buffer (pH 7.2, 50 mM, 100 μL). A similar protocol was utilized for synthesizing the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus.

4.4.11 Solid-phase synthesis of the peptide trp-cage$_{3\text{-}20}$ containing 2-thiophenylalanine

Following a previously described procedure on peptides with an internal 4-iodophenylalanine,[404] the resin containing the peptide Boc-(2-iodophenylalanine)IQWLKDGGPSSGRPPPS (40 mg, 15 μmol) was placed in a glass vial with a small stir bar. Copper(I) iodide (7.6 mg, 40 μmol, 44 μM final concentration), 1,10-phenanthroline (14.4 mg, 80 μmol, 88 μM final concentration), toluene (800 μL), DIPEA (79 μL, 0.45 mmol, 0.49 M final concentration), and thiolacetic acid (34.0 μL, 0.48 mmol, 0.52 M final concentration) were added sequentially. The vial was sealed and the mixture was stirred in an oil bath set to 100 °C for 8 h. The resin was then washed with DMF, CH_2Cl_2, and dried with ether.

Due to the rapid *S*→*N* acyl transfer between the aryl thioacetyl and the unprotected N-terminus in peptides that had been deprotected, a solid-phase deacetylation reaction on the protected peptide on solid phase was required for efficient synthesis of the peptide trp-cage$_{3\text{-}20}$ containing 2-thiolphenylalanine. Following the previously described procedure,[169] the resin was placed into a fritted disposable tube containing MeOH (3 mL) immediately following the cross-coupling reaction and washing steps. 2-Aminoethanethiol hydrochloride (34 mg, 0.3 mmol), thiophenol (30.5 μL, 0.3 mmol), and sodium borohydride (11 mg, 0.3 mmol) were added, and the resultant mixture was stirred at room temperature for 90 min. The reaction solution was removed from the resin, the reagents were replenished, and the reaction to effect thiolysis on solid phase was repeated once more. The reaction solution was removed from the resin, and the resin was then washed with MeOH and dried with ether. The resin was subjected to cleavage and deprotection reactions using standard conditions with TFA. The peptide was precipitated from ether. The mixture was centrifuged, the ether was removed, and the precipitate was dried. The resultant precipitate was dissolved in phosphate buffer (pH 7.2, 50 mM) containing DTT (25 mM, for reduction of disulfides) to generate the peptide trp-cage$_{3\text{-}20}$ containing 2-thiolphenylalanine.

4.4.12 Solid-phase synthesis of the peptide cVHP$_{17\text{-}35}$ containing 2-thiophenylalanine at the N-terminus via cross-coupling reaction on solid phase

Initial reaction attempts using a previously described procedure on peptides containing an internal 4-iodophenylalanine[404] yielded similar results as in the peptide trp-cage$_{3\text{-}20}$ containing N-terminal Boc-2-iodophenylalanine, where the major product observed was consistent (by molecular weight) with a peptide that did not contain the

N-terminal phenylalanine and contained an additional acetyl group (described above). The cross-coupling reaction conditions were optimized, examining the roles of the thiolacetic acid, base, and reaction solvent. Resin with the protected peptide $cVHP_{17\text{-}35}$ containing N-terminal Boc-2-iodophenylalanine (40 mg, 10 μmol) was placed in a glass vial with a small stir bar. Copper(I) iodide (7.6 mg, 40 μmol, 46 μM final concentration), 1,10-phenanthroline (14.4 mg, 80 μmol, 91 μM final concentration), *tert*-amyl alcohol (800 μL), DIPEA (66 μL, 0.40 mmol, 0.46 M final concentration), and thiolacetic acid (11.4 μL, 0.16 mmol, 0.18 M final concentration) were added sequentially. The vial was sealed and the mixture was stirred in an oil bath set to 100 °C for 18 h. The resin was then washed with DMF, CH_2Cl_2, and dried with ether. The protected peptide was subjected to solid-phase deacetylation reaction conditions, which were required for efficient synthesis of the peptide $cVHP_{17\text{-}35}$ containing 2-thiolphenylalanine (due to the rapid *S*→*N* acyl transfer reaction between the aryl thioacetate and the unprotected N-terminus in peptides that had been subjected to TFA cleavage and deprotection reaction conditions). Following the previously described procedure,[404] the resin with the protected peptide $cVHP_{17\text{-}35}$ containing Boc-(2-thioacetyl-phenylalanine) was placed into a fritted disposable tube containing MeOH (3 mL). 2-Aminoethanethiol (34 mg, 0.3 mmol), thiophenol (30.5 μL, 0.3 mmol), and sodium borohydride (11 mg, 0.3 mmol) were added to the resin, and the mixture was stirred at room temperature for 90 min. The solution was removed from the resin, the reagents were replenished, and resin was subjected to the solid-phase deacetylation reaction conditions again. The resin was then washed with MeOH and dried with ether. The resin was subjected to cleavage and deprotection reaction using standard cleavage/deprotection conditions. TFA was removed by evaporation with nitrogen,

and the peptide was precipitated in ether. The mixture was centrifuged, the ether was removed, and the precipitate was dissolved in water containing phosphate buffer (pH 7.2, 50 mM), urea (8 M), and DTT (25 mM, for reduction of disulfides) to generate the peptide $cVHP_{17\text{-}35}$ containing N-terminal 2-thiolphenylalanine. The peptide was purified via HPLC under conditions previously described.

4.4.13 Alternative synthesis of the peptide $cVHP_{17\text{-}35}$ containing 2-thiophenylalanine via coupling Boc-2-S(S-*tert*-butyl)-thio-L-phenylalanine

In order to avoid any potential side-reaction during the solid-phase cross-coupling reaction on the protected peptide, an alternative synthesis of the peptide $cVHP_{17\text{-}35}$ containing N-terminal 2-thiolphenylalanine was utilized, via amide coupling reaction of the synthesized peptide using Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine. In a disposable, fritted reaction vessel, the resin with the protected peptide Fmoc-ANLPLWKQQNLKKEKGLF (130 mg, 43 μmol) was swelled in DMF at room temperature for 20 minutes. The N-terminal amine was subjected to Fmoc-deprotection reaction using a solution of 20% piperidine in DMF (3 mL × 5 min × 3). The reaction solution was removed from the resin, and the resin was washed with DMF (3 mL × 3). The resin was transferred to a 2 mL Eppendorf tube. To a solution of DIPEA in DMF (10% v/v, 1.5 mL), Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine (25 mg, 65 μmol) and HATU (33 mg, 87 μmol) were dissolved and then incubated at room temperature for 15 minutes in order to pre-activate the Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine. The solution containing the activated of Boc-2-S(S-*tert*-butyl)-thiol-L-phenylalanine was added to the resin, and the resultant mixture was agitated on a rotary shaker for 8 hours at room temperature. The reaction solution was removed from the resin, and the resin was washed with DMF and CH_2Cl_2, and then dried with

ether. The resin was subjected to cleavage and deprotection reaction conditions using 85% TFA/5% H_2O/5% thioanisole/5% phenol (2 mL) at room temperature for 4 hours. TFA was removed via evaporation under nitrogen, and the peptide was precipitated using cold ether. The precipitate was centrifuged, the ether was removed, and then the precipitate was dried. The crude precipitate was dissolved in a solution of DTT (25 mM final concentration) in phosphate buffer (pH 7.2, 50 mM, 100 μL) to generate the peptide $cVHP_{17\text{-}35}$ containing N-terminal 2-thiolphenylalanine. The peptide was purified via HPLC under conditions previously described.

4.4.14 Synthesis of peptides containing 2*S*,4*R*-mercaptoproline on solid phase

Following the procedure of Pandey *et al.* for the synthesis of 4*S*-bromoproline within a peptide,[373] triphenylphosphine (262 mg, 1 mmol) was dissolved in dry THF (2 mL) in a disposable fritted reaction vessel and chilled on ice. Diisoprolylazodicarboxylate (197 μL, 1 mmol) was added to the solution under nitrogen and the solution was mixed and chilled on ice for 5 minutes. When a white precipitate became visible, the resin containing the Boc-(4*R*-hydroxyproline)RAFS peptide (10 mg, 27 μmol) and dried carbon tetrabromide (332 mg, 1 mmol) were added to the solution (carbon tetrabromide was dried under high vacuum at room temperature for at least 8 hours prior to use). The mixture was subjected to vigorous shaking via vortex until all of the precipitate had dissolved, and then stirred on ice for 15 minutes. The mixture was then agitated on a rotary shaker for 24 h at room temperature while protected from light to generate the peptide Boc-(4*S*-bromoproline)RAFS peptide on solid phase in 77% yield. The resin was then washed with CH_2Cl_2 (4 mL × 3) and dried with ether.

To the resin containing the protected peptide (4*S*-bromoproline)RAFS (10 mg, 27 μmol) in a disposable fritted reaction vessel with a small stir bar, DMF (1 mL) and 15-crown-5 (22 μL, 0.11 mmol) were added. Sodium hydrosulfide (5.6 mg, 0.10 mmol) was added and the mixture stirred for 3 hours at room temperature to generate the peptide (4*R*-mercaptoproline)RAFS on solid phase. The resin was then washed with DMF (4 mL × 3) and CH_2Cl_2 (4 mL × 3) and dried with ether. The resin was subjected to cleavage and deprotection reaction using standard TFA conditions. The TFA was removed by evaporation with nitrogen, and the peptide was precipitated with ether. The precipitate was dissolved in phosphate buffer (pH 7.2, 50 mM) containing DTT (25 mM, for reduction of disulfides) to generate the peptide (4*R*-mercaptoproline)RAFS-NH_2 in 59% overall yield over 2 steps on solid phase from the peptide containing 4*R*-hydroxyproline (with an 11-18% product loss due to an elimination side-product formed during the S_N2 reaction and 16% unreacted (4*R*-hydroxyproline)RAFS).

4.4.15 General procedure for native chemical ligation reactions

The purified, lyophilized peptide (2-SH-Phe)RAFS-NH_2 (0.038-0.040 μmol, 1.9-2.0 mM final reaction concentration) was dissolved in 10 μL ligation buffer (100 mM phosphate buffer, pH 7.2, 20 mM TCEP, 2 mM phenol). The purified, lyophilized peptide Ac-LYRAZ-SPh (0.010-0.024 μmol, 0.5-1.2 mM final reaction concentration; Z = Gly, 1.1-1.2 mM; Ala, 0.9-1.0 mM; Val, 1.0 mM; Leu, 0.7-0.8 mM; Phe, 0.5-1.0 mM; Pro, 0.5-0.7 mM) was dissolved in 10 μL ligation buffer (100 mM phosphate buffer, pH 7.2, 20 mM TCEP, 2 mM phenol). The dissolved peptide containing 2-thiolphenylalanine was added to the solution containing the peptide Ac-LYRAZ-SPh, and the resultant solution was mixed via pipette for 30 sec, and then incubated on a

solid heating block at 23 °C. At indicated time intervals, aliquots of the ligation reaction were quenched using buffer A (98% water, 2% MeCN, 0.06% TFA). Reaction progress was monitored via HPLC. For ligation reactions with the trp cage and the cVHP-35 peptides, 6 M guanidinium was required in the reaction in order to dissolve the peptides. 2-Naphthol was used as an internal standard for the ligation reaction with cVHP-35 peptides. Purification methods and ESI-MS data are shown for all peptides in Table. 4.5.

Table 4.5. Purification and ESI-MS of the products of the ligation reactions

Peptide	**HPLC Conditions**	**Retention time (min)**	**Expected Mass**	**Observed Mass**
Ac-LYRAG(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	33.2	1259.6	631.1 $(M+2H)^{2+}$
Ac-LYRAA(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	41.8	1273.6	638.1 $(M+2H)^{2+}$
Ac-LYRAV(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	49.9	1301.7	652.1 $(M+2H)^{2+}$
Ac-LYRAL(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	52.7	1315.7	659.1 $(M+2H)^{2+}$
Ac-LYRAF(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	51.3	1349.7	676.2 $(M+2H)^{2+}$
Ac-LYRAP(2-SH-Phe)RAFS-NH_2	60 minutes 15-65% buffer B	36.6	1299.7	651.2 $(M+2H)^{2+}$
Ac-LYRAACRAFS-NH_2	60 minutes 15-65% buffer B	34.6	1198.4	600.0 $(M+2H)^{2+}$
Ac-LYRAA(Mpt)RAFS-NH_2	60 minutes 15-65% buffer B	24.5	1224.4	613.0 $(M+2H)^{2+}$
Y3(2-SH-Phe) trp cage miniprotein Ac-NL(2-SH-Phe) IQWLKDGGPSSGRPPPS-NH_2	60 minutes 15-65% buffer B	47.8	2226.5	1114.3 $(M+2H)^{2+}$
F17(2-SH-Phe) cVHP-35 LSDEDFKAVFGMTRSA(2-SH-Phe) ANLPLWKQQNLKKEKGLF-NH_2	80 minutes 15-65% buffer B	72.3	4089.7	819.1 $(M+5H)^{5+}$

4.4.16 General procedure for desulfurization reaction of ligated peptides

The purified, lyophilized peptide Ac-LYRAZ(2-SH-Phe)RAFS-NH_2 (Z = Gly, Ala, or Leu, 0.02 μmol) was dissolved in 10 μL phosphate buffer (100 mM, pH 7.2). Sodium borohydride (1 mg, 26 μmol, 1.3 M final concentration) was added to 10 μL phosphate buffer (100 mM, pH 7.2) and the resultant solution was placed on ice. Nickel(II) chloride hexahydrate (1 μL of a 185 mM solution in water, 9 mM final concentration) was added to the solution containing sodium borohydride, and the resultant solution was vortexed. The dissolved peptide was added to the solution containing nickel(II) chloride and sodium borohydride, vortexed, and incubated on ice for 20 minutes. During this time, the reaction mixture resulted in a grey precipitate that generated hydrogen gas. The mixture was quenched with 3 drops of trifluoroacetic acid (30 μL) and diluted with buffer A (500 μL), and then the solids were filtered to generate the desulfurized peptides in solution. Purification methods and ESI-MS data are shown for peptides in Table 4.6. The recovered sample from the desulfurization reaction to generate the peptide cVHP-35 was too low to obtain conclusive data from mass spectrometry, although the UV-Vis was consistent with expected product.

Table 4.6. Purification and ESI-MS of ligated and desulfurized peptides

Peptide	HPLC Conditions	Retention time (min)	Expected Mass	Observed Mass
Ac-LYRAGFRAFS-NH_2	60 minutes 15-65% buffer B	32.8	1227.7	615.1 $(M+2H)^{2+}$
Ac-LYRAAFRAFS-NH_2	60 minutes 15-65% buffer B	41.8	1273.6	638.1 $(M+2H)^{2+}$
Ac-LYRALFRAFS-NH_2	60 minutes 15-65% buffer B	52.7	1315.7	659.1 $(M+2H)^{2+}$
Y3F trp cage miniprotein Ac-NLFIQWLKDGGPSSGRPPPS-NH_2	60 minutes 15-65% buffer B	48.4	2194.5	1098.5 $(M+2H)^{2+}$
cVHP-35 LSDEDFKAVFGMTRSA- -FANLPLWKQQNLKKEKGLF-NH_2	60 minutes 15-65% buffer B	53.7	4057.7	Not confirmed

4.4.17 UV-Vis Spectroscopy for peptides containing N-terminal 2-thiophenylalanine

Purified (2-SH-Phe)RAFS-NH_2 was dissolved in buffer containing 50 mM phosphate and 1 mM TCEP. Peptide concentrations were 40-60 µM, as determined via NMR. Absorbance spectra were collected on a Perkin-Elmer Lambda 25 UV-Vis spectrometer in a 1 cm cell. Absorbance scans were taken from 350 nm to 220 nm with a slit width of 1 nm. After measurements, the pH of each sample was verified with a pH electrode (Mettler Toledo) and the absence of disulfide formation was confirmed via HPLC. Data were the average of at least three independent trials.

The pK_a of the peptide (2-SH-Phe)RAFS-NH_2 was measured via pH dependence of the molar extinction coefficient at λ_{max} wavelengths for the thiolate (272 nm). The data were fit to equation (6) using a non-linear least squares fitting algorithm (Kaleidagraph, version 4.1.1, Synergy Software), where ε = observed molar extinction coefficient, pH = pH of the sample, pK_a = acidity constant.

$$\varepsilon_{272} = \frac{\varepsilon_{272,thiolate} 10^{(pH-pKa)} + \varepsilon_{272,thiol}}{1+10^{(pH-pKa)}} \quad (6)$$

4.4.18 NMR Spectroscopy

Standard NMR spectra were recorded on a Brüker 400 MHz NMR spectrometer equipped with a cryogenic QNP probe or a Brüker 600 MHz NMR spectrometer equipped with a 5-mm Brüker SMART probe. NMR spectra on peptide samples were collected at 298 K on a Brüker AVN 600 MHz NMR spectrometer equipped with a triple resonance cryoprobe or a TXI probe. Spectra were internally referenced with TSP. 1-D spectra were collected with a Brüker w5 watergate pulse sequence and a relaxation delay of 1.7-2 s. 2-D spectra were collected with a watergate TOCSY pulse sequence.eptides were dissolved in buffer containing 5 mM phosphate (pH 4.0 or as indicated), 25 mM NaCl, 90% H_2O/10% D_2O, and 100 µM TSP. 100 µM TCEP was added for peptides containing thiophenylalanine to prevent disulfide formation during NMR experiments. Peptide concentrations were 10 µM-200 µM.

REFERENCES

1. Pace, N. J.; Weerapana, E., Diverse Functional Roles of Reactive Cysteines. *ACS Chemical Biology* **2013,** *8* (2), 283-296.
2. Giles, N. M.; Giles, G. I.; Jacob, C., Multiple roles of cysteine in biocatalysis. *Biochemical and Biophysical Research Communications* **2003,** *300* (1), 1-4.
3. Jacob, C.; Giles, G. I.; Giles, N. M.; Sies, H., Sulfur and selenium: the role of oxidation state in protein structure and function. *Angewandte Chemie International Edition* **2003,** *42* (39), 4742-4758.
4. Paulsen, C. E.; Carroll, K. S., Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery. *Chemical Reviews* **2013,** *113* (7), 4633-4679.
5. Reddie, K. G.; Carroll, K. S., Expanding the functional diversity of proteins through cysteine oxidation. *Current Opinion in Chemical Biology* **2008,** *12* (6), 746-54.
6. Paulsen, C. E.; Carroll, K. S., Orchestrating redox signaling networks through regulatory cysteine switches. *ACS Chemical Biology* **2010,** *5* (1), 47-62.
7. Beno, B. R.; Yeung, K.-S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A., A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. *Journal of Medicinal Chemistry* **2015,** *58* (11), 4383-4438.
8. Biswas, S.; Chida, A. S.; Rahman, I., Redox modifications of protein-thiols: emerging roles in cell signaling. *Biochemical Pharmacology* **2006,** *71* (5), 551-564.
9. Poole, L. B.; Karplus, P. A.; Claiborne, A., Protein sulfenic acids in redox signaling. *Annual Review of Pharmacology and Toxicology* **2004,** *44*, 325-347.
10. Leonard, S. E.; Carroll, K. S., Chemical 'omics' approaches for understanding protein cysteine oxidation in biology. *Current Opinion in Chemical Biology* **2011,** *15* (1), 88-102.
11. Leonard, S. E.; Reddie, K. G.; Carroll, K. S., Mining the Thiol Proteome for Sulfenic Acid Modifications Reveals New Targets for Oxidation in Cells. *ACS Chemical Biology* **2009,** *4* (9), 783-799.
12. Allison, W. S., Formation and Reactions of Sulfenic Acids in Proteins. *Accounts of Chemical Research* **1976,** *9* (8), 293-299.
13. Boutureira, O.; Bernardes, G. J. L., Advances in Chemical Protein Modification. *Chemical Reviews* **2015,** *115* (5), 2174-2195.
14. Baslé, E.; Joubert, N.; Pucheault, M., Protein chemical modification on endogenous amino acids. *Chemistry & Biology* **2010,** *17* (3), 213-227.
15. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical modification of proteins at cysteine: Opportunities in chemistry and biology. *Chemistry, An Asian Journal* **2009,** *4* (5), 630-640.

16. Johnson, T. B.; Brautlecht, C. A., Hydantoins: The Synthesis of Thiotyrosine. *Journal of Biological Chemistry* **1912,** *12*, 175-196.
17. Chen, S.; Zhang, Y.; Hecht, S. M., p-Thiophenylalanine-induced DNA cleavage and religation activity of a modified vaccinia topoisomerase IB. *Biochemistry* **2011,** *50* (43), 9340-9351.
18. Meyer, E. A.; Castellano, R. K.; Diederich, F., Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angewandte Chemie International Edition* **2003,** *42* (11), 1210-1250.
19. Hunter, C.; Lawson, K. R.; Perkins, J.; Urch, C. J., Aromatic interactions. *Journal of the Chemical Society, Perkin Transactions 2* **2001,** (5), 651-669.
20. Elliot, D. F.; Harington, C., 292. Bacteriostasis in the Amino-acid series. Part II. Further studies with alanine derivatives. *Journal of the Chemical Society* **1949**, 1374-1378.
21. Colescott, R. L.; Herr, R. R.; Dailey, J. P., Analogs of Phenylalanine Containing Sulfur. *Journal of the American Chemical Society* **1957,** *79* (15), 4232-4235.
22. Bernier, M.; Parent, P.; Escher, E., Angiotensin II Analogues. Part II. Synthesis and Incorporation of the Sulfur-Containing Aromatic Amino Acids: L-(4'-SH)Phe, L-(4'-SO_2NH_2)Phe, and L-(4'-S-CH_3)Phe. *Helvetica Chimica Acta* **1983,** *66* (128), 1355-1365.
23. Rajagopalan, S.; Radke, G.; Evans, M.; Tomich, J. M., Synthesis of N-*t*-Boc-4-S-*t*-Butyl-L-thiophenylalanine via Palladium Catalyzed Cross-Coupling Reaction of N-*t*-Boc-4-Iodo-L-phenylalanine with *t*-Butylthiol or Sodium *t*-Butylthiolate. *Synthetic Communications* **1996,** *26* (7), 1431--1440.
24. Seechurn, C.; Kitching, M. O.; Colacot, T. J.; Snieckus, V., Palladium-Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize. *Angewandte Chemie-International Edition* **2012,** *51* (21), 5062-5085.
25. Kunz, K.; Scholz, U.; Ganzer, D., Renaissance of Ullmann and Goldberg Reactions - Progress in Copper Catalyzed C-N-, C-O- and C-S-Coupling. *Synlett* **2003,** (15), 2428-2439.
26. Beletskaya, I. P.; Ananikov, V. P., Transition-metal-catalyzed C-S, C-Se, and C-Te bond formation via cross-coupling and atom-economic addition reactions. *Chemical Reviews* **2011,** *111* (3), 1596-1636.
27. Jordan, I. K.; Kondrashov, F. A.; Adzhubei, I. A.; Wolf, Y. I.; Koonin, E. V.; Kondrashov, A. S.; Sunyaev, S., A universal trend of amino acid gain and loss in protein evolution. *Nature* **2005,** *433* (7026), 633-638.
28. Brooks, D. J.; Fresco, J. R., Increased frequency of cysteine, tyrosine, and phenylalanine residues since the last universal ancestor. *Molecular & Cellular Proteomics* **2002,** *1* (2), 125-131.
29. Wu, H.; Ma, B. G.; Zhao, J. T.; Zhang, H. Y., How similar are amino acid mutations in human genetic diseases and evolution. *Biochemical and Biophysical Research Communications* **2007,** *362* (2), 233-237.

30. Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M., Analysis of catalytic residues in enzyme active sites. *Journal of Molecular Biology* **2002,** *324* (1), 105-121.
31. Marino, S. M.; Gladyshev, V. N., Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *Journal of Molecular Biology* **2010,** *404* (5), 902-916.
32. Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R.; Sacchettini, J. C., Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. *Nature Structural Biology* **1998,** *5* (1), 31-36.
33. Moroder, L., Isosteric replacement of sulfur with other chalcogens in peptides and proteins. *Journal of Peptide Science* **2005,** *11* (4), 187--214.
34. Shannon, D. A.; Weerapana, E., Orphan PTMs: Rare, yet functionally important modifications of cysteine. *Biopolymers* **2014,** *101* (2), 156-164.
35. Barrett, A. J.; Rawlings, N. D., Evolutionary lines of cysteine peptidases. *Biological Chemistry* **2001,** *382* (5), 727-733.
36. Shah, N. H.; Muir, T. W., Inteins: nature's gift to protein chemists. *Chemical Science* **2014,** *5* (2), 446-461.
37. Paulus, H., Protein splicing and related forms of protein autoprocessing. *Annual Review of Biochemistry* **2000,** *69*, 447-496.
38. Xu, M. Q.; Perler, F. B., The mechanism of protein splicing and its modulation by mutation. *The EMBO journal* **1996,** *15* (19), 5146-5153.
39. Hunt, M. C.; Alexson, S. E. H., The role acyl-CoA thioesterases play in mediating intracellular lipid metabolism. *Progress in Lipid Research* **2002,** *41* (2), 99-130.
40. Brass, E. P., Overview of Coenzyme-A Metabolism and its Role in Cellular Toxicity. *Chemico-Biological Interactions* **1994,** *90* (3), 203-214.
41. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annual Review of Biochemistry* **2003,** *72*, 249-289.
42. Paulsen, C. E.; Truong, T. H.; Garcia, F. J.; Homann, A.; Gupta, V.; Leonard, S. E.; Carroll, K. S., Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nature Chemical Biology* **2012,** *8*, 57-64.
43. Jeong, J.; Jung, Y.; Na, S.; Jeong, J.; Lee, E.; Kim, M.-S.; Choi, S.; Shin, D.-H.; Paek, E.; Lee, H.-Y.; Lee, K.-J., Novel oxidative modifications in redox-active cysteine residues. *Molecular & Cellular Proteomics* **2011,** *10* (3), M110.00078 - M110.00078.
44. Reddie, K. G.; Seo, Y. H.; Muse, W. B.; Leonard, S. E.; Carroll, K. S., A chemical approach for detecting sulfenic acid-modified proteins in living cells. *Molecular Biosystems* **2008,** *4* (6), 521-531.
45. DeCollo, T. V.; Lees, W. J., Effects of Aromatic Thiols on Thiol−Disulfide Interchange Reactions That Occur during Protein Folding. *The Journal of Organic Chemistry* **2001,** *66* (12), 4244-4249.
46. Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A., Disulfide bonds and protein folding. *Biochemistry* **2000,** *39* (15), 4207-4216.

47. Jacques, A.; Clemancey, M.; Blondin, G.; Fourmond, V.; Latour, J.-M.; Seneque, O., A cyclic peptide-based redox-active model of rubredoxin. *Chemical Communications* **2013,** *49* (28), 2915-2917.
48. Foster, M. W.; McMahon, T. J.; Stamler, J. S., S-nitrosylation in health and disease. *Trends in Molecular Medicine* **2003,** *9* (4), 160-168.
49. Rissing, C.; Son, D. Y., Application of Thiol-Ene Chemistry to the Preparation of Carbosilane-Thioether Dendrimers. *Organometallics* **2009,** *28* (11), 3167-3172.
50. Lo Conte, M.; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A., Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. *Chemical Communications* **2011,** *47* (39), 11086-11088.
51. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angewandte Chemie International Edition* **2009,** *48* (38), 6974-6998.
52. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *Journal of the American Chemical Society* **2002,** *124* (21), 6063-6076.
53. Bolotina, V. M.; Najibi, S.; Palacino, J. J.; Pagano, P. J.; Cohen, R. A., Nitric oxide directly activates calcium-dependent potassium channels in vascular smooth muscle. *Nature* **1994,** *368*, 850-853.
54. Claiborne, A.; Mallett, T. C.; Yeh, J. I.; Luba, J.; Parsonage, D., Structural, redox, and mechanistic parameters for cysteine-sulfenic acid function in catalysis and regulation. *Advances in Protein Chemistry* **2001,** *58*, 215-276.
55. Conte, M. L.; Carroll, K. S., The Chemistry of Thiol Oxidation and Detection. In *Oxidative Stress and Redox Regulation*, Jakob, U.; Reichmann, D., Eds. Springer Netherlands: Dordrecht, 2013; pp 1-42.
56. Kice, J. L., Mechanisms and Reactivity in Reactions of Organic Oxyacids of Sulfur and their Anhydrides. In *Advances in Physical Organic Chemistry*, Gold, V.; Bethell, D., Eds. Academic Press: 1981; Vol. Volume 17, pp 65-181.
57. Bennett, E. L.; Niemann, C., The Preparation and Resolution of the 3 Isomeric Nuclear Substituted Monofluoro-D,L-phenylalanines. *Journal of the American Chemical Society* **1950,** *72* (4), 1800-1803.
58. Guillemette, G.; Bernier, M.; Parent, P.; Leduc, R.; Escher, E., Angiotensin II: dependence of hormone affinity on the electronegativity of a single side chain. *Journal of Medicinal Chemistry* **1984,** *27*, 315-320.
59. Isidro-Llobet, A.; Alvarez, M.; Albericio, F., Amino acid-protecting groups. *Chemical Reviews* **2009,** *109* (6), 2455-2504.
60. Hobbs, D. W.; Still, C. W., Synthesis of a Thioether Analog of the Macrocyclic Tripeptide K-13. *Tetrahedron Letters* **1989,** *30* (40), 5405-5408.
61. Hobbs, D. W.; Still, C. W., Synthesis of Diphenyl Thioether Derivatives of Peptides and Amino Acids. *Tetrahedron Letters* **1987,** *28* (25), 2805-2808.
62. Lu, H. S. M.; Volk, M.; Kholodenko, Y.; Gooding, E.; Hochstrasser, R. M.; Degrado, W. F., Aminothiotyrosine Disulfide, an Optical Trigger for Initiation of

Protein Folding. *Journal of the American Chemical Society* **1997,** *119* (31), 7173-7180.
63. Gao, R.; Zhang, Y.; Choudhury, A. K.; Dedkova, L. M.; Hecht, S. M., Analogues of vaccinia virus DNA topoisomerase I modified at the active site tyrosine. *Journal of the American Chemical Society* **2005,** *127* (10), 3321-3331.
64. Gao, R.; Zhang, Y.; Dedkova, L. M.; Choudhury, A. K.; Rahier, N. J.; Hecht, S. M., Effects of modification of the active site tyrosine of human DNA topoisomerase I. *Biochemistry* **2006,** *45* (27), 8402-8410.
65. Rudolf, J. D.; Poulter, C. D., Tyrosine O-Prenyltransferase SirD Catalyzes S-, C-, and N-Prenylations on Tyrosine and Tryptophan Derivatives. *ACS Chemical Biology* **2013,** *8* (12), 2707-2714.
66. Ullmann, F., A new presentation method of phenyl ether salic acid. *Berichte Der Deutschen Chemischen Gesellschaft* **1904,** *37*, 853-854.
67. Ullmann, F., On a new formation-manner of diphenyl-amine-derivatives. Preliminary announcement. *Berichte Der Deutschen Chemischen Gesellschaft* **1903,** *36*, 2382-2384.
68. Ullmann, F.; Bielecki, J., Synthesis in the Biphenyl series. (I. Announcement). *Berichte Der Deutschen Chemischen Gesellschaft* **1901,** *34*, 2174-2185.
69. Nicolaou, K. C.; Bulger, P. G.; Sarlah, D., Palladium-catalyzed cross-coupling reactions in total synthesis. *Angewandte Chemie-International Edition* **2005,** *44* (29), 4442-4489.
70. Kondo, T.; Mitsudo, T.-a., Metal-catalyzed carbon-sulfur bond formation. *Chemical Reviews* **2000,** *100* (8), 3205-3220.
71. Bichler, P.; Love, J. A., Organometallic Approaches to Carbon-Sulfur Bond Formation. *Topics in Organometallic Chemistry* **2010,** *31*, 39-64.
72. Kosgui, M.; Shimzu, T. M., T., Reactions of aryl halides with thiolate anions in the presence of catalytic amounts of tetrakis(triphenylphosphine)palladium preparation of aryl sulfides. *Chemistry Letters* **1978**, 13-14.
73. Migita, T.; Shimizu, T.; Asami, Y.; Shiobara, J.; Kato, Y.; Kosugi, M., The palladium catalyzed nucleophilic substitution of aryl halides by thiolate anions. *The Chemical Society of Japan* **1980,** *53*, 1385-1389.
74. Heck, R. F., Arylation, Methylation and Carboxyalkylation of Olefins by Group VIII Metal Derivatives. *Journal of the American Chemical Society* **1968,** *90* (20), 5518-5531.
75. Mori, K.; Mizoroki, T.; Ozaki, A., Arylation of Olefin with Iodobenzene Catalyzed by Palladium. *Bulletin of the Chemical Society of Japan* **1973,** *46* (5), 1505-1508.
76. Julia, M.; Duteil, M., Condensation of Aromatic Halides with Olefins, Catalyzed by Palladium. *Bulletin De La Societe Chimique De France Partie II-Chimie Moleculaire Organique Et Biologique* **1973,** (9-10), 2790-2790.
77. Takagi, K.; Okamoto, T.; Sakakiba.Y; Oka, S., Palladium(II) catalyzed synthesis of aryl cyanides from aryl halides. *Chemistry Letters* **1973,** (5), 471-474.

78. Bunnett, J. F.; Creary, X., Arylation of Arenethiolate Ions by SRN1 Mechanism: Convenient Synthesis of Diaryl Sulfides. *Journal of Organic Chemistry* **1974,** *39* (21), 3173-3174.
79. Moreau, X.; Campagne, J. M.; Meyer, G.; Jutand, A., Palladium-catalyzed C-S bond formation: Rate and mechanism of the coupling of aryl or vinyl halides with a thiol derived from a cysteine. *European Journal of Organic Chemistry* **2005,** (17), 3749-3760.
80. Moreau, X.; Campagne, J. M., Palladium catalyzed thiol cross-coupling of cystein derivatives with aryl and alkenyl halides. *Journal of Organometallic Chemistry* **2003,** *687* (2), 322-326.
81. Ley, S. V.; Thomas, A. W., Modern synthetic methods for copper-mediated C(aryl)-O, C(aryl)-N, and C(aryl)-S bond formation. *Angewandte Chemie International Edition* **2003,** *42* (44), 5400-5449.
82. Kwong, F. Y.; Buchwald, S. L., A General, Efficient, and Inexpensive Catalyst System for the Coupling of Aryl Iodides and Thiols. *Organic Letters* **2002,** *4* (20), 3517-3520.
83. Wu, Y.-J.; He, H., Copper-Catalyzed Cross-Coupling of Aryl Halides and Thiols Using Microwave Heating. *Synlett* **2003,** (12), 1789-1790.
84. Deng, W.; Zou, Y.; Wang, Y.-F.; Liu, L.; Guo, Q.-X., CuI-Catalyzed Coupling Reactions of Aryl Iodides and Bromides with Thiols Promoted by Amino Acid Ligands. *Synlett* **2004,** (7), 1254-1258.
85. Sperotto, E.; van Klink, G. P. M.; de Vries, J. G.; van Koten, G., Ligand-free copper-catalyzed C-S coupling of aryl iodides and thiols. *Journal of Organic Chemistry* **2008,** *73* (14), 5625-5628.
86. Jiang, Y.; Qin, Y.; Xie, S.; Zhang, X.; Dong, J.; Ma, D., A general and efficient approach to aryl thiols: CuI-catalyzed coupling of aryl iodides with sulfur and subsequent reduction. *Organic Letters* **2009,** *11* (22), 5250-5253.
87. Martinek, M.; Korf, M.; Srogl, J., Ascorbate mediated copper catalyzed reductive cross–coupling of disulfides with aryl iodides. *Chemical Communications* **2010,** *46*, 4387-4389.
88. Sawada, N.; Itoh, T.; Yasuda, N., Efficient copper-catalyzed coupling of aryl iodides and thiobenzoic acid. *Tetrahedron Letters* **2006,** *47* (37), 6595-6597.
89. Kabir, M. S.; Lorenz, M.; Van Linn, M. L.; Namjoshi, O. A.; Ara, S.; Cook, J. M., A Very Active Cu-Catalytic System for the Synthesis of Aryl, Heteroaryl, and Vinyl Sulfides. *Journal of Organic Chemistry* **2010,** *75* (11), 3626-3643.
90. Meng, H. Y.; Thomas, K. M.; Lee, A. E.; Zondlo, N. J., Effects of *i* and *i*+3 Residue Identity on *Cis–Trans* Isomerism of the Aromatic$_{i+1}$-Proyl$_{i+2}$ Amide Bond: Implications for Type VI beta-turn Formation. *Biopolymers* **2006,** *84*, 192-204.
91. Thomas, K. M.; Naduthambi, D.; Zondlo, N. J., Electronic Control of Amide cis-trans Isomerism via the Aromatic-Prolyl Interaction. *Journal of the American Chemical Society* **2006,** *128* (7), 2216-2217.

92. Cereda, A.; Hitchcock, A.; Symes, M. D.; Cronin, L.; Bibby, T. S.; Jones, A. K., A bioelectrochemical approach to characterize extracellular electron transfer by Synechocystis sp. PCC6803. *PLoS One* **2014,** *9* (3), e91484.
93. Zondlo, N. J., Tunable CH-pi Interactions. *Accounts of Chemical Research* **2013,** *46*, 1039-1049.
94. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H., Designing a 20-residue protein. *Nature Structural Biology* **2002,** *9* (6), 425-30.
95. Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H., The Trp-cage: optimizing the stability of a globular miniprotein. *Protein engineering* **2008,** *21* (3), 171--85.
96. Naduthambi, D.; Zondlo, N. J., Stereoelectronic Tuning of the Structure and Stability of the trp cage Miniprotein. *Journal of the American Chemical Society* **2006,** *128*, 12430-12431.
97. Kakinoki, S.; Hirano, Y.; Oka, M., On the stability of polyproline-I and II structures of proline oligopeptides. *Polymer Bulletin* **2005,** *53* (2), 109-115.
98. Horng, J.-C.; Raines, R. T., Stereoelectronic effects on polyproline conformation. *Protein Science* **2006,** *15* (1), 74-83.
99. Lin, Y. J.; Chu, L. K.; Horng, J. C., Effects of the Terminal Aromatic Residues on Polyproline Conformation: Thermodynamic and Kinetic Studies. *Journal of Physical Chemistry B* **2015,** *119* (52), 15796-15806.
100. Brown, A. M.; Zondlo, N. J., A Propensity Scale for Type II Polyproline Helices: Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline-Aromatic Interactions. *Biochemistry* **2012,** *51* (25), 5041-5051.
101. Bernardes, G. J. L.; Marston, J. P.; Batsanov, A. S.; Howard, J. A. K.; Davis, B. G., A trisulfide-linked glycoprotein. *Chemical Communications* **2007,** (30), 3145-3147.
102. Rovo, P.; Farkas, V.; Straner, P.; Szabo, M.; Jermendy, A.; Hegyi, O.; Toth, G. K.; Perczel, A., Rational Design of alpha-Helix-Stabilized Exendin-4 Analogues. *Biochemistry* **2014,** *53* (22), 3540-3552.
103. Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J., Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. *Journal of the American Chemical Society* **2002,** *124*, 12952--12953.
104. Paschek, D.; Nymeyer, H.; Garcia, A. E., Replica Exchange Simulation of Reversible Folding/Unfolding of the Trp-cage Miniprotein in Explicit Solvent: On the Structure and Possible Role of Internal Water. *Journal of Structural Biology* **2007,** *157*, 524-533.
105. Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P., CH-pi hydrogen bonds in biological macromolecules. *Physical Chemistry Chemical Physics : PCCP* **2014,** *16* (25), 12648-12683.
106. Nishio, M.; Hirota, M.; Umezawa, Y., *The CH/π interaction: Evidence, Nature, and Consequences*. John Wiley & Sons: 1998; Vol. 21.
107. Skoog, D. A.; Holler, F. J.; Nieman, T. A., *Principles of Instrumental Analysis*. Saunders College Pub.: 1998.

108. Danehy, J. P.; Parameswaran, K. N., Acidic dissociation constants of thiols. *Journal of Chemical & Engineering Data* **1968,** *13* (3), 386-389.
109. Ellman, G. L., Tissue Sulfhydryl Groups. *Archives of Biochemistry and Biophysics* **1959,** *82* (1), 70-77.
110. *CRC Handbook of Chemistry and Physics*. CRC press: Boca Raton, FL, 2006-2007.
111. Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G., Allyl sulfides are privileged substrates in aqueous cross-metathesis: application to site-selective protein modification. *Journal of the American Chemical Society* **2008,** *130* (30), 9642-9643.
112. Lin, Y. A.; Chalker, J. M.; Davis, B. G., Olefin cross-metathesis on proteins: investigation of allylic chalcogen effects and guiding principles in metathesis partner selection. *Journal of the American Chemical Society* **2010,** *132* (47), 16805-16811.
113. Lin, Y. A.; Chalker, J. M.; Davis, B. G., Olefin metathesis for site-selective protein modification. *Chembiochem: A European Journal Of Chemical Biology* **2009,** *10* (6), 959-969.
114. Dondoni, A.; Marra, A., Recent applications of thiol-ene coupling as a click process for glycoconjugation. *Chemical Society Reviews* **2012,** *41* (2), 573-586.
115. Dondoni, A.; Massi, A.; Nanni, P.; Roda, A., A new ligation strategy for peptide and protein glycosylation: photoinduced thiol-ene coupling. *Chemistry - A European Journal* **2009,** *15* (43), 11444-11449.
116. Best, M. D., Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **2009,** *48* (28), 6571--6584.
117. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted 3 + 2 azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society* **2004,** *126* (46), 15046-15047.
118. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G., Bioconjugation by copper(I)-catalyzed azide-alkyne 3 + 2 cycloaddition. *Journal of the American Chemical Society* **2003,** *125* (11), 3192-3193.
119. Li, N.; Lim, R. K. V.; Edwardraja, S.; Lin, Q., Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *Journal of the American Chemical Society* **2011,** *133* (39), 15316-15319.
120. Philipson, K. D.; Gallivan, J. P.; Brandt, G. S.; Dougherty, D. A.; Lester, H. A., Incorporation of caged cysteine and caged tyrosine into a transmembrane segment of the nicotinic ACh receptor. *American Journal of Physiology-Cell Physiology* **2001,** *281* (1), C195-C206.
121. McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W., Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008,** *453* (7196), 812-816.
122. Haines, L. A.; Rajagopal, K.; Ozbas, B.; Salick, D. A.; Pochan, D. J.; Schneider, J. P., Light-activated hydrogel formation via the triggered folding and self-

assembly of a designed peptide. *Journal of the American Chemical Society* **2005,** *127* (48), 17025-17029.
123. Wang, Y.-S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R., A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum. *Journal of the American Chemical Society* **2012,** *134* (6), 2950-2953.
124. James Link, A.; Mock, M. L.; Tirrell, D. A., Non-canonical amino acids in protein engineering. *Current Opinion in Biotechnology* **2003,** *14* (6), 603-609.
125. Chapman, R. N.; Dimartino, G.; Arora, P. S., A highly stable short alpha-helix constrained by a main-chain hydrogen-bond surrogate. *Journal of the American Chemical Society* **2004,** *126* (39), 12252-12253.
126. Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G., Thiyl Glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis. *Angewandte Chemie International Edition* **2009,** *121* (42), 7938-7942.
127. Bernal, F.; Tyler, A. F.; Korsmeyer, S. J.; Walensky, L. D.; Verdine, G. L., Reactivation of the p53 tumor suppressor pathway by a stapled p53 peptide. *Journal of the American Chemical Society* **2007,** *129* (9), 2456-2457.
128. de Graaf, A. J.; Kooijman, M.; Hennink, W. E.; Mastrobattista, E., Nonnatural amino acids for site-specific protein conjugation. *Bioconjugate Chemistry* **2009,** *20* (7), 1281-1295.
129. Binder, J. B.; Raines, R. T., Olefin metathesis for chemical biology. *Current Opinion in Chemical Biology* **2008,** *12* (6), 767-773.
130. Noren, C.; Anthony-Cahill, S.; Griffith, M.; Schultz, P., A general method for site-specific incorporation of unnatural amino acids into proteins *Science* **1989,** *244* (4901), 182-188.
131. Lim, J. C.; Kim, G.; Levine, R. L., Stereospecific oxidation of calmodulin by methionine sulfoxide reductase A. *Free Radical Biology & Medicine* **2013,** *61*, 257-264.
132. Vogt, W., Oxidation of Methionyl Residues in Proteins - Tools, Targets, and Reversal. *Free Radical Biology and Medicine* **1995,** *18* (1), 93-105.
133. Verzele, D.; Madder, A., Patchwork Protein Chemistry: A Practitioner's Treatise on the Advances in Synthetic Peptide Stitchery. *Chembiochem* **2013,** *14* (9), 1032-1048.
134. Lim, J. C.; You, Z.; Kim, G.; Levine, R. L., Methionine sulfoxide reductase A is a stereospecific methionine oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (26), 10472-10477.
135. Pal, D.; Chakrabarti, P., Non-hydrogen bond interactions involving the methionine sulfur atom. *Journal of Biomolecular Structure & Dynamics* **2001,** *19* (1), 115-128.
136. Valley, C. C.; Cembran, A.; Perlmutter, J. D.; Lewis, A. K.; Labello, N. P.; Gao, J.; Sachs, J. N., The methionine-aromatic motif plays a unique role in stabilizing protein structure. *The Journal of Biological Chemistry* **2012,** *287* (42), 34979-34991.

137. Jiang, L.; Burgess, K., Oxidative Induction of beta-Turn Conformations in Cyclic Peptidomimetics: Conformational Analyses As Indicators of Configuration. *Journal of the American Chemical Society* **2002,** *124* (31), 9028-9029.
138. Druhan, L. J.; Swenson, R. P., Role of methionine 56 in the control of the oxidation-reduction potentials of the *Clostridium beijerinckii* flavodoxin: Effects of substitutions by aliphatic amino acids and evidence for a role of sulfur-flavin interactions. *Biochemistry* **1998,** *37* (27), 9668-9678.
139. Forbes, C. R.; Pandey, A. K.; Ganguly, H. K.; Yap, G. P. A.; Zondlo, N. J., 4*R*- and 4*S*-iodophenyl hydroxyproline, 4*R*-pentynoyl hydroxyproline, and S-propargyl-4-thiolphenylalanine: conformationally biased and tunable amino acids for bioorthogonal reactions. *Organic & Biomolecular Chemistry* **2016,** *14* (7), 2327-2346.
140. Barglow, K. T.; Knutson, C. G.; Wishnok, J. S.; Tannenbaum, S. R.; Marletta, M. A., Site-specific and redox-controlled S-nitrosation of thioredoxin. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (35), E600-E606.
141. Radi, R., Nitric oxide, oxidants, and protein tyrosine nitration. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (12), 4003--8.
142. Seo, Y. H.; Carroll, K. S., Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (38), 16163-16168.
143. Pollack, S. J.; Schultz, P. G., A Semisynthetic Catalytic Antibody. *Journal of the American Chemical Society* **1989,** *111* (5), 1929-1931.
144. Singh, R.; Whitesides, G. M., Thiol-disulfide interchange. In *Supplement S: The Chemistry of Sulphur-Containing Functional Groups*, Patai, S.; Rappoport, Z., Eds. 1993.
145. Fu, X.; Kassim, S. Y.; Parks, W. C.; Heinecke, J. W., Hypochlorous acid oxygenates the cysteine switch domain of pro-matrilysin (MMP-7). *The Journal of biological chemistry* **2001,** *276* (44), 41279-41287.
146. Truce, W. E.; Murphy, A. M., The Preparation of Sulfinic Acids. *Chemical Reviews* **1951,** *48* (1), 69-124.
147. Ballistreri, F. P.; Tomaselli, G. A.; Toscano, R. M., Selective oxidation reactions of diaryl- and dialkyldisulfides to sulfonic acids by CH_3ReO_3/hydrogen peroxide. *Tetrahedron Letters* **2009,** *50* (46), 6231-6232.
148. Lazzaro, F.; Crucianelli, M. a., A novel oxidative side-chain transformation of alpha-amino acids and peptides by methyltrioxorhenium/H2O2 system. *Tetrahedron Letters* **2004,** *45* (50), 9237-9240.
149. Bartberger, M. D.; Mannion, J. D.; Powell, S. C.; Stamler, J. S.; Houk, K. N.; Toone, E. J., S-N dissociation energies of S-nitrosothiols: on the origins of nitrosothiol decomposition rates. *Journal of the American Chemical Society* **2001,** *123* (36), 8868-8869.
150. Tsang, A. H. K.; Chung, K. K. K., Oxidative and nitrosative stress in Parkinson's disease. *Biochimica et Biophysica Acta* **2009,** *1792* (7), 643-650.

151. Forrester, M. T.; Benhar, M.; Stamler, J. S., Nitrosative Stress in the ER: A New Role for S-Nitrosylation in Neurodegenerative Diseases. *ACS Chemical Biology* **2006,** *1* (6), 355-358.
152. Tornoe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on solid phase: 1,2,3 -triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *Journal of Organic Chemistry* **2002,** *67* (9), 3057-3064.
153. Amblard, F.; Cho, J. H.; Schinazi, R. F., Cu(I)-Catalyzed Huisgen Azide-Alkyne 1,3-Dipolar Cycloaddition Reaction in Nucleoside, Nucleotide, and Oligonucleotide Chemistry. *Chemical Reviews* **2009,** *109* (9), 4207-4220.
154. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie-International Edition* **2002,** *41* (14), 2596-2599.
155. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie-International Edition* **2001,** *40* (11), 2004-2021.
156. Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A., Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *Journal of the American Chemical Society* **2012,** *134* (6), 2898-2901.
157. Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *Journal of the American Chemical Society* **2008,** *130* (41), 13518-13519.
158. Binder, W. H.; Sachsenhofer, R., 'Click' chemistry in polymer and materials science. *Macromolecular Rapid Communications* **2007,** *28* (1), 15-54.
159. Savarin, C.; Srogl, J.; Liebeskind, L. S., Substituted Alkyne Synthesis under Nonbasic Conditions: Copper Carboxylate-Mediated, Palladium-Catalyzed Thioalkyne−Boronic Acid Cross-Coupling. *Organic Letters* **2001,** *3* (1), 91-93.
160. Riley, K. E.; Hobza, P., On the Importance and Origin of Aromatic Interactions in Chemistry and Biodisciplines. *Accounts of Chemical Research* **2012,** *46* (4), 927-936.
161. Salonen, L. M.; Ellermann, M.; Diederich, F., Aromatic rings in chemical and biological recognition: Energetics and structures. *Angewandte Chemie International Edition* **2011,** *50* (21), 4808-4842.
162. Dougherty, D. A., Cation-π Interactions Involving Aromatic Amino Acids. *Journal of Nutrition* **2007,** *137* (6 Suppl 1), 1504S-1508S.
163. Waters, M. L., Aromatic interactions in peptides: Impact on structure and function. *Biopolymers* **2004,** *76* (5), 435--445.
164. Biswal, H. S.; Wategaonkar, S., Sulfur, not too far behind O, N, and C: SH···π hydrogen bond. *Journal of Physical Chemistry A* **2009,** *113*, 12774-12782.
165. Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J., Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **2000,** *53* (3), 233-248.

166. Lebl, M.; Sugg, E. E.; Hruby, V. J., Proton NMR Spectroscopic Evidence for Sulfur-Aromatic Interactions in Peptides. *International Journal of Peptide and Protein Research* **1987,** *29* (1), 40-45.
167. Bundy, B. C.; Swartz, J. R., Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein-Protein Click Conjugation. *Bioconjugate Chemistry* **2010,** *21* (2), 255-263.
168. Bergel, F.; Stock, J. A., 19. Cyto-active Amino-acids and Peptides Part V. Derivatives of p-Amino- and p-Mercapto-phenylalanine. *Journal of the Chemical Society* **1959**, 90-97.
169. Forbes, C. R.; Zondlo, N. J., Synthesis of Thiophenylalanine-containing Peptides via Cu(I)-mediated Cross-Coupling. *Organic Letters* **2012,** *14* (2), 464-467.
170. Bordwell, F. G.; Boutan, P. J., Conjugative Effects in Divalent Sulfur Groupings. *Journal of the American Chemical Society* **1956,** *78* (6), 854-860.
171. Yamasaki, R.; Tanatani, A.; Azumaya, I.; Saito, S.; Yamaguchi, K.; Kagechika, H., Amide Conformational Switching Induced by Protonation of Aromatic Substituent. *Organic Letters* **2003,** *5* (8), 1265-1267.
172. Pandey, A. K.; Thomas, K. M.; Forbes, C. R.; Zondlo, N. J., Tunable Control of Polyproline Helix (PPII) Structure via Aromatic Electronic Effects: An Electronic Switch of Polyproline Helix. *Biochemistry* **2014,** *53* (32), 5307-5314.
173. Reid, K. S. C.; Lindley, P. F.; Thornton, J. M., Sulfur-aromatic interactions in proteins. *FEBS Letters* **1985,** *190* (2), 209-213.
174. Daeffler, K. N. M.; Lester, H. A.; Dougherty, D. A., Functionally Important Aromatic-Aromatic and Sulfur-π Interactions in the D2 Dopamine Receptor. *Journal of the American Chemical Society* **2012,** *134* (36), 14890-14896.
175. De Filippis, V.; Draghi, A.; Frasson, R.; Grandi, C.; Musi, V.; Fontana, A.; Pastore, A., o-Nitrotyrosine and p-iodophenylalanine as spectroscopic probes for structural characterization of SH3 complexes. *Protein Science* **2007,** *16*, 1257-1265.
176. Sinha, M. K.; Reany, O.; Parvari, G.; Karmakar, A.; Keinan, E., Switchable Cucurbituril-Bipyridine Beacons. *Chemistry-a European Journal* **2010,** *16* (30), 9056-9067.
177. Zhang, B.; Zavalij, P. Y.; Isaacs, L., Acyclic CB n -type molecular containers: effect of solubilizing group on their function as solubilizing excipients. *Organic & Biomolecular Chemistry* **2014,** *12* (15), 2413-2422.
178. Whitesides, G. M.; Simanek, E. E.; Mathias, J. P.; Seto, C. T.; Chin, D. N.; Mammen, M.; Gordon, D. M., Noncovalent Synthesis: Using Physical-Organic Chemistry to Make Aggregates. *Accounts of Chemical Research* **1995,** *28* (1), 37-44.
179. Zondlo, N. J., Non-covalent interactions: Fold globally, bond locally. *Nature Chemical Biology* **2010,** *6* (8), 567-568.
180. Pauling, L., *The Nature of the Chemical Bond*. First ed.; Cornell University Press: New York, NY: 1939.
181. Mahadevi, A. S.; Sastry, G. N., Cation-π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chemical Reviews* **2013,** *113* (3), 2100-2138.

182. Dougherty, D. A., The Cation-π Interaction. *Accounts of Chemical Research* **2012,** *46* (4), 885-893.
183. Ma, J. C.; Dougherty, D. A., The Cation−π Interaction. *Chemical Reviews* **1997,** *97* (5), 1303-1324.
184. Caldwell, J. W.; Kollman, P. A., Cation-π Interactions: Nonadditive effects are critical in their accurate representation. *Journal of the American Chemical Society* **1995,** *117* (14), 4177-4178.
185. Newberry, R. W.; VanVeller, B.; Guzei, I. A.; Raines, R. T., *n*→π* Interactions of Amides and Thioamides: Implications for Protein Stability. *Journal of the American Chemical Society* **2013,** *135* (21), 7843-7846.
186. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N., n→π* Interactions in Proteins. *Nature Chemical Biology* **2010,** *6*, 615-620.
187. Hodges, J. A.; Raines, R. T., Energetics of an n→π* interaction that impacts protein structure. *Organic Letters* **2006,** *8* (21), 4695-4697.
188. Pinter, B.; Nagels, N.; Herrebout, W. A.; De Proft, F., Halogen Bonding from a Hard and Soft Acids and Bases Perspective: Investigation by Using Density Functional Theory Reactivity Indices. *Chemistry: A European Journal* **2013,** *19* (2), 518-529.
189. Politzer, P.; Murray, J. S.; Clark, T., Halogen bonding and other σ-hole interactions: A perspective. *Physical Chemistry Chemical Physics* **2013,** *15* (27), 11178-11189.
190. Desiraju, G. R.; Steiner, T., *The Weak Hydrogen Bond: Applications to Structural Chemistry and Biology*. Clarendon Press: New York, 1999; Vol. 9.
191. Steiner, T.; Desiraju, G. R., Distinction between the weak hydrogen bond and the van der Waals interaction. *Chemical Communications* **1998,** (8), 891-892.
192. Zhou, P.; Tian, F.; Lv, F.; Shang, Z., Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins-Structure Function and Bioinformatics* **2009,** *76* (1), 151-163.
193. Politzer, P.; Murray, J. S.; Clark, T., Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Physical Chemistry Chemical Physics* **2010,** *12* (28), 7748-7757.
194. Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G., The Halogen Bond. *Chemical Reviews* **2016,** *116* (4), 2478-2601.
195. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General atomic and molecular electronic-structure system. *Journal of Computational Chemistry* **1993,** *14* (11), 1347-1363.
196. Bode, B. M.; Gordon, M. S., MacMolPlt: A graphical user interface for GAMESS. *Journal of Molecular Graphics & Modelling* **1998,** *16* (3), 133-+.
197. Pauling, L., The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *Journal of the American Chemical Society* **1935,** *57*, 2680-2684.

198. Senti, F.; Harker, D., The crystal structure of rhombohedral acetamide. *Journal of the American Chemical Society* **1940,** *62*, 2008-2019.
199. Albrecht, G.; Corey, R. B., The crystal structure of glycine. *Journal of the American Chemical Society* **1939,** *61*, 1087-1103.
200. Hendricks, S. B.; Wulf, O. R.; Hilbert, G. E.; Liddel, U., Hydrogen bond formation between hydroxyl groups and nitrogen atoms in some organic compounds. *Journal of the American Chemical Society* **1936,** *58*, 1991-1996.
201. Hilbert, G. E.; Wulf, O. R.; Hendricks, S. B.; Liddel, U., The hydrogen bond between oxygen atoms in some organic compounds. *Journal of the American Chemical Society* **1936,** *58*, 548-555.
202. Pauling, L., *The Nature of the Chemical Bond*. Third ed.; Cornell University Press: New York, NY: 1960.
203. Gilli, G.; Gilli, P., *The Nature of the Hydrogen Bond: Outline of a Comprehensive Hydrogen Bond Theory*. Oxford University Press: New York, 2009; Vol. 23.
204. Pauling, L.; Corey, R. B.; Branson, H. R., The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 205-211.
205. Pauling, L.; Corey, R. B., The pleated sheet: A new layer of configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37* (5), 251-256.
206. Baker, E. N.; Hubbard, R. E., Hydrogen-bonding in globular-proteins. *Progress in Biophysics & Molecular Biology* **1984,** *44* (2), 97-179.
207. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-7155.
208. Wiberley, S. E.; Colthup, N.; Daly, L., *Introduction to Infrared and Raman Spectroscopy*. Academic Press, Inc. : San Diego, 1990.
209. Jackson, M.; Mantsch, H. H., The use and misuse of FTIR spectroscopy in the determination of protein-structure. *Critical Reviews in Biochemistry and Molecular Biology* **1995,** *30* (2), 95-120.
210. Elliott, A.; Ambrose, E. J., Structure of synthetic polypeptides. *Nature* **1950,** *165* (4206), 921-922.
211. Huggins, C. M.; Pimentel, G. C.; Shoolery, J. N., Proton magnetic resonance studies of the hydrogen bonding of phenol, substituted phenols and acetic acid. *Journal of Physical Chemistry* **1956,** *60*, 1311-1314.
212. Huggins, C. M.; Pimentel, G. C.; Shoolery, J. N., Proton magnetic resonance studies of chloroform in solution: Evidence for hydrogen bonding. *Journal of Chemical Physics* **1955,** *23* (7), 1244-1247.
213. Andersen, N. H.; Neidigh, J. W.; Harris, S. M.; Lee, G. M.; Liu, Z. H.; Tong, H., Extracting information from the temperature gradients of polypeptide NH chemical shifts .1. The importance of conformational averaging. *Journal of the American Chemical Society* **1997,** *119* (36), 8547-8561.
214. Tamaoki, H.; Kobayashi, Y.; Nishimura, S.; Ohkubo, T.; Kyogoku, Y.; Nakajima, K.; Kumagaye, S.; Kimura, T.; Sakakibara, S., Solution conformation of

endothelin determined by means of ^{1}H-NMR spectroscopy and distance geometry calculations. *Protein Engineering* **1991,** *4* (5), 509-18.
215. Wang, C. K.; Hu, S.-H.; Martin, J. L.; Sjogren, T.; Hajdu, J.; Bohlin, L.; Claeson, P.; Goransson, U.; Rosengren, K. J.; Tang, J.; Tan, N.-H.; Craik, D. J., Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. *The Journal of Biological Chemistry* **2009,** *284* (16), 10672-10683.
216. Cordier, F.; Nisius, L.; Dingley, A. J.; Grzesiek, S., Direct detection of N-H···O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nature Protocols* **2008,** *3* (2), 235-241.
217. Yoshida, Z. I.; Osawa, E., Hydrogen bonding of phenol to π electrons of aromatic, polyolefins, heteroaromatics, fulvenes, and azulenes. *Journal of the American Chemical Society* **1966,** *88* (17), 4019-4026.
218. Burley, S. K.; Petsko, G. A., Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985,** *229* (4708), 23-28.
219. Sunner, J.; Nishizawa, K.; Kebarle, P., Ion-solvent molecule interactions in the gas-phase. The potassium-ion and benzene. *Journal of Physical Chemistry* **1981,** *85* (13), 1814-1820.
220. Kumpf, R. A.; Dougherty, D. A., A Mechanism for ion selectivity in potassium channels: Computational studies of cation-π interactions. *Science* **1993,** *261* (5129), 1708-1710.
221. Dougherty, D. A., Cation–π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **1996,** *271* (5246), 163-168.
222. Steiner, T.; Koellner, G., Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *Journal of Molecular Biology* **2001,** *305* (3), 535-557.
223. Hunter, C. A.; Sanders, J. K. M., The nature of π-π interactions. *Journal of the American Chemical Society* **1990,** *112* (14), 5525-5534.
224. Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van der Oost, J.; Smit, A. B.; Sixma, T. K., Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001,** *411* (6835), 269-276.
225. Malone, J. F.; Murray, C. M.; Charlton, M. H.; Docherty, R.; Lavery, A. J., X-H···π (phenyl) interactions: Theoretical and crystallographic observations. *Journal of the Chemical Society: Faraday Transactions* **1997,** *93* (19), 3429-3436.
226. Marcus, S. H.; Miller, S. I., Hydrogen bonding in thiols. *Journal of the American Chemical Society* **1966,** *88* (16), 3719-3724.
227. Pal, D.; Chakrabarti, P., Different types of interactions involving cysteine sulfhydryl group in proteins. *Journal of Biomolecular Structure & Dynamics* **1998,** *15* (6), 1059-1072.
228. Duan, G. L.; Smith, V. H.; Weaver, D. F., Characterization of aromatic-thiol π-type hydrogen bonding and phenylalanine-cysteine side chain interactions through ab initio calculations and protein databa*se analys*es. *Molecular Physics* **2001,** *99* (19), 1689-1699.

229. Colebrook, L.; Tarbell, D. S., Evidence for hydrogen bonding in thiols from NMR measurements. *Proc. Natl. Acad. Sci. U. S. A.* **1961,** *47* (7), 993-996.
230. David, J. G.; Hallam, H. E., Hydrogen-bonding studies of thiophenols. *Spectrochimica Acta* **1965,** *21* (4), 841-850.
231. David, J. G.; Hallam, H. E., Infra-red solvent shifts and molecular interactions: Part 8. Acidity of thiophenols. *Transactions of the Faraday Society* **1964,** *60* (503P), 2013-2016.
232. Bicca de Alencastro, R.; Sandorfy, C., A Low Temperature Infrared Study of Self-association in Thiols. *Canadian Journal of Chemistry* **1972,** *50* (22), 3594-3600.
233. Saggu, M.; Levinson, N. M.; Boxer, S. G., Experimental Quantification of Electrostatics in X−H···π Hydrogen Bonds. *Journal of the American Chemical Society* **2012,** *134*, 18986-18997.
234. Rozenberg, M. S.; Nishio, T.; Steiner, T., Structural and IR-spectroscopic evidence of S-H···Ph hydrogen bonding in the solid state. *New Journal of Chemistry* **1999,** *23* (6), 585-586.
235. Viguera, A. R.; Serrano, L., Side-chain interactions between sulfur-containing amino-acids and phenylalanine in alpha-helices. *Biochemistry* **1995,** *34* (27), 8771-8779.
236. Tatko, C. D.; Waters, M. L., Investigation of the nature of the methionine-pi-interaction in beta-hairpin peptide model systems. *Protein Science* **2004,** *13* (9), 2515-2522.
237. Ludwig, M. L.; Pattridge, K. A.; Metzger, A. L.; Dixon, M. M.; Eren, M.; Feng, Y. C.; Swenson, R. P., Control of oxidation-reduction potentials in flavodoxin from *Clostridium beijerinckii*: The role of conformation changes. *Biochemistry* **1997,** *36* (6), 1259-1280.
238. Zhou, F. F.; Liu, R. R.; Li, P.; Zhang, H. Y., On the properties of S···O and S···π noncovalent interactions: The analysis of geometry, interaction energy and electron density. *New Journal of Chemistry* **2015,** *39* (3), 1611-1618.
239. Konda, Y.; Takahashi, Y.; Arima, S.; Sato, N.; Takeda, K.; Dobashi, K.; Baba, M.; Harigaya, Y., First total synthesis of Mer-N5075A and a diastereomeric mixture of alpha and beta-MAPI, new HIV-I protease inhibitors from a species of Streptomyces. *Tetrahedron* **2001,** *57* (20), 4311-4321.
240. Wallace, O. B.; Springer, D. M., Mild, selective deprotection of thioacetates using sodium thiomethoxide. *Tetrahedron Letters* **1998,** *39* (18), 2693-2694.
241. Reichenbach, G.; Santini, S.; Mazzucato, U., Charge transfer complexes of iodine with mercaptans and related sulfur derivatives. *Journal of the Chemical Society: Faraday Transactions 1* **1973,** *69* (1), 143-150.
242. Good, M.; Major, A.; Nag-Chaudhuri, J.; McGlynn, S. P., Iodine complexes of ethyl mercaptan, diethyl sulfide and diethyl disulfide. *Journal of the American Chemical Society* **1961,** *83* (21), 4329-4333.
243. Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P., Halogen bonding: The σ-hole. *Journal of Molecular Modeling* **2007,** *13* (2), 291-296.

244. Chen, C.; Weng, Z.; Hartwig, J. F., Synthesis of Copper(I) Thiolate Complexes in the Thioetherification of Aryl Halides. *Organometallics* **2012,** *31* (22), 8031-8036.
245. Stange, A. F.; Klein, A.; Klinkhammer, K. W.; Kaim, W., Aggregation control of copper(I) thiolates through substituent size and ancillary chelate ligands: closely related mono-, di-, tri- and tetranuclear complexes. *Inorganica Chimica Acta* **2001,** *324* (1-2), 336-341.
246. Li, X., Methyl 2-[(*tert*-but-oxy-carbon-yl)amino]-3-(4-hy-droxy-phen-yl)propano-ate. *Acta Crystallographica. Section E* **2013,** *43* (9), o1372.
247. Hussain, H. B.; Wilson, K. A.; Wetmore, S. D., Serine and Cysteine pi-Interactions in Nature: A Comparison of the Frequency, Structure, and Stability of Contacts Involving Oxygen and Sulfur. *Australian Journal of Chemistry* **2015,** *68* (3), 385-395.
248. Laszlo, P., Solvent effects and nuclear magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1967,** *3*, 231-402.
249. van Rossum, B. J.; Forster, H.; de Groot, H. J. M., High-field and high-speed CP-MAS C-13 NMR heteronuclear dipolar-correlation spectroscopy of solids with frequency-switched Lee-Goldburg homonuclear decoupling. *Journal of Magnetic Resonance* **1997,** *124* (2), 516-519.
250. Jabłońska, A.; Ponikiewski, Ł.; Ejsmont, K.; Herman, A.; Dołęga, A., Syntheses, spectroscopic and structural properties of phenoxysilyl compounds: X-ray structures, FT-IR and DFT calculations. *Journal of Molecular Structure* **2013,** *1054*, 359-366.
251. Fukin, G. K.; Linderman, S. V.; Kochi, J. K., Molecular structures of cation···π(arene) interactions for alkali metals with π- and σ-modalities. *Journal of the American Chemical Society* **2002,** *124* (28), 8329-8336.
252. Ringer, A. L.; Senenko, A.; Sherrill, C. D., Models of S/π interactions in protein structures: Comparison of the H2S-benzene c$_o$mplex with PDB data. *Protein Science* **2007,** *16* (10), 2216-2223.
253. Sencanski, M.; Dosen-Micovic, L.; Sukalovic, V.; Kostic-Rajacic, S., Theoretical insight into sulfur aromatic interactions with extension to D-2 receptor activation mechanism. *Structural Chemistry* **2015,** *26* (4), 1139-1149.
254. Biswal, H. S.; Bhattacharyya, S.; Bhattacherjee, A.; Wategaonkar, S., Nature and strength of sulfur-centred hydrogen bonds: laser spectroscopic investigations in the gas phase and quantum-chemical calculations. *International Reviews in Physical Chemistry* **2015,** *34* (1), 99-160.
255. Mielcarek, A.; Dolega, A., Weak hydrogen bonding interaction S-H···O=C studied by FT-IR spectroscopy and DFT calculations. *Journal of Molecular Structure* **2016,** *1103*, 217-223.
256. Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO*, Version 3.1.
257. Reed, A. E.; Curtiss, L. A.; Weinhold, F., Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chemical Reviews* **1988,** *88* (6), 899-926.

258. Lu, X.; Shi, H.; Chen, J.; Ji, D., Theoretical study of different substituent benzenes and benzene dimers blue-shifted hydrogen bonds. *Computational and Theoretical Chemistry* **2012,** *982*, 34-39.
259. Berg, L.; Mishra, B. K.; Andersson, C. D.; Ekstrom, F.; Linusson, A., The Nature of Activated Non-classical Hydrogen Bonds: A Case Study on Acetylcholinesterase-Ligand Complexes. *Chemistry-a European Journal* **2016,** *22* (8), 2672-2681.
260. Umezawa, Y.; Sei, T.; Hiroki, T.; Jun, U.; Motohiro, N., CH/π interaction in the conformation of peptides: A database study. *Bioorganic and Medicinal Chemistry* **1999,** *7*, 2021-2026.
261. Umezawa, Y.; Tsuboyama, S.; Honda, K.; Uzawa, J.; Nishio, M., CH/π interaction in the crystal structure of organic compounds. A database study. *Bulletin of the Chemical Society of Japan* **1998,** *71* (5), 1207-1213.
262. Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A., Nicotine binding to brain receptors requires a strong cation-π interaction. *Nature* **2009,** *458* (7237), 534-537.
263. Gadre, S. R.; Sen, K. D., Radii of monopositive atomic ions. *Journal of Physical Chemistry* **1993,** *99* (4), 3149-3150.
264. D'Souza, V. T.; Nanjundiah, R.; Baeza, J.; Szmant, H. H., Thiol-olefin cooxidation (TOCO) reaction: 7. A ^{1}H-NMR study of thiol solvation. *Journal of Organic Chemistry* **1987,** *52* (9), 1720-1725.
265. Samanta, U.; Chakrabarti, P.; Chandrasekhar, J., Ab initio study of energetics of X-H···π (X = N, O, and C)
interactions involving a heteroaromatic ring. *Journal of Physical Chemistry A* **1998,** *102* (45), 8964-8969.
266. Tauer, T. P.; Derrick, M. E.; Sherrill, C. D., Estimates of the ab initio limit for sulfur-π interactions: The H_2S-benzene dimer. *Journal of Physical Chemistry A* **2005,** *109* (1), 191-196.
267. Crittenden, D. L., A Systematic CCSD(T) Study of Long-Range and Noncovalent Interactions between Benzene and a Series of First- and Second-Row Hydrides and Rare Gas Atoms. *Journal of Physical Chemistry A* **2009,** *113* (8), 1663-1669.
268. Pearson, R. G., Chemical hardness and bond-dissociation energies. *Journal of the American Chemical Society* **1988,** *110* (23), 7684-7690.
269. Pearson, R. G., Recent advances in the concept of hard and soft acids and bases. *Journal of Chemical Education* **1987,** *64* (7), 561-567.
270. Pearson, R. G., Hard and Soft Acids and Bases. *Journal of the American Chemical Society* **1963,** *85* (22), 3533-3539.
271. Ho, T. L., The hard soft acids bases (HSAB) principle and organic chemistry. *Chemical Reviews* **1975,** *75* (1), 1-20.
272. LoPachin, R. M.; Gavin, T.; DeCaprio, A.; Barber, D. S., Application of the hard and soft, acids and bases (HSAB) theory to toxicant-target interactions. *Chemical Research in Toxicology* **2012,** *25* (2), 239-251.

273. Steiner, T., The hydrogen bond in the solid state. *Angewandte Chemie-International Edition* **2002,** *41* (1), 48-76.
274. Chung, W. J.; Ammam, M.; Gruhn, N. E.; Nichol, G. S.; Singh, W. P.; Wilson, G. S.; Glass, R. S., Interactions of Arenes and Thioethers Resulting in Facilitated Oxidation. *Organic Letters* **2009,** *11* (2), 397-400.
275. Koch, S.; Schollmeyer, D.; Lowe, H.; Kunst, H., C-Glycosyl Amino Acids through Hydroboration-Cross-Coupling of *exo*-Glycals and their Application in Automated Solid-Phase Synthesis. *Chemistry - A European Journal* **2013,** *19*, 7020-7041.
276. Bruker *Apex2*, Bruker AXS, Inc.: Madison, WI, 2007.
277. Sheldrick, G. M., A short history of SHELX. *Acta Crystallographica Section A* **2008,** *64* (1), 112-122.
278. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A., *Mercury CSD 2.0* - new features for the visualization and investigation of crystal structures. *Journal of Applied Crystallography* **2008,** *41*, 466-470.
279. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *GAUSSIAN 09*, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2009.
280. Breneman, C. M.; Wiberg, K. B., Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry* **1990,** *11* (3), 361-373.
281. Bolen, D. W.; Rose, G. D., Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annual Review of Biochemistry* **2008,** *77*, 339-362.
282. Andreotti, A. H., Native State Proline Isomerization : An Intrinsic Molecular Switch. *Biochemistry* **2003,** *42* (32), 9515--9524.
283. Schmid, F. X., Prolyl isomerase: enzymatic catalysis of slow protein-folding reactions. *Annual Review of Biophysics and Biomolecular Structure* **1993,** *22*, 123-142.
284. Brandts, J. F.; Halvorson, H. R.; Brennan, M., Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **1975,** *14* (1973), 4953-4963.

285. Pal, D.; Chakrabarti, P., Cis Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations. *Journal of Molecular Biology* **1999,** *294*, 271-288.
286. MacArthur, M. W.; Thornton, J. M., Influence of proline residues on protein conformation. *Journal of molecular biology* **1991,** *218* (2), 397-412.
287. Lu, K. P.; Finn, G.; Lee, T. H.; Nicholson, L. K., Prolyl cis-trans isomerization as a molecular timer. *Nature Chemical Biology* **2007,** *3* (10), 619-629.
288. Wu, W.-J.; Raleigh, D. P., Local Control of Peptide Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions. *Biopolymers* **1998,** *45*, 381-394.
289. Yao, J.; Dyson, H. J.; Wright, P. E., Three-dimensional structure of a type VI turn in a linear peptide in water solution. Evidence for stacking of aromatic rings as a major stabilizing factor. *Journal of Molecular Biology* **1994,** *243* (4), 754-766.
290. Nishio, M., The CH/pi hydrogen bond: Implication in chemistry. *Journal of Molecular Structure* **2012,** *1018*, 2-7.
291. Mammen, M.; Choi, S.-K.; Whitesides, G. M., Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angewandte Chemie International Edition* **1998,** *37* (20), 2754-2794.
292. Wedemeyer, W. J.; Welker, E.; Scheraga, H. A., Proline Cis-Trans Isomerization and Protein Folding. *Biochemistry* **2002,** *41* (50), 14637-14644.
293. Stewart, D. E.; Sarkar, A.; Wampler, J. E., Occurrence and role of cis peptide bonds in protein structures. *Journal of Mmolecular Biology* **1990,** *214* (1), 253-260.
294. Weiss, M. S.; Jabs, A.; Hilgenfeld, R., Peptide bonds revisited. *Nature Structural Biology* **1998,** *5* (8), 676.
295. Schmidpeter, P. A. M.; Koch, J. R.; Schmid, F. X., Control of protein function by prolyl isomerization. *Biochimica Et Biophysica Acta-General Subjects* **2015,** *1850* (10), 1973-1982.
296. Grathwohl, C.; Wuthrich, K., The X-Pro Peptide Bond as an NMR Probe for Conformational Studies of Flexible Linear Peptides. *Biopolymers* **1976,** *15*, 2025-2041.
297. Wathen, B.; Jia, Z., Local and Nonlocal Environments around Cis Peptides. *Journal of Proteome Research* **2008,** *7*, 145-153.
298. Jabs, A.; Weiss, M. S.; Hilgenfeld, R., Non-proline cis peptide bonds in proteins. *Journal of Molecular Biology* **1999,** *286* (1), 291-304.
299. vanScheltinga, A. C. T.; Hennig, M.; Dijkstra, B. W., The 1.8 angstrom resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. *Journal of Molecular Biology* **1996,** *262* (2), 243-257.
300. Dominguez, R.; Souchon, H.; Spinelli, S.; Dauter, Z.; Wilson, K. S.; Chauvaux, S.; Beguin, P.; Alzari, P. M., A common protein fold and similar active-site in 2 distinct families of beta-glycanases. *Nature Structural Biology* **1995,** *2* (7), 569-576.

301. Harrison, R. K.; Stein, R. L., Mechanistic studies of peptidyl prolyl cis trans isomerase - Evidence for catalysis by distortion. *Biochemistry* **1990,** *29* (7), 1684-1689.
302. Su, D.; Berndt, C.; Fomenko, D. E.; Holmgren, A.; Gladyshev, V. N., A conserved cis-proline precludes metal binding by the active site thiolates in members of the thioredoxin family of proteins. *Biochemistry* **2007,** *46* (23), 6903-6910.
303. Rees, D. C.; Lewis, M.; Honzatko, R. B.; Lipscomb, W. N.; Hardman, K. D., Zinc environment and cis peptide-bonds in carboxypeptidase-A at 1.75 angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1981,** *78* (6), 3408-3412.
304. Smajlovi, Essential role of Pro 74 in stefin B amyloid-fibril formation: dual action of cyclophilin A on the process. *FEBS letters* **2009,** *583* (7), 1114--20.
305. Torbeev, V. Y.; Hilvert, D., Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate beta 2-microglobulin amyloid assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (50), 20051-20056.
306. Hutchinson, E. G.; Thornton, J. M., A revised set of potentials for beta-turn formation in proteins. *Protein Science* **1994,** *3*, 2207-2216.
307. Brazin, K. N.; Mallis, R. J.; Fulton, D. B.; Andreotti, A. H., Regulation of the tyrosine kinase Itk by the peptidyl-prolyl isomerase cyclophilin A. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (4), 1899-1904.
308. Pletneva, E. V.; Sundd, M.; Fulton, D. B.; Andreotti, A. H., Molecular details of Itk activation by prolyl isomerization and phospholigand binding: the NMR structure of the Itk SH2 domain bound to a phosphopeptide. *Journal of Molecular Biology* **2006,** *357* (2), 550-561.
309. Severin, A.; Joseph, R. E.; Boyken, S.; Fulton, D. B.; Andreotti, A. H., Proline isomerization preorganizes the Itk SH2 domain for binding to the Itk SH3 domain. *Journal of Molecular Biology* **2009,** *387* (3), 726-743.
310. Sansom, M. S.; Weinstein, H., Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices. *Trends in Pharmacological Sciences* **2000,** *21* (11), 445-451.
311. Brandl, C. J. a. D. C. M., Hypothesis about the function of membrane-buried proline residues in transport proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83* (4), 917--21.
312. Lummis, S. C. R., 5-HT3 Receptors. *Journal of Biological Chemistry* **2012,** *287* (48), 40239-40245.
313. Melis, C.; Bussi, G.; Lummis, S. C. R.; Molteni, C., Trans-cis Switching Mechanisms in Proline Analogues and Their Relevance for the Gating of the 5-HT3 Receptor. *Journal of Physical Chemistry B* **2009,** *113* (35), 12148-12153.
314. Lummis, S. C. R.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A., Cis–trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **2005,** *438*, 248-252.
315. Hassaine, G.; Deluz, C.; Grasso, L.; Wyss, R.; Tol, M. B.; Hovius, R.; Graff, A.; Stahlberg, H.; Tomizaki, T.; Desmyter, A.; Moreau, C.; Li, X.-D.; Poitevin, F.;

Vogel, H.; Nury, H., X-ray structure of the mouse serotonin 5-HT3 receptor. *Nature* **2014,** *512* (7514), 276-+.
316. Limapichat, W.; Lester, H. A.; Dougherty, D. A., Chemical scale studies of the Phe-Pro conserved motif in the cys loop of Cys loop receptors. *The Journal of Biological Chemistry* **2010,** *285* (12), 8976-8984.
317. Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G., Side-chain Effects on Peptidyl-prolyl cis/trans Isomerisation. *Journal of Molecular Biology* **1998,** *279*, 449-460.
318. Yao, J.; Feher, V. A.; Espejo, B. F.; Reymond, M. T.; Wright, P. E.; Dyson, H. J., Stabilization of a Type VI Turn in a Family of Linear Peptides in Water Solution. *Journal of Molecular Biology* **1994,** *243*, 736-753.
319. Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E., Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *Journal of Molecular Biology* **1988,** *201* (1), 161-200.
320. Ganguly, H. K.; Majumder, B.; Chattopadhyay, S.; Chakrabarti, P.; Basu, G., Direct Evidence for CH···pi Interaction Mediated Stabilization of Pro-cisPro Bond in Peptides with Pro-Pro-Aromatic motifs. *Journal of the American Chemical Society* **2012,** *134* (10), 4661-4669.
321. Sonti, R.; Rao, K. N. S.; Chidanand, S.; Gowd, K. H.; Raghothama, S.; Balaram, P., Conformational Analysis of a 20-Membered Cyclic Peptide Disulfide from Conus virgo with a WPW Segment: Evidence for an Aromatic-Proline Sandwich. *Chemistry-A European Journal* **2014,** *20* (17), 5075-5086.
322. Leitgeb, B.; Toth, G., Aromatic-aromatic and proline-aromatic interactions in endomorphin-1 and endomorphin-2. *European Journal of Medicinal Chemistry* **2005,** *40*, 674-686.
323. Nishio, M., CH/pi hydrogen bonds in crystals. *Crystengcomm* **2004,** *6*, 130-158.
324. Paliwal, S.; Geib, S.; Wilcox, C. S., Chemistry of synthetic receptors and functional-group arrays .24. Molecular torsion balance for weak molecular recognition forces - Effects of tilted-T edge-to-face aromatic interactions on conformational selection and solid state structure. *Journal of the American Chemical Society* **1994,** *116* (10), 4497-4498.
325. Chowdhury, S. K.; Joshi, V. S.; Samuel, A. G.; Puranik, V. G.; Tavale, S. S.; Sarkar, A., Conformational preferences in molybdenum(II) pi-allyl complexes - Role of CH/pi interaction. *Organometallics* **1994,** *13* (10), 4092-4096.
326. Kolatkar, A. R.; Leung, A. K.; Isecke, R.; Brossmer, R.; Drickamer, K.; Weis, W. I., Mechanism of N-acetylgalactosamine binding to a C-type animal lectin carbohydrate-recognition domain. *Journal of Biological Chemistry* **1998,** *273* (31), 19502-19508.
327. Huggins, C. M.; Pimentel, G. C., Systematics of the infrared spectral properties of hydrogen bonding systems - Frequency shift, half width and intensity. *Journal of Physical Chemistry* **1957,** *60* (12), 1615-1619.

328. Asensio, J. L.; Ardá, A.; Cañada, F. J.; Jiménez-Barbero, J., Carbohydrate-Aromatic Interactions. *Accounts of Chemical Research* **2013,** *46* (4), 946-954.
329. Vyas, N. K.; Vyas, M. N.; Quiocho, F. A., Sugar and signal-transducer binding-cites of the Escherichia coli galactose chemoreceptor protein. *Science* **1988,** *242* (4883), 1290-1295.
330. Bolton, W.; Perutz, M. F., Three dimensional fourier synthesis of Horse Deoxyhaemoglobin at 2.8 angstrom resolution. *Nature* **1970,** *228* (5271), 551-552.
331. Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R., C-H pi-Interactions in Proteins. *Journal of Molecular Biology* **2001,** *307*, 357-377.
332. Sammelson, R. E.; Kurth, M. J., Carbon-carbon bond-forming solid-phase reactions. Part II. *Chemical Reviews* **2001,** *101* (1), 137-202.
333. Spicer, C. D.; Davis, B. G., Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chemical cCommunications (Cambridge, England)* **2011,** *47* (6), 1698-1700.
334. Chalker, J. M.; Wood, C. S. C.; Davis, B. G., A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *Journal of the American Chemical Society* **2009,** *131* (45), 16346-16347.
335. Wang, F.; Niu, W.; Guo, J.; Schultz, P. G., Unnatural Amino Acid Mutagenesis of Fluorescent Proteins. *Angewandte Chemie International Edition* **2012,** *51* (40), 10132–10135.
336. Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G., A Genetically Encoded Boronate-Containing Amino Acid. *Angewandte Chemie International Edition* **2008,** *120* (43), 8344-8347.
337. Zhao, J.; Davidson, M. G.; Mahon, M. F.; Gabriele, K.-K.; James, T. D., An enantioselective fluorescent sensor for sugar acids. *Journal of the American Chemical Society* **2004,** *126* (49), 16179-16186.
338. Nishiyabu, R.; Kubo, Y.; James, T. D.; Fossey, J. S., Boronic acid building blocks: tools for self assembly. *Chemical Communications (Cambridge, England)* **2011,** *47* (4), 1124-1150.
339. Nishiyabu, R.; Kubo, Y.; James, T. D.; Fossey, J. S., Boronic acid building blocks: tools for sensing and separation. *Chemical Communications (Cambridge, England)* **2011,** *47* (4), 1106-1123.
340. Bull, S. D.; Davidson, M. G.; van den Elsen, J. M.; Fossey, J. S.; Jenkins, T. A.; Jiang, Y.-B.; Kubo, Y.; Marken, F.; Sakurai, K.; Zhao, J.; James, T. D., Exploiting the Reversible Covalent Bonding of Boronic Acids: Recognition, Sensing, and Assembly. *Accounts of Chemical Research* **2013,** *46* (2), 312-326.
341. Westmark, P. R.; Gardiner, S. J.; Smith, B. D., Selective monosaccharide transport through lipid bilayers using boronic acid carriers. *Journal of the American Chemical Society* **1996,** *118* (45), 11093-11100.
342. Sivaev, I. B.; Bregadze, V. I., L-4-Boronophenylalanine (all around the one molecule). *Arkivoc* **2008,** *2008* (iv), 47-61.

343. Afonso, A.; Roses, C.; Planas, M.; Feliu, L., Biaryl Peptides from 4-Iodophenylalanine by Solid-Phase Borylation and Suzuki-Miyaura Cross-Coupling. *European Journal of Organic Chemistry* **2010,** (8), 1461-1468.
344. Bamoharram, F. F.; Heravi, M. M.; Tavakoli-hoseini, N., Organic Salts of Polyoxometalates: Novel and Efficient Catalysts for the Synthesis of Pyridine N-oxide Derivatives. *Synthesis and Reactivity in Inorganic, Metal-Organic, and Nano-Metal Chemistry* **2010,** *40* (10), 912-915.
345. Sharma, V.; Jain, S.; Sain, B., Methyltrioxorhenium catalyzed aerobic oxidation of organonitrogen compounds. *Tetrahedron Letters* **2003,** *44* (16), 3235-3237.
346. Laus, G., Kinetics of acetonitrile-assisted oxidation of tertiary amines by hydrogen peroxide. *Journal of the Chemical Society, Perkin Transactions 2* **2001,** (6), 864--868.
347. Vitaku, E.; Smith, D. T.; Njardarson, J. T., Analysis of the Structural Diversity, Substitution Patterns, and Frequency of Nitrogen Heterocycles among US FDA Approved Pharmaceuticals. *Journal of Medicinal Chemistry* **2014,** *57* (24), 10257-10274.
348. Wolfe, J. P.; Tomori, H.; Sadighi, J. P.; Yin, J.; Buchwald, S. L., Simple, efficient catalyst system for the palladium-catalyzed amination of aryl chlorides, bromides, and triflates. *The Journal of Organic Chemistry* **2000,** *65* (4), 1158-1174.
349. Cano, R.; Ramón, D. J.; Yus, M., Transition-metal-free O-, S-, and N-arylation of alcohols, thiols, amides, amines, and related heterocycles. *The Journal of Organic Chemistry* **2011,** *76* (2), 654-660.
350. Lundgren, R. J.; Sappong-Kumankumah, A.; Stradiotto, M., A highly versatile catalyst system for the cross-coupling of aryl chlorides and amines. *Chemistry: A European Journal* **2010,** *16* (6), 1983-1991.
351. Kwong, F. Y.; Klapars, A.; Buchwald, S. L., Copper-catalyzed coupling of alkylamines and aryl iodides: an efficient system even in an air atmosphere. *Organic Letters* **2002,** *4* (4), 581-584.
352. Yong, F.-F.; Teo, Y.-C., Efficient Ligand-Free Copper-Catalyzed Arylation of Aliphatic Amines. *Synlett* **2010,** *2010* (20), 3068-3072.
353. Radi, R., Protein Tyrosine Nitration: Biochemical Mechanisms and Structural Basis of Functional Effects. *Accounts of Chemical Research* **2013,** *46* (2), 550–559.
354. Dalsin, J. L.; Hu, B.-H.; Lee, B. P.; Messersmith, P. B., Mussel Adhesive Protein Mimetic Polymers for the Preparation of Nonfouling Surfaces. *Journal of the American Chemical Society* **2003,** *125*, 4253-4258.
355. Souza, J. M.; Peluffo, G.; Radi, R., Protein tyrosine nitration--functional alteration or just a biomarker? *Free Radical Biology & Medicine* **2008,** *45* (4), 357-366.
356. Tietze, F.; Kohn, L. D.; Kohn, A. D.; Bernardini, I.; Andersson, H. C.; Adamson, M. D.; Harper, G. S.; Gahl, W. A., Carrier-mediated transport of monoiodotyrosine out of thyroid cell lysosomes. *Journal of Biological Chemistry* **1989,** *264* (9), 4762-4765.

357. Tatko, C. D.; L., W. M., Effect of Halogenation on Edge − Face Aromatic Interactions in a beta-Hairpin Peptide: Enhanced Affinity with Iodo-substituents. *Organic Letters* **2004,** *6* (22), 3969-3972.
358. Hansch, C.; Leo, A.; Taft, R. W., A survey of Hammett substituent constants and resonance and field parameters. *Chemical Reviews* **1991,** *91* (2), 165-195.
359. Liu, Z.; Hu, B.-H.; Messersmith, P. B., Convenient synthesis of acetonide-protected 3, 4-dihydroxyphenylalanine (DOPA) for Fmoc solid-phase peptide synthesis. *Tetrahedron Letters* **2008,** *49*, 5519-5521.
360. Hu, B.-H.; Messersmith, P. B., Protection of 3,4-dihydroxyphenylalanine (DOPA) for Fmoc solid-phase peptide synthesis. *Tetrahedron Letters* **2000,** *41*, 5795-5798.
361. Bernini, R.; Barontini, M.; Crisante, F.; Ginnasi, M. C.; Saladino, R., A novel and efficient synthesis of DOPA and DOPA peptides by oxidation of tyrosine residues with IBX. *Tetrahedron Letters* **2009,** *50* (47), 6519-6521.
362. Hooker, J. M.; Kovacs, E. W.; Francis, M. B., Interior Surface Modification of Bacteriophage MS2. *Journal of the American Chemical Society* **2004,** *126* (12), 3718-3719.
363. Wisastra, R.; Poelstra, K.; Bischoff, R.; Maarsingh, H.; Haisma, H. J.; Dekker, F. J., Antibody-Free Detection of Protein Tyrosine Nitration in Tissue Sections. *Chembiochem: A European Journal Of Chemical Biology* **2011,** *12*, 2016-2020.
364. Roy, S.; Mazinani, S. K. S.; Groy, T. L.; Gan, L.; Tarakeshwar, P.; Mujica, V.; Jones, A. K., Catalytic hydrogen evolution by Fe(II) carbonyls featuring a dithiolate and a chelating phosphine. *Inorganic Chemistry* **2014,** *53* (17), 8919-8929.
365. Colorado-Peralta, R.; Sanchez-Vazquez, M.; Hernández-Ahuactzi, I. F.; Sánchez-Ruiz, S. a.; Contreras, R.; Flores-Parra, A.; Castillo-Blum, S. E., Structural study of molybdenum(VI) complexes containing bidentate ligands: Synthesis, characterization and DFT calculations. *Polyhedron* **2012,** *48*, 72-79.
366. Jacques, J. G. J.; Fourmond, V.; Arnoux, P.; Sabaty, M.; Etienne, E.; Grosse, S.; Biaso, F.; Bertrand, P.; Pignol, D.; Leger, C.; Guigliarelli, B.; Burlat, B., Reductive activation in periplasmic nitrate reductase involves chemical modifications of the Mo-cofactor beyond the first coordination sphere of the metal ion. *Biochimica et Biophysica Acta-Bioenergetics* **2014,** *1837* (2), 277-286.
367. Braude, E. A.; Nachod, F. C., *Determination of Organic Structures by Physical Methods*. Elsevier: 2013.
368. McDaniel, D. H.; Brown, H. C., An Extended Table of Hammett Substituent Constants Based on the Ionization of Substituted Benzoic Acids. *Journal of Organic Chemistry* **1958,** *23*, 420-427.
369. Anslyn, E. V.; Dougherty, D. A., *Modern Physical Organic Chemistry*. University Science Books.
370. Wüthrich, K., *NMR of Proteins and Nucleic Acids*. Wiley: 1986.
371. Sayyed, F. B.; Suresh, C. H., Quantification of substituent effects using molecular electrostatic potentials: additive nature and proximity effects. *New Journal of Chemistry* **2009,** *33* (12), 2465-2471.

372. Taylor, C. M.; Hardre, R.; Edwards, P. J. B.; Park, J. H., Factors affecting conformation in proline-containing peptides. *Organic Letters* **2003,** *5* (23), 4413-4416.
373. Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J., Proline Editing: A General and Practical Approach to the Synthesis of Functionally and Structurally Diverse Peptides. Analysis of Steric versus Stereoelectronic Effects of 4-Substituted Prolines on Conformation within Peptides. *Journal of the American Chemical Society* **2013,** *135* (11), 4333-4363.
374. Thomas, K. M.; Naduthambi, D.; Tririya, G.; Zondlo, N. J., Proline Editing: A Divergent Strategy for the Synthesis of Conformationally Diverse Peptides. *Organic Letters* **2005,** *7* (12), 2397-2400.
375. Hodges, J. A.; Raines, R. T., Stereoelectronic Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers. *Journal of the American Chemical Society* **2003,** *125* (31), 9262-9263.
376. Ganguly, H. K.; Kaur, H.; Basu, G., Local Control of cis-Peptidyl−Prolyl Bonds Mediated by CH···pi Interactions: The Xaa-Pro-Tyr Motif. *Biochemistry* **2013,** *52*, 6348-6357.
377. Vuister, G. W.; Bax, A., Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond $J_{(HNHa)}$ Coupling Constants in ^{15}N-Enriched Proteins. *Journal of the American Chemical Society* **1993,** *115* (17), 7772-7777.
378. Gilli, P.; Gilli, G.; Borea, P. A.; Varani, K.; Scatturin, A.; Dalpiaz, A., Binding thermodynamics as a tool to investigate the mechanisms of drug-receptor interactions: Thermodynamics of cytoplasmic steroid/nuclear receptors in comparison with membrane receptors. *Journal of Medicinal Chemistry* **2005,** *48* (6), 2026-2035.
379. Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L., Fluoroprolines as tools for protein design and engineering. *Angewandte Chemie-International Edition* **2001,** *40* (5), 923-925.
380. Chodera, J. D.; Mobley, D. L., Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annual Review of Biophysics, Vol 42* **2013,** *42*, 121-142.
381. Pandey, A. K.; Yap, G. P. A.; Zondlo, N. J., (2S,4R)-4-Hydroxyproline(4-nitrobenzoate): Strong Induction of Stereoelectronic Effects via a Readily Synthesized Proline Derivative. Crystallographic Observation of a Correlation between Torsion Angle and Bond Length in a Hyperconjugative Interaction. *Journal of Organic Chemistry* **2014,** *79*, 4174-4179.
382. Mukherjee, A.; Tothadi, S.; Desiraju, G. R., Halogen Bonds in Crystal Engineering: Like Hydrogen Bonds yet Different. *Accounts of Chemical Research* **2014,** *47* (8), 2514-2524.
383. Bassegoda, A.; Madden, C.; Wakerley, D. W.; Reisner, E.; Hirst, J., Reversible lnterconversion of CO_2 and Formate by a Molybdenum-Containing Formate Dehydrogenase. *Journal of the American Chemical Society* **2014,** *136* (44), 15473-15476.

384. Reda, T.; Plugge, C. M.; Abram, N. J.; Hirst, J., Reversible interconversion of carbon dioxide and formate by an electroactive enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (31), 10654-10658.
385. Boyington, J. C.; Gladyshev, V. N.; Khangulov, S. V.; Stadtman, T. C.; Sun, P. D., Crystal Structure of Formate Dehydrogenase H: Catalysis Involving Mo, Molybdopterin, Selenocysteine, and an Fe4S4 Cluster. *Science* **1997,** *275*, 1305-1308.
386. Isaacs, N. S., *Physical organic chemistry*. Longman Scientific & Technical: 1995.
387. Krygowski, T. M.; Stepien, B. T., Sigma- and Pi-electron delocalization: Focus on substituent effects. *Chemical Reviews* **2005,** *105* (10), 3482-3512.
388. Stasyuk, O. A.; Szatylowicz, H.; Krygowskib, T. M.; Guerra, C. F., How amino and nitro substituents direct electrophilic aromatic substitution in benzene: An explanation with Kohn–Sham molecular orbital theory and Voronoi deformation density analysis. *Phys. Chem. Chem. Phys* **2016,** *18*, 11624-11633.
389. Wannere, C. S.; Schleyer, P. V., How do ring currents affect H-1 NMR chemical shifts? *Organic Letters* **2003,** *5* (5), 605-608.
390. Fasting, C.; Schalley, C. A.; Weber, M.; Seitz, O.; Hecht, S.; Koksch, B.; Dernedde, J.; Graf, C.; Knapp, E.-W.; Haag, R., Multivalency as a Chemical Organization and Action Principle. *Angewandte Chemie International Edition* **2012**, 10472-10498.
391. Nataro, C.; Fosbenner, S. M., Synthesis and Characterization of Transition-Metal Complexes Containing 1,1´-Bis(diphenylphosphino) ferrocene. *Journal of Chemical Education* **2009,** *86* (12), 1412-1415.
392. Scheuermann, M. J. Design of SH3 Domain Ligands Using Novel Arginine Mimetics and Development of Tunable EF-Hand Sensors of Post Translational Modifications. University of Delaware, 2011.
393. Peng, J. B.; Clive, D. L. J., Synthesis of dihydrooxepin models related to the antitumor antibiotic MPC1001. *Organic Letters* **2007,** *9* (15), 2938-2941.
394. Kent, S.; Sohma, Y.; Liu, S.; Bang, D.; Pentelute, B.; Mandal, K., Through the Looking Glass - A New World of Proteins Enabled by Chemical Synthesis. *Journal of Peptide Science* **2012,** *18* (7), 428-436.
395. Kent, S. B. H., Total Chemical Synthesis of Proteins. *Chemical Society reviews* **2009,** *38* (2), 338-351.
396. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H., Synthesis of Proteins by Native Chemical Ligation. *Science* **1994,** *266* (5186), 776-779.
397. Wang, P.; Dong, S. W.; Shieh, J. H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J., Erythropoietin derived by chemical synthesis. *Science* **2013,** *342* (6164), 1357-1360.
398. Wang, P.; Dong, S.; Brailsford, J. A.; Iyer, K.; Townsend, S. D.; Zhang, Q.; Hendrickson, R. C.; Shieh, J.; Moore, M. A. S.; Danishefsky, S. J., At Last: Erythropoietin as a Single Glycoform. *Angewandte Chemie International Edition* **2012,** *51* (46), 11576-11584.

399. Rohde, H.; Seitz, O., Ligation-Desulfurization: A Powerful Combination in the Synthesis of Peptides and Glycopeptides. *Biopolymers* **2010,** *94* (4), 551-559.
400. Yan, L. Z.; Dawson, P. E., Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *Journal of the American Chemical Society* **2001,** *123* (4), 526-533.
401. Crich, D.; Banerjee, A., Native Chemical Ligation at Phenylalanine. *Journal of the American Chemical Society* **2007,** *129* (33), 10064-10065.
402. Crich, D.; Banerjee, A., Expedient Synthesis of threo-beta-Hydroxy-alpha-amino Acid Derivatives: Phenylalanine, Tyrosine, Histidine, and Tryptophan. *Journal of Organic Chemistry* **2006,** *71* (18), 7106-7109.
403. Tchertchian, S.; Opligger, F.; Paolini, M.; Raimondi, S.; Depresle, B.; Dafflon, N.; Gaertner, H.; Botti, P., Native Chemical Ligation at Aromatic Residues. *American Peptide Symposia* **2006,** *9* (3), 61-63.
404. Forbes, C. R.; Zondlo, N. J., 4-Thiophenylalanine: Aromatic Electronic Properties and Structural Control of Peptides. In *50th Japanese Peptide Symposium*, Nishiuchi, Y.; Teshima, T., Eds. The Japanese Peptide Society: Osaka, Japan, 2013; Vol. 2014, pp 85-88.
405. Merrifield, R. B., Solid Phase Peptide Synthesis. 1. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **1963,** *85* (14), 2149-2154.
406. Kent, S. B. H., Chemical Synthesis of Peptides and Proteins. *Annual Review of Biochemistry* **1988,** *57*, 957-989.
407. Beyerman, H. C.; Bontekoe, J. S.; Koch, A. C., Synthesis of a Tyrosine Homologue of Lysine-Vasopressin. *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society* **1960,** *79* (8), 1034-1038.
408. Vigneaud, V. D.; Gish, D. T.; Katsoyannis, P. G.; Hess, G. P., Synthesis of the Pressor-Antidiuretic Hormone, Arginine-Vasopressin. *Journal of the American Chemical Society* **1958,** *80* (13), 3355-3358.
409. Duvigneaud, V.; Bartlett, M. F.; Johl, A., The synthesis of Lysine-Vasopressin. *Journal of the American Chemical Society* **1957,** *79* (20), 5572-5575.
410. Roeske, R.; Stewart, F. H. C.; Stedman, R. J.; Duvigneaud, V., Synthesis of a Protected Tetrapeptide Amide Containing the Carboxyl Terminal Sequence of Lysine-Vasopressin. *Journal of the American Chemical Society* **1956,** *78* (22), 5883-5887.
411. Duvigneaud, V.; Ressler, C.; Swan, J. M.; Roberts, C. W.; Katsoyannis, P. G., The Synthesis of Oxytocin. *Journal of the American Chemical Society* **1954,** *76* (12), 3115-3121.
412. Kimmerlin, T.; Seebach, D., '100 Years of Peptide Synthesis': Ligation Methods for Peptide and Protein Synthesis With Applications To Beta-Peptide Assemblies. *The Journal of Peptide Research: Official Journal of the American Peptide Society* **2005,** *65* (2), 229-260.
413. Gutte, B.; Merrifield, R. B., Synthesis of Ribonuclease-A. *Journal of Biological Chemistry* **1971,** *246* (6), 1922-1941.

414. Milton, R. C.; Milton, S. C. F.; Kent, S. B. H., Total Chemical Synthesis of a D-Enzyme : The Enantiomers of HIV-1 Protease Show Demonstration of Reciprocal Chiral Substrate Specificity. *Science* **1990,** *256*, 1445-1448.
415. Nishiuchi, Y.; Inui, T.; Nishio, H.; Bodi, J.; Kimura, T.; Tsuji, F. I.; Sakakibara, S., Chemical synthesis of the precursor molecule of the Aequorea green fluorescent protein, subsequent folding, and development of fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (23), 13549-13554.
416. Dawson, P. E.; Kent, S. B. H., Synthesis of Native Proteins by Chemical Ligation. *Annual Review of Biochemistry* **2000,** *69*, 923-960.
417. Lemieux, G. A.; Bertozzi, C. R., Chemoselective ligation reactions with proteins, oligosaccharides and cells. *Trends in biotechnology* **1998,** *16* (12), 506-513.
418. Stephanopoulos, N.; Francis, M. B., Choosing an effective protein bioconjugation strategy. *Nature chemical biology* **2011,** *7* (12), 876-884.
419. Francis, M. B.; Carrico, I. S., New frontiers in protein bioconjugation. *Current opinion in chemical biology* **2010,** *14* (6), 771-773.
420. Schnolzer, M.; Kent, S. B. H., Constructing proteins by dovetailing unprotected synthetic peptides - Backbone-engineered HIV protease. *Science* **1992,** *256* (5054), 221-225.
421. Rose, K., Facile synthesis of homogeneous artificial proteins. *Journal of the American Chemical Society* **1994,** *116*, 30--33.
422. Tiefenbrunn, T. K.; Dawson, P. E., Chemoselective ligation techniques: modern applications of time-honored chemistry. *Biopolymers* **2010,** *94* (1), 95-106.
423. Joshi, N. S.; Whitaker, L. R.; Francis, M. B., A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *Journal of the American Chemical Society* **2004,** *126* (49), 15942-15943.
424. Witus, L. S.; Francis, M. B., Using Synthetically Modified Proteins to Make New Materials. *Accounts of chemical research* **2011,** *44* (9), 774-783.
425. Farkas, M. E.; Aanei, I. L.; Behrens, C. R.; Tong, G. J.; Murphy, Stephanie T.; O'Neil, J. P.; Francis, M. B., PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Molecular pharmaceutics* **2013,** *10* (1), 69-76.
426. Talbert, J. N.; Goddard, J. M., Enzymes on material surfaces. *Colloids and Surfaces B-Biointerfaces* **2012,** *93*, 8-19.
427. Kemp, D. S.; Leung, S. L.; Kerkman, D. J., Models that Demonstrate Peptide-Bond Formation by prior Thiol Capture: 1. Capture by Disulfide Formation. *Tetrahedron Letters* **1981,** *22* (3), 181-184.
428. Kemp, D. S.; Kerkman, D. J., Models that Demonstrate Peptide-Bond Formation by prior Thiol Capture:2. Capture by Organomercury Derivatives. *Tetrahedron Letters* **1981,** *22* (3), 185-186.
429. Kemp, D. S.; Grattan, J. A.; Reczek, J., Peptide-Bond Formation by prior Amine Capture Principle. *Journal of Organic Chemistry* **1975,** *40* (23), 3465-3466.
430. Kemp, D. S., The Amine Capture Strategy for Peptide-Bond Formation - An Outline of Progress. *Biopolymers* **1981,** *20* (9), 1793-1804.

431. Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H., Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives. *Journal of the American Chemical Society* **1997,** *119* (19), 776-779.
432. Johnson, E. C. B.; Kent, S. B. H., Insights into the mechanism and catalysis of the native chemical ligation reaction. *Journal of the American Chemical Society* **2006,** *128* (20), 6640-6646.
433. Bang, D.; Pentelute, B. L.; Kent, S. B. H., Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angewandte Chemie International Edition* **2006,** *45* (24), 3985-3988.
434. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (12), 6705-6710.
435. Mootz, H. D.; Muir, T. W., Protein splicing triggered by a small molecule. *Journal of the American Chemical Society* **2002,** *124* (31), 9044-9045.
436. Arnaudo, A. M.; Link, A. J.; Garcia, B. A., Bioorthogonal Chemistry for the Isolation and Study of Newly Synthesized Histones and Their Modifications. *ACS Chemical Biology* **2016,** *11* (3), 782-791.
437. Mueller, M. M.; Muir, T. W., Histones: At the Crossroads of Peptide and Protein Chemistry. *Chemical Reviews* **2015,** *115* (6), 2296-2349.
438. Kouzarides, T., Chromatin modifications and their function. *Cell* **2007,** *128* (4), 693-705.
439. Chatterjee, C.; Muir, T. W., Chemical Approaches for Studying Histone Modifications. *Journal of Biological Chemistry* **2010,** *285* (15), 11045-11050.
440. Hofmann, R. M.; Muir, T. W., Recent advances in the application of expressed protein ligation to protein engineering. *Current Opinion in Biotechnology* **2002,** *13*, 297-303.
441. Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T., Protein prosthesis: a semisynthetic enzyme with a beta-peptide reverse turn. *Journal of the American Chemical Society* **2002,** *124* (29), 8522-8523.
442. Focke, P. J.; Valiyaveetil, F. I., Studies of ion channels using expressed protein ligation. *Current opinion in chemical biology* **2010,** *14* (6), 797-802.
443. Evans, T. C.; Benner, J.; Xu, M. Q., The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *Journal of Biological Chemistry* **1999,** *274* (26), 18359-18363.
444. Muralidharan, V.; Muir, T. W., Protein ligation : an enabling technology for the biophysical analysis of proteins. *Nature Methods* **2006,** *3* (6), 429-438.
445. Pollock, S. B.; Kent, S. B. H., An investigation into the origin of the dramatically reduced reactivity of peptide-prolyl-thioesters in native chemical ligation. *Chemical communications (Cambridge, England)* **2011,** *47* (8), 2342-2344.
446. Wan, Q.; Danishefsky, S. J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angewandte Chemie International Edition* **2007,** *46* (48), 9248-9252.

447. Tam, J. P.; Yu, Q., Methionie Ligation Strategy in the Biomimetic Synthesis of Parathyroid Hormones. *Biopolymers* **1998,** *46*, 319-327.
448. Dawson, P. E., Native Chemical Ligation Combined with Desulfurization and Deselenization: A General Strategy for Chemical Protein Synthesis. *Israel Journal of Chemistry* **2011,** *51* (8-9), 862--867.
449. Haase, C.; Rohde, H.; Seitz, O., Native Chemical Ligation at Valine. *Angewandte Chemie International Edition* **2008,** *47* (36), 6807-6810.
450. Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J., Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angewandte Chemie International Edition* **2008,** *120* (44), 8649-8652.
451. Harpaz, Z.; Siman, P.; Kumar, K. S. A.; Brik, A., Protein synthesis assisted by native chemical ligation at leucine. *Chembiochem: A European Journal Of Chemical Biology* **2010,** *11* (9), 1232-1235.
452. Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J., A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. *Tetrahedron* **2010,** *66* (13), 2277-2283.
453. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F., Dual native chemical ligation at lysine. *Journal of the American Chemical Society* **2009,** *131* (38), 13592-13593.
454. Shang, S.; Tan, Z.; Dong, S.; Danishefsky, S. J., An Advance in Proline Ligation. *Journal of the American Chemical Society* **2011,** *133* (28), 10784-10786.
455. Thomas, G. L.; Payne, R. J., Phosphate-assisted peptide ligation. *Chemical communications (Cambridge, England)* **2009,** (28), 4260-4262.
456. Chen, G.; Warren, J. D.; Chen, J.; Wu, B.; Wan, Q.; Danishefsky, S. J., Studies related to the relative thermodynamic stability of C-terminal peptidyl esters of O-hydroxy thiophenol: emergence of a doable strategy for non-cysteine ligation applicable to the chemical synthesis of glycopeptides. *Journal of the American Chemical Society* **2006,** *128* (23), 7460-7462.
457. Li, J.; Cui, H.-K.; Liu, L., Peptide ligation assisted by an auxiliary attached to amidyl nitrogen. *Tetrahedron Letters* **2010,** *51* (13), 1793-1796.
458. Offer, J., Native chemical ligation with Nalpha acyl transfer auxiliaries. *Peptide Science* **2010,** *94* (4), 530--41.
459. Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E., An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications. *Bioorganic & medicinal chemistry* **2004,** *12* (10), 2749-2757.
460. Kawakami, T.; Aimoto, S., A photoremovable ligation auxiliary for use in polypeptide synthesis. *Tetrahedron Letters* **2003,** *44* (32), 6059-6061.
461. Tchertchian, S.; Hartley, O.; Botti, P., Synthesis of N alpha-(1-phenyl-2-mercaptoethyl) amino acids, new building blocks for ligation and cyclization at non-cysteine sites: scope and limitations in peptide synthesis. *The Journal of organic chemistry* **2004,** *69* (26), 9208-9214.
462. Brik, A.; Yang, Y.-Y.; Ficht, S.; Wong, C.-H., Sugar-assisted glycopeptide ligation. *Journal of the American Chemical Society* **2006,** *128* (17), 5626-5627.

463. Kumar, K. S. A.; Brik, A., Accessing posttranslationally modified proteins through thiol positioning. *Journal of peptide science : an official publication of the European Peptide Society* **2010,** *16* (10), 524-529.
464. Blanco-Canosa, J. B.; Dawson, P. E., An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angewandte Chemie International Edition* **2008,** *47* (36), 6851-6855.
465. Benesch, R. E.; Benesch, R., The Acid Strength of the -SH group in cysteine and related compounds. *Journal of the American Chemical Society* **1955,** *77*, 5877.
466. Huber, R. E. J.; Criddle, R. S., Comparison of the Chemical Properties of Selenocysteine and Selenocystine with their Sulfur Analogs. *Archives of Biochemistry and Biophysics* **1967,** *122*, 164-173.
467. Hondal, R. J., Incorporation of selenocysteine into proteins using peptide ligation. *Protein and Peptide Letters* **2005,** *12* (8), 757-764.
468. Hondal, R. J.; Nilsson, B. L.; Raines, R. T., Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *Journal of the American Chemical Society* **2001,** *123* (21), 5140-5141.
469. Gieselman, M. D.; Xie, L.; van Der Donk, W. A., Synthesis of a selenocysteine-containing peptide by native chemical ligation. *Organic letters* **2001,** *3* (9), 1331-1334.
470. Hackeng, T. M.; Griffin, J. H.; Dawson, P. E., Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (18), 10068-10073.
471. Siman, P.; Blatt, O.; Moyal, T.; Danieli, T.; Lebendiker, M.; Lashuel, H. A.; Friedler, A.; Brik, A., Chemical synthesis and expression of the HIV-1 Rev protein. *Chembiochem: A European Journal Of Chemical Biology* **2011,** *12* (7), 1097-1104.
472. Malins, L. R.; Cergol, K. M.; Payne, R. J., Chemoselective sulfenylation and peptide ligation at tryptophan. *Chemical Science* **2014,** *5* (1), 260-266.
473. Canne, L. E.; Bark, S. J.; Kent, S. B. H., Extending the Applicability of Native Chemical Ligation. *Journal of the American Chemical Society* **1996,** *118* (25), 5891-5896.
474. Williamson, M. P., The structure and function of proline-rich regions in proteins. *Biochemistry Journal* **1994,** *297*, 249--260.
475. Yaron, A.; Naider, F., Proline-dependent structural and biological properties of peptides and proteins. *Critical reviews in biochemistry and molecular biology* **1993,** *28* (1), 31-81.
476. Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. A.; Danishefsky, S. J., Advances in proline ligation. *Journal of the American Chemical Society* **2012,** *134* (8), 3912-3916.
477. Goodman, M.; Stueben, K. C., Peptide Synthesis via Amino Acid Active Esters: 2. Some Abnormal Reactions During Peptide Synthesis. *Journal of the American Chemical Society* **1962,** *84* (7), 1279-1283.
478. Lu, W. Y.; Qasim, M. A.; Kent, S. B. H., Comparative total syntheses of turkey ovomucoid third domain by both stepwise solid phase peptide synthesis and

native chemical ligation. *Journal of the American Chemical Society* **1996,** *118* (36), 8518-8523.
479. Thapa, P.; Zhang, R.-Y.; Menon, V.; Bingham, J.-P., Native Chemical Ligation: A Boon to Peptide Chemistry. *Molecules* **2014,** *19* (9), 14461-14483.
480. Back, T. G.; Baron, D. L.; Yang, K. X., Desulfurization with Nickal and Cobalt Boride - Scope, Selectivity, Stereochemistry, and Deuterium-Labeling Studies. *Journal of Organic Chemistry* **1993,** *58* (9), 2407-2413.
481. Bonner, W. A.; Grimm, R. A., Chapter 2 - Mechanisms of Raney Nickel Desulfurization. In *The Chemistry of Organic Sulfur Compounds*, Pergamon: 1966; pp 35-71.
482. Kardos, J.; Kiss, B.; Micsonai, A.; Rovo, P.; Menyhard, D. K.; Kovacs, J.; Varadi, G.; Toth, G. K.; Perczel, A., Phosphorylation as Conformational Switch from the Native to Amyloid State: Trp-Cage as a Protein Aggregation Model. *Journal of Physical Chemistry B* **2015,** *119* (7), 2946-2955.
483. Herman, R. E.; Badders, D.; Fuller, M.; Makienko, E. G.; Houston, M. E., Jr.; Quay, S. C.; Johnson, P. H., The Trp cage motif as a scaffold for the display of a randomized peptide library on bacteriophage T7. *Journal of Biological Chemistry* **2007,** *282* (13), 9813-9824.
484. McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S., A thermostable 35-residue subdomain within villin headpiece. *Journal of Molecular Biology* **1996,** *260* (2), 126-134.
485. Vugmeyster, L.; Ostroysky, D.; Khadjinoya, A.; Ellden, J.; Hoatson, G. L.; Vold, R. L., Slow Motions in the Hydrophobic Core of Chicken Villin Headpiece Subdomain and Their Contributions to Configurational Entropy and Heat Capacity from Solid-State Deuteron NMR Measurements. *Biochemistry* **2011,** *50* (49), 10637-10646.
486. Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R., High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (21), 7517-7522.
487. Vermeulen, W.; Vanhaesebrouck, P.; Van Troys, M.; Verschueren, M.; Fant, F.; Goethals, M.; Ampe, C.; Martins, J. C.; Borremans, F. A. M., Solution structures of the C-terminal headpiece subdomains of human villin and advillin, evaluation of headpiece F-actin-binding requirements. *Protein Science* **2004,** *13* (5), 1276-1287.
488. Vardar, D.; Chishti, A. H.; Frank, B. S.; Luna, E. J.; Noegel, A. A.; Oh, S. W.; Schleicher, M.; McKnight, C. J., Villin-type headpiece domains show a wide range of F-actin-binding affinities. *Cell Motility and the Cytoskeleton* **2002,** *52* (1), 9-21.
489. Frank, B. S.; Vardar, D.; Buckley, D. A.; McKnight, C. J., The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein science : a publication of the Protein Society* **2002,** *11*, 680-687.
490. Meng, W.; Shan, B.; Tang, Y.; Raleigh, D. P., Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein. *Protein Science* **2009,** *18* (8), 1692-1701.

491. Frank, B. S.; Vardar, D.; Chishti, A. H.; McKnight, C. J., The NMR structure of dematin headpiece reveals a dynamic loop that is conformationally altered upon phosphorylation at a distal site. *Journal of Biological Chemistry* **2004,** *279* (9), 7909-7916.
492. Offer, J.; Boddy, C. N. C.; Dawson, P. E., Extending Synthetic Access to Proteins with a Removable Acyl Transfer Auxiliary. *Journal of the American Chemical Society* **2002,** *124* (17), 4642-4646.
493. Malins, L. R.; Payne, R. J., Synthesis and Utility of beta-Selenol-Phenylalanine for Native Chemical Ligation-Deselenization Chemistry. *Organic letters* **2012,** *14* (12), 3142-3145.
494. Wagner, A. M.; Fegley, M. W.; Warner, J. B.; Grindley, C. L. J.; Marotta, N. P.; Petersson, E. J., N-terminal protein modification using simple aminoacyl transferase substrates. *Journal of the American Chemical Society* **2011,** *133* (38), 15139-15147.
495. Bodanszky, M., *Principles of peptide synthesis*. Springer Science & Business Media: 2012; Vol. 16.
496. El-Faham, A.; Albericio, F., Peptide coupling reagents, more than a letter soup. *Chemical Reviews* **2011,** *111* (11), 6557-6602.
497. Offer, J.; Dawson, P. E., N-alpha-2-Mercaptobenzylamine-Assisted Chemical Ligation. *Organic Letters* **2000,** *2* (1), 23-26.
498. Williams, D. V.; Barua, B.; Andersen, N. H., Hyperstable miniproteins: additive effects of D- and L-Ala mutations. *Organic & biomolecular chemistry* **2008,** *6* (23), 4287-4289.
499. Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R., A self-replicating peptide. *Nature* **1996,** *382* (6591), 525-528.
500. Severin, K.; Lee, D. H.; Martinez, J. A.; Ghadiri, M. R., Peptide self-replication via template-directed ligation. *Chemistry-a European Journal* **1997,** *3* (7), 1017-1024.
501. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R., Fmoc-Based Synthesis of Peptide-a Thioesters: Application to the Total Chemical Synthesis of a Glycoprotein by Native Chemical Ligation. *Journal of the American Chemical Society* **1999,** *121* (50), 11684-11689.
502. Wilkinson, B. L.; Chun, C. K. Y.; Payne, R. J., Synthesis of MUC1 glycopeptide thioesters and ligation via direct aminolysis. *Biopolymers* **2011,** *96* (2), 137-146.
503. Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J., Native Chemical Ligation of Thioamide-Containing Peptides: Development and Application to the Synthesis of Labeled alpha-Synuclein for Misfolding Studies. *Journal of the American Chemical Society* **2012,** *134* (22), 9172-9182.
504. Xu, Z. R.; Hu, W. M.; Liu, Q. A.; Zhang, L. H.; Jia, Y. X., Total Synthesis of Clavicipitic Acid and Aurantioclavine: Stereochemistry of Clavicipitic Acid Revisited. *Journal of Organic Chemistry* **2010,** *75* (22), 7626-7635.

Appendix A

X-RAY CRYSTALLOGRAPHIC INFORMATION

Compounds were described and characterized in Chapter 2: Insights into S–H/π Aromatic Interactions: Studies on Boc-4-Thiol-L-Phenylalanine-*tert*-Butyl Ester via IR Spectroscopy, X-Ray Crystallography, and *ab initio* Calculations

Crystallographic data for Boc-4-thiol-L-phenylalanine-*tert*-butyl ester

Table A1. Crystallographic data and refinement details for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-tert-butyl ester.

empirical formula	$C_{18}H_{27}NO_4S$	
formula weight	353.46	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, $P2_12_12_1$	
Unit cell dimensions (Å, °)	a = 5.4629(6)	α = 90
	b = 10.1217(11)	β = 90
	c = 35.898(4)	γ = 90
Volume (Å^3)	1984.9(4)	
Z, Z′, calcd density (g/cm^3)	4, 1, 1.183	
absorption coefficient (mm^{-1})	0.182	
F(000)	760	
crystal size (mm)	0.188 x 0.195 x 0.328	
□ range for data collection	2.09 to 27.47°	
Index ranges	$-6 \le h \le 7$, $-13 \le k \le 13$, $-46 \le l \le 46$	
Reflections collected/ unique	15122/4518 [R(int) = 0.0378]	
Coverage of independent reflections	100.0%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7456 and 0.6789	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	4518 / 0 / 231	
Goodness-of-fit on F^2	1.015	
Final *R* indices	3579 data; I>2σ(I)	R1 = 0.0403, wR2 = 0.0870
	all data	R1 = 0.0578, wR2 = 0.0962
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0425P)^2+0.2479P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.161 and -0.232 eÅ^{-3}	
R.M.S. deviation from mean	0.34 $^{-3}$	

Table A2. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
S1	0.34521(19)	0.52173(9)	0.52548(2)	0.0622(3)
N1	0.5840(4)	0.5118(2)	0.34791(6)	0.0304(5)
O1	0.8164(4)	0.62103(16)	0.28823(5)	0.0436(5)
O2	0.6554(3)	0.82223(15)	0.30105(4)	0.0313(4)
O3	0.1861(3)	0.48255(17)	0.36414(5)	0.0403(4)
O4	0.4708(3)	0.31949(16)	0.37059(5)	0.0403(5)
C1	0.6943(5)	0.7273(2)	0.37559(7)	0.0393(6)
C2	0.3915(5)	0.7269(2)	0.42901(7)	0.0383(6)
C3	0.3109(5)	0.6810(3)	0.46314(7)	0.0397(6)
C4	0.4424(5)	0.5864(3)	0.48204(7)	0.0387(6)
C5	0.6586(5)	0.5389(3)	0.46688(7)	0.0427(6)
C6	0.7381(5)	0.5859(3)	0.43259(7)	0.0405(7)
C7	0.6071(5)	0.6801(2)	0.41336(7)	0.0346(6)
C8	0.5698(5)	0.6527(2)	0.34302(7)	0.0305(5)
C9	0.3936(5)	0.4416(2)	0.36141(6)	0.0309(6)
C10	0.3022(5)	0.2244(2)	0.38818(8)	0.0414(6)
C11	0.2005(10)	0.2795(3)	0.42364(10)	0.0920(15)
C12	0.4681(7)	0.1081(3)	0.39590(12)	0.0736(11)
C13	0.1084(7)	0.1843(3)	0.36078(10)	0.0678(10)
C14	0.6941(5)	0.6938(2)	0.30706(6)	0.0301(5)
C15	0.7963(5)	0.8954(2)	0.27215(7)	0.0320(5)
C16	0.0670(5)	0.8846(3)	0.28079(8)	0.0436(7)
C17	0.7318(6)	0.8458(3)	0.23376(7)	0.0483(8)
C18	0.7083(5)	0.0368(2)	0.27747(8)	0.0437(7)

Table A3. Bond lengths [Å] and angles [°] for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

Bond lengths

S1-C4	1.773(3)	S1-H1S	1.26(4)
N1-C9	1.350(3)	N1-C8	1.439(3)
N1-H1N	0.75(3)	O1-C14	1.202(3)
O2-C14	1.335(3)	O2-C15	1.489(3)
O3-C9	1.211(3)	O4-C9	1.346(3)
O4-C10	1.474(3)	C1-C7	1.514(3)
C1-C8	1.549(3)	C1-H1A	0.99
C1-H1B	0.99	C2-C3	1.383(4)
C2-C7	1.388(4)	C2-H2	0.95
C3-C4	1.376(4)	C3-H3	0.95
C4-C5	1.386(4)	C5-C6	1.389(4)
C5-H5	0.95	C6-C7	1.378(4)
C6-H6	0.95	C8-C14	1.517(3)
C8-H8	1.0	C10-C11	1.497(4)
C10-C13	1.501(4)	C10-C12	1.511(4)
C11-H11A	0.98	C11-H11B	0.98
C11-H11C	0.98	C12-H12A	0.98
C12-H12B	0.98	C12-H12C	0.98
C13-H13A	0.98	C13-H13B	0.98
C13-H13C	0.98	C15-C17	1.509(4)
C15-C16	1.515(4)	C15-C18	1.522(4)
C16-H16A	0.98	C16-H16B	0.98
C16-H16C	0.98	C17-H17A	0.98
C17-H17B	0.98	C17-H17C	0.98
C18-H18A	0.98	C18-H18B	0.98
C18-H18C	0.98		

Table A3 continued

Bond angles

C4-S1-H1S	99.4(15)	C9-N1-C8	121.6(2)
C9-N1-H1N	118.(2)	C8-N1-H1N	117.(2)
C14-O2-C15	121.01(19)	C9-O4-C10	120.6(2)
C7-C1-C8	112.6(2)	C7-C1-H1A	109.1
C8-C1-H1A	109.1	C7-C1-H1B	109.1
C8-C1-H1B	109.1	H1A-C1-H1B	107.8
C3-C2-C7	120.9(3)	C3-C2-H2	119.5
C7-C2-H2	119.5	C4-C3-C2	120.3(3)
C4-C3-H3	119.8	C2-C3-H3	119.8
C3-C4-C5	119.5(2)	C3-C4-S1	122.3(2)
C5-C4-S1	118.2(2)	C4-C5-C6	119.7(3)
C4-C5-H5	120.1	C6-C5-H5	120.1
C7-C6-C5	121.2(2)	C7-C6-H6	119.4
C5-C6-H6	119.4	C6-C7-C2	118.3(2)
C6-C7-C1	120.2(2)	C2-C7-C1	121.4(2)
N1-C8-C14	110.54(19)	N1-C8-C1	111.6(2)
C14-C8-C1	108.2(2)	N1-C8-H8	108.8
C14-C8-H8	108.8	C1-C8-H8	108.8
O3-C9-O4	126.0(2)	O3-C9-N1	124.7(2)
O4-C9-N1	109.3(2)	O4-C10-C11	110.6(2)
O4-C10-C13	109.7(2)	C11-C10-C13	113.3(3)
O4-C10-C12	102.3(2)	C11-C10-C12	110.9(3)
C13-C10-C12	109.4(3)	C10-C11-H11A	109.5
C10-C11-H11B	109.5	H11A-C11-H11B	109.5
C10-C11-H11C	109.5	H11A-C11-H11C	109.5
H11B-C11-H11C	109.5	C10-C12-H12A	109.5
C10-C12-H12B	109.5	H12A-C12-H12B	109.5
C10-C12-H12C	109.5	H12A-C12-H12C	109.5
H12B-C12-H12C	109.5	C10-C13-H13A	109.5
C10-C13-H13B	109.5	H13A-C13-H13B	109.5
C10-C13-H13C	109.5	H13A-C13-H13C	109.5
H13B-C13-H13C	109.5	O1-C14-O2	126.4(2)
O1-C14-C8	124.0(2)	O2-C14-C8	109.46(19)
O2-C15-C17	110.5(2)	O2-C15-C16	109.0(2)
C17-C15-C16	113.0(2)	O2-C15-C18	102.55(19)
C17-C15-C18	110.7(2)	C16-C15-C18	110.5(2)

Table A3 continued

C15-C16-H16A	109.5	C15-C16-H16B	109.5
H16A-C16-H16B	109.5	C15-C16-H16C	109.5
H16A-C16-H16C	109.5	H16B-C16-H16C	109.5
C15-C17-H17A	109.5	C15-C17-H17B	109.5
H17A-C17-H17B	109.5	C15-C17-H17C	109.5
H17A-C17-H17C	109.5	H17B-C17-H17C	109.5
C15-C18-H18A	109.5	C15-C18-H18B	109.5
H18A-C18-H18B	109.5	C15-C18-H18C	109.5
H18A-C18-H18C	109.5	H18B-C18-H18C	109.5

Table A4. Anisotropic atomic displacement parameters (Å2) for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12].

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	0.0795(6)	0.0668(5)	0.0404(4)	0.0099(4)	0.0190(4)	0.0032(5)
N1	0.0320(12)	0.0243(10)	0.0349(11)	0.0032(9)	0.0036(10)	0.0016(10)
O1	0.0578(12)	0.0335(9)	0.0395(10)	0.0021(8)	0.0175(10)	0.0070(9)
O2	0.0321(9)	0.0254(7)	0.0365(9)	0.0080(7)	0.0065(8)	-0.0009(7)
O3	0.0353(10)	0.0314(9)	0.0542(11)	0.0077(8)	0.0049(9)	0.0006(8)
O4	0.0386(11)	0.0255(8)	0.0568(11)	0.0109(8)	0.0052(9)	-0.0013(8)
C1	0.0486(16)	0.0307(12)	0.0387(14)	-0.0031(11)	0.0077(13)	-0.0099(12)
C2	0.0447(16)	0.0321(13)	0.0380(14)	-0.0035(11)	-0.0004(12)	0.0013(12)
C3	0.0405(15)	0.0406(14)	0.0380(14)	-0.0077(12)	0.0074(13)	0.0022(13)
C4	0.0468(16)	0.0400(14)	0.0292(13)	-0.0035(11)	0.0033(12)	-0.0054(13)
C5	0.0454(16)	0.0431(14)	0.0398(14)	0.0018(12)	-0.0004(13)	0.0087(14)
C6	0.0378(16)	0.0448(15)	0.0389(14)	-0.0040(12)	0.0058(12)	0.0041(12)
C7	0.0410(15)	0.0315(12)	0.0314(13)	-0.0069(11)	0.0030(11)	-0.0088(12)
C8	0.0357(14)	0.0216(11)	0.0341(13)	0.0022(10)	0.0047(11)	0.0000(10)
C9	0.0413(16)	0.0234(11)	0.0280(12)	0.0019(10)	-0.0004(11)	-0.0026(10)
C10	0.0463(16)	0.0280(12)	0.0499(16)	0.0096(12)	0.0049(14)	-0.0100(12)
C11	0.166(4)	0.0501(19)	0.060(2)	0.0021(17)	0.054(3)	-0.022(3)
C12	0.063(2)	0.0437(18)	0.113(3)	0.038(2)	0.004(2)	-0.0029(16)

Table A4 continued

C13	0.062(2)	0.0511(18)	0.090(3)	0.0147(18)	-0.011(2)	-0.0226(17)
C14	0.0311(13)	0.0282(11)	0.0308(12)	0.0013(10)	0.0014(11)	-0.0024(11)
C15	0.0266(13)	0.0333(12)	0.0361(13)	0.0107(10)	0.0054(11)	-0.0057(11)
C16	0.0296(15)	0.0446(15)	0.0564(17)	0.0082(14)	0.0025(13)	-0.0044(13)
C17	0.059(2)	0.0519(17)	0.0345(14)	0.0085(13)	0.0024(13)	-0.0062(14)
C18	0.0398(16)	0.0316(13)	0.0595(17)	0.0125(13)	0.0065(14)	-0.0016(12)

Table A5. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

	x/a	y/b	z/c	U(eq)
H1S	0.173(7)	0.601(3)	0.5324(9)	0.083(11)
H1N	0.709(5)	0.482(3)	0.3493(7)	0.032(8)
H1A	0.8737	0.7150	0.3739	0.047
H1B	0.6597	0.8230	0.3732	0.047
H2	0.2980	0.7915	0.4161	0.046
H3	0.1640	0.7149	0.4736	0.048
H5	0.7519	0.4746	0.4799	0.051
H6	0.8856	0.5524	0.4222	0.049
H8	0.3936	0.6792	0.3419	0.037
H11A	0.3347	0.3130	0.4391	0.138
H11B	0.1129	0.2099	0.4371	0.138
H11C	0.0874	0.3518	0.4179	0.138
H12A	0.5273	0.0716	0.3723	0.11
H12B	0.3769	0.0401	0.4095	0.11
H12C	0.6078	0.1373	0.4109	0.11
H13A	-0.0109	0.2559	0.3581	0.102
H13B	0.0254	0.1046	0.3698	0.102
H13C	0.1843	0.1661	0.3366	0.102
H16A	1.0969	0.9134	0.3064	0.065
H16B	1.1597	0.9408	0.2636	0.065
H16C	1.1197	0.7926	0.2779	0.065
H17A	0.7926	0.7553	0.2307	0.072
H17B	0.8072	0.9030	0.2150	0.072
H17C	0.5536	0.8467	0.2306	0.072
H18A	0.5309	1.0408	0.2736	0.065
H18B	0.7900	1.0945	0.2594	0.065
H18C	0.7474	1.0663	0.3028	0.065

Table A6. Torsion angles (°) for orthorhombic crystals of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester.

C7-C2-C3-C4	-0.8(4)	C2-C3-C4-C5	1.0(4)
C2-C3-C4-S1	-178.9(2)	C3-C4-C5-C6	-0.9(4)
S1-C4-C5-C6	179.0(2)	C4-C5-C6-C7	0.6(4)
C5-C6-C7-C2	-0.4(4)	C5-C6-C7-C1	-178.5(2)
C3-C2-C7-C6	0.5(4)	C3-C2-C7-C1	178.6(2)
C8-C1-C7-C6	95.4(3)	C8-C1-C7-C2	-82.6(3)
C9-N1-C8-C14	-140.0(2)	C9-N1-C8-C1	99.6(3)
C7-C1-C8-N1	-50.5(3)	C7-C1-C8-C14	-172.3(2)
C10-O4-C9-O3	-4.7(4)	C10-O4-C9-N1	176.4(2)
C8-N1-C9-O3	14.1(4)	C8-N1-C9-O4	-167.0(2)
C9-O4-C10-C11	-57.8(4)	C9-O4-C10-C13	68.0(3)
C9-O4-C10-C12	-176.0(3)	C15-O2-C14-O1	-9.5(4)
C15-O2-C14-C8	167.5(2)	N1-C8-C14-O1	-9.7(4)
C1-C8-C14-O1	112.7(3)	N1-C8-C14-O2	173.2(2)
C1-C8-C14-O2	-64.4(3)	C14-O2-C15-C17	66.5(3)
C14-O2-C15-C16	-58.3(3)	C14-O2-C15-C18	-175.4(2)

Crystallographic data for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester

Table A7. Crystallographic data and refinement details for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

empirical formula	$C_{18}H_{27}NO_4S$	
formula weight	353.46	
T (K)	143(2)	
wavelength (Å)	1.54178	
crystal system, space group	Monoclinic, P 1 21/c 1	
Unit cell dimensions (Å, °)	a = 10.3648(5)	α = 90
	b = 9.4888(4)	β = 96.902(3)
	c = 19.8874(10)	γ = 90
Volume (Å^3)	1941.74(16)	
Z, Z′, calcd density (g/cm^3)	4, 1, 1.209	
absorption coefficient (mm^{-1})	1.647	
F(000)	760	
crystal size (mm)	0.050 x 0.069 x 0.364	
□ range for data collection	4.48 to 73.17°	
Index ranges	$-12 \le h \le 11$, $-11 \le k \le 11$, $-24 \le l \le 24$	
Reflections collected/ unique	27332/3866 [R(int) = 0.0889]	
Coverage of independent reflections	99.2%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7536 and 0.5791	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	3866 / 0 / 231	
Goodness-of-fit on F^2	1.080	
Final *R* indices	2754 data; I>2σ(I)	R1 = 0.0630, wR2 = 0.1597
	all data	R1 = 0.0878, wR2 = 0.1718
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0687P)^2+0.0782P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.457 and -0.343 eÅ^{-3}	
R.M.S. deviation from mean	0.058 eÅ^{-3}	

Table A8. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
S1	0.47804(8)	0.13536(9)	0.19027(4)	0.0502(2)
N1	0.0319(2)	0.3369(2)	0.05287(12)	0.0388(6)
O1	0.2338(2)	0.2473(2)	0.07541(10)	0.0477(5)
O2	0.09943(19)	0.18397(19)	0.98091(9)	0.0383(5)
O3	0.1741(2)	0.5993(2)	0.06091(10)	0.0444(5)
O4	0.23193(19)	0.53581(19)	0.17015(9)	0.0391(5)
C1	0.9265(3)	0.5047(3)	0.12215(15)	0.0449(7)
C2	0.7121(3)	0.3826(3)	0.08977(16)	0.0463(7)
C3	0.6107(3)	0.2983(3)	0.10464(15)	0.0454(7)
C4	0.6078(3)	0.2436(3)	0.16929(15)	0.0404(6)
C5	0.7101(3)	0.2733(3)	0.21865(16)	0.0456(7)
C6	0.8125(3)	0.3570(3)	0.20330(16)	0.0458(7)
C7	0.8148(3)	0.4131(3)	0.13829(15)	0.0428(7)
C8	0.0499(3)	0.4228(3)	0.11361(15)	0.0408(7)
C9	0.1311(3)	0.2542(3)	0.03949(14)	0.0373(6)
C10	0.1922(3)	0.0845(3)	0.95573(15)	0.0431(7)
C11	0.1141(3)	0.0266(3)	0.89213(16)	0.0525(8)
C12	0.3100(3)	0.1640(4)	0.93791(18)	0.0548(8)
C13	0.2262(4)	0.9673(3)	0.00684(17)	0.0537(8)
C14	0.1596(3)	0.5282(3)	0.11102(14)	0.0383(6)
C15	0.3388(3)	0.6390(3)	0.18422(15)	0.0395(7)
C16	0.4342(3)	0.6270(4)	0.13315(19)	0.0616(9)
C17	0.4002(3)	0.5916(3)	0.25351(17)	0.0566(9)
C18	0.2800(3)	0.7834(3)	0.18696(18)	0.0516(8)

Table A9. Bond lengths [Å] and angles [°] for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

Bond lengths

S1-C4	1.781(3)	S1-H1S	0.99(3)
N1-C9	1.345(4)	N1-C8	1.451(3)
N1-H1N	0.90(3)	O1-C9	1.211(3)
O2-C9	1.348(3)	O2-C10	1.477(3)
O3-C14	1.228(3)	O4-C14	1.319(3)
O4-C15	1.480(3)	C1-C7	1.513(4)
C1-C8	1.523(4)	C1-H1A	0.99
C1-H1B	0.99	C2-C7	1.379(4)
C2-C3	1.380(4)	C2-H2	0.95
C3-C4	1.390(4)	C3-H3	0.95
C4-C5	1.385(4)	C5-C6	1.389(4)
C5-H5	0.95	C6-C7	1.401(4)
C6-H6	0.95	C8-C14	1.519(4)
C8-H8	1.0	C10-C12	1.513(4)
C10-C13	1.519(4)	C10-C11	1.520(4)
C11-H11A	0.98	C11-H11B	0.98
C11-H11C	0.98	C12-H12A	0.98
C12-H12B	0.98	C12-H12C	0.98
C13-H13A	0.98	C13-H13B	0.98
C13-H13C	0.98	C15-C18	1.504(4)
C15-C16	1.504(4)	C15-C17	1.515(4)
C16-H16A	0.98	C16-H16B	0.98
C16-H16C	0.98	C17-H17A	0.98
C17-H17B	0.98	C17-H17C	0.98
C18-H18A	0.98	C18-H18B	0.98
C18-H18C	0.98		

Table A9 continued

Bond angles

C4-S1-H1S	97.9(19)	C9-N1-C8	117.9(2)
C9-N1-H1N	122.5(19)	C8-N1-H1N	118.3(19)
C9-O2-C10	120.5(2)	C14-O4-C15	122.5(2)
C7-C1-C8	113.9(2)	C7-C1-H1A	108.8
C8-C1-H1A	108.8	C7-C1-H1B	108.8
C8-C1-H1B	108.8	H1A-C1-H1B	107.7
C7-C2-C3	121.0(3)	C7-C2-H2	119.5
C3-C2-H2	119.5	C2-C3-C4	121.1(3)
C2-C3-H3	119.5	C4-C3-H3	119.5
C5-C4-C3	118.6(3)	C5-C4-S1	119.1(2)
C3-C4-S1	122.2(2)	C4-C5-C6	120.1(3)
C4-C5-H5	120.0	C6-C5-H5	120.0
C5-C6-C7	121.2(3)	C5-C6-H6	119.4
C7-C6-H6	119.4	C2-C7-C6	118.0(3)
C2-C7-C1	121.2(3)	C6-C7-C1	120.8(3)
N1-C8-C14	111.3(2)	N1-C8-C1	110.7(2)
C14-C8-C1	108.0(2)	N1-C8-H8	108.9
C14-C8-H8	108.9	C1-C8-H8	108.9
O1-C9-N1	123.6(3)	O1-C9-O2	126.1(3)
N1-C9-O2	110.3(2)	O2-C10-C12	109.7(2)
O2-C10-C13	110.4(2)	C12-C10-C13	113.1(3)
O2-C10-C11	102.1(2)	C12-C10-C11	110.5(3)
C13-C10-C11	110.5(3)	C10-C11-H11A	109.5
C10-C11-H11B	109.5	H11A-C11-H11B	109.5
C10-C11-H11C	109.5	H11A-C11-H11C	109.5
H11B-C11-H11C	109.5	C10-C12-H12A	109.5
C10-C12-H12B	109.5	H12A-C12-H12B	109.5
C10-C12-H12C	109.5	H12A-C12-H12C	109.5
H12B-C12-H12C	109.5	C10-C13-H13A	109.5
C10-C13-H13B	109.5	H13A-C13-H13B	109.5
C10-C13-H13C	109.5	H13A-C13-H13C	109.5
H13B-C13-H13C	109.5	O3-C14-O4	125.5(3)
O3-C14-C8	123.7(3)	O4-C14-C8	110.7(2)
O4-C15-C18	108.2(2)	O4-C15-C16	111.1(2)
C18-C15-C16	113.3(3)	O4-C15-C17	101.4(2)
C18-C15-C17	111.4(3)	C16-C15-C17	110.7(3)
C15-C16-H16A	109.5	C15-C16-H16B	109.5

Table A9 continued

H16A-C16-H16B	109.5	C15-C16-H16C	109.5
H16A-C16-H16C	109.5	H16B-C16-H16C	109.5
C15-C17-H17A	109.5	C15-C17-H17B	109.5
H17A-C17-H17B	109.5	C15-C17-H17C	109.5
H17A-C17-H17C	109.5	H17B-C17-H17C	109.5
C15-C18-H18A	109.5	C15-C18-H18B	109.5
H18A-C18-H18B	109.5	C15-C18-H18C	109.5
H18A-C18-H18C	109.5	H18B-C18-H18C	109.5

Table A10. Anisotropic atomic displacement parameters (Å2) for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	0.0415(5)	0.0539(5)	0.0528(5)	0.0165(4)	-0.0037(3)	-0.0077(3)
N1	0.0416(15)	0.0373(12)	0.0356(13)	-0.0081(10)	-0.0030(11)	0.0001(10)
O1	0.0444(12)	0.0533(12)	0.0416(12)	-0.0052(10)	-0.0105(9)	0.0055(10)
O2	0.0441(12)	0.0366(10)	0.0328(10)	-0.0069(8)	-0.0009(8)	0.0001(8)
O3	0.0477(13)	0.0472(11)	0.0366(11)	0.0022(9)	-0.0015(9)	-0.0050(9)
O4	0.0453(12)	0.0353(10)	0.0355(10)	-0.0012(8)	0.0001(8)	-0.0081(8)
C1	0.0507(19)	0.0396(15)	0.0444(17)	-0.0023(13)	0.0052(13)	-0.0026(13)
C2	0.056(2)	0.0451(16)	0.0385(16)	-0.0012(14)	0.0078(14)	-0.0053(14)
C3	0.0492(19)	0.0455(16)	0.0403(16)	-0.0024(14)	0.0001(13)	-0.0004(14)
C4	0.0411(17)	0.0374(14)	0.0422(16)	0.0007(13)	0.0025(12)	0.0016(12)
C5	0.0452(18)	0.0488(17)	0.0418(16)	0.0011(14)	0.0016(13)	0.0001(13)
C6	0.0451(18)	0.0462(16)	0.0443(17)	-0.0049(14)	-0.0018(13)	-0.0047(13)
C7	0.0460(18)	0.0399(15)	0.0439(17)	-0.0015(13)	0.0108(13)	0.0017(13)
C8	0.0445(18)	0.0391(14)	0.0382(15)	-0.0077(12)	0.0023(12)	-0.0060(12)
C9	0.0446(17)	0.0327(13)	0.0332(14)	-0.0004(12)	-0.0011(12)	-0.0023(12)
C10	0.0481(19)	0.0381(14)	0.0431(16)	-0.0047(13)	0.0050(13)	0.0022(13)
C11	0.064(2)	0.0491(17)	0.0435(17)	-0.0113(15)	0.0037(15)	0.0000(16)
C12	0.048(2)	0.063(2)	0.0532(19)	-0.0072(16)	0.0079(15)	-0.0059(16)
C13	0.066(2)	0.0403(16)	0.0536(19)	-0.0008(15)	0.0010(16)	0.0088(15)
C14	0.0415(17)	0.0360(14)	0.0366(15)	-0.0028(12)	0.0017(12)	-0.0002(12)
C15	0.0390(17)	0.0366(14)	0.0414(16)	0.0020(13)	-0.0008(12)	-0.0053(12)
C16	0.050(2)	0.073(2)	0.064(2)	0.0005(19)	0.0130(17)	-0.0067(18)
C17	0.058(2)	0.0517(18)	0.054(2)	0.0070(16)	-0.0173(16)	-0.0105(15)
C18	0.062(2)	0.0333(15)	0.057(2)	0.0001(14)	-0.0048(16)	-0.0036(14)

Table A11. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

	x/a	y/b	z/c	U(eq)
H1S	-0.591(3)	0.172(3)	0.1561(17)	0.054(9)
H1N	-0.037(3)	0.355(3)	0.0220(16)	0.039(8)
H1A	-0.0538	0.5744	0.1590	0.054
H1B	-0.1009	0.5574	0.0799	0.054
H2	-0.2889	0.4201	0.0454	0.056
H3	-0.4581	0.2774	0.0701	0.055
H5	-0.2898	0.2364	0.2631	0.055
H6	-0.1177	0.3766	0.2375	0.055
H8	0.0713	0.3600	0.1538	0.049
H11A	0.0370	-0.0225	-0.0956	0.079
H11B	0.1678	-0.0395	-0.1302	0.079
H11C	0.0870	0.1044	-0.1388	0.079
H12A	0.2817	0.2447	-0.0909	0.082
H12B	0.3635	0.1013	-0.0864	0.082
H12C	0.3610	0.1975	-0.0205	0.082
H13A	0.2787	0.0057	0.0469	0.081
H13B	0.2757	-0.1060	-0.0135	0.081
H13C	0.1462	-0.0735	0.0201	0.081
H16A	0.4553	0.5276	0.1269	0.092
H16B	0.5136	0.6788	0.1494	0.092
H16C	0.3958	0.6669	0.0899	0.092
H17A	0.3351	0.5948	0.2854	0.085
H17B	0.4726	0.6543	0.2693	0.085
H17C	0.4324	0.4950	0.2506	0.085
H18A	0.2329	0.8067	0.1427	0.077
H18B	0.3491	0.8528	0.1987	0.077
H18C	0.2197	0.7849	0.2213	0.077

Table A12. Torsion angles (°) for Boc-4-thiol-D,L-phenylalanine-*tert*-butyl ester.

C7-C2-C3-C4	1.1(5)	C2-C3-C4-C5	-1.1(5)
C2-C3-C4-S1	179.7(2)	C3-C4-C5-C6	0.5(4)
S1-C4-C5-C6	179.7(2)	C4-C5-C6-C7	0.1(5)
C3-C2-C7-C6	-0.6(4)	C3-C2-C7-C1	179.7(3)
C5-C6-C7-C2	-0.1(5)	C5-C6-C7-C1	179.7(3)
C8-C1-C7-C2	-104.3(3)	C8-C1-C7-C6	76.0(4)
C9-N1-C8-C14	62.1(3)	C9-N1-C8-C1	-177.7(2)
C7-C1-C8-N1	68.4(3)	C7-C1-C8-C14	-169.4(2)
C8-N1-C9-O1	0.3(4)	C8-N1-C9-O2	-178.8(2)
C10-O2-C9-O1	1.8(4)	C10-O2-C9-N1	-179.1(2)
C9-O2-C10-C12	-65.6(3)	C9-O2-C10-C13	59.7(3)
C9-O2-C10-C11	177.2(2)	C15-O4-C14-O3	2.9(4)
C15-O4-C14-C8	-174.5(2)	N1-C8-C14-O3	44.6(4)
C1-C8-C14-O3	-77.2(3)	N1-C8-C14-O4	-138.0(2)
C1-C8-C14-O4	100.3(3)	C14-O4-C15-C18	69.9(3)
C14-O4-C15-C16	-55.1(3)	C14-O4-C15-C17	-172.8(3)

Crystallographic data for *p*-thiocresol

Table A13. Crystallographic data and refinement details for *p*-thiocresol.

empirical formula	C_7H_8S	
formula weight	124.19	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Monoclinic, P1 $2_1/c1$	
Unit cell dimensions (Å, °)	a = 7.742(3)	α = 90
	b = 7.223(3)	β = 92.953(6)
	c = 5.956(2)	γ = 90
Volume ($Å^3$)	332.6(2)	
Z, Z′, calcd density (g/cm^3)	2, 0.5, 1.240	
absorption coefficient (mm^{-1})	0.371	
F(000)	132	
crystal size (mm)	0.141 x 0.204 x 0.230	
θ range for data collection	2.63 to 27.68°	
Index ranges	$-10 \le h \le 10$, $-9 \le k \le 9$, $-7 \le l \le 7$	
Reflections collected/ unique	4782/774 [R(int) = 0.0501]	
Coverage of independent reflections	99.7%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7456 and 0.6534	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	774 / 1 / 41	
Goodness-of-fit on F^2	1.077	
Final *R* indices	560 data; I>2σ(I)	R1 = 0.0575, wR2 = 0.1597
	all data	R1 = 0.0809, wR2 = 0.1745
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0974P)^2+0.0719P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.368 and -0.186 $eÅ^{-3}$	
R.M.S. deviation from mean	0.064 $eÅ^{-3}$	

Table A14. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for *p*-thiocresol.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
S1	0.3507(3)	0.4458(6)	0.7745(6)	0.0533(10)
C1	0.6804(14)	0.5443(17)	0.2107(19)	0.0323(11)
C2	0.839(3)	0.533(4)	0.370(5)	0.0323(11)
C3	0.839(3)	0.450(4)	0.580(5)	0.0341(12)
C4	0.993(3)	0.424(4)	0.706(5)	0.0325(11)
C5	0.148(3)	0.482(4)	0.621(5)	0.0323(11)
C6	0.149(3)	0.565(4)	0.411(5)	0.0341(12)
C7	0.994(3)	0.591(4)	0.285(5)	0.0325(11)

Table A15. Bond lengths [Å] and angles [°] for *p*-thiocresol.

Bond lengths

S1-C5	1.793(12)	S1-H1S	1.07(7)
C1-C2	1.514(11)	C1-H1A	0.98
C1-H1B	0.98	C1-H1C	0.98
C2-C3	1.39	C2-C7	1.39
C3-C4	1.39	C3-H3	0.95
C4-C5	1.39	C4-H4	0.95
C5-C6	1.39	C6-C7	1.39
C6-H6	0.95	C7-H7	0.95

Bond angles

C5-S1-H1S	109.(4)	C2-C1-H1A	109.5
C2-C1-H1B	109.5	H1A-C1-H1B	109.5
C2-C1-H1C	109.5	H1A-C1-H1C	109.5
H1B-C1-H1C	109.5	C3-C2-C7	120.0
C3-C2-C1	123.1(17)	C7-C2-C1	116.4(17)
C4-C3-C2	120.0	C4-C3-H3	120.0
C2-C3-H3	120.0	C3-C4-C5	120.0

Table A15 continued

C3-C4-H4	120.0	C5-C4-H4	120.0
C4-C5-C6	120.0	C4-C5-S1	121.4(12)
C6-C5-S1	118.6(13)	C5-C6-C7	120.0
C5-C6-H6	120.0	C7-C6-H6	120.0
C6-C7-C2	120.0	C6-C7-H7	120.0
C2-C7-H7	120.0		

Table A16. Anisotropic atomic displacement parameters (Å2) for *p*-thiocresol.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
S1	0.0358(12)	0.0682(15)	0.0550(14)	-0.0026(9)	-0.0068(9)	0.0079(10)
C1	0.038(3)	0.029(3)	0.0291(19)	-0.0008(19)	-0.0022(14)	0.0039(19)
C2	0.038(3)	0.029(3)	0.0291(19)	-0.0008(19)	-0.0022(14)	0.0039(19)
C3	0.037(3)	0.033(3)	0.033(2)	-0.001(2)	0.0075(15)	-0.001(2)
C4	0.045(3)	0.030(3)	0.022(2)	0.003(2)	0.0024(15)	0.002(2)
C5	0.038(3)	0.029(3)	0.0291(19)	-0.0008(19)	-0.0022(14)	0.0039(19)
C6	0.037(3)	0.033(3)	0.033(2)	-0.001(2)	0.0075(15)	-0.001(2)
C7	0.045(3)	0.030(3)	0.022(2)	0.003(2)		

Table A17. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for *p*-thiocresol.

	x/a	y/b	z/c	U(eq)
H1S	1.448(9)	0.529(8)	0.702(11)	0.064
H1A	0.7074	0.6171	0.0781	0.048
H1B	0.5862	0.6041	0.2872	0.048
H1C	0.6450	0.4193	0.1640	0.048
H3	0.7325	0.4100	0.6378	0.041
H4	0.9925	0.3673	0.8496	0.039
H6	1.2547	0.6048	0.3531	0.041
H7	0.9947	0.6476	0.1413	0.039

Table A18. Torsion angles (°) for *p*-thiocresol.

C7-C2-C3-C4	0	C1-C2-C3-C4	172.3(17)
C2-C3-C4-C5	0	C3-C4-C5-C6	0
C3-C4-C5-S1	-178.2(12)	C4-C5-C6-C7	0
S1-C5-C6-C7	178.2(11)	C5-C6-C7-C2	0
C3-C2-C7-C6	0	C1-C2-C7-C6	-172.8(15)

Appendix B

FULL IR SPECTRA OF COMPOUNDS IN ALL SOLVENTS

Compounds were described and characterized in Chapter 2: Insights into S–H/π Aromatic Interactions: Studies on Boc-4-Thiol-L-Phenylalanine-*tert*-Butyl Ester via IR Spectroscopy, X-Ray Crystallography, and *ab initio* Calculations

All spectra in solution samples were taken in triplicate, background subtracted, baseline corrected, and averaged. Error bars represent standard error. Data is shown as absorbance (top) and transmittance (bottom).

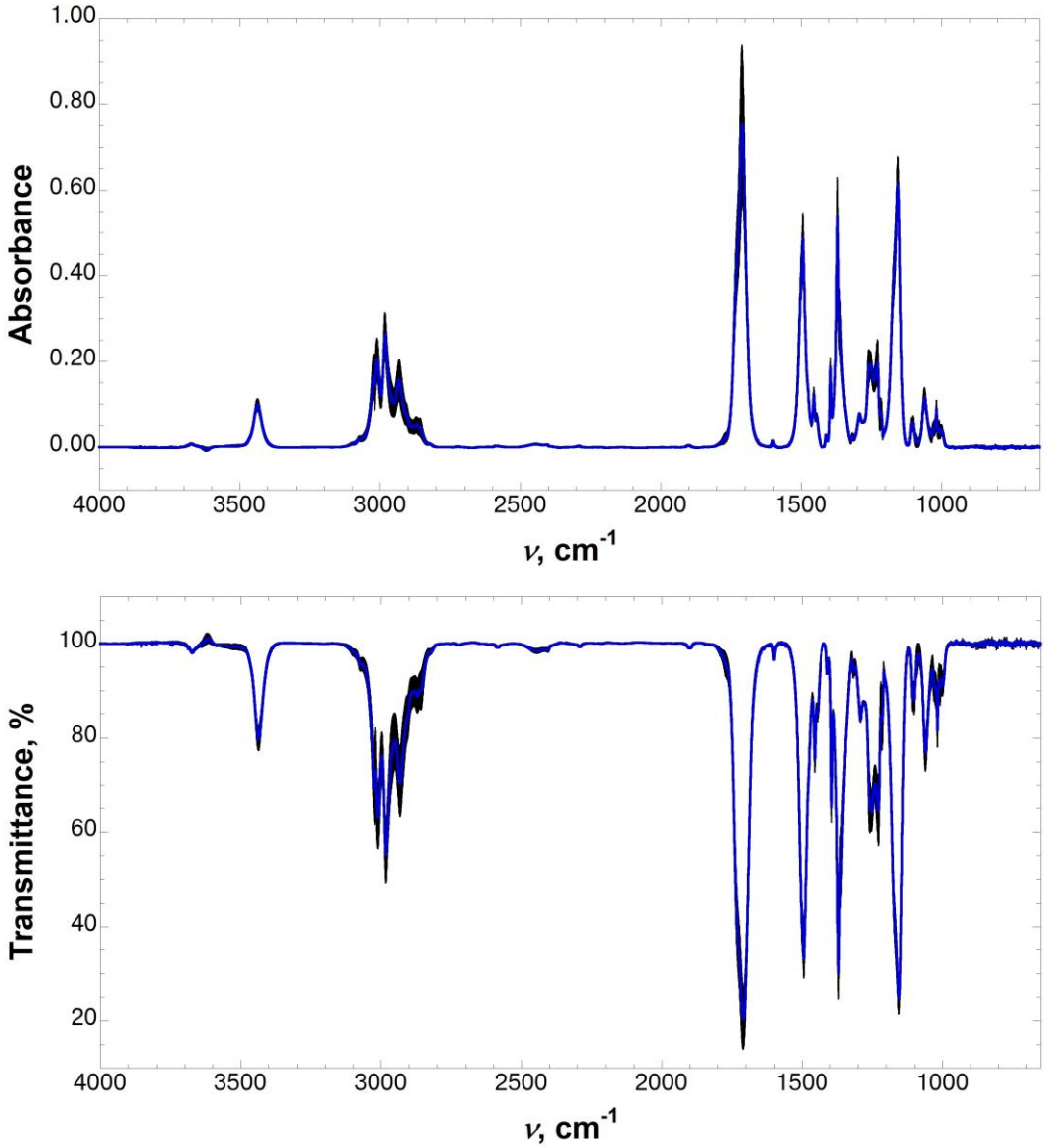


Figure B1 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in $CHCl_3$.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2585 cm^{-1}.

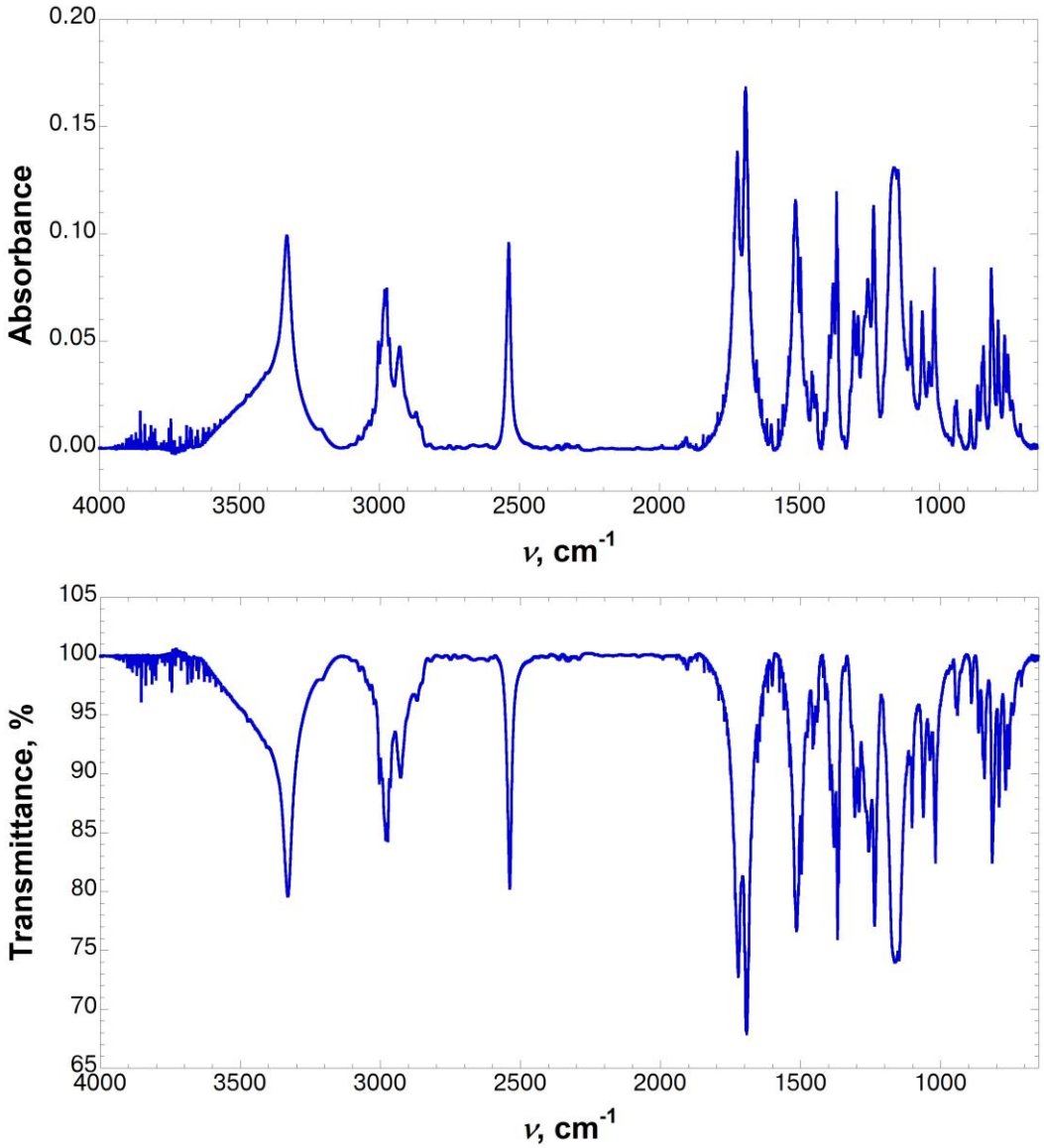


Figure B2 FT-IR spectrum of crystalline Boc-4-thiol-L-phenylalanine-*tert*-butyl ester in a pressed KBr pellet, crystallized from 25% ethyl acetate in hexanes (v/v) via slow evaporation at room temperature.

The spectrum was baseline corrected. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2538 cm^{-1}.

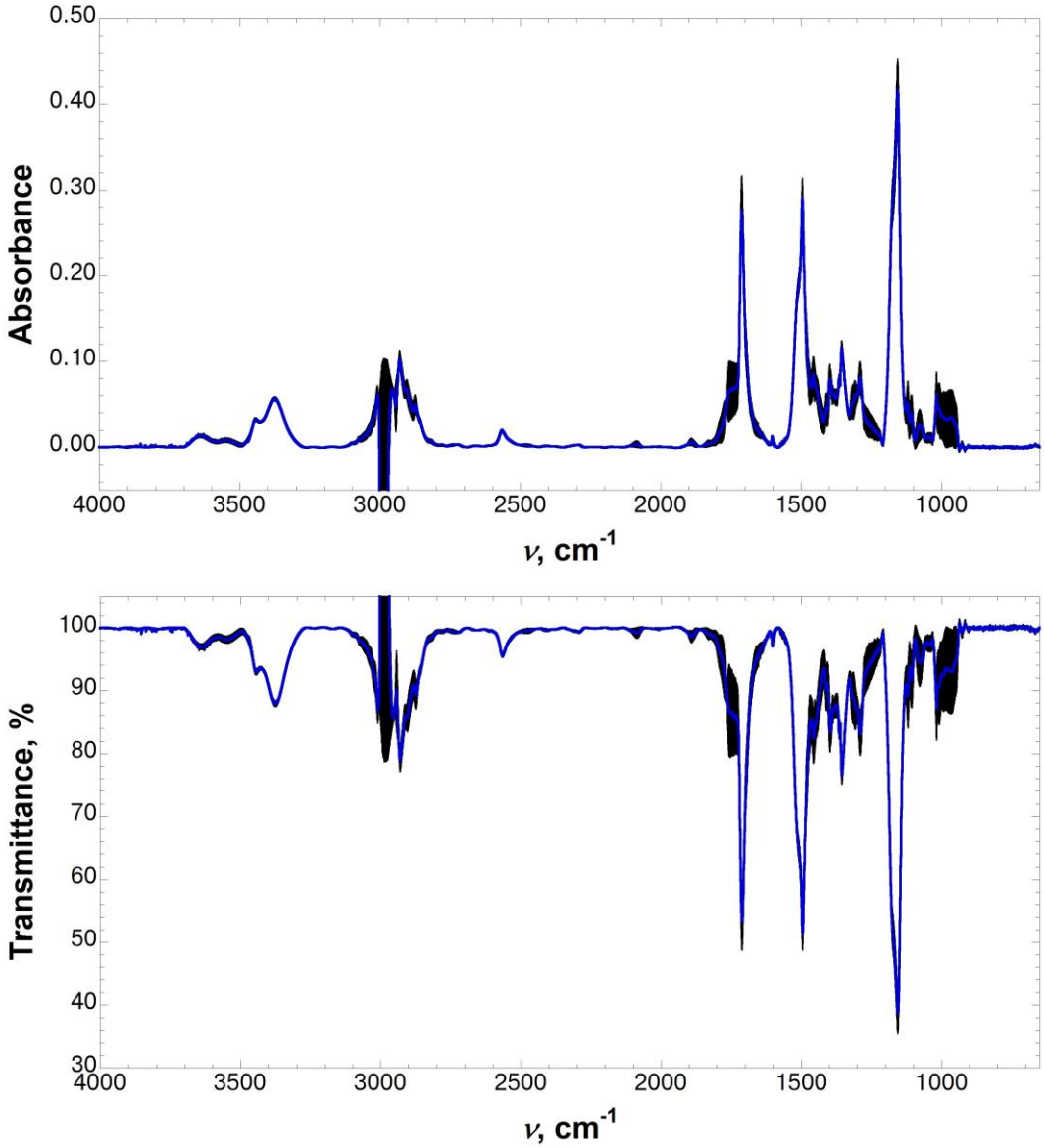


Figure B3 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in ethyl acetate.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2566 cm^{-1}.

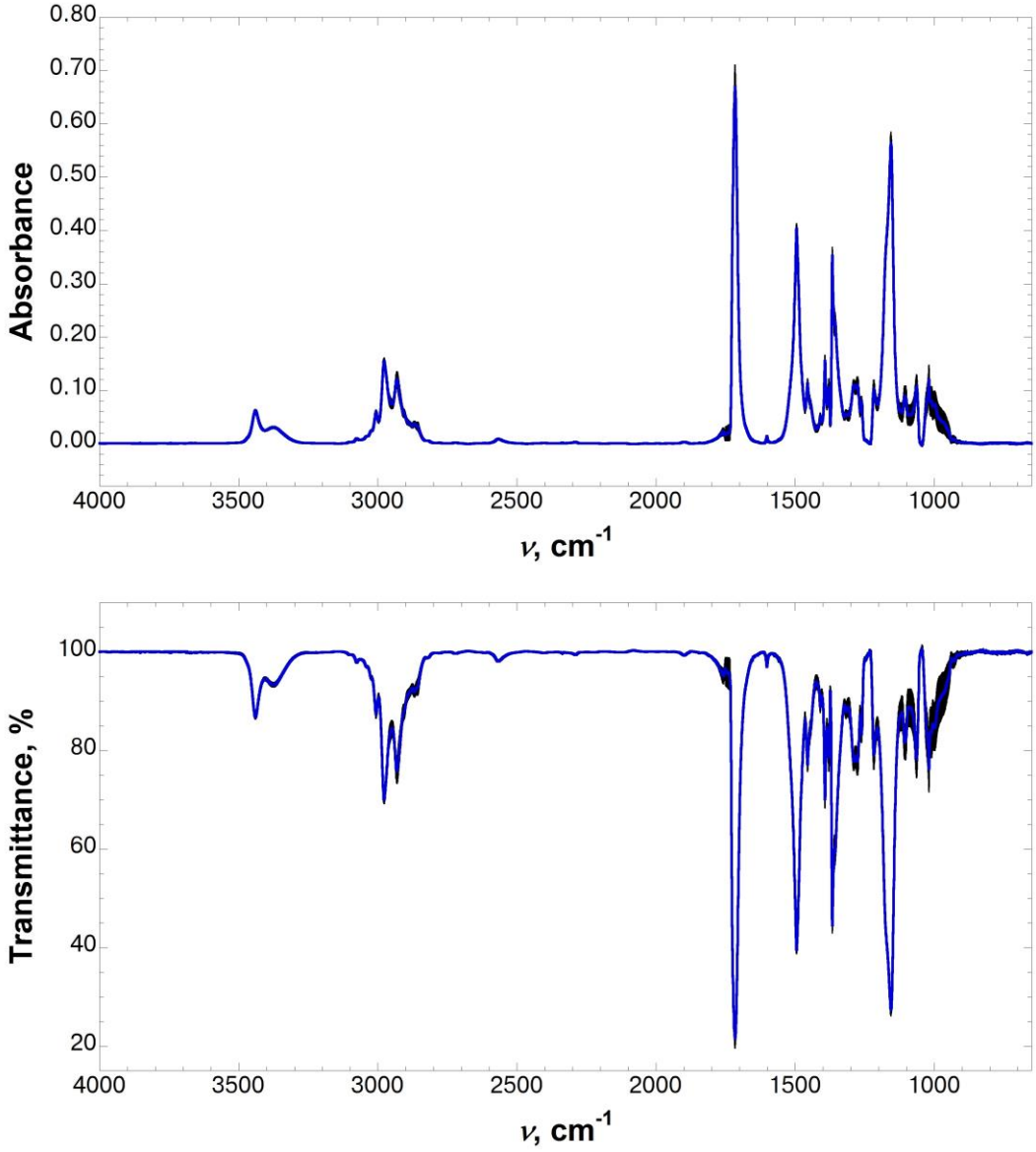


Figure B4 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 25% ethyl acetate in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2567 cm^{-1}.

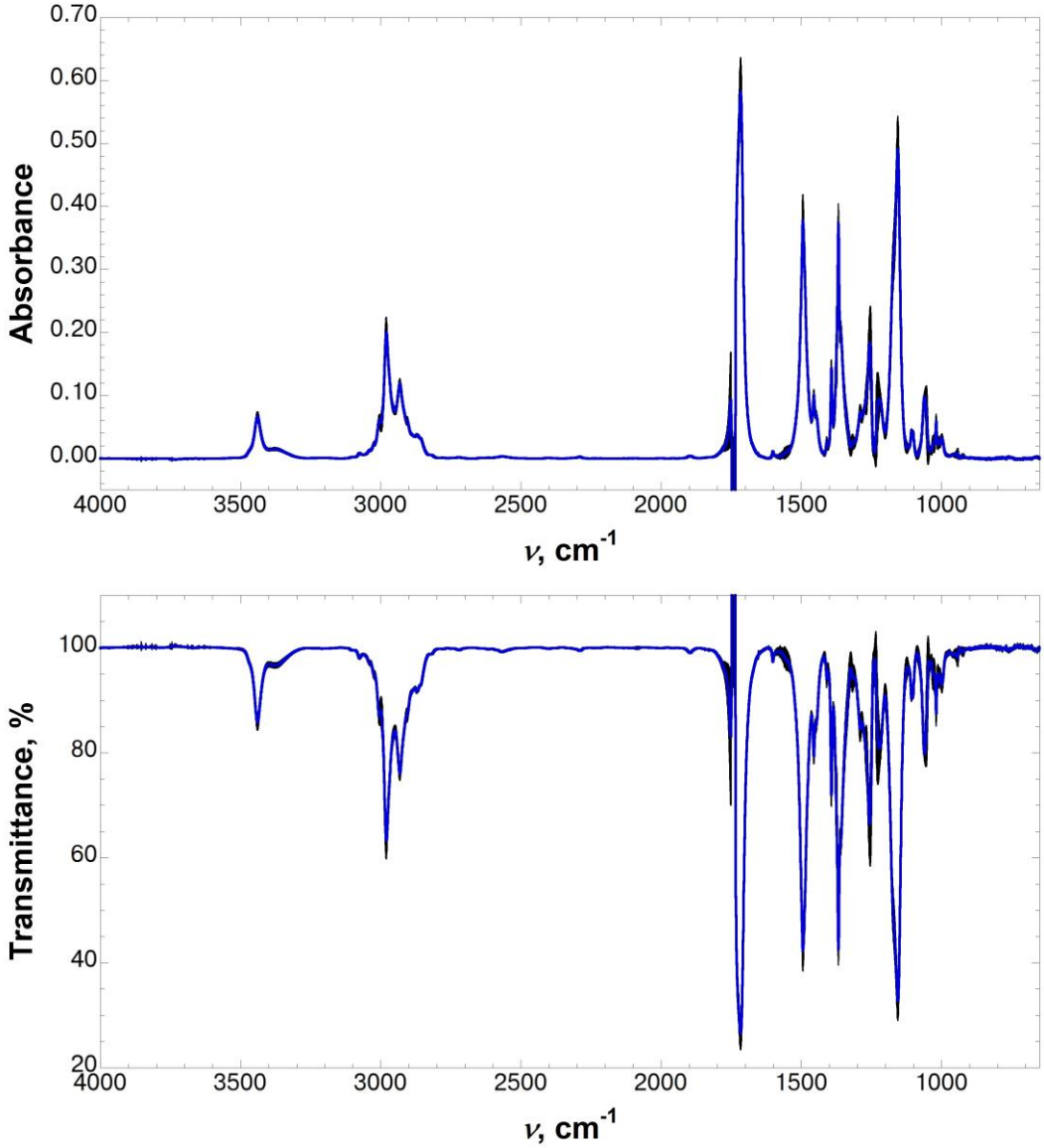


Figure B5 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 10% ethyl acetate in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2569 cm^{-1}.

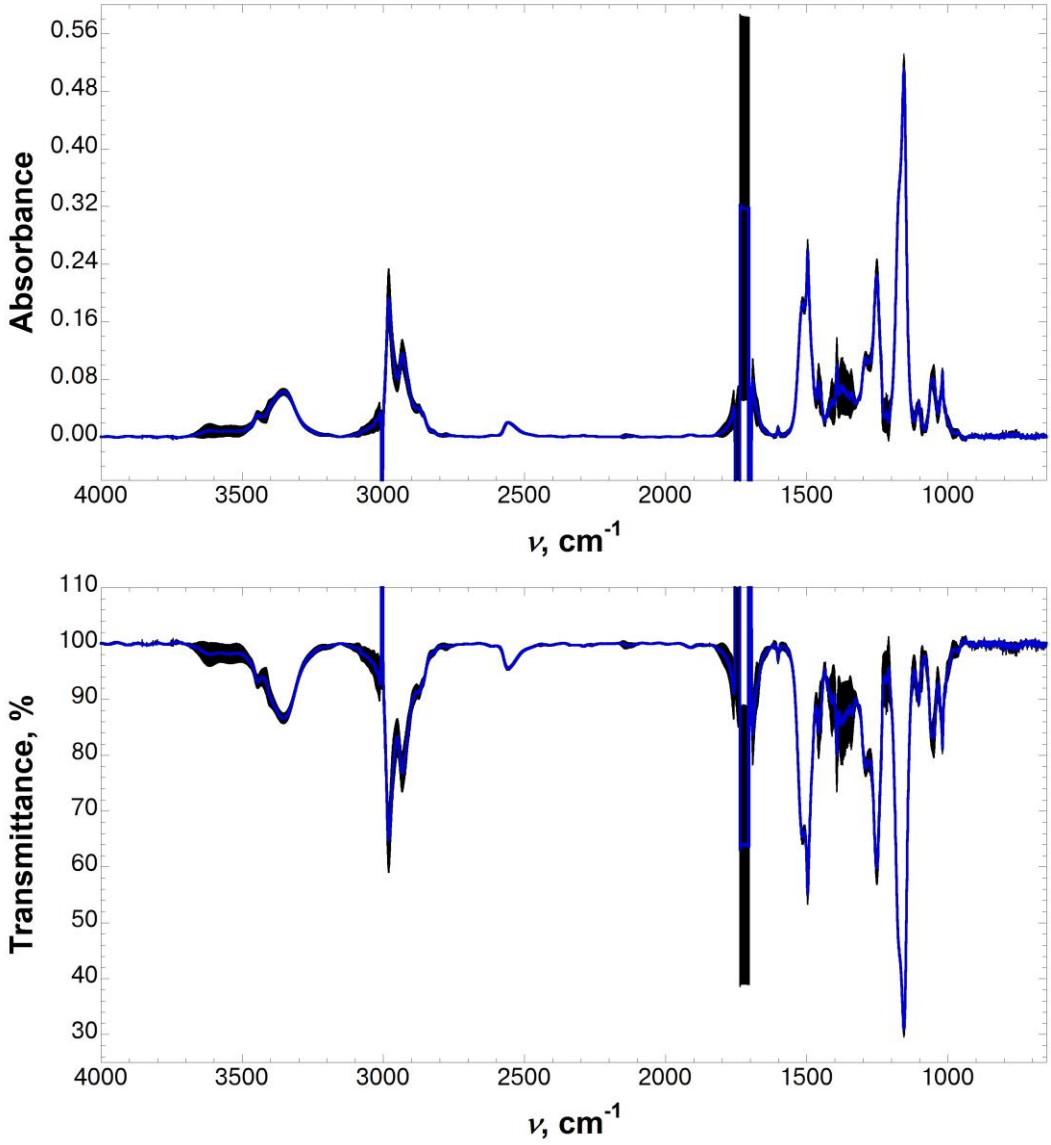


Figure B6 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in acetone.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2558 cm^{-1}.

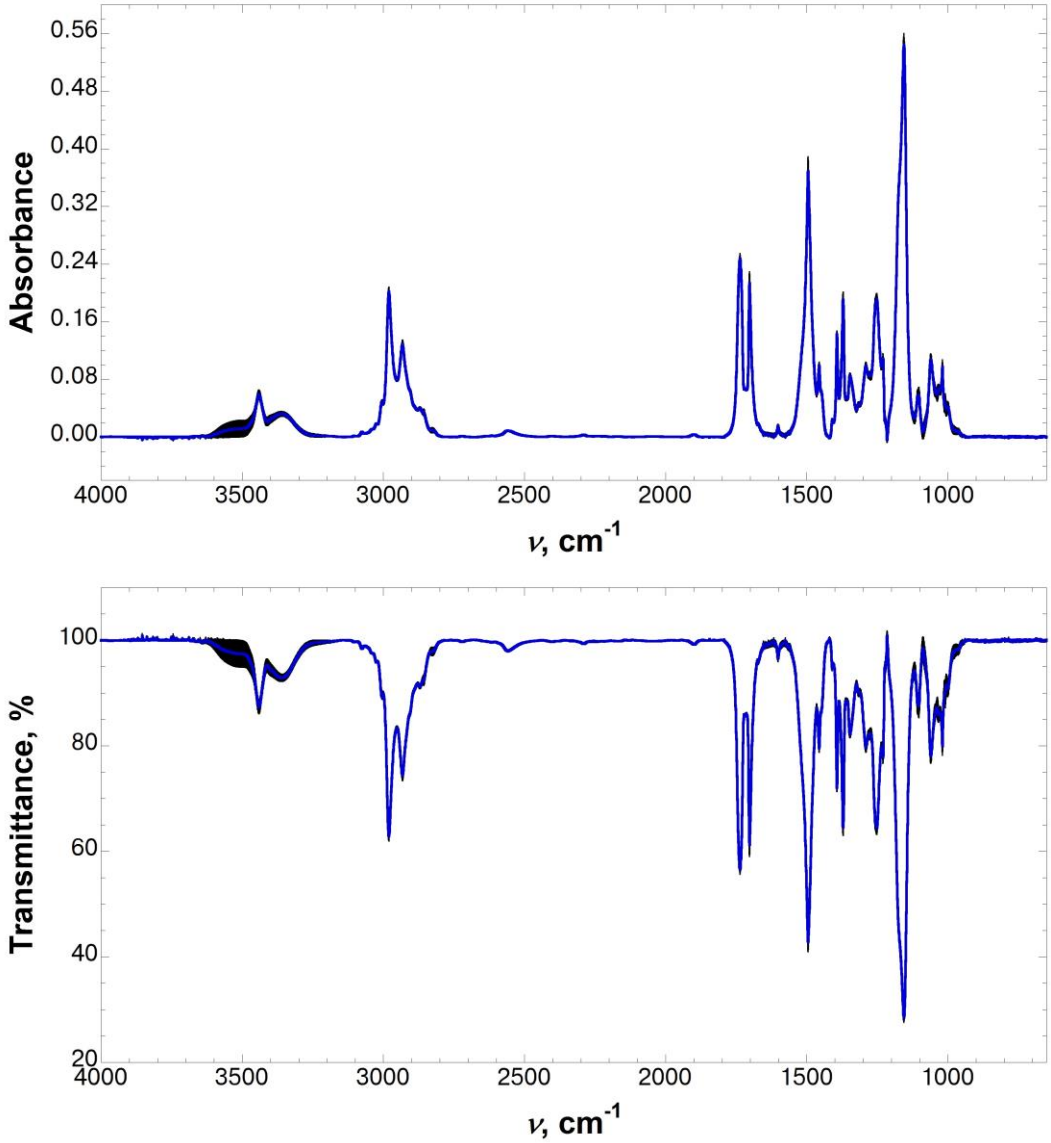


Figure B7 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 25% acetone in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2559 cm^{-1}.

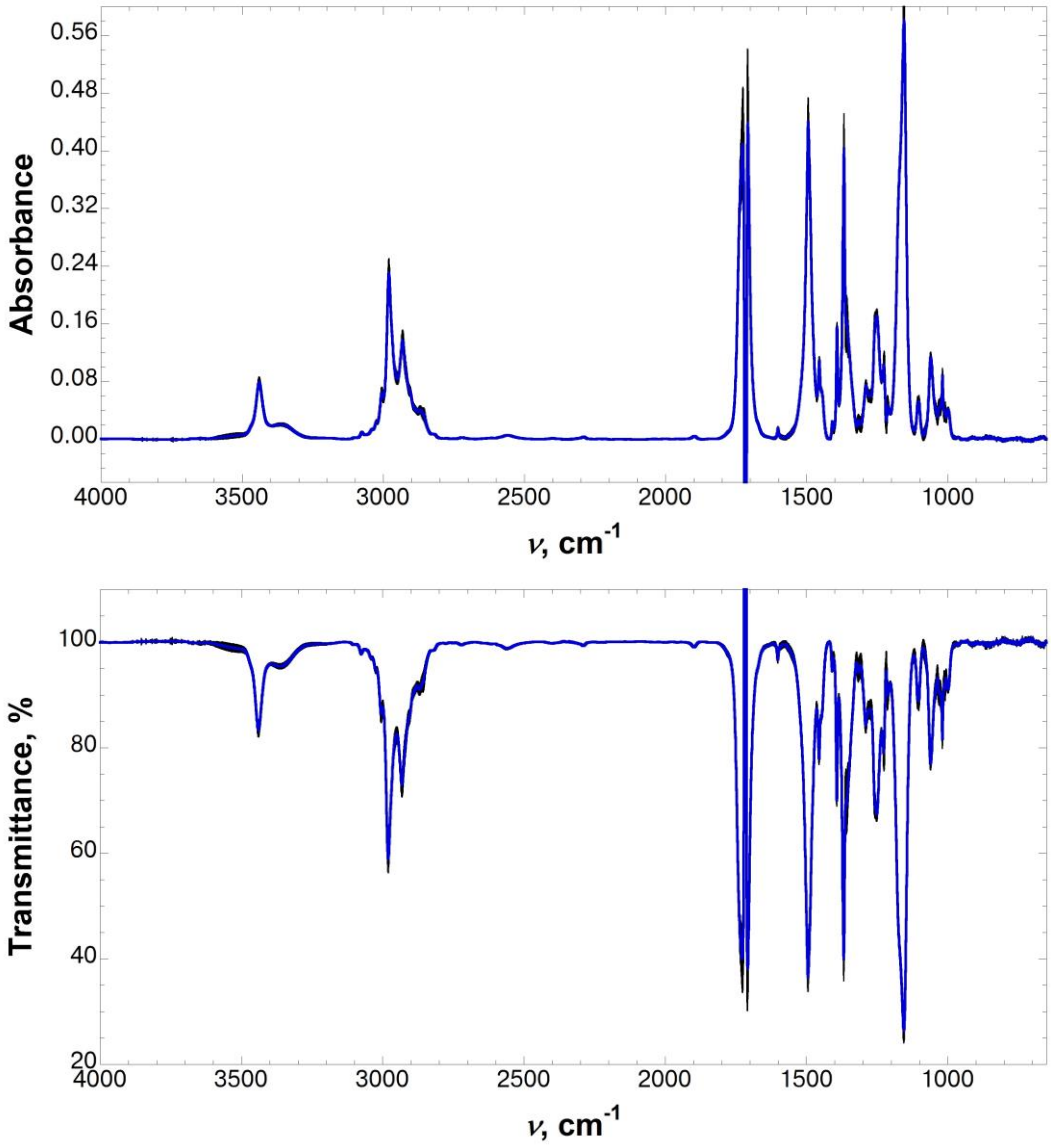


Figure B8 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 10% acetone in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2559 cm^{-1}.

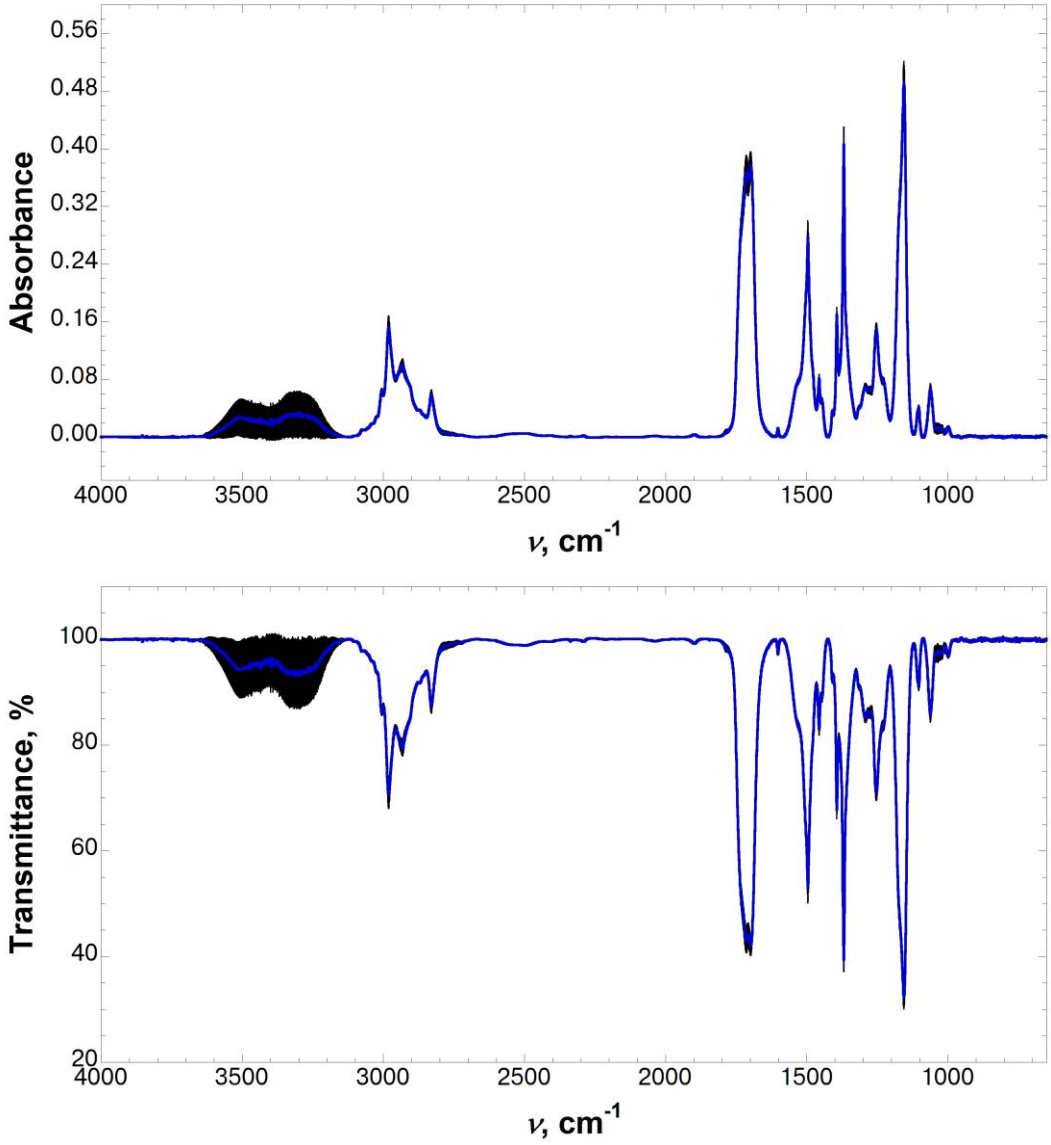


Figure B9 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 25% methanol in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2497 cm^{-1}.

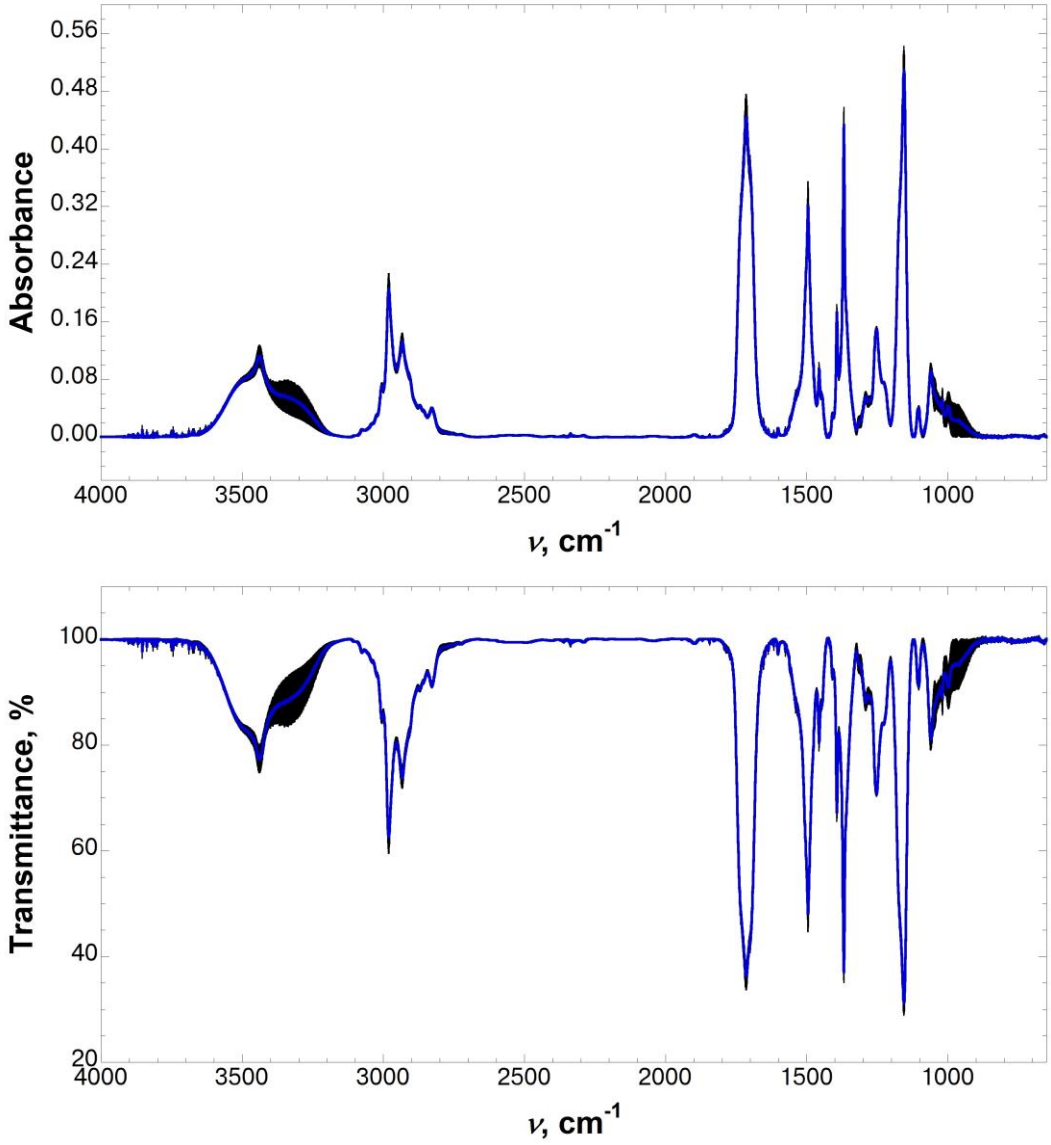


Figure B10 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 10% methanol in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum.

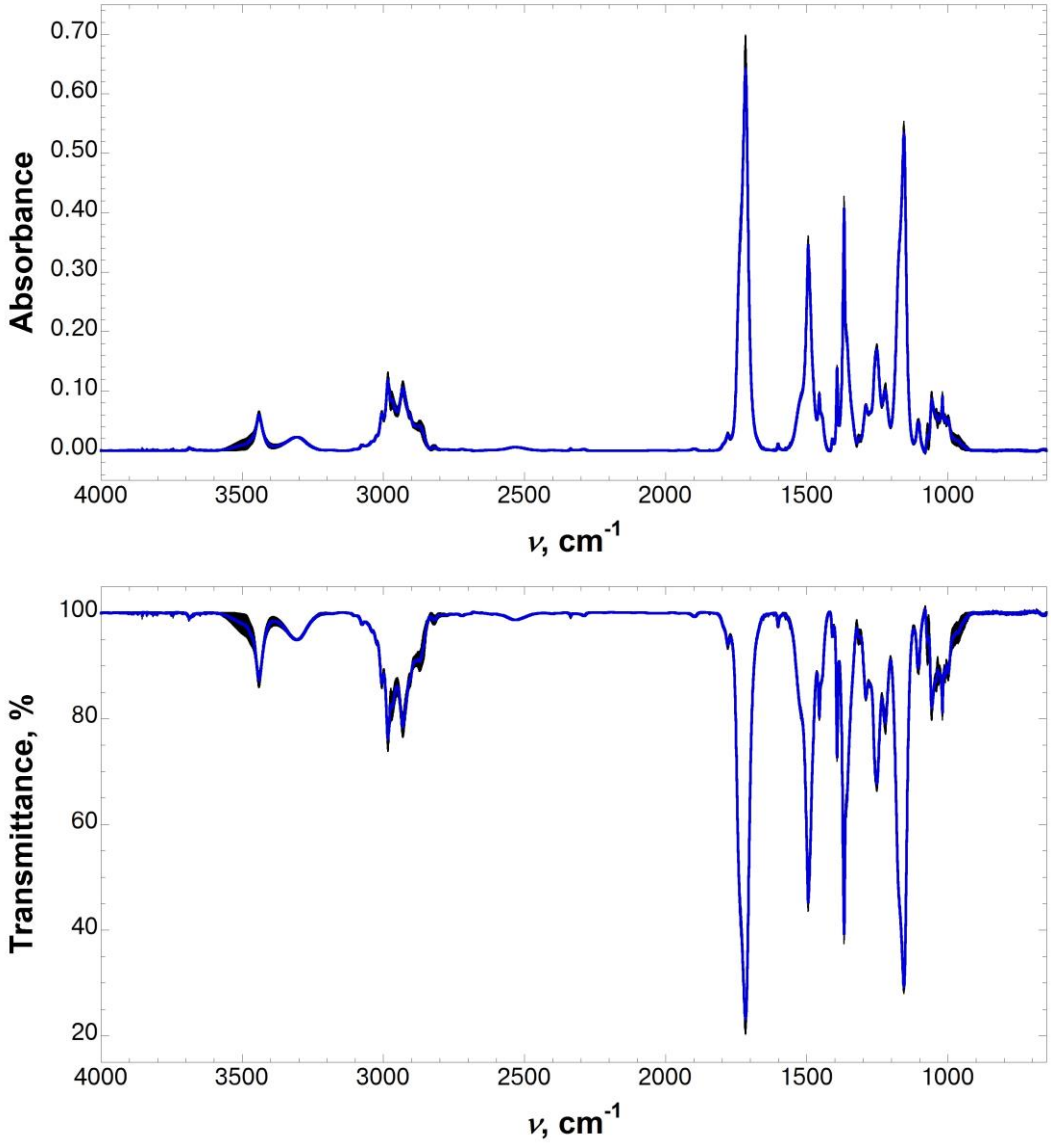


Figure B11 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 25% THF in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2534 cm^{-1}.

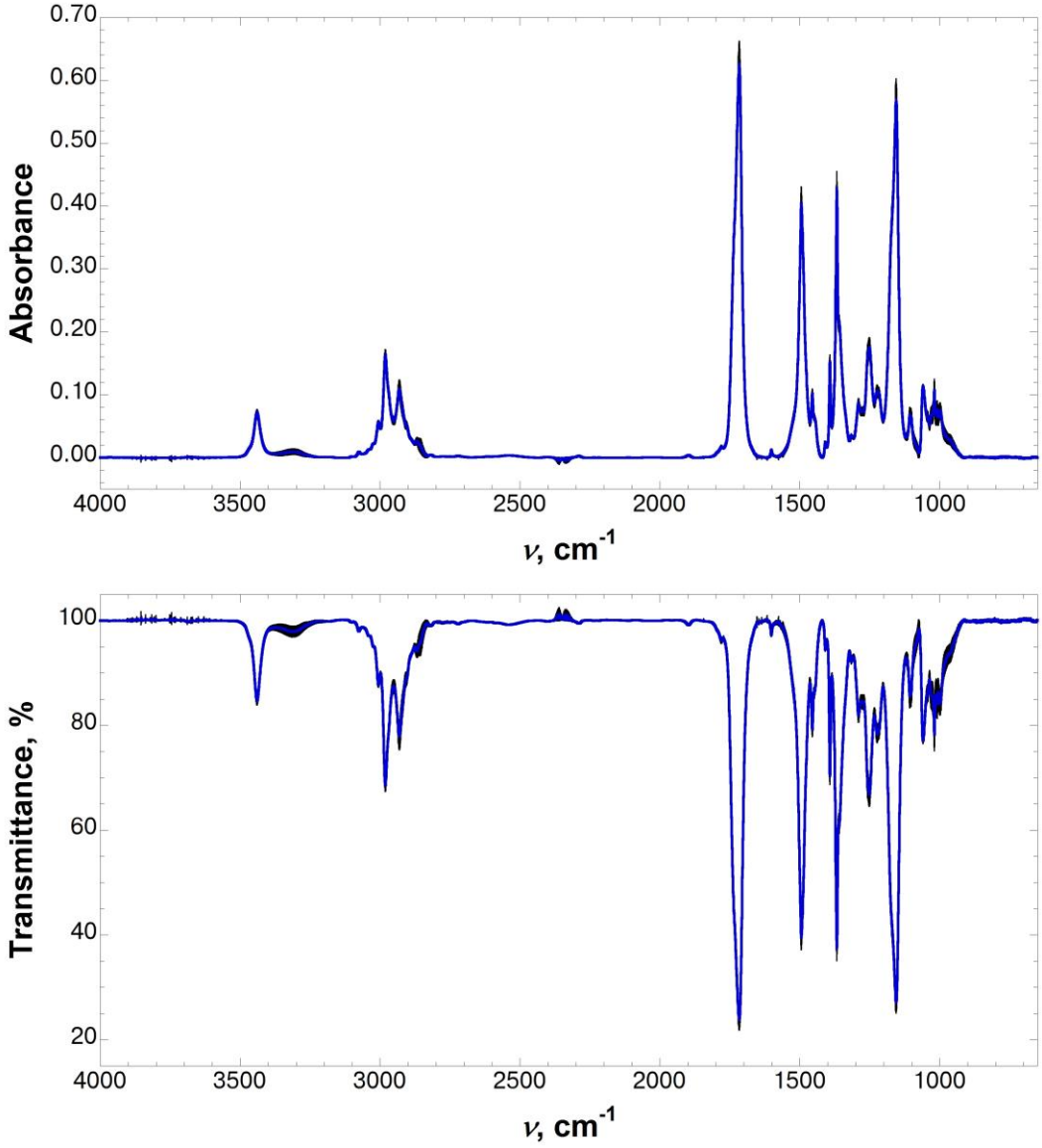


Figure B12 FT-IR spectrum of Boc-4-thiol-L-phenylalanine-*tert*-butyl ester (200 mM) in 10% THF in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2541 cm^{-1}.

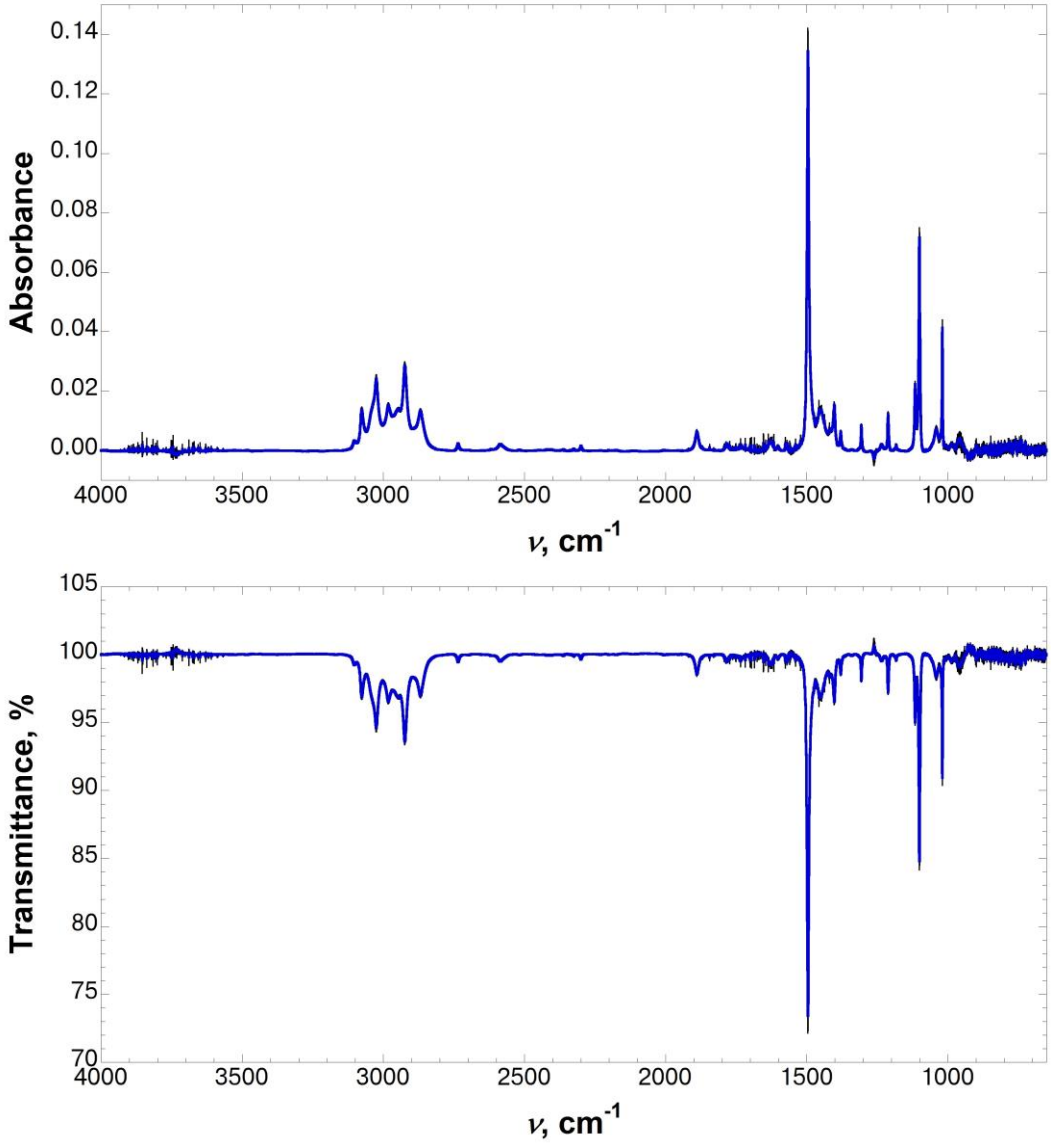


Figure B13 FT-IR spectra of *p*-thiocresol (200 mM) in CCl_4.
Four independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2586 cm^{-1}.

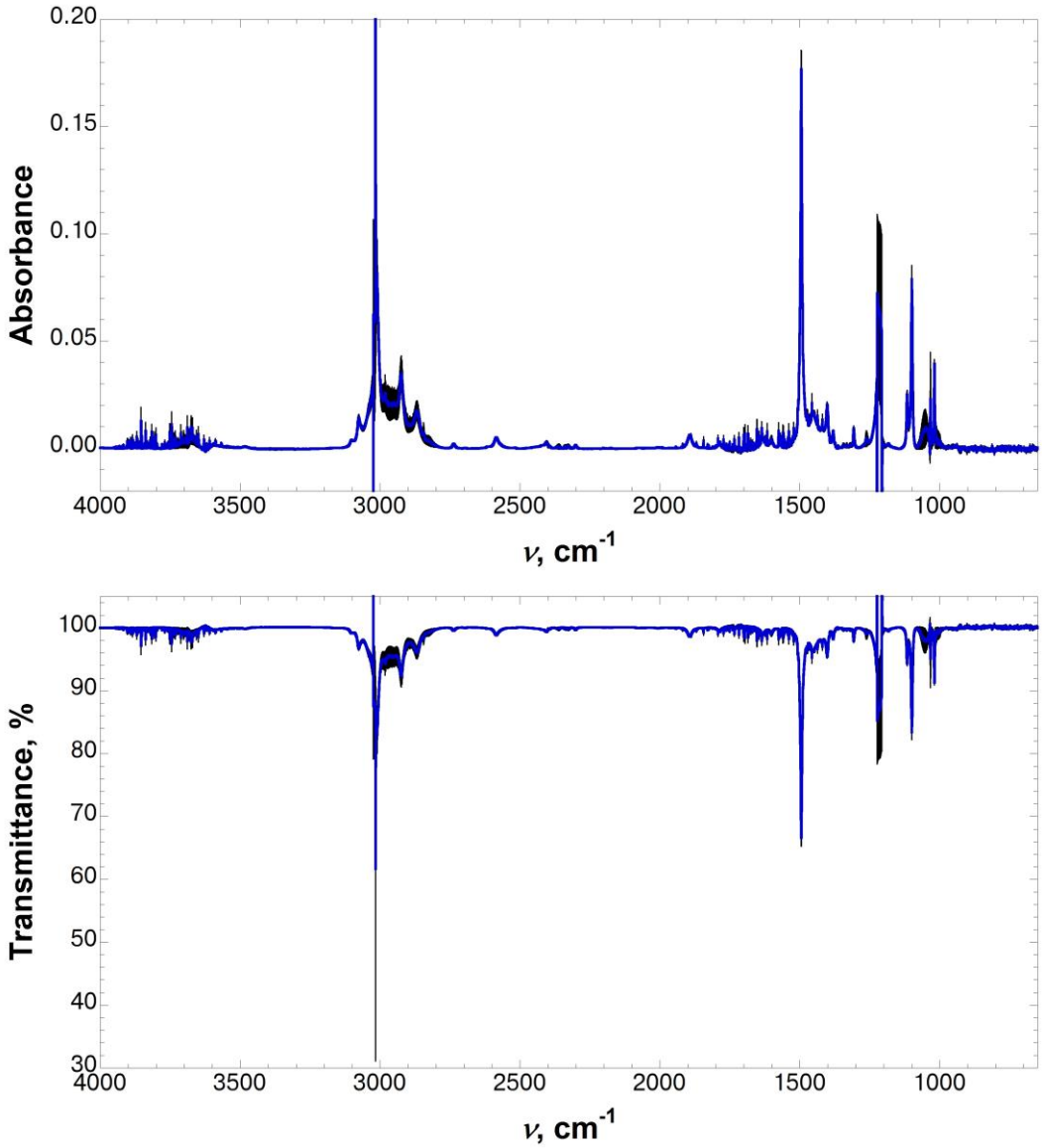


Figure B14 FT-IR spectrum of *p*-thiocresol (200 mM) in $CHCl_3$.
Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2585 cm^{-1}.

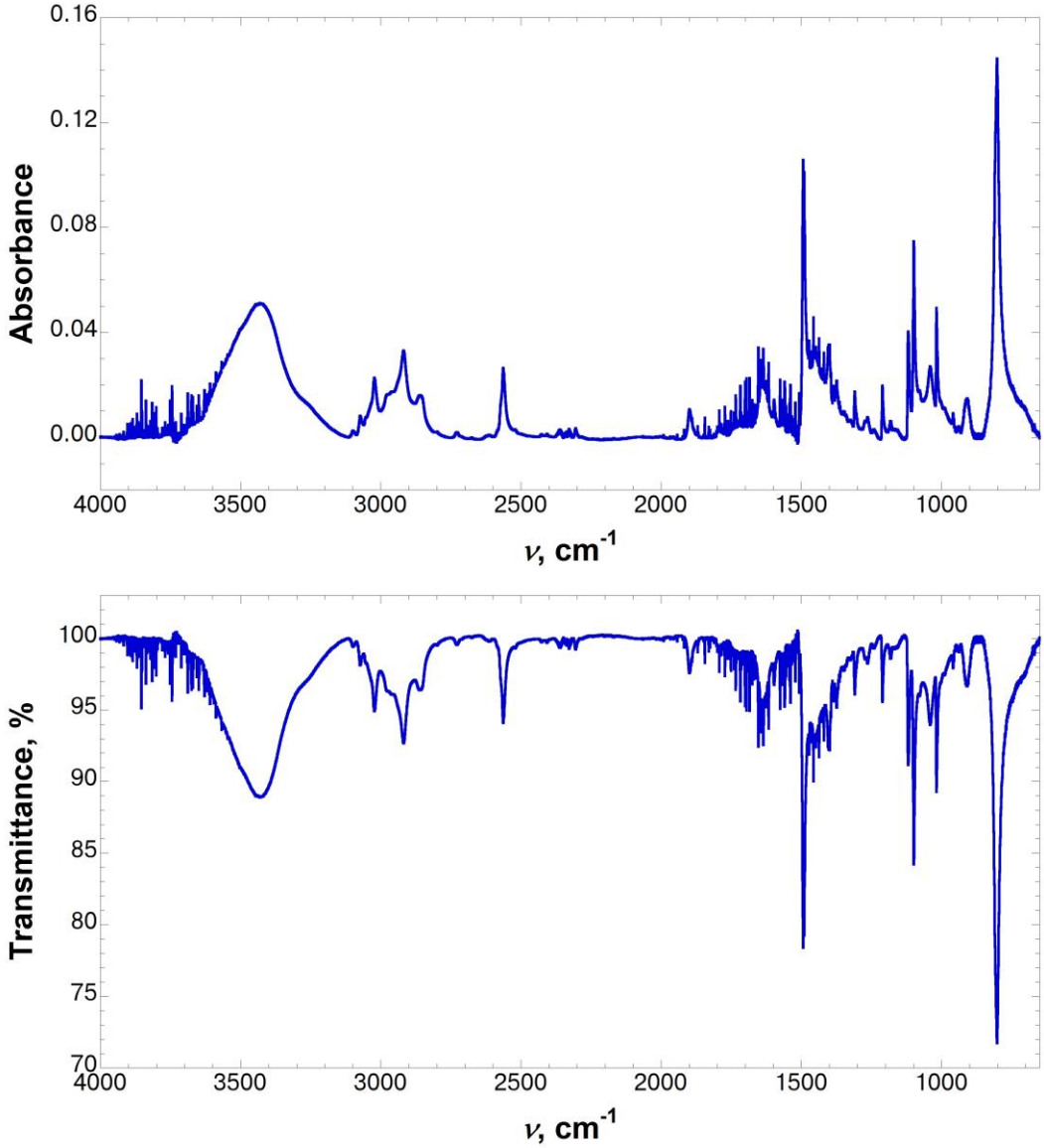


Figure B15 FT-IR spectrum of crystalline *p*-thiocresol in a pressed KBr pellet. Crystalline *p*-thiocresol was not recrystallized, and was used as purchased. The spectrum was baseline corrected. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2563 cm^{-1}.

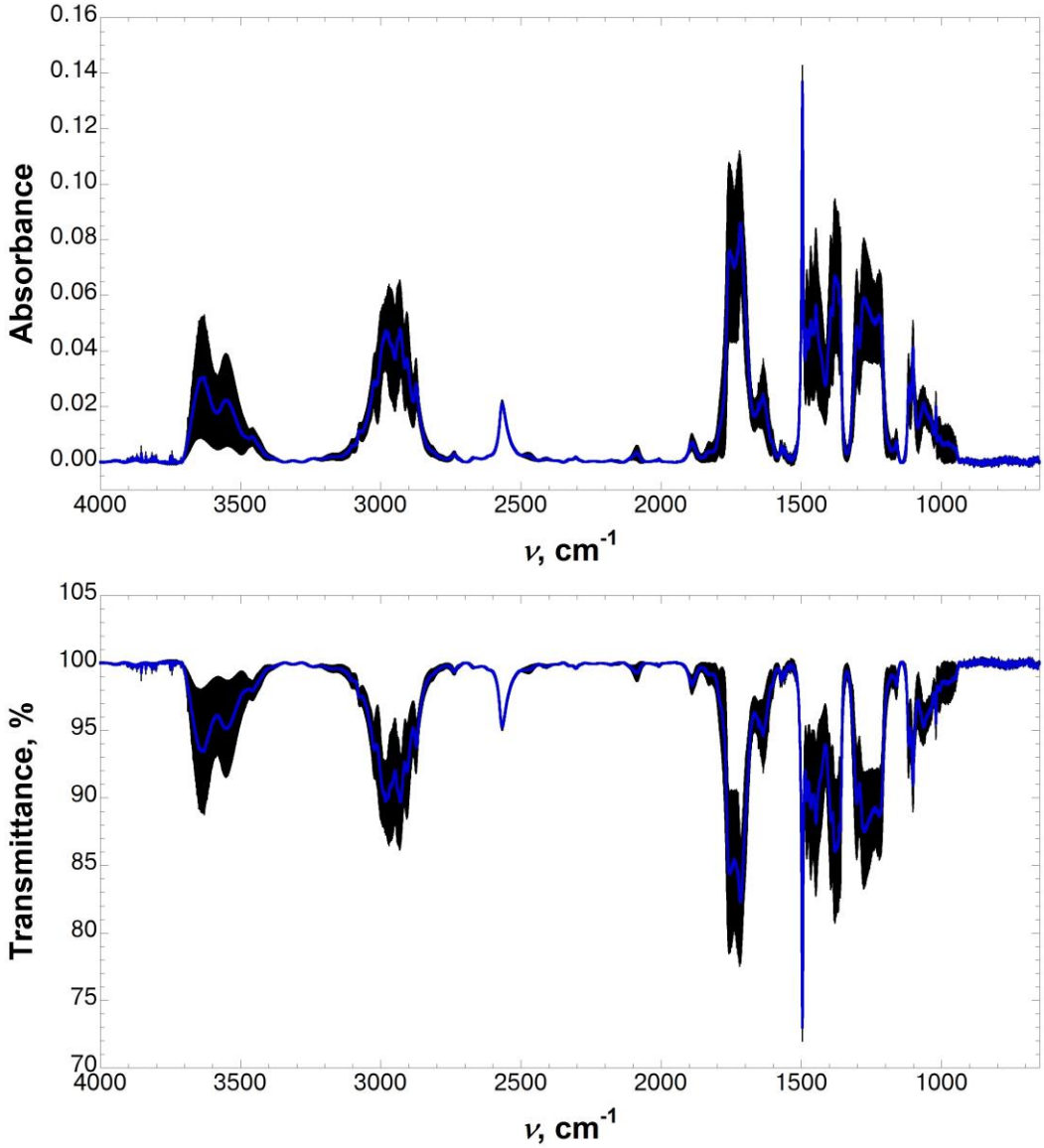


Figure B16 FT-IR spectrum of *p*-thiocresol (200 mM) in ethyl acetate.
Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2567 cm^{-1}.

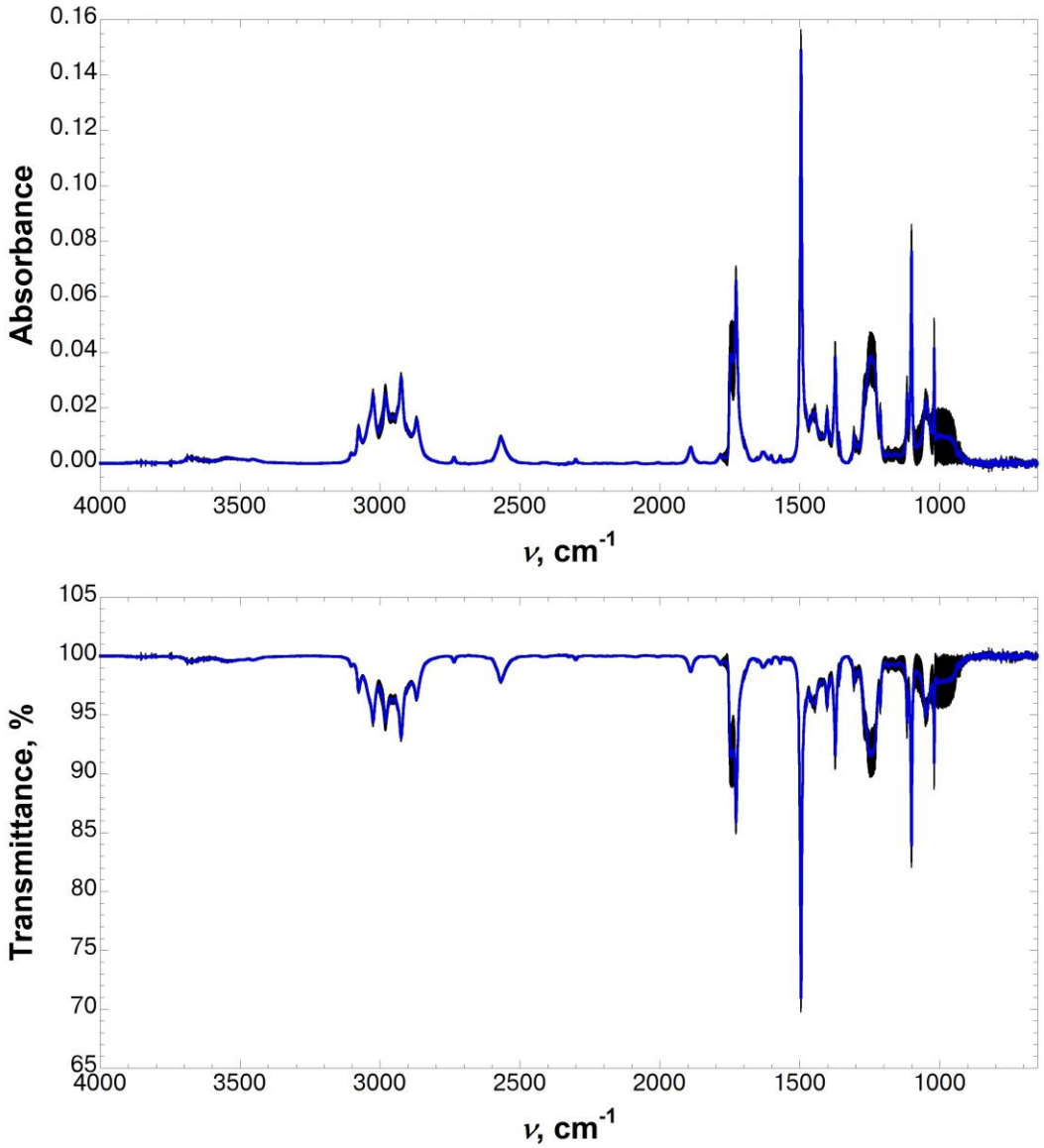


Figure B17 FT-IR spectrum of *p*-thiocresol (200 mM) in 25% ethyl acetate in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2569 cm^{-1}.

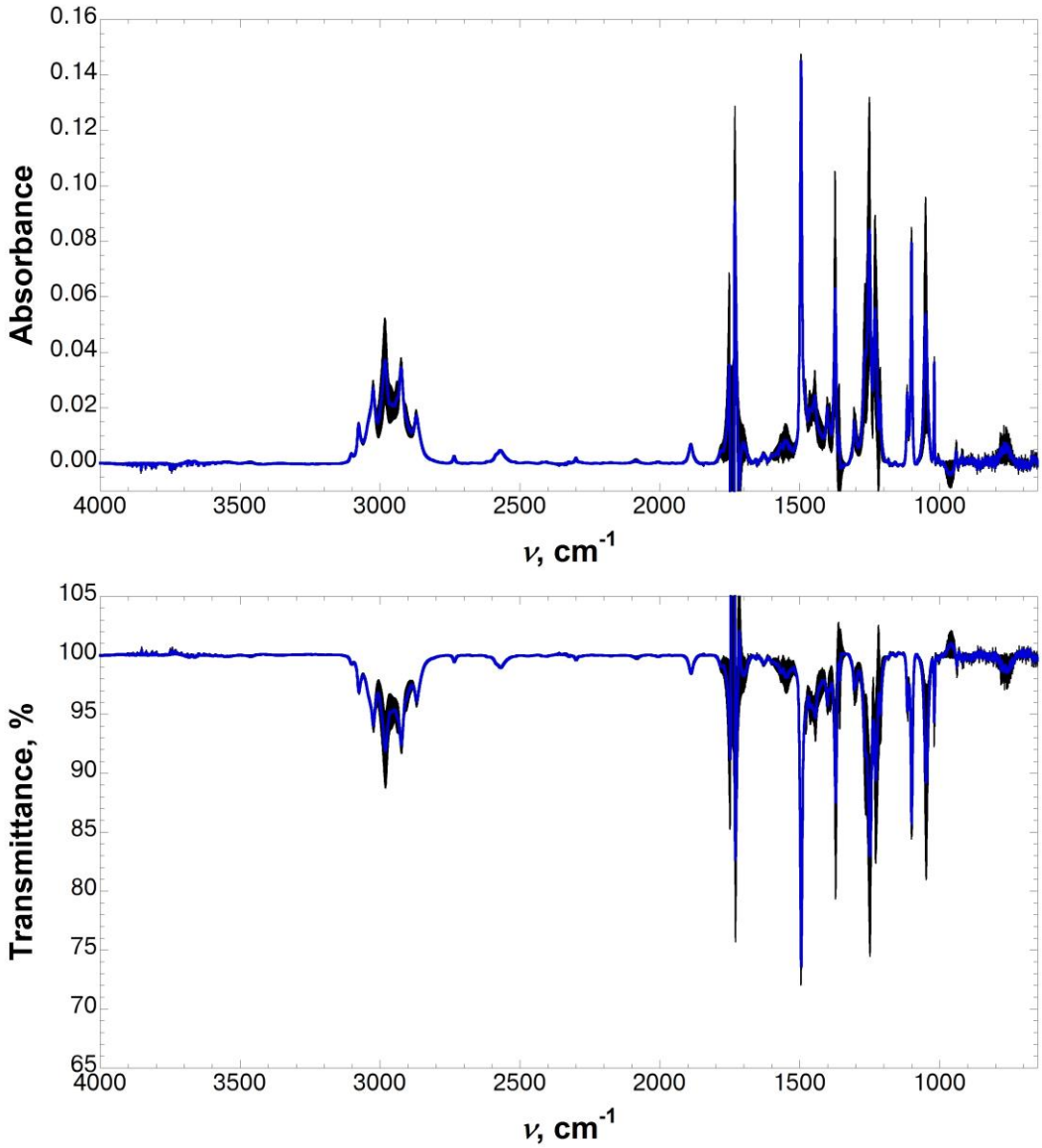


Figure B18 FT-IR spectrum of *p*-thiocresol (200 mM) in 10% ethyl acetate in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2571 cm^{-1}.

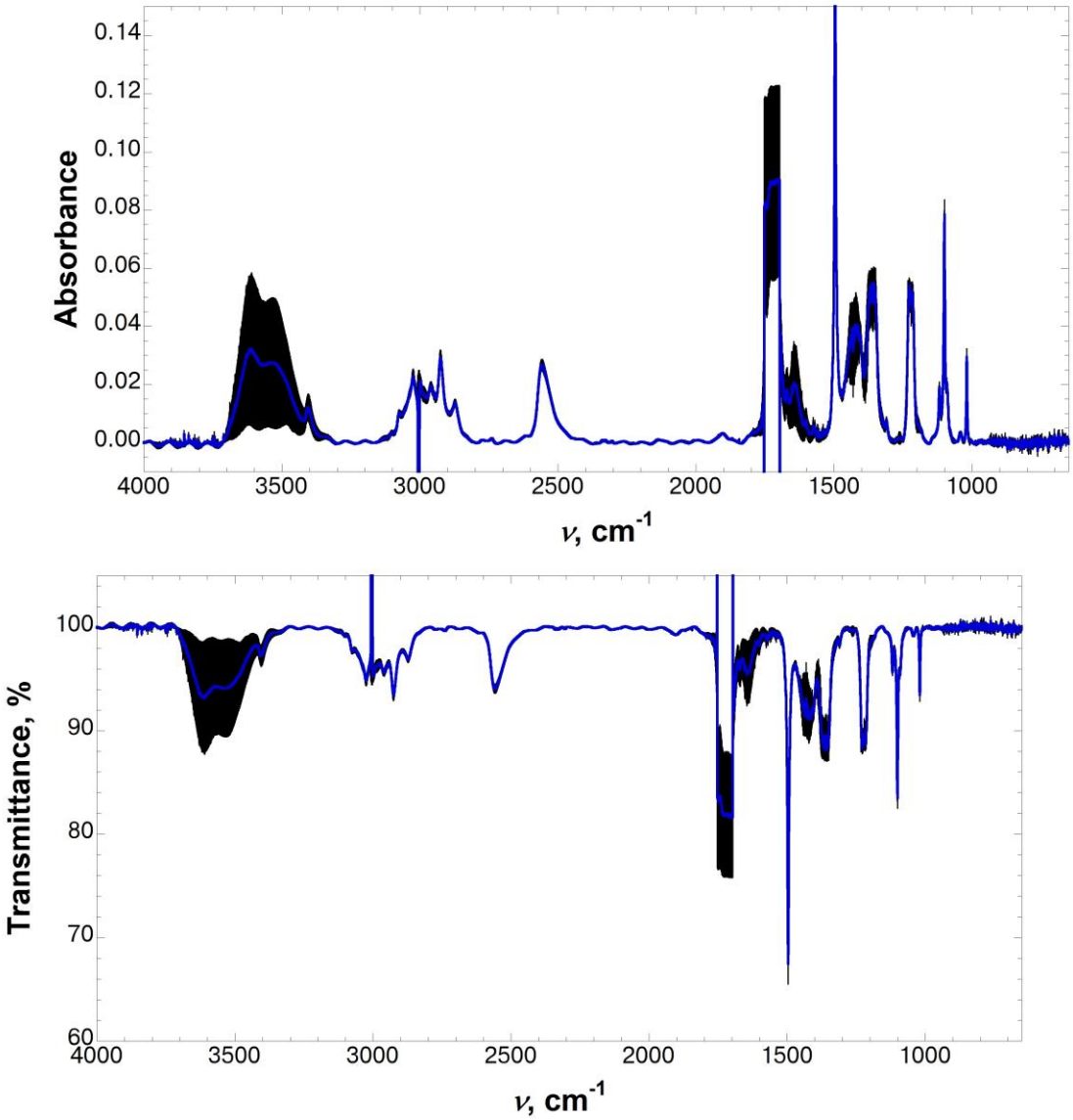


Figure B19 FT-IR spectrum of *p*-thiocresol (200 mM) in acetone.
Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2558 cm^{-1}.

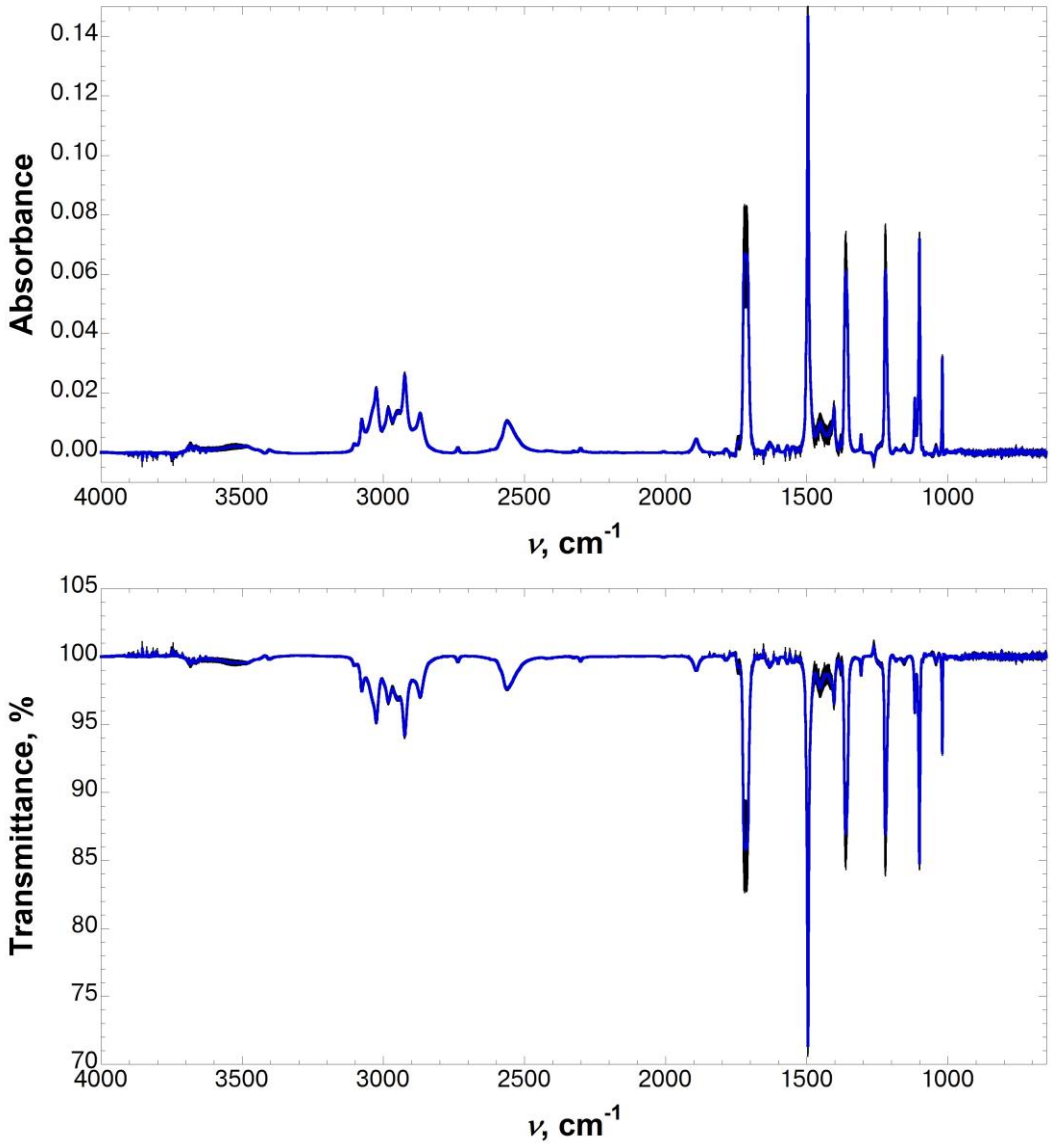


Figure B20 FT-IR spectrum of *p*-thiocresol (200 mM) in 25% acetone in CCl_4. Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2561 cm^{-1}.

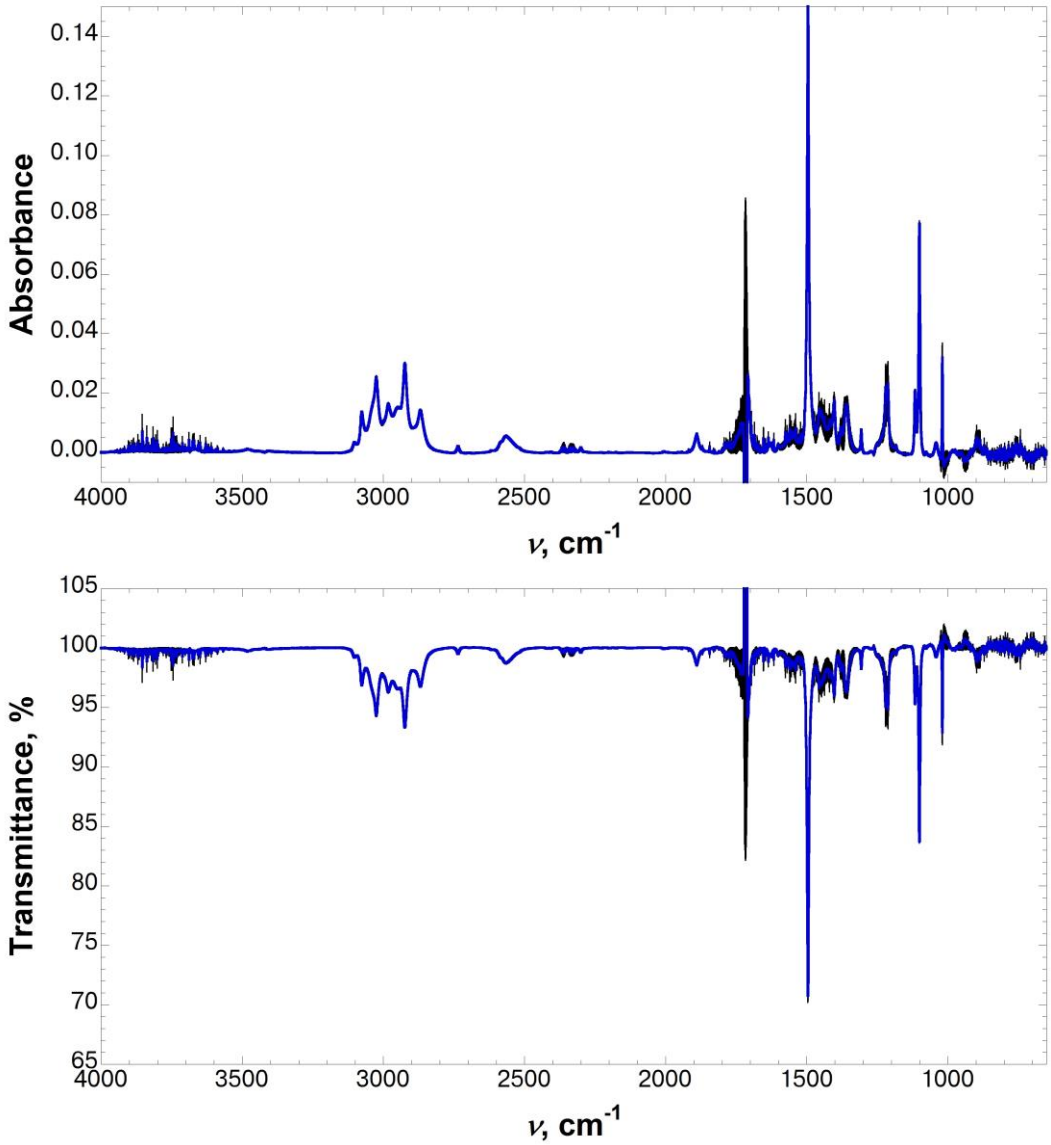


Figure B21 FT-IR spectrum of *p*-thiocresol (200 mM) in 10% acetone in CCl_4. Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2565 cm^{-1}.

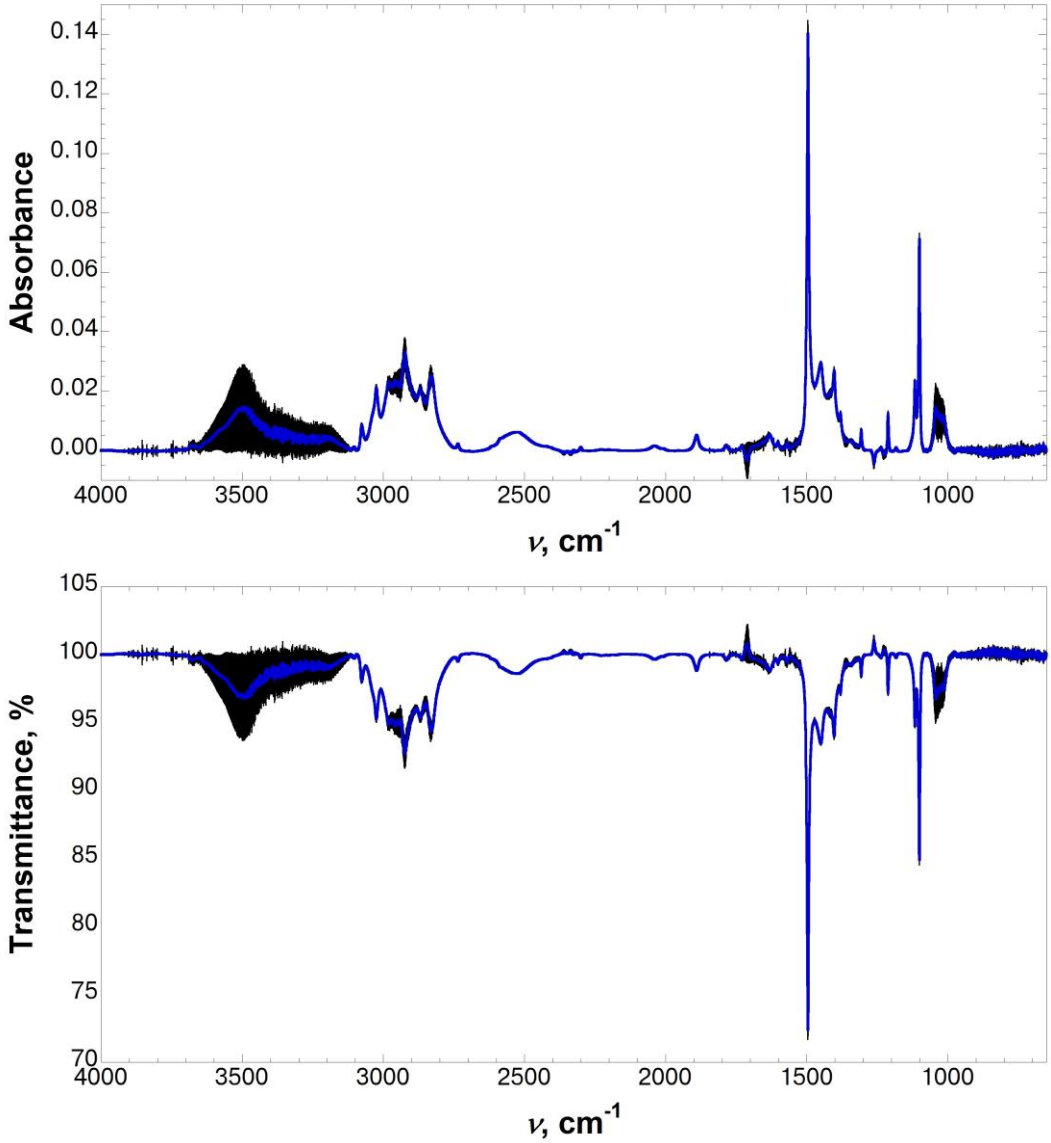


Figure B22 FT-IR spectrum of *p*-thiocresol (200 mM) in 25% methanol in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2521 cm^{-1}.

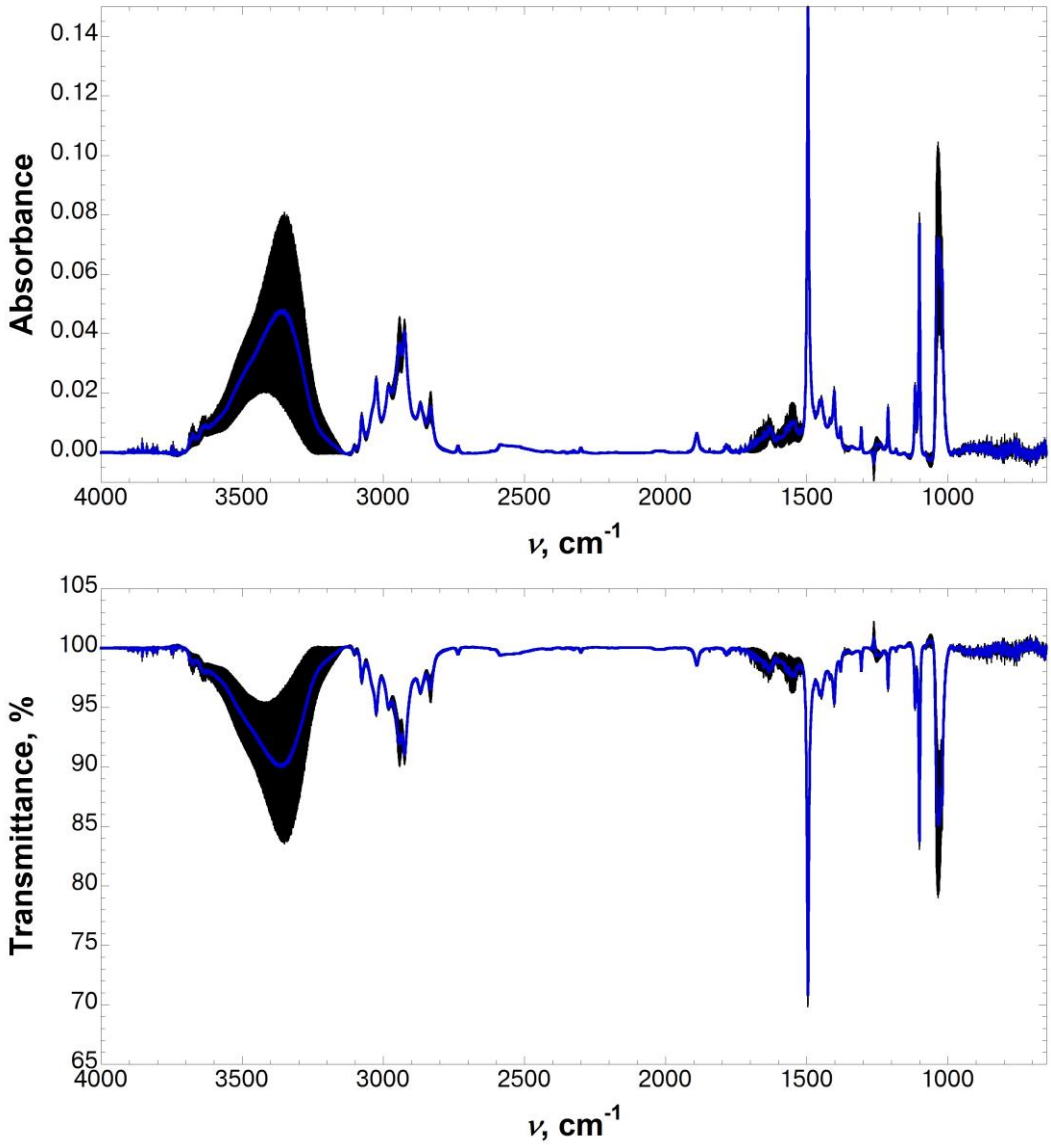


Figure B23 FT-IR spectrum of *p*-thiocresol (200 mM) in 10% methanol in CCl_4.

Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum.

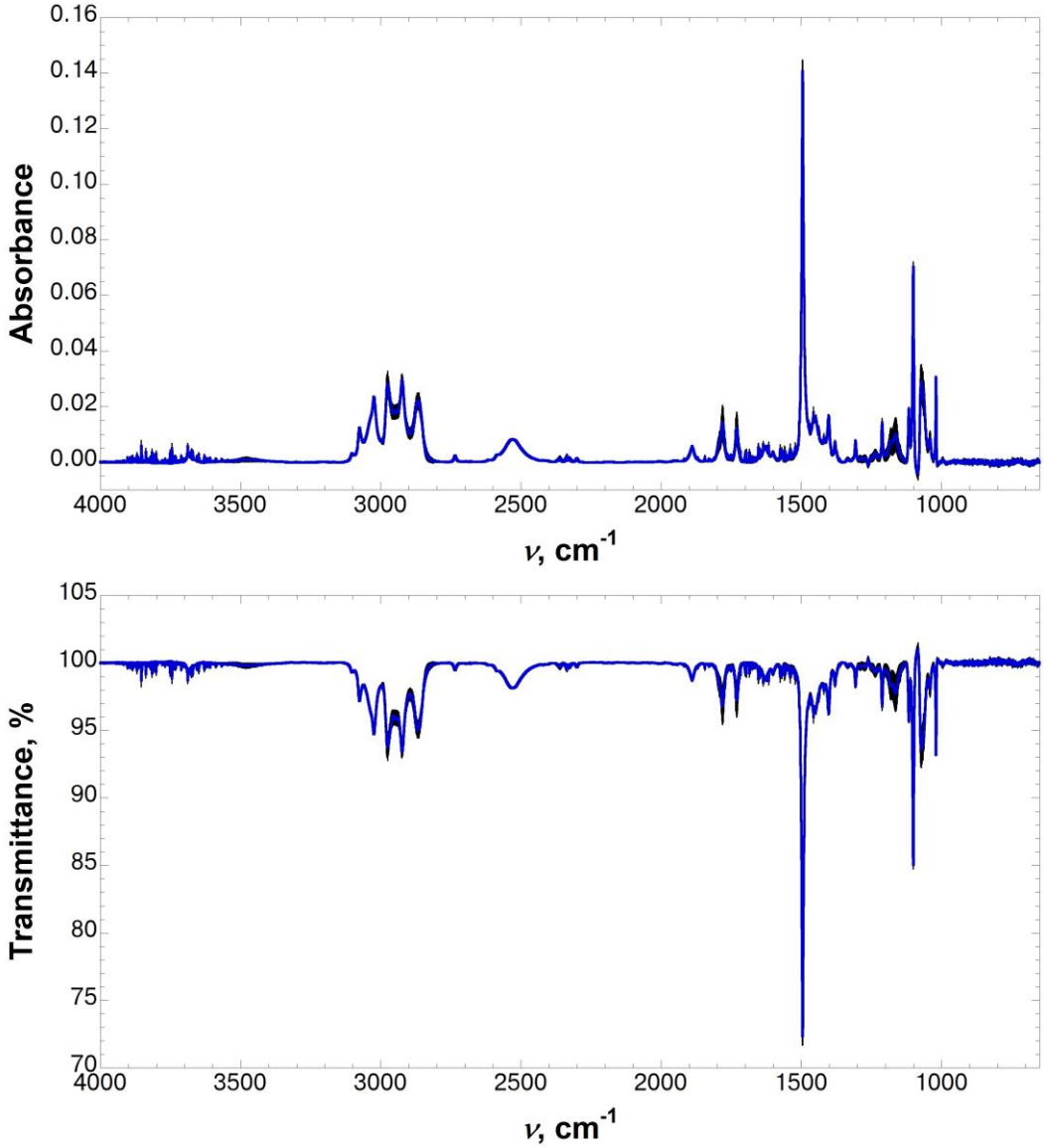


Figure B24 FT-IR spectrum of *p*-thiocresol (200 mM) in 25% THF in CCl_4. Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2530 cm^{-1}.

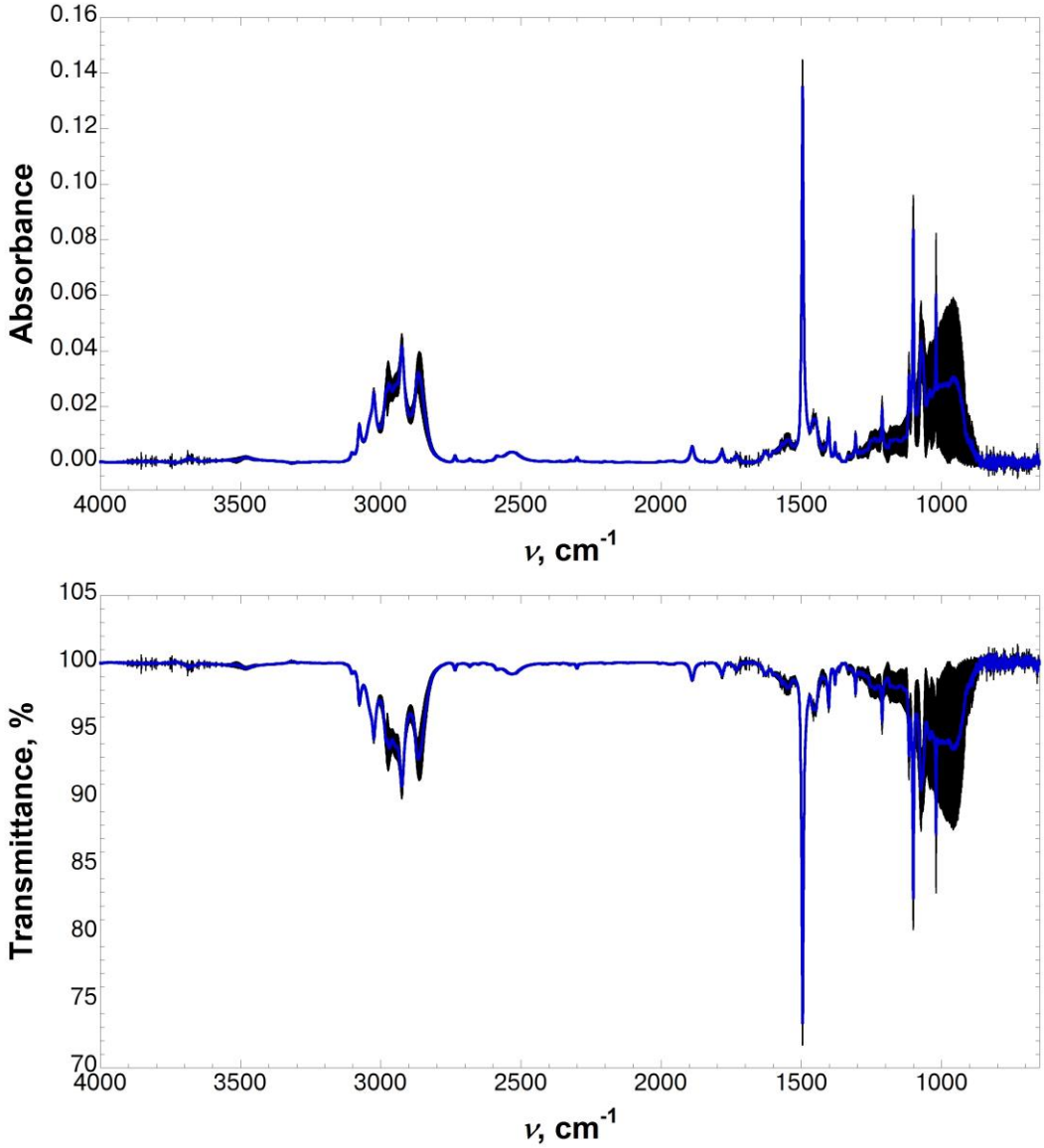


Figure B25 FT-IR spectrum of *p*-thiocresol (200 mM) in 10% THF in CCl_4. Three independent spectra were solvent-subtracted, baseline corrected, and averaged. Error bars indicate standard error. Top: full IR absorbance spectrum; bottom: full IR transmittance spectrum. ν_{max} of the S–H stretching frequency is 2531 cm^{-1}.

Appendix C

CRYSTALLOGRAPHIC ENTRIES FROM THE CAMBRIDGE STRUCTURAL DATABASE

Parameters for the search criteria were described in Chapter 2: Insights into S–H/π Aromatic Interactions: Studies on Boc-4-Thiol-L-Phenylalanine-*tert*-Butyl Ester via IR Spectroscopy, X-Ray Crystallography, and *ab initio* Clculations, and discussed in section 2.2.6 and 2.4.10.

Table C1. Structure entries for Li^+-aromatic interactions.

All entries in analysis of Li^+-aromatic interactions, following initial search parameters of the CSD and cylinder restrictions. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the Li^+-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55.
Annotations: (A) polycyclic aromatic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings

Refcode	$d_{Li+–centroid}$, Å	$d_{Li+–Cmin}$, Å	Li^+ d_{plane}, Å	Li^+ r, Å	$\Delta d_{centroid–Cmin}$, Å	Parsing Rationale
JENREA	2.593	2.125	2.130	1.479	0.468	
JENREA	2.739	2.125	2.086	1.775	0.614	
BUJWOS	2.885	2.141	1.795	2.259	0.744	
JENREA	2.584	2.158	2.160	1.418	0.426	
GAGFAV	2.889	2.169	2.092	1.992	0.720	
JENREA	3.021	2.182	1.759	2.456	0.839	
JENREA	3.004	2.185	1.790	2.412	0.819	
JENREA	2.667	2.198	2.205	1.500	0.469	
GAGFAV	2.923	2.198	2.112	2.021	0.725	
JENREA	3.022	2.210	1.784	2.439	0.812	
JENREA	3.049	2.246	1.817	2.448	0.803	
PALWII	3.017	2.248	2.062	2.202	0.769	
PALWII	3.002	2.252	1.982	2.255	0.750	
JIRVAH	1.947	2.253	1.937	0.197	-0.306	
YOJKIR	2.905	2.253	2.069	2.039	0.652	
YIFWUH	3.072	2.283	1.758	2.519	0.789	
YIFWUH	3.103	2.297	1.722	2.581	0.806	
YOJKIR	3.048	2.311	2.029	2.275	0.737	
JENREA	2.991	2.327	2.085	2.144	0.664	
JENREA	3.043	2.333	1.944	2.341	0.710	
JENREA	2.998	2.334	1.999	2.234	0.664	
JIRVAH	1.967	2.338	1.964	0.109	-0.371	
GIJWOL	3.022	2.346	2.282	1.981	0.676	
YOJKIR	3.016	2.348	2.035	2.226	0.668	
JENREA	3.083	2.380	2.020	2.329	0.703	

Table C1 continued

PHMGLI	3.130	2.419	2.328	2.092	0.711	
CROBLI	3.256	2.443	1.915	2.633	0.813	
BAHFUM	3.502	2.454	1.968	2.897	1.048	
EGAQUY	3.188	2.471	2.309	2.198	0.717	
PHMGLI	3.252	2.472	2.336	2.262	0.780	
RAKSAX	3.145	2.481	2.422	2.006	0.664	
CROBLI	3.364	2.493	1.842	2.815	0.871	
RAKSIF	3.276	2.495	2.306	2.327	0.781	
BAHFUM	3.510	2.517	2.047	2.851	0.993	
RAKSIF	3.420	2.563	2.270	2.558	0.857	
LIBPNI	2.704	2.596	2.517	0.988	0.108	
LOJFAR	3.614	2.618	2.139	2.913	0.996	
RAKSIF	3.549	2.747	2.569	2.449	0.802	
PAUELI10	3.459	2.926	2.880	1.916	0.533	
RAKSIF	3.849	2.938	2.494	2.932	0.911	
LOVLAJ	4.139	3.250	2.852	3.000	0.889	
UHUQAR	2.087	2.359	2.066	0.295	-0.272	A1
UHUQAR	3.128	2.507	2.253	2.170	0.621	A1
PUKROC	1.667	2.197	1.667	0.000	-0.530	A
PUKROC	1.719	2.213	1.717	0.083	-0.494	A
ADEBES	2.220	2.332	2.071	0.800	-0.112	A
ADEBES	2.914	2.357	2.385	1.674	0.557	A
ADEBES	2.243	2.382	2.191	0.480	-0.139	A
LIPHPR	2.450	2.640	2.424	0.356	-0.190	A
LIRVEN	3.529	2.748	2.365	2.619	0.781	A
ADEBES	3.572	2.834	2.545	2.506	0.738	A
ADEBES	3.759	2.836	2.389	2.902	0.923	A
QUKCEE	3.794	2.949	2.466	2.883	0.845	A
COKVON	3.084	2.223	1.956	2.384	0.861	B
ZONJIV	3.094	2.261	1.750	2.552	0.833	B
CALKOP	3.353	2.278	1.860	2.790	1.075	B
PADCON	1.990	2.293	1.974	0.252	-0.303	B
PADCON	2.005	2.315	1.993	0.219	-0.310	B
BECWEL	3.088	2.317	1.869	2.458	0.771	B
PADCON	3.137	2.337	1.817	2.557	0.800	B
COKVON	3.193	2.358	1.811	2.630	0.835	B
COGFOT	3.304	2.377	2.088	2.561	0.927	B
ZONJIV	2.050	2.380	2.038	0.221	-0.330	B

Table C1 continued

BECWEL	3.207	2.383	1.819	2.641	0.824	B
PADCON	3.200	2.391	1.830	2.625	0.809	B
GAFBUK	3.246	2.403	2.220	2.368	0.843	B
GAFBUK	3.286	2.419	2.208	2.434	0.867	B
GAFBUK	3.342	2.450	2.145	2.563	0.892	B
GAFBUK	3.354	2.458	2.198	2.533	0.896	B
PADCON	2.884	2.498	2.416	1.575	0.386	B
PADCON	2.940	2.519	2.410	1.684	0.421	B
AXOCEV	3.821	3.062	2.493	2.896	0.759	B
AXOCEV	3.886	3.152	2.552	2.931	0.734	B
LIANTR	2.003	2.272	1.967	0.378	-0.269	A, B
RACSAP	2.499	2.308	2.277	1.030	0.191	A, B
LIANTR	2.000	2.331	1.997	0.110	-0.331	A, B
CIFDOK	1.973	2.336	1.971	0.089	-0.363	A, B
CIFDOK	3.208	2.428	1.904	2.582	0.780	A, B
LIANTR	3.190	2.453	1.979	2.502	0.737	A, B
CIFDOK	3.400	2.507	1.897	2.822	0.893	A, B
BFLULI	3.501	2.549	1.966	2.897	0.952	A, B
BFLULI	3.487	2.564	2.040	2.828	0.923	A, B
LIANTR	3.368	2.569	2.024	2.692	0.799	A, B
CIFDOK	3.464	2.616	1.934	2.874	0.848	A, B
RACSAP	3.582	2.652	2.261	2.778	0.930	A, B
LIANTR	3.501	2.666	2.110	2.794	0.835	A, B
PABYAT	1.569	1.707	1.482	0.515	-0.138	A1, B

Table C2. Structure entries for Na^+-aromatic interactions.

All entries in analysis of Na^+-aromatic interactions, following initial search parameters of the CSD and cylinder restrictions. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the Na^+-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55.
Annotations: (A) polycyclic aromatic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (G) the cation was potentially interacting with an adjacent crown ether.

Refcode	$d_{Na+-centroid}$, Å	$d_{Na+-Cmin}$, Å	Na^+ d_{plane}, Å	Na^+ r, Å	$\Delta d_{centroid-Cmin}$, Å	Parsing Rationale
IDADEY	2.698	2.618	2.511	0.987	0.080	
BAHFUM	3.484	2.631	2.197	2.704	0.853	
BAHFUM	3.591	2.691	2.191	2.845	0.900	
BAHFUM	2.696	2.721	2.609	0.679	-0.025	
SEBJUE	3.382	2.730	2.583	2.183	0.652	
SPHIND	3.219	2.744	2.587	1.916	0.475	
SEBJUE	3.349	2.748	2.592	2.121	0.601	
BAHFUM	2.666	2.771	2.608	0.553	-0.105	
SEBJUE	3.613	2.788	2.503	2.606	0.825	
LEXDIB	3.095	2.802	2.698	1.517	0.293	
LEXDIB	2.974	2.809	2.765	1.095	0.165	
GUPSAM	3.633	2.809	2.571	2.567	0.824	
LEXDIB	3.165	2.810	2.754	1.560	0.355	
LEXDIB	3.012	2.820	2.767	1.190	0.192	
LEXDIB	2.932	2.832	2.792	0.895	0.100	
GUPSAM	3.569	2.834	2.438	2.607	0.735	
DILLUH	2.721	2.835	2.668	0.534	-0.114	
BEMNIQ	3.705	2.840	2.508	2.727	0.865	
PURKAP	3.322	2.841	2.783	1.814	0.481	
GUPSAM	3.721	2.843	2.453	2.798	0.878	
PIPGEA	3.808	2.846	2.392	2.963	0.962	
LEXDIB	2.819	2.865	2.750	0.620	-0.046	
TIYYUX	3.213	2.874	2.870	1.444	0.339	
SEBJUE	3.732	2.891	2.535	2.739	0.841	

Table C2 continued

JUBYOU	3.245	2.895	2.864	1.526	0.350
CAMBIC	2.776	2.932	2.749	0.386	-0.156
CAMBOI	2.774	2.937	2.750	0.364	-0.163
NODBOZ	2.700	2.945	2.691	0.220	-0.245
PURKAP	3.709	2.954	2.607	2.638	0.755
UCOYIW	2.700	2.960	2.690	0.232	-0.260
EHIFUX	3.612	2.965	2.826	2.250	0.647
UCOYIW	2.765	2.969	2.742	0.356	-0.204
PURKAP	3.281	2.969	2.927	1.482	0.312
NODBOZ	3.098	2.980	2.914	1.052	0.118
HIKLIW	3.287	2.981	2.974	1.400	0.306
CEQDIL	3.390	2.988	2.941	1.686	0.402
ZZZUPI02	2.910	2.995	2.855	0.563	-0.085
UCOYIW	2.815	3.008	2.801	0.280	-0.193
TIYYUX	3.138	3.014	2.957	1.050	0.124
SUFVAS	3.455	3.024	2.964	1.775	0.431
PURKAP	3.652	3.025	2.821	2.319	0.627
PABYIB	2.803	3.031	2.790	0.270	-0.228
DIWWIP	3.944	3.033	2.668	2.905	0.911
BACGUH	3.744	3.101	2.938	2.321	0.643
FOWFUU	4.019	3.120	2.680	2.995	0.899
HAGYOD	2.814	3.130	2.814	0.000	-0.316
CAYCOU	3.944	3.158	2.627	2.942	0.786
HEFYOH	3.765	3.162	2.871	2.436	0.603
ZUKLAU	3.850	3.175	2.948	2.476	0.675
PIPGIE	3.803	3.180	3.049	2.273	0.623
CIYJAX	3.820	3.188	3.019	2.341	0.632
HORSUC	4.033	3.220	2.764	2.937	0.813
CUPPUZ	3.938	3.229	2.967	2.589	0.709
DOFNES	3.971	3.250	2.932	2.678	0.721
OGOTUA	3.498	3.252	3.248	1.299	0.246
ZZZUPI01	3.028	3.253	3.020	0.220	-0.225
PIPGIE	3.929	3.259	3.068	2.454	0.670
ZUKLAU	3.944	3.264	2.935	2.635	0.680
CIYJAX	3.817	3.268	3.108	2.216	0.549
ZAGSIJ	3.752	3.270	3.166	2.013	0.482
TEVFAD	3.264	3.272	3.185	0.714	-0.008
ZUQMED	3.985	3.273	2.850	2.785	0.712

Table C2 continued

JUBYOU	3.944	3.281	3.117	2.416	0.663	
JUBYOU	3.661	3.305	3.298	1.589	0.356	
ZUQMED	4.052	3.319	2.894	2.836	0.733	
EFIWIZ01	4.095	3.319	2.953	2.837	0.776	
ZUQMED	4.101	3.322	2.828	2.970	0.779	
ZUQMED	3.971	3.331	2.994	2.609	0.640	
BACGUH	4.172	3.366	2.958	2.942	0.806	
YAKVOX	3.992	3.400	3.262	2.301	0.592	
PIGROO	2.838	2.620	2.596	1.147	0.218	A1
TABROE	2.707	2.783	2.654	0.533	-0.076	A1
YERFIK	3.016	3.238	3.013	0.134	-0.222	A1
YERFIK	3.183	3.315	3.166	0.329	-0.132	A1
FINHUF	3.820	3.106	2.557	2.838	0.714	G
XILFIH	3.250	3.227	3.157	0.772	0.023	G
UZIHIV	4.079	3.235	2.771	2.993	0.844	G
XILFIH	3.443	3.352	3.284	1.034	0.091	G
KAMHEL	4.056	3.354	2.811	2.924	0.702	G
NEMNAL	2.572	2.888	2.570	0.101	-0.316	A
SANMAL	2.816	2.938	2.782	0.436	-0.122	A
VIHNUV	3.856	2.965	2.533	2.907	0.891	A
VIHNUV	3.854	2.970	2.546	2.893	0.884	A
SANMAL	2.830	3.007	2.810	0.336	-0.177	A
KOFCIR	3.887	3.032	2.615	2.876	0.855	A
DOLGAO	3.505	3.045	2.864	2.021	0.460	A
DOLGAO	3.706	3.045	2.805	2.422	0.661	A
QEJYEK	3.278	3.368	3.237	0.517	-0.090	A
QEJYEK	3.317	3.477	3.299	0.345	-0.160	A
JEVPUV	2.469	2.745	2.463	0.172	-0.276	B
JEVPUV	2.588	2.753	2.559	0.386	-0.165	B
NAPHBZ	3.610	2.814	2.593	2.512	0.796	B
JECCAW	3.696	2.837	2.284	2.906	0.859	B
YANRIO	3.337	2.859	2.784	1.840	0.478	B
TPMNAE	3.686	2.885	2.418	2.782	0.801	B
GIZBUO	3.403	2.899	2.847	1.864	0.504	B
TPMNAE	3.705	2.959	2.741	2.493	0.746	B
TONYAW	3.147	2.669	2.604	1.767	0.478	A, B
TONYAW	2.693	2.785	2.631	0.575	-0.092	A, B
YUGDOT	3.632	2.793	2.404	2.723	0.839	A, B

Table C2 continued

XACSID	3.492	2.838	2.605	2.326	0.654	A, B
XACSID	3.506	2.858	2.628	2.321	0.648	A, B
XACSID	3.746	2.894	2.593	2.703	0.852	A, B
XICMED	4.172	3.345	2.985	2.915	0.827	A, B
XICMED	4.113	3.376	3.076	2.730	0.737	A, B

Table C3. Structure entries for K^+-aromatic interactions.

All entries in analysis of K^+-aromatic interactions, following initial search parameters of the CSD and cylinder restrictions. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the K^+-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55.
Annotations: (A) polycyclic aromatic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (E) redundant structures or entries; (G) the cation was potentially interacting with an adjacent crown ether.

Refcode	$d_{K+-centroid}$, Å	$d_{K+-Cmin}$, Å	K^+ d_{plane}, Å	K^+ r, Å	$\Delta d_{centroid-Cmin}$, Å	Parsing Rationale
ZECXUC	2.934	2.250	2.139	2.008	0.684	
UFEFIW	3.506	2.799	2.314	2.634	0.707	
NOFCAO	3.004	2.880	2.841	0.976	0.124	
RUNSAU	3.044	2.884	2.892	0.950	0.160	
QEJFUH	3.318	2.884	2.854	1.692	0.434	
KEKLUH	3.548	2.903	2.652	2.357	0.645	
QEJFUH	3.582	2.938	2.693	2.362	0.644	
NOFCAO	2.961	2.955	2.857	0.778	0.006	
NOFCAO	2.904	2.964	2.823	0.681	-0.060	
SOFHIG	3.043	2.971	2.948	0.754	0.072	
UFEFIW	3.179	2.973	2.958	1.165	0.206	
BUGMUN	3.673	2.978	2.803	2.374	0.695	
RIVQIX	3.521	2.979	2.894	2.006	0.542	
YEMYIA	2.853	2.984	2.823	0.413	-0.131	
PEBSAQ	2.926	2.992	2.870	0.570	-0.066	
NOFCAO	3.485	2.998	2.936	1.878	0.487	
PEBRUJ	2.826	3.004	2.805	0.344	-0.178	
KMEPBZ	3.250	3.004	2.997	1.257	0.246	
XUDCUW	3.468	3.004	2.978	1.777	0.464	
KMEPBZ	3.443	3.013	2.970	1.742	0.430	
HAKWIB	3.138	3.015	2.982	0.977	0.123	
BUGMUN	3.787	3.016	2.758	2.595	0.771	
GEBCUL	2.882	3.018	2.854	0.401	-0.136	

Table C3 continued

PEBRUJ	2.943	3.018	2.887	0.571	-0.075
SUGXEX	3.283	3.018	3.025	1.276	0.265
PEBROD	3.288	3.020	2.986	1.376	0.268
PEBSAQ	2.947	3.023	2.893	0.562	-0.076
PEBSAQ	3.321	3.023	3.010	1.403	0.298
PEBROD	3.093	3.029	2.966	0.877	0.064
HAKWIB	2.990	3.035	2.916	0.661	-0.045
YOLDAG	2.898	3.040	2.867	0.423	-0.142
NOVQOG	3.755	3.041	2.830	2.468	0.714
WIZJIY	2.908	3.044	2.877	0.423	-0.136
FEJPAN	2.965	3.044	2.907	0.584	-0.079
NOFCAO	2.854	3.047	2.832	0.354	-0.193
FEJPAN	3.076	3.051	2.976	0.778	0.025
OBIJOA	3.118	3.053	2.998	0.857	0.065
PAGYUR	2.998	3.057	2.940	0.587	-0.059
KCATCR	3.632	3.058	2.963	2.100	0.574
PEBSAQ	3.438	3.059	3.024	1.636	0.379
BOBWIY	3.599	3.062	2.994	1.997	0.537
YEMYIA	2.846	3.064	2.838	0.213	-0.218
KCATFE	3.577	3.067	3.000	1.948	0.510
BUGMUN	3.785	3.067	2.841	2.501	0.718
XUDCUW	3.650	3.073	2.988	2.096	0.577
KCATCR	3.737	3.073	2.939	2.308	0.664
PAGYUR	2.878	3.074	2.860	0.321	-0.196
BAXSUO	2.809	3.075	2.806	0.130	-0.266
HAKWEX	3.182	3.077	3.037	0.950	0.105
KCATFE	3.319	3.077	3.072	1.256	0.242
BAXSUO	2.853	3.078	2.847	0.185	-0.225
ANUDUJ	2.849	3.080	2.842	0.200	-0.231
RINYOE	3.088	3.080	2.996	0.748	0.008
DARGEK	3.259	3.081	3.052	1.143	0.178
BOBWIY	3.799	3.081	2.797	2.571	0.718
GICHEH	2.911	3.083	2.888	0.365	-0.172
BOBWIY	3.790	3.083	2.920	2.416	0.707
BOBWIY	3.327	3.084	3.082	1.253	0.243
KCATFE	3.789	3.084	2.927	2.406	0.705
BUGMUN	2.887	3.085	2.873	0.284	-0.198
KCATCR	3.781	3.085	2.859	2.474	0.696

Table C3 continued

BUGMUN	2.957	3.086	2.920	0.466	-0.129
KCATCR	3.876	3.086	2.483	2.976	0.790
GICHUX	2.969	3.087	2.928	0.492	-0.118
YOLDAG	2.781	3.088	2.781	0.000	-0.307
KCATFE	3.810	3.088	2.811	2.572	0.722
KCATCR	3.424	3.089	3.078	1.500	0.335
GICHEH	3.012	3.090	2.955	0.583	-0.078
TABRUK	3.084	3.090	2.987	0.767	-0.006
VAMSOR	2.875	3.091	2.864	0.251	-0.216
VAMSOR	2.853	3.093	2.848	0.169	-0.240
GICHUX	3.765	3.096	2.939	2.353	0.669
YAKWIS	3.242	3.097	3.082	1.006	0.145
YEMWUK	2.867	3.099	2.861	0.185	-0.232
EREMIY	3.006	3.099	2.968	0.476	-0.093
NOVQOG	3.322	3.100	3.054	1.307	0.222
WUBRIW	2.995	3.103	2.954	0.494	-0.108
CUDXAA	3.830	3.104	2.655	2.760	0.726
OJAGUC	2.892	3.106	2.881	0.252	-0.214
FORFUP	3.101	3.107	3.018	0.713	-0.006
WIZJIY	2.905	3.109	2.890	0.295	-0.204
WIZJIY	2.888	3.110	2.877	0.252	-0.222
PAGYUR	3.060	3.115	2.992	0.642	-0.055
KIYZIB	3.188	3.115	3.047	0.938	0.073
OJAGUC	2.953	3.118	2.933	0.343	-0.165
DAMMEJ	3.080	3.118	3.023	0.590	-0.038
OJAGUC	2.933	3.119	2.921	0.265	-0.186
OJAGUC	3.024	3.120	2.982	0.502	-0.096
LOCXAC	2.975	3.124	2.951	0.377	-0.149
BUBGIQ	3.862	3.126	2.879	2.574	0.736
PEBRUJ	3.656	3.127	3.049	2.017	0.529
GICHEH	2.915	3.129	2.903	0.264	-0.214
HAKWEX	3.008	3.129	2.976	0.438	-0.121
YAKVUD	3.234	3.130	3.108	0.894	0.104
NOFCAO	2.873	3.132	2.870	0.131	-0.259
NOVQOG01	3.221	3.132	3.066	0.987	0.089
WIZJIY	2.897	3.133	2.890	0.201	-0.236
OBIJUG	3.126	3.134	3.031	0.765	-0.008
MIYKUC	3.297	3.134	3.112	1.089	0.163

Table C3 continued

HAKWEX	3.606	3.134	3.075	1.884	0.472
GEBCUL01	2.887	3.135	2.878	0.228	-0.248
NOMKEF	3.964	3.135	2.697	2.905	0.829
NOVQOG01	3.941	3.136	2.708	2.863	0.805
EREMIY	3.098	3.137	3.028	0.655	-0.039
GEBCUL	2.882	3.139	2.876	0.186	-0.257
GEBCUL	2.900	3.140	2.893	0.201	-0.240
GEBCUL	2.847	3.141	2.847	0.000	-0.294
PAGYUR	2.926	3.144	2.919	0.202	-0.218
WUBRIW	2.963	3.144	2.947	0.308	-0.181
BEJDEB	3.219	3.146	3.115	0.812	0.073
GICHUX	3.151	3.147	3.060	0.752	0.004
MUTYUW	3.145	3.149	3.063	0.713	-0.004
SUGXEX	3.765	3.149	2.970	2.314	0.616
SUGXEX	3.873	3.149	2.708	2.769	0.724
PEBRUJ	3.695	3.150	3.065	2.064	0.545
LEYCUP	2.959	3.152	2.949	0.243	-0.193
BEJDEB	3.244	3.152	3.135	0.834	0.092
KMEPBZ	3.541	3.152	3.138	1.641	0.389
VEPBID	2.987	3.153	2.967	0.345	-0.166
SUGXEX	3.801	3.153	2.974	2.367	0.648
OJAGUC	2.971	3.155	2.955	0.308	-0.184
ZIWZIO	2.874	3.156	2.872	0.107	-0.282
OVODAF	3.930	3.157	2.677	2.877	0.773
OJAGUC	3.022	3.160	2.988	0.452	-0.138
XOMXAA	3.003	3.161	2.982	0.355	-0.158
GEBCUL01	2.869	3.162	2.867	0.107	-0.293
KTPHEB02	2.960	3.162	2.946	0.288	-0.202
RIVQIX	3.807	3.162	2.954	2.401	0.645
XOCGIH	3.126	3.163	3.056	0.658	-0.037
VIXQEY	2.881	3.164	2.880	0.076	-0.283
OJAGUC	2.921	3.165	2.916	0.171	-0.244
KCATCR	3.957	3.167	2.878	2.716	0.790
YAKWIS	3.326	3.168	3.161	1.035	0.158
FUTCIH	3.403	3.168	3.162	1.258	0.235
TABRUK	2.883	3.170	2.882	0.076	-0.287
YEZFOY	3.427	3.170	3.130	1.396	0.257
CIGVAP	3.796	3.171	3.012	2.310	0.625

Table C3 continued

OJAGUC	2.875	3.172	2.875	0.000	-0.297
HAKWIB	2.977	3.172	2.960	0.318	-0.195
SELLIF	3.046	3.175	3.026	0.348	-0.129
TUVYEP	2.917	3.179	2.912	0.171	-0.262
GICHEH	2.933	3.183	2.927	0.188	-0.250
MIYJOV	3.001	3.183	2.980	0.354	-0.182
WIZJIY	2.988	3.184	2.973	0.299	-0.196
PEBROD	3.693	3.184	3.076	2.044	0.509
OJAGUC	3.009	3.188	3.004	0.173	-0.179
KIBMAL	3.883	3.189	2.823	2.666	0.694
ZOSCER	3.081	3.190	3.048	0.450	-0.109
KCATFE	4.025	3.190	2.835	2.857	0.835
KTPHEB	2.986	3.191	2.975	0.256	-0.205
VEPBOJ	2.950	3.196	2.944	0.188	-0.246
MIYKUC	3.071	3.196	3.043	0.414	-0.125
ANUFEV	2.937	3.197	2.935	0.108	-0.260
VEPBOJ	3.357	3.198	3.145	1.174	0.159
LADGIH	3.060	3.203	3.034	0.398	-0.143
GICGUW	4.051	3.204	2.770	2.956	0.847
PEBRUJ	3.800	3.205	2.971	2.369	0.595
MIYKUC	3.039	3.206	3.023	0.311	-0.167
UCIRAB	4.021	3.206	2.713	2.968	0.815
BOBWIY	4.031	3.207	2.865	2.836	0.824
NOMKEF	3.102	3.214	3.072	0.430	-0.112
KPHTNB10	3.420	3.214	3.171	1.281	0.206
RARKAY	3.171	3.216	3.110	0.619	-0.045
DAMMEJ	2.957	3.217	2.954	0.133	-0.260
BDTASH	3.936	3.217	3.133	2.383	0.719
RIZZOP	3.093	3.218	3.061	0.444	-0.125
YEXJIW	3.183	3.220	3.128	0.589	-0.037
VOHXUL	3.698	3.220	3.126	1.976	0.478
XOCGIH	3.871	3.222	2.981	2.469	0.649
OJAGUC	2.945	3.223	2.944	0.077	-0.278
GICGUW	3.087	3.223	3.062	0.392	-0.136
QIVWEX	3.088	3.223	3.064	0.384	-0.135
HAKWEX	3.227	3.224	3.148	0.710	0.003
OJAGUC	3.182	3.225	3.103	0.705	-0.043
BUGMUN	4.076	3.226	2.806	2.956	0.850

Table C3 continued

AVETUQ	3.862	3.227	3.013	2.416	0.635
ANUFEV	2.951	3.229	2.950	0.077	-0.278
KMEPBZ	3.989	3.232	2.975	2.657	0.757
KACSAK	3.820	3.237	3.109	2.220	0.583
RINYOE	3.897	3.240	3.132	2.319	0.657
RIWMEQ	2.989	3.241	2.985	0.155	-0.252
VEPBID	3.313	3.244	3.198	0.865	0.069
POWNAS	3.232	3.245	3.153	0.710	-0.013
KEKLUH	4.071	3.247	2.818	2.938	0.824
XOCGIH	3.592	3.248	3.147	1.732	0.344
PEBRUJ	3.880	3.253	2.831	2.653	0.627
KBZILT	3.078	3.254	3.062	0.313	-0.176
BATWEY	4.074	3.256	2.856	2.905	0.818
ANUFEV	3.053	3.257	3.042	0.259	-0.204
LADGIH	3.053	3.257	3.041	0.270	-0.204
LEYCUP	3.049	3.258	3.041	0.221	-0.209
KNPHEH01	3.697	3.260	3.129	1.969	0.437
ANUFEV	2.990	3.262	2.989	0.077	-0.272
PEGNEW	2.989	3.265	2.988	0.077	-0.276
BACWEH10	3.034	3.266	3.031	0.135	-0.232
VEPBID	3.052	3.268	3.044	0.221	-0.216
ZIWZIO	2.974	3.270	2.974	0.000	-0.296
FEJPAN	3.605	3.270	3.274	1.509	0.335
SOJTOC	2.996	3.272	2.994	0.109	-0.276
NAYXEQ	3.165	3.274	3.131	0.463	-0.109
NAYXAM	4.073	3.275	2.905	2.855	0.798
ZOSCER	3.123	3.284	3.102	0.362	-0.161
BAGLOM	3.023	3.285	3.022	0.078	-0.262
VISNUI	3.063	3.288	3.058	0.175	-0.225
YEZFOY	4.079	3.289	2.993	2.771	0.790
VEPBEZ	3.063	3.290	3.056	0.207	-0.227
TMALNK	3.096	3.292	3.088	0.222	-0.196
CEYLEX	3.746	3.295	3.219	1.916	0.451
RINYOE	4.017	3.299	3.115	2.536	0.718
QIVWIB	3.213	3.300	3.178	0.473	-0.087
WIZJIY	3.103	3.303	3.095	0.223	-0.200
WOKYOM	3.989	3.304	3.077	2.539	0.685
UCIRAB	4.092	3.307	2.892	2.895	0.785

Table C3 continued

VEPBEZ	3.058	3.314	3.056	0.111	-0.256
BERMAO	3.323	3.322	3.230	0.781	0.001
WARHUU	3.215	3.323	3.186	0.431	-0.108
XEBVEF	3.840	3.323	3.124	2.233	0.517
BACWAD10	3.281	3.325	3.234	0.553	-0.044
KIBMAL	3.281	3.327	3.212	0.669	-0.046
VEXVAX	3.528	3.329	3.270	1.324	0.199
DOCCIH	3.630	3.330	3.288	1.538	0.300
SEYHOV	3.722	3.330	3.298	1.725	0.392
KOTJUA	3.323	3.333	3.256	0.664	-0.010
GADNAB	3.910	3.334	3.049	2.448	0.576
VAZZAZ01	3.260	3.335	3.220	0.509	-0.075
ELECIH	3.190	3.336	3.173	0.329	-0.146
BUBGIQ	3.133	3.337	3.124	0.237	-0.204
JOLVUD	3.054	3.342	3.054	0.000	-0.288
WARHUU	3.223	3.342	3.199	0.393	-0.119
YEXJIW	3.135	3.343	3.128	0.209	-0.208
NAYXAM	3.658	3.347	3.343	1.485	0.311
NOSGOT	3.814	3.347	3.299	1.914	0.467
SUGXEX	4.100	3.347	3.055	2.734	0.753
YEZFOY	3.550	3.349	3.327	1.238	0.201
YEBDIS	3.459	3.350	3.323	0.960	0.109
VICCOZ	4.032	3.355	2.928	2.772	0.677
FEJPAN	3.190	3.356	3.176	0.299	-0.166
CIGVAP	4.154	3.360	2.971	2.903	0.794
KCATCR	4.192	3.370	2.993	2.935	0.822
NOYKET	3.686	3.373	3.362	1.511	0.313
XUNWAF	4.132	3.373	3.015	2.825	0.759
REKVAG	3.250	3.380	3.225	0.402	-0.130
ELECIH	3.321	3.382	3.279	0.526	-0.061
YACVIH	3.344	3.385	3.299	0.547	-0.041
PEBSAQ	4.064	3.387	2.993	2.749	0.677
REKVAG	3.150	3.392	3.148	0.112	-0.242
UFEJUL	3.525	3.396	3.365	1.050	0.129
TUVPAC	3.501	3.406	3.368	0.956	0.095
FUTCIH	4.070	3.410	3.252	2.447	0.660
WOKYOM	3.329	3.423	3.299	0.446	-0.094
GICHUX	4.182	3.426	3.112	2.794	0.756

Table C3 continued

DIBYUI	4.132	3.430	3.192	2.624	0.702	
XUXVUI	3.486	3.431	3.374	0.877	0.055	
TUVPIK	3.429	3.434	3.363	0.670	-0.005	
BERMAO	3.730	3.437	3.426	1.475	0.293	
BERLUH	3.514	3.438	3.386	0.940	0.076	
SENHUO	3.667	3.443	3.383	1.415	0.224	
TUVPAC	3.607	3.445	3.425	1.131	0.162	
VEPFAX10	4.079	3.449	3.041	2.719	0.630	
BERMAO	4.094	3.450	3.021	2.763	0.644	
XOXBIW	3.597	3.455	3.425	1.099	0.142	
VIHMII	3.197	3.467	3.197	0.000	-0.270	
TEZJEN	3.807	3.470	3.395	1.723	0.337	
QOYLIA	4.056	3.475	3.343	2.297	0.581	
KPHTHT	4.037	3.507	3.329	2.284	0.530	
CEYLEX	3.546	3.511	3.443	0.848	0.035	
OGEPAT	4.133	3.523	3.331	2.447	0.610	
SIPSUF	3.893	3.533	3.494	1.717	0.360	
KTYSUH10	3.558	3.552	3.473	0.773	0.006	
NOVQOG01	4.040	3.552	3.387	2.202	0.488	
ZAPQOW	4.098	3.567	3.304	2.424	0.531	
SEZQUK	4.172	3.568	3.166	2.717	0.604	
GIFXEA	4.110	3.573	3.492	2.167	0.537	
PAVXAL	3.712	3.581	3.549	1.088	0.131	
VEPFAX10	3.975	3.600	3.572	1.744	0.375	
GADNAB	4.082	3.602	3.514	2.077	0.480	
NOSGOT	4.167	3.627	3.528	2.217	0.540	
BERLUH	3.460	3.637	3.452	0.235	-0.177	
JOLVUD	4.114	3.652	3.540	2.096	0.462	
QOYLOG	4.071	3.673	3.627	1.849	0.398	
PDNMET	4.113	3.712	3.602	1.986	0.401	
LADGIH	4.122	3.787	3.749	1.713	0.335	
WIZJIY	3.833	3.843	3.773	0.676	-0.010	
CEVZUA	3.115	2.904	2.949	1.003	0.211	A1
CEVZUA	2.792	2.941	2.791	0.075	-0.149	A1
CEVZUA	2.920	2.984	2.785	0.878	-0.064	A1
CEVZUA	2.967	3.095	2.960	0.204	-0.128	A1
CEVZUA	3.512	3.097	2.964	1.884	0.415	A1
SUDPIQ	3.137	3.118	3.045	0.754	0.019	A1

Table C3 continued

SUCCUO	3.763	3.133	2.923	2.370	0.630	A1
CEVZUA	3.597	3.134	3.153	1.731	0.463	A1
HEVWOW	2.880	3.157	2.877	0.131	-0.277	A1
CEVZUA	3.531	3.161	3.031	1.811	0.370	A1
FOVDOL	3.111	3.187	3.061	0.556	-0.076	A1
FOVDOL	3.104	3.194	3.058	0.532	-0.090	A1
CEVZUA	3.059	3.196	3.013	0.528	-0.137	A1
CEVZUA	3.514	3.204	3.105	1.645	0.310	A1
FEHWOF	4.048	3.213	2.776	2.946	0.835	A1
FOVDOL	3.062	3.214	3.037	0.390	-0.152	A1
FEHWOF	4.031	3.220	2.834	2.867	0.811	A1
FOVDOL	3.132	3.263	3.101	0.440	-0.131	A1
FEHWOF	4.025	3.263	2.941	2.748	0.762	A1
SUDPIQ	3.359	3.282	3.211	0.986	0.077	A1
WUHVOK	3.216	3.286	3.159	0.603	-0.070	A1
SUCCUO	4.037	3.374	3.066	2.626	0.663	A1
MIQTAJ	3.140	3.160	3.062	0.696	-0.020	G
VIXJOD	3.882	3.167	2.856	2.629	0.715	G
GENCEH	3.232	3.168	3.108	0.887	0.064	G
SIKVOZ	3.727	3.171	3.007	2.202	0.556	G
WIMCAW	3.933	3.178	2.854	2.706	0.755	G
XOFNAH	2.995	3.180	2.984	0.256	-0.185	G
MIQTAJ	3.085	3.188	3.045	0.495	-0.103	G
XOFLUB	3.568	3.190	3.165	1.647	0.378	G
VIXJOD	3.729	3.192	2.969	2.256	0.537	G
EZEYIR	3.326	3.196	3.145	1.082	0.130	G
XESWEX	3.890	3.203	2.843	2.655	0.687	G
CETGEP	3.534	3.207	3.198	1.504	0.327	G
GEJNOY01	3.734	3.208	3.122	2.048	0.526	G
CETGEP	3.704	3.211	3.150	1.949	0.493	G
GEJNOY01	3.533	3.212	3.203	1.491	0.321	G
ZITSUQ	3.184	3.235	3.117	0.650	-0.051	G
FAJNIO	3.200	3.241	3.130	0.666	-0.041	G
YARMUZ	3.207	3.242	3.140	0.652	-0.035	G
MIQTAJ	3.103	3.253	3.079	0.385	-0.150	G
FEBSUC	3.369	3.257	3.185	1.098	0.112	G
ASENUI	3.752	3.279	3.207	1.947	0.473	G
HOODKI01	3.277	3.281	3.196	0.724	-0.004	G

Table C3 continued

MIQTAJ	3.169	3.286	3.130	0.496	-0.117	G
GEJNOY	3.833	3.286	3.188	2.128	0.547	G
HOODKI10	3.273	3.287	3.191	0.728	-0.014	G
BACVAC10	3.119	3.288	3.107	0.273	-0.169	G
CETGOZ	4.024	3.290	2.847	2.844	0.734	G
GEJNOY	3.617	3.294	3.276	1.533	0.323	G
GANSIX	3.919	3.302	3.037	2.477	0.617	G
FAJNIO	4.076	3.302	2.878	2.886	0.774	G
BAPGAA	3.481	3.309	3.293	1.128	0.172	G
YAVCUU	3.590	3.336	3.330	1.341	0.254	G
WAJGET	3.942	3.343	3.136	2.388	0.599	G
GUNFIF	3.283	3.347	3.228	0.598	-0.064	G
BACTUU10	3.369	3.348	3.252	0.880	0.021	G
DAKTOZ	4.060	3.353	2.982	2.755	0.707	G
CANSOA	3.293	3.357	3.239	0.594	-0.064	G
TIFDAN	3.688	3.358	3.241	1.760	0.330	G
YAVCUU	3.469	3.360	3.319	1.009	0.109	G
CAZHEQ	3.485	3.381	3.352	0.954	0.104	G
YAVCUU	3.587	3.385	3.346	1.293	0.202	G
YARMUZ	3.274	3.387	3.236	0.497	-0.113	G
QOZYOT	3.184	3.393	3.176	0.226	-0.209	G
GENKAN	4.019	3.400	3.260	2.350	0.619	G
BEBFAP	3.702	3.408	3.363	1.548	0.294	G
BACTOO10	3.639	3.442	3.419	1.246	0.197	G
HOODKI10	3.929	3.453	3.401	1.967	0.476	G
VEQTIU01	3.439	3.457	3.376	0.655	-0.018	G
HOODKI01	3.924	3.467	3.385	1.985	0.457	G
CETGUF	4.185	3.472	3.207	2.689	0.713	G
ROTPUN	3.251	3.475	3.249	0.114	-0.224	G
OCABEZ	3.428	3.484	3.384	0.547	-0.056	G
YAVCUU	3.784	3.496	3.471	1.507	0.288	G
JOYJUC	4.043	3.513	3.318	2.310	0.530	G
GENKAN	3.563	3.569	3.491	0.713	-0.006	G
AHIPEN	3.411	3.581	3.400	0.274	-0.170	G
PAJHEN	4.155	3.829	3.766	1.755	0.326	G
DAMHIK	3.826	3.064	2.814	2.592	0.762	A
EKUNEE	2.864	3.080	2.855	0.227	-0.216	A
EKUNEE	2.903	3.131	2.898	0.170	-0.228	A

Table C3 continued

DAMHIK	3.999	3.222	3.022	2.619	0.777	A
FIWTOU	3.661	3.244	3.233	1.718	0.417	A
CUTFUS	3.185	3.256	3.133	0.573	-0.071	A
SICHIX	3.267	3.266	3.190	0.705	0.001	A
SORNEU	4.080	3.270	2.941	2.828	0.810	A
LEYZIY	3.425	3.289	3.254	1.069	0.136	A
LAQXOQ	3.860	3.401	3.286	2.025	0.459	A
LAQXUW	3.987	3.416	3.320	2.208	0.571	A
CUTFUS	4.151	3.432	3.060	2.805	0.719	A
TIWSIC	3.874	3.532	3.502	1.656	0.342	A
VIZCUE	2.920	2.981	2.862	0.579	-0.061	B
VIZCUE	3.079	3.088	2.991	0.731	-0.009	B
VIZCUE	2.872	3.114	2.866	0.186	-0.242	B
EPUVIV	2.896	3.129	2.891	0.170	-0.233	B
VIZCUE	3.836	3.159	3.025	2.359	0.677	B
WIZJIY	3.497	3.459	3.410	0.775	0.038	E
XACSAV	3.036	3.001	2.923	0.821	0.035	A, B
XACSAV	3.611	3.022	2.940	2.097	0.589	A, B
FLMENK	3.445	3.044	2.995	1.702	0.401	A, B
XACSAV	2.946	3.083	2.915	0.426	-0.137	A, B
FLMENK	3.790	3.177	2.985	2.335	0.613	A, B
BIZGIC	3.948	3.200	2.813	2.770	0.748	A, B
FLMENK	3.961	3.251	2.998	2.589	0.710	A, B
FLMENK	4.101	3.350	3.014	2.781	0.751	A, B
CUTFUS10	3.185	3.256	3.133	0.573	-0.071	A, E
CUTFUS10	4.151	3.432	3.060	2.805	0.719	A, E
CEDXAM	3.115	3.165	3.045	0.657	-0.050	A, G
MEPPAZ	2.952	3.225	2.950	0.109	-0.273	A, G
MEPPED	3.649	3.295	3.299	1.559	0.354	A, G
MEPPAZ	3.829	3.305	3.253	2.020	0.524	A, G
YEMYEW	3.030	3.312	3.030	0.000	-0.282	A, G
YEMYEW	3.843	3.373	3.363	1.860	0.470	A, G
TEJTIL	4.172	3.408	3.035	2.863	0.764	A, G
MEPPAZ01	4.000	3.427	3.263	2.314	0.573	A, G
MEPPED	3.980	3.455	3.520	1.857	0.525	A, G
TUGWUO	3.112	3.119	3.029	0.714	-0.007	A1, G
TUGWUO	3.192	3.220	3.122	0.665	-0.028	A1, G
YARMUZ10	3.207	3.242	3.140	0.652	-0.035	E, G
YARMUZ10	3.274	3.387	3.236	0.497	-0.113	E, G

Table C4. Structure entries for thiol-aromatic interactions, with the cylinder restriction on the reported hydrogen atom.

All entries in analysis of thiol-aromatic interactions, following initial search parameters of the CSD and cylinder restriction on the hydrogen atom, as reported in the CSD. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the thiol-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55. Z refers to the atom or hybridized carbon that is bound to the sulfur atom.
Annotations: (A) polycyclic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (C) thiol groups were "pointed away" from the plane of the aromatic ring (as defined by ($d_{S–plane} - d_{H–plane}$) ≥ –0.10 Å); (D) where thiol groups where the sulfur atom is apparently tri-coordinate; (E) redundant structures or entries; (F) unusual structures with questionable features.

Refcode	S–H bond length, Å	Angle θ, Z–S–H bond angle, degrees	$d_{H–centroid}$, Å	$d_{H–Cmin}$, Å	H d_{plane}, Å	H r, Å	$d_{S–centroid}$, Å	$d_{S–Cmin}$, Å	S d_{plane}, Å	S r, Å	Δ S $d_{plane–Hplane}$, Å	Angle S–H–centroid, degrees	$\Delta d_{centroid–Cmin}$, Å	Parsing Rationale	Z
WANCIX	1.251	107.05	2.610	2.200	2.197	1.409	3.729	3.152	3.099	2.074	0.902	147.84	0.410		sp2
FULGIE	1.214	95.37	2.875	2.327	2.291	1.737	3.932	3.087	2.684	2.873	0.393	145.08	0.548		sp2
NUMWUP	1.236	100.29	2.823	2.392	2.396	1.493	3.875	3.089	2.837	2.640	0.441	142.31	0.431		sp2
ODOSAD	1.312	110.13	2.641	2.472	2.428	1.039	3.746	3.644	3.612	0.993	1.184	140.34	0.169		sp2
HOMPOP	1.020	94.95	3.108	2.481	2.402	1.972	3.992	3.138	2.762	2.882	0.360	145.68	0.627		sp2
NUZLAW	1.200	109.44	3.531	2.484	1.886	2.985	3.955	3.215	2.777	2.816	0.891	101.79	1.047		Pd

Table C4 continued

COLDEN	1.322	103.47	2.902	2.490	2.464	1.533	4.204	3.668	3.568	2.223	1.104	167.89	0.412	sp2
DEBYUG	1.251	102.78	3.196	2.518	2.290	2.229	3.812	3.220	3.072	2.257	0.782	110.14	0.678	sp2
COLDEN	1.356	100.51	3.235	2.525	2.399	2.170	4.273	3.737	3.537	2.398	1.138	132.68	0.710	sp2
YULZUA	1.303	113.44	2.960	2.535	2.515	1.561	4.026	3.764	3.690	1.610	1.175	138.27	0.425	sp2
GAGTUF	1.375	92.92	2.833	2.536	2.504	1.325	3.721	3.722	3.652	0.713	1.148	120.26	0.297	sp2
COLDEN	1.343	99.16	2.690	2.549	2.494	1.008	4.014	3.801	3.796	1.305	1.302	168.33	0.141	sp2
DUCPIB	1.200	109.49	3.434	2.560	2.145	2.682	4.604	3.704	3.025	3.471	0.880	164.99	0.874	sp2
KOYJIT	1.240	98.99	3.091	2.569	2.533	1.771	4.127	3.767	3.737	1.751	1.204	140.79	0.522	sp3
VECXIL	1.226	96.64	2.740	2.600	2.522	1.071	3.593	3.621	3.539	0.621	1.017	125.46	0.140	sp2
KOYJIT	1.257	96.83	2.990	2.613	2.590	1.494	4.169	3.839	3.749	1.824	1.159	156.01	0.377	sp3
YUCYUR	1.200	109.45	3.266	2.653	2.397	2.218	4.443	3.838	3.397	2.864	1.000	166.88	0.613	sp2
VECXIL	1.251	97.16	2.648	2.668	2.560	0.677	3.551	3.581	3.497	0.617	0.937	127.57	-0.020	sp2
VECXIL01	1.195	96.13	2.853	2.679	2.608	1.157	3.624	3.639	3.562	0.667	0.954	121.53	0.174	sp2
YULZUA	1.186	104.69	2.892	2.684	2.643	1.174	4.010	3.782	3.725	1.485	1.082	156.84	0.208	sp2
COLDAJ	1.243	96.30	2.780	2.694	2.632	0.895	3.779	3.693	3.629	1.054	0.997	136.41	0.086	sp2
COLDEN	1.344	87.89	2.637	2.696	2.574	0.573	3.878	3.625	3.619	1.393	1.045	152.21	-0.059	sp2
ODOSAD	1.300	100.68	2.743	2.696	2.608	0.850	3.805	3.933	3.791	0.326	1.183	137.62	0.047	sp2
YULZOU	1.160	99.53	2.692	2.705	2.608	0.667	3.684	3.649	3.598	0.791	0.990	142.92	-0.013	sp2
VECXIL01	1.184	100.93	2.726	2.722	2.632	0.710	3.599	3.607	3.537	0.665	0.905	129.70	0.004	sp2
PESPIN	1.231	97.82	3.163	2.724	2.552	1.869	3.927	3.139	2.832	2.720	0.280	120.07	0.439	sp3
QAXPEM01	1.198	102.81	3.196	2.729	2.602	1.856	4.076	3.315	3.018	2.740	0.416	130.35	0.467	Si
KILCUE	1.200	109.41	2.698	2.731	2.628	0.611	3.715	3.558	3.506	1.228	0.878	141.64	-0.033	sp2
EXUVUO	1.131	94.93	3.058	2.733	2.731	1.376	4.061	3.767	3.723	1.622	0.992	147.83	0.325	Ge
LIGDOU	1.255	106.57	3.361	2.752	2.528	2.215	4.445	3.690	2.891	3.376	0.363	144.66	0.609	Ru
MIBSAR	1.199	104.36	2.774	2.758	2.655	0.804	3.960	3.862	3.828	1.014	1.173	169.90	0.016	sp2
YUKDIR	1.280	116.38	3.254	2.767	2.653	1.884	4.315	3.607	3.313	2.765	0.660	140.07	0.487	W

Table C4 continued

HIYSAK	1.108	100.54	2.642	2.771	2.602	0.458	3.750	3.784	3.697	0.628	1.095	178.53	-0.129	sp3
GUJNOP	1.246	111.77	3.119	2.776	2.681	1.594	3.903	3.413	3.351	2.001	0.670	120.56	0.343	Mo
TAXMUA	1.365	96.41	2.537	2.783	2.527	0.225	3.725	3.810	3.689	0.517	1.162	143.66	-0.246	sp3
SIZBAE	1.312	100.26	2.619	2.786	2.592	0.375	3.733	3.873	3.717	0.345	1.125	141.10	-0.167	sp3
VECXIL	1.229	95.88	3.161	2.787	2.684	1.670	4.229	3.778	3.720	2.011	1.036	145.23	0.374	sp2
GIQDER	1.259	96.21	3.186	2.797	2.702	1.688	4.407	4.021	3.851	2.143	1.149	163.22	0.389	sp2
XUDQOE	1.199	109.50	3.287	2.802	2.751	1.799	4.159	3.919	3.892	1.466	1.141	129.92	0.485	sp2
WAKHOF02	1.257	94.04	3.546	2.806	2.453	2.561	4.764	4.035	3.362	3.375	0.909	163.40	0.740	Si
GAGTUF	1.380	86.42	3.543	2.808	2.539	2.471	4.260	3.857	3.803	1.920	1.264	112.13	0.735	sp2
SOSPAT	1.254	92.96	2.832	2.810	2.737	0.727	3.559	3.737	3.552	0.223	0.815	115.69	0.022	Al
HUDFIW	1.200	109.51	3.676	2.811	2.461	2.731	4.166	3.583	3.315	2.523	0.854	105.80	0.865	sp2
WAQFEZ	1.362	104.07	2.975	2.814	2.769	1.088	3.912	3.816	3.757	1.090	0.988	124.55	0.161	sp3
QAXPEM	1.107	94.08	3.444	2.823	2.663	2.184	4.186	3.367	2.939	2.981	0.276	125.56	0.621	Si
MEDCII	1.235	99.89	3.673	2.823	2.159	2.971	4.195	3.568	3.221	2.688	1.062	106.59	0.850	sp2
ICOPUN	1.186	97.77	2.575	2.830	2.568	0.190	3.573	3.667	3.538	0.499	0.970	140.63	-0.255	sp2
NAMQEY	1.017	102.78	3.714	2.843	2.471	2.773	4.573	3.737	3.177	3.289	0.706	143.71	0.871	sp2
FAMZIC01	1.210	104.99	3.534	2.850	2.594	2.400	4.688	3.893	3.196	3.430	0.602	159.73	0.684	S
KUKGAZ	1.190	96.11	3.590	2.856	2.583	2.493	4.620	3.650	2.837	3.646	0.254	145.50	0.734	sp3
ESABAB	1.211	94.47	3.483	2.862	2.729	2.164	4.573	4.045	3.830	2.499	1.101	150.23	0.621	sp3
ZZZLWW01	1.295	95.39	3.653	2.867	2.599	2.567	4.337	3.871	3.704	2.256	1.105	113.65	0.786	sp2
FAMZIC01	1.210	104.99	3.574	2.872	2.676	2.369	4.114	3.682	3.518	2.133	0.842	108.09	0.702	S
PESPIN	1.231	97.82	2.822	2.891	2.757	0.602	3.489	3.565	3.445	0.552	0.688	112.82	-0.069	sp3
VECXIL01	1.190	98.86	3.270	2.892	2.768	1.741	4.259	3.823	3.763	1.995	0.995	140.68	0.378	sp2
HUCJOF	1.200	109.45	3.378	2.902	2.872	1.778	3.630	3.577	3.539	0.808	0.667	92.30	0.476	sp2
FAMZIC01	1.210	104.99	3.431	2.910	2.821	1.953	3.922	3.264	3.016	2.507	0.195	104.97	0.521	S
WONHAK	1.288	96.52	3.798	2.929	2.577	2.790	4.403	3.622	2.892	3.320	0.315	109.73	0.869	sp2

Table C4 continued

KECXIA	1.221	94.28	3.695	2.944	2.632	2.593	4.798	4.125	3.602	3.170	0.970	150.87	0.751	sp2
DPDHPS10	1.272	101.22	3.290	2.964	2.927	1.502	3.702	3.707	3.638	0.685	0.711	98.71	0.326	P
KUKGAZ	1.201	97.21	3.461	2.969	2.940	1.826	4.366	3.557	3.227	2.941	0.287	132.77	0.492	sp3
POHJON01	1.200	109.37	3.630	2.978	2.812	2.296	4.372	3.765	3.433	2.707	0.621	121.06	0.652	sp3
HOMPOP	1.016	105.66	3.552	2.980	2.792	2.196	4.246	3.617	3.442	2.486	0.650	127.33	0.572	sp2
DPDHPS01	1.247	98.51	3.327	2.983	2.944	1.550	3.651	3.645	3.579	0.721	0.635	94.87	0.344	P
XAFVEH	1.236	98.77	2.840	2.986	2.813	0.391	3.625	3.674	3.575	0.600	0.762	120.36	-0.146	sp3
JURJAH	0.990	83.80	3.730	2.994	2.802	2.462	3.990	3.366	2.991	2.641	0.189	98.00	0.736	Ru
ICOPUN	1.158	97.37	3.018	3.002	2.928	0.732	3.846	3.895	3.806	0.553	0.878	128.34	0.016	sp2
AZIKEZ	1.181	97.05	3.644	3.013	2.625	2.527	4.452	3.676	3.257	3.035	0.632	126.70	0.631	sp2
HUDFIW	1.200	109.48	3.449	3.016	2.847	1.947	4.592	4.110	3.923	2.387	1.076	159.55	0.433	sp2
MELFAK	1.200	109.49	3.197	3.019	2.988	1.137	3.324	3.590	3.323	0.082	0.335	85.40	0.178	sp2
KECXIA	1.221	94.28	3.802	3.028	2.712	2.665	3.716	3.378	3.343	1.623	0.631	76.70	0.774	sp2
DIXFAS	1.180	99.20	3.733	3.040	2.843	2.419	3.930	3.626	3.620	1.530	0.777	90.78	0.693	sp3
COLDEN	1.327	83.11	3.500	3.055	2.948	1.887	4.579	3.929	3.391	3.077	0.443	138.45	0.445	sp2
DUCPIB	1.200	109.46	3.139	3.058	3.011	0.887	3.486	3.666	3.478	0.236	0.467	96.56	0.081	sp2
GUJNOP	1.246	111.77	3.568	3.074	3.026	1.890	4.161	3.508	3.187	2.675	0.161	109.90	0.494	Mo
QUSSUS	1.200	109.50	3.886	3.084	2.541	2.940	4.221	3.572	3.356	2.560	0.815	97.88	0.802	Zn
XUGPEU	0.905	106.19	2.890	3.108	2.881	0.228	3.731	3.762	3.674	0.650	0.793	155.26	-0.218	sp2
SIZBAE	1.396	97.94	3.827	3.110	2.901	2.496	4.781	4.213	3.927	2.727	1.026	125.92	0.717	sp3
WESQER	1.334	90.12	2.950	3.122	2.930	0.343	3.873	3.760	3.727	1.053	0.797	125.03	-0.172	sp3
WUJWAZ	1.314	93.55	3.848	3.139	2.774	2.667	4.154	3.612	3.464	2.293	0.690	94.10	0.709	sp2
LUVKOD	1.286	95.53	3.450	3.147	3.058	1.597	4.527	3.926	3.594	2.753	0.536	141.45	0.303	Si
MOCGIW	1.156	102.82	3.784	3.179	3.023	2.276	4.755	3.924	3.336	3.388	0.313	142.66	0.605	Ti
HUCMIC	1.200	109.47	4.046	3.196	2.729	2.987	3.905	3.396	3.337	2.028	0.608	74.71	0.850	sp2
GUJNOP	1.246	111.77	3.621	3.200	3.179	1.734	4.121	4.072	4.037	0.828	0.858	104.92	0.421	Mo

Table C4 continued

WIXTII	1.200	109.46	3.240	3.213	3.129	0.841	4.253	3.911	3.896	1.706	0.767	142.27	0.027	sp2
VECXIL01	1.185	99.28	3.773	3.216	2.897	2.417	4.622	3.793	2.914	3.588	0.017	129.81	0.557	sp2
WAWQOC	0.897	108.76	3.567	3.219	3.205	1.566	3.924	3.583	3.518	1.738	0.313	106.97	0.348	sp2
VECXIL	1.255	96.43	3.761	3.221	2.931	2.357	4.581	3.742	2.834	3.599	-0.097	123.91	0.540	sp2
GIQDER	1.259	96.21	3.846	3.251	3.127	2.239	4.371	3.755	3.345	2.814	0.218	106.41	0.595	sp2
WIXTII	1.200	109.46	3.543	3.275	3.190	1.542	4.245	4.190	4.151	0.888	0.961	118.26	0.268	sp2
SOSPAT	1.203	92.20	3.043	3.277	3.038	0.174	3.740	3.639	3.614	0.963	0.576	116.62	-0.234	Al
MEDBUT	1.120	99.64	4.009	3.284	2.909	2.759	4.370	3.950	3.824	2.115	0.915	101.37	0.725	sp2
XUDQOE	1.200	109.50	3.517	3.305	3.250	1.344	4.434	4.407	4.349	0.864	1.099	133.94	0.212	sp2
VECXIL	1.251	97.16	4.016	3.309	2.757	2.920	4.653	3.662	2.659	3.818	-0.098	113.17	0.707	sp2
VECXIL01	1.184	100.93	4.016	3.310	2.768	2.910	4.698	3.710	2.723	3.828	-0.045	118.52	0.706	sp2
FULGIE	1.214	95.37	3.417	3.329	3.283	0.948	4.132	3.638	3.556	2.104	0.273	118.20	0.088	sp2
HUCMIC	1.201	109.51	3.681	3.351	3.336	1.556	4.595	4.135	4.013	2.238	0.677	133.86	0.330	sp2
LAWKIE	0.959	110.63	3.181	3.354	3.170	0.264	3.725	3.609	3.562	1.090	0.392	117.65	-0.173	sp3
BOZTOB	1.339	107.26	3.622	3.358	3.296	1.502	4.711	4.208	4.119	2.286	0.823	138.69	0.264	sp3
PEDXIF	0.998	99.92	3.228	3.368	3.202	0.409	4.027	3.948	3.900	1.003	0.698	138.21	-0.140	Ir
COLDAJ	1.243	96.30	3.928	3.396	3.283	2.157	4.637	3.855	3.214	3.342	-0.069	117.63	0.532	sp2
COLDEN	1.322	103.47	4.175	3.474	3.172	2.715	4.452	3.700	3.226	3.068	0.054	93.36	0.701	sp2
XIGPUA	1.158	97.46	4.172	3.491	3.289	2.567	3.906	3.619	3.613	1.484	0.324	68.79	0.681	sp2
HUSNEP	1.317	95.72	4.178	3.496	3.268	2.603	4.703	3.986	3.376	3.274	0.108	105.43	0.682	sp2
BEJTAM	0.956	109.17	4.160	3.508	3.334	2.488	4.502	4.096	4.061	1.943	0.727	104.91	0.652	sp3
ILOQAB	1.007	98.37	4.204	3.523	3.141	2.794	4.623	3.883	3.558	2.952	0.417	108.49	0.681	sp2
FATQIC	1.200	109.50	4.170	3.536	3.215	2.656	4.127	3.550	3.430	2.295	0.215	79.65	0.634	sp2
BEJTAM	0.956	109.17	3.953	3.542	3.510	1.818	4.063	3.969	3.946	0.968	0.436	89.75	0.411	sp3
YIRGEM	1.196	97.55	4.214	3.595	3.163	2.784	3.725	3.424	3.355	1.619	0.192	58.17	0.619	sp2
RONVAR01	1.403	93.42	3.943	3.628	3.593	1.624	4.738	4.063	3.647	3.025	0.054	116.47	0.315	sp2

Table C4 continued

ICOPUN	1.186	97.77	3.689	3.660	3.600	0.805	3.531	3.662	3.510	0.385	-0.090	73.06	0.029		sp2
RONVAR01	1.403	93.42	4.393	3.759	3.559	2.575	4.525	3.932	3.504	2.863	-0.055	86.33	0.634		sp2
MURMER	1.220	99.59	4.432	3.779	3.489	2.733	4.792	3.944	3.441	3.335	-0.048	99.76	0.653		sp3
RUFDAZ	1.200	109.51	4.415	3.806	3.508	2.681	4.025	3.712	3.660	1.675	0.152	63.48	0.609		sp3
HELXUR01	0.961	109.12	3.170	2.484	2.209	2.274	3.820	3.266	3.027	2.330	0.818	126.41	0.686	A1	sp2
ZORQUU	0.950	109.52	2.889	2.553	2.474	1.492	3.724	3.123	2.927	2.302	0.453	147.31	0.336	A1	sp2
HELXUR01	0.963	109.20	3.290	2.566	2.178	2.466	3.901	3.263	2.979	2.519	0.801	123.15	0.724	A1	sp2
HELXUR01	0.959	109.15	3.002	2.634	2.508	1.650	3.827	3.192	2.926	2.467	0.418	145.04	0.368	A1	sp2
HELXUR01	0.959	109.31	3.220	2.750	2.596	1.905	4.022	3.340	2.992	2.688	0.396	142.39	0.470	A1	sp2
VEZLOC	1.281	97.94	2.821	2.780	2.696	0.830	4.056	3.994	3.947	0.934	1.251	161.44	0.041	A1	sp3
GOCGUB	1.327	102.12	3.477	2.786	2.394	2.522	4.662	4.101	3.667	2.879	1.273	148.70	0.691	A1	sp2
ZORQUU	1.180	117.01	3.408	3.026	2.946	1.713	4.060	3.348	2.973	2.765	0.027	115.63	0.382	A1	sp2
VEZLOC	1.281	97.94	3.300	3.040	2.981	1.415	3.640	3.771	3.623	0.351	0.642	94.86	0.260	A1	sp3
YOMWAY01	1.136	107.93	2.948	3.182	2.941	0.203	3.783	3.645	3.620	1.098	0.679	130.31	-0.234	A1	sp2
KASJIZ	1.388	109.86	3.570	3.533	3.470	0.839	3.504	3.517	3.434	0.697	-0.036	76.02	0.037	A1	sp2
HELXUR01	0.959	109.31	3.921	3.554	3.438	1.885	3.924	3.292	2.897	2.647	-0.541	83.19	0.367	A1	sp2
HELXUR01	0.959	109.15	3.863	3.591	3.526	1.578	3.848	3.273	3.019	2.386	-0.507	81.97	0.272	A1	sp2
CEDYEQ	1.129	113.94	3.638	2.802	2.076	2.988	4.707	3.878	3.001	3.626	0.925	158.45	0.836	A	sp3
VOPBEH	1.102	106.39	3.214	2.927	2.827	1.529	4.192	3.638	3.420	2.424	0.593	148.44	0.287	A	sp3
PUMDUX	1.201	109.45	3.753	2.943	2.438	2.853	3.957	3.319	3.103	2.455	0.665	90.83	0.810	A	Re
XORTOP	0.979	99.94	3.485	2.963	2.834	2.028	4.054	3.786	3.703	1.650	0.869	119.23	0.522	A	sp2
VOPBEH	1.199	96.12	2.753	2.979	2.742	0.246	3.466	3.402	3.356	0.866	0.614	117.04	-0.226	A	sp3
XORTOP	0.980	102.96	3.506	2.981	2.742	2.185	4.064	3.732	3.633	1.821	0.891	118.38	0.525	A	sp2
KAPNIZ	0.986	104.86	3.488	3.008	2.809	2.068	3.837	3.525	3.510	1.550	0.701	103.28	0.480	A	sp2
XORTIJ	1.205	89.96	3.538	3.020	2.966	1.929	4.597	4.089	3.988	2.287	1.022	147.15	0.518	A	sp2
XORTIJ	1.197	98.15	3.514	3.051	3.001	1.828	4.506	4.093	4.035	2.006	1.034	140.86	0.463	A	sp2

Table C4 continued

BAJNUW	1.247	95.97	2.953	3.059	2.917	0.460	3.679	3.630	3.591	0.800	0.674	116.24	-0.106	A	Se
BAPBOK	1.265	91.32	3.461	3.059	3.033	1.667	3.377	3.488	3.344	0.471	0.311	75.60	0.402	A	sp2
WOGRIT01	1.137	97.56	3.679	3.091	3.000	2.130	4.709	4.184	4.042	2.416	1.042	151.49	0.588	A	sp3
XORTOP	0.981	91.71	3.992	3.242	2.671	2.967	4.582	3.789	3.284	3.195	0.613	121.54	0.750	A	sp2
XORTOP	0.980	102.96	3.252	3.244	3.178	0.690	3.620	3.372	3.356	1.357	0.178	104.24	0.008	A	sp2
XORTOP	0.980	102.96	3.671	3.244	3.190	1.817	4.130	3.525	3.351	2.414	0.161	111.28	0.427	A	sp2
XORTOP	0.981	91.71	3.978	3.262	2.740	2.884	4.149	3.529	3.326	2.480	0.586	93.14	0.716	A	sp2
IKEGOV	1.200	109.50	3.496	3.402	3.370	0.930	4.345	4.403	4.311	0.542	0.941	128.42	0.094	A	sp3
XORTIJ	1.197	98.15	4.125	3.483	3.042	2.786	4.699	3.792	3.096	3.535	0.054	111.63	0.642	A	sp2
XORTOP	0.980	105.91	4.304	3.512	3.207	2.870	3.895	3.407	3.333	2.015	0.126	59.23	0.792	A	sp2
KECXIA	1.221	94.28	3.417	2.544	2.228	2.591	3.971	2.769	1.598	3.635	-0.630	108.18	0.873	C	sp2
KOYJIT	1.240	98.99	3.622	2.699	2.123	2.935	4.023	2.794	1.273	3.816	-0.850	99.75	0.923	C	sp3
KOYJIT	1.257	96.83	3.604	2.704	2.289	2.784	3.992	2.787	1.456	3.717	-0.833	98.69	0.900	C	sp3
SIZBAE	1.396	97.94	3.652	2.745	2.300	2.837	3.945	2.745	1.274	3.734	-1.026	91.58	0.907	C	sp3
XAFVEH	1.236	98.77	3.831	2.873	2.402	2.984	3.987	2.789	1.639	3.635	-0.763	88.13	0.958	C	sp3
ADELOL	1.075	87.31	4.064	3.238	2.837	2.910	3.743	2.915	2.488	2.796	-0.349	65.25	0.826	C	sp2
JURJAH	1.180	99.19	3.990	3.388	3.173	2.419	4.531	3.727	2.897	3.484	-0.276	110.05	0.602	C	Ru
NAMROI	1.200	109.49	4.090	3.443	3.253	2.479	4.481	3.504	2.747	3.540	-0.506	101.25	0.647	C	Zn
GUJNOP	0.972	102.79	3.851	3.483	3.461	1.689	3.750	3.276	3.127	2.070	-0.334	76.77	0.368	C	Mo
GADNIJ	1.199	109.49	3.902	3.619	3.521	1.682	3.574	3.435	3.408	1.077	-0.113	65.43	0.283	C	sp2
QUSSUS	1.200	109.50	4.243	3.630	3.481	2.426	4.534	3.581	2.907	3.479	-0.574	96.30	0.613	C	Zn
HUCGES	1.200	109.48	4.223	3.701	3.581	2.238	4.385	3.696	3.153	3.047	-0.428	89.73	0.522	C	sp2
JIVNAD	1.171	105.99	4.482	3.738	3.415	2.903	4.650	3.709	2.873	3.656	-0.542	90.89	0.744	C	sp2
GAQQAR	1.226	97.35	4.499	3.912	3.518	2.804	4.414	3.603	3.176	3.065	-0.342	78.19	0.587	C	sp2
GAYKUO	1.059	102.96	3.152	2.472	2.305	2.150	3.816	3.336	3.140	2.168	0.835	121.73	0.680	D	Ru, sp3
WORNAS	1.200	110.39	3.161	2.583	2.502	1.932	3.712	3.274	3.193	1.893	0.691	108.00	0.578	D	Ru, H2S

Table C4 continued

WORNEW	1.242	100.71	2.917	2.585	2.548	1.420	3.906	3.328	3.014	2.484	0.466	135.86	0.332	D	Ru, H2S
SULTOI	1.187	115.03	3.244	2.653	2.544	2.013	4.177	3.274	2.728	3.163	0.184	135.68	0.591	D	Ru, sp3
DIRVAB	1.225	104.52	3.072	2.661	2.673	1.514	4.031	3.545	3.445	2.093	0.772	134.87	0.411	D	Ru, Ru
DIRVAB	1.170	109.79	3.057	2.671	2.581	1.638	4.114	3.565	3.232	2.545	0.651	150.32	0.386	D	Ru, Ru
JIHHUD	1.349	108.14	3.389	2.718	2.526	2.259	4.163	3.382	2.895	2.992	0.369	116.09	0.671	D	Ru, sp3
CEMROB	1.037	104.36	3.561	2.760	2.429	2.604	4.042	3.331	2.811	2.904	0.382	110.42	0.801	D	Rh, Rh
WORNAS	1.200	110.39	2.951	2.803	2.734	1.111	3.671	3.185	3.145	1.893	0.411	117.96	0.148	D	Ru, H_2S
SULTOI	1.187	115.03	3.361	2.819	2.576	2.159	4.258	3.429	2.651	3.332	0.075	132.84	0.542	D	Ru, sp3
GAYKUO	1.059	102.96	3.197	2.846	2.829	1.489	3.734	3.127	2.990	2.237	0.161	112.58	0.351	D	Ru, sp3
RODBIV	1.149	113.16	3.534	2.942	2.848	2.092	4.257	3.682	3.464	2.474	0.616	122.09	0.592	D	Mn, Mn
CEMROB	1.037	104.36	3.489	3.129	3.120	1.562	3.655	3.184	3.014	2.068	-0.106	90.89	0.360	D	Rh, Rh
SIZCAF	1.197	107.39	3.949	3.138	2.807	2.778	3.904	3.302	2.936	2.573	0.129	79.10	0.811	D	Ru, H_2S
VETKAH	1.303	102.18	4.048	3.177	2.741	2.979	3.719	3.274	3.120	2.024	0.379	66.24	0.871	D	Rh, sp2
RODBIV	1.149	113.16	3.816	3.194	3.037	2.311	4.556	4.079	3.984	2.210	0.947	123.74	0.622	D	Mn, Mn
YUKDIR01	1.280	116.38	3.254	2.767	2.653	1.884	4.315	3.607	3.313	2.765	0.660	140.07	0.487	E	W
WEKYAO	1.314	100.24	3.813	3.040	2.675	2.717	3.670	3.226	3.175	1.841	0.500	73.79	0.773	F	Ru^+
VOPBEH	1.199	96.12	3.647	2.705	2.278	2.848	3.985	2.787	1.625	3.639	-0.653	97.50	0.942	A, C	sp3
ZORQUU	1.180	117.01	3.607	3.254	3.194	1.676	3.947	3.316	2.900	2.677	-0.294	97.96	0.353	A1, C	sp2
ZORQUU	0.950	109.52	3.862	3.590	3.536	1.553	3.813	3.257	3.009	2.342	-0.527	80.00	0.272	A1, C	sp2
SIZCAF10	1.197	107.39	3.949	3.138	2.807	2.778	3.904	3.302	2.936	2.573	0.129	79.10	0.811	D, E	Ru, H_2S

Table C5. Structure entries for thiol-aromatic interactions, with the cylinder restriction on the sulfur atom and reported S–H bond lengths.

All entries in analysis of thiol-aromatic interactions, following initial search parameters of the CSD and cylinder restriction on the sulfur atom, with S–H bond lengths as reported in the CSD. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the thiol-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55. Z refers to the atom or hybridized carbon that is bound to the sulfur atom.
Annotations: (A) polycyclic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (C) thiol groups were "pointed away" from the plane of the aromatic ring (as defined by ($d_{S–plane} - d_{H–plane}$) ≥ –0.10 Å); (D) where thiol groups where the sulfur atom is apparently tri-coordinate; (E) redundant structures or entries; (F) unusual structures with questionable features.

Refcode	S–H bond length, Å	Angle θ, Z–S–H bond angle, degrees	$d_{H–centroid}$, Å	$d_{H–Cmin}$, Å	H d_{plane}, Å	H r, Å	$d_{S–centroid}$, Å	$d_{S–Cmin}$, Å	S d_{plane}, Å	S r, Å	Δ S $d_{plane–Hplane}$, Å	Angle S–H–centroid, degrees	$\Delta d_{centroid–Cmin}$, Å	Parsing Rationale	Z
WANCIX	1.251	107.05	2.610	2.200	2.197	1.409	3.729	3.152	3.099	2.074	0.902	147.84	0.410		sp2
FULGIE	1.214	95.37	2.875	2.327	2.291	1.737	3.932	3.087	2.684	2.873	0.393	145.08	0.548		sp2
NUMWUP	1.236	100.29	2.823	2.392	2.396	1.493	3.875	3.089	2.837	2.640	0.441	142.31	0.431		sp2
ODOSAD	1.312	110.13	2.641	2.472	2.428	1.039	3.746	3.644	3.612	0.993	1.184	140.34	0.169		sp2
HOMPOP	1.020	94.95	3.108	2.481	2.402	1.972	3.992	3.138	2.762	2.882	0.360	145.68	0.627		sp2
NUZLAW	1.200	109.44	3.531	2.484	1.886	2.985	3.955	3.215	2.777	2.816	0.891	101.79	1.047		Pd

Table C5 continued

COLDEN	1.322	103.47	2.902	2.490	2.464	1.533	4.204	3.668	3.568	2.223	1.104	167.89	0.412	sp2
DEBYUG	1.251	102.78	3.196	2.518	2.290	2.229	3.812	3.220	3.072	2.257	0.782	110.14	0.678	sp2
COLDEN	1.356	100.51	3.235	2.525	2.399	2.170	4.273	3.737	3.537	2.398	1.138	132.68	0.710	sp2
YULZUA	1.303	113.44	2.960	2.535	2.515	1.561	4.026	3.764	3.690	1.610	1.175	138.27	0.425	sp2
GAGTUF	1.375	92.92	2.833	2.536	2.504	1.325	3.721	3.722	3.652	0.713	1.148	120.26	0.297	sp2
COLDEN	1.343	99.16	2.690	2.549	2.494	1.008	4.014	3.801	3.796	1.305	1.302	168.33	0.141	sp2
DUCPIB	1.200	109.49	3.434	2.560	2.145	2.682	4.604	3.704	3.025	3.471	0.880	164.99	0.874	sp2
KOYJIT	1.240	98.99	3.091	2.569	2.533	1.771	4.127	3.767	3.737	1.751	1.204	140.79	0.522	sp3
TASBIY	1.200	84.40	3.722	2.594	1.470	3.419	4.466	3.510	2.652	3.593	1.182	121.36	1.128	Mg
VECXIL	1.226	96.64	2.740	2.600	2.522	1.071	3.593	3.621	3.539	0.621	1.017	125.46	0.140	sp2
KOYJIT	1.257	96.83	2.990	2.613	2.590	1.494	4.169	3.839	3.749	1.824	1.159	156.01	0.377	sp3
SIZBAE	1.312	100.26	3.638	2.615	1.565	3.284	3.917	2.754	1.605	3.573	0.040	92.32	1.023	sp3
YUCYUR	1.200	109.45	3.266	2.653	2.397	2.218	4.443	3.838	3.397	2.864	1.000	166.88	0.613	sp2
VECXIL	1.251	97.16	2.648	2.668	2.560	0.677	3.551	3.581	3.497	0.617	0.937	127.57	-0.020	sp2
VECXIL01	1.195	96.13	2.853	2.679	2.608	1.157	3.624	3.639	3.562	0.667	0.954	121.53	0.174	sp2
YULZUA	1.186	104.69	2.892	2.684	2.643	1.174	4.010	3.782	3.725	1.485	1.082	156.84	0.208	sp2
COLDAJ	1.243	96.30	2.780	2.694	2.632	0.895	3.779	3.693	3.629	1.054	0.997	136.41	0.086	sp2
COLDEN	1.344	87.89	2.637	2.696	2.574	0.573	3.878	3.625	3.619	1.393	1.045	152.21	-0.059	sp2
ODOSAD	1.300	100.68	2.743	2.696	2.608	0.850	3.805	3.933	3.791	0.326	1.183	137.62	0.047	sp2
YULZOU	1.160	99.53	2.692	2.705	2.608	0.667	3.684	3.649	3.598	0.791	0.990	142.92	-0.013	sp2
VECXIL01	1.184	100.93	2.726	2.722	2.632	0.710	3.599	3.607	3.537	0.665	0.905	129.70	0.004	sp2
PESPIN	1.231	97.82	3.163	2.724	2.552	1.869	3.927	3.139	2.832	2.720	0.280	120.07	0.439	sp3
QAXPEM01	1.198	102.81	3.196	2.729	2.602	1.856	4.076	3.315	3.018	2.740	0.416	130.35	0.467	Si
KILCUE	1.200	109.41	2.698	2.731	2.628	0.611	3.715	3.558	3.506	1.228	0.878	141.64	-0.033	sp2
EXUVUO	1.131	94.93	3.058	2.733	2.731	1.376	4.061	3.767	3.723	1.622	0.992	147.83	0.325	Ge
LIGDOU	1.255	106.57	3.361	2.752	2.528	2.215	4.445	3.690	2.891	3.376	0.363	144.66	0.609	Ru

Table C5 continued

MIBSAR	1.199	104.36	2.774	2.758	2.655	0.804	3.960	3.862	3.828	1.014	1.173	169.90	0.016	sp2
YUKDIR	1.280	116.38	3.254	2.767	2.653	1.884	4.315	3.607	3.313	2.765	0.660	140.07	0.487	W
HIYSAK	1.108	100.54	2.642	2.771	2.602	0.458	3.750	3.784	3.697	0.628	1.095	178.53	-0.129	sp3
GUJNOP	1.246	111.77	3.119	2.776	2.681	1.594	3.903	3.413	3.351	2.001	0.670	120.56	0.343	Mo
TAXMUA	1.365	96.41	2.537	2.783	2.527	0.225	3.725	3.810	3.689	0.517	1.162	143.66	-0.246	sp3
SIZBAE	1.312	100.26	2.619	2.786	2.592	0.375	3.733	3.873	3.717	0.345	1.125	141.10	-0.167	sp3
VECXIL	1.229	95.88	3.161	2.787	2.684	1.670	4.229	3.778	3.720	2.011	1.036	145.23	0.374	sp2
GIQDER	1.259	96.21	3.186	2.797	2.702	1.688	4.407	4.021	3.851	2.143	1.149	163.22	0.389	sp2
XUDQOE	1.199	109.50	3.287	2.802	2.751	1.799	4.159	3.919	3.892	1.466	1.141	129.92	0.485	sp2
WAQFEZ	1.362	104.07	3.915	2.803	1.699	3.527	4.749	3.806	3.014	3.670	1.315	120.25	1.112	sp3
WAKHOF02	1.257	94.04	3.546	2.806	2.453	2.561	4.764	4.035	3.362	3.375	0.909	163.40	0.740	Si
GAGTUF	1.380	86.42	3.543	2.808	2.539	2.471	4.260	3.857	3.803	1.920	1.264	112.13	0.735	sp2
SOSPAT	1.254	92.96	2.832	2.810	2.737	0.727	3.559	3.737	3.552	0.223	0.815	115.69	0.022	AL
HUDFIW	1.200	109.51	3.676	2.811	2.461	2.731	4.166	3.583	3.315	2.523	0.854	105.80	0.865	sp2
WAQFEZ	1.362	104.07	2.975	2.814	2.769	1.088	3.912	3.816	3.757	1.090	0.988	124.55	0.161	sp3
QAXPEM	1.107	94.08	3.444	2.823	2.663	2.184	4.186	3.367	2.939	2.981	0.276	125.56	0.621	Si
MEDCII	1.235	99.89	3.673	2.823	2.159	2.971	4.195	3.568	3.221	2.688	1.062	106.59	0.850	sp2
ICOPUN	1.186	97.77	2.575	2.830	2.568	0.190	3.573	3.667	3.538	0.499	0.970	140.63	-0.255	sp2
NAMQEY	1.017	102.78	3.714	2.843	2.471	2.773	4.573	3.737	3.177	3.289	0.706	143.71	0.871	sp2
FAMZIC01	1.210	104.99	3.534	2.850	2.594	2.400	4.688	3.893	3.196	3.430	0.602	159.73	0.684	S
KUKGAZ	1.190	96.11	3.590	2.856	2.583	2.493	4.620	3.650	2.837	3.646	0.254	145.50	0.734	sp3
ESABAB	1.211	94.47	3.483	2.862	2.729	2.164	4.573	4.045	3.830	2.499	1.101	150.23	0.621	sp3
ZZZLWW01	1.295	95.39	3.653	2.867	2.599	2.567	4.337	3.871	3.704	2.256	1.105	113.65	0.786	sp2
FAMZIC01	1.210	104.99	3.574	2.872	2.676	2.369	4.114	3.682	3.518	2.133	0.842	108.09	0.702	S
PESPIN	1.231	97.82	2.822	2.891	2.757	0.602	3.489	3.565	3.445	0.552	0.688	112.82	-0.069	sp3
VECXIL01	1.190	98.86	3.270	2.892	2.768	1.741	4.259	3.823	3.763	1.995	0.995	140.68	0.378	sp2

Table C5 continued

HUCJOF	1.200	109.45	3.378	2.902	2.872	1.778	3.630	3.577	3.539	0.808	0.667	92.30	0.476	sp2
FAMZIC01	1.210	104.99	3.431	2.910	2.821	1.953	3.922	3.264	3.016	2.507	0.195	104.97	0.521	S
SIZBAE	1.396	97.94	4.119	2.923	0.372	4.102	3.999	2.778	1.088	3.848	0.716	75.28	1.196	sp3
WONHAK	1.288	96.52	3.798	2.929	2.577	2.790	4.403	3.622	2.892	3.320	0.315	109.73	0.869	sp2
KECXIA	1.221	94.28	3.695	2.944	2.632	2.593	4.798	4.125	3.602	3.170	0.970	150.87	0.751	sp2
SUMLAN	1.100	89.40	3.979	2.956	1.353	3.742	4.333	3.148	1.643	4.009	0.290	101.44	1.023	P
TAPZIU	1.337	96.76	4.094	2.956	0.810	4.013	3.957	2.769	1.391	3.704	0.581	74.66	1.138	sp3
DPDHPS10	1.272	101.22	3.290	2.964	2.927	1.502	3.702	3.707	3.638	0.685	0.711	98.71	0.326	P
KUKGAZ	1.201	97.21	3.461	2.969	2.940	1.826	4.366	3.557	3.227	2.941	0.287	132.77	0.492	sp3
POHJON01	1.200	109.37	3.630	2.978	2.812	2.296	4.372	3.765	3.433	2.707	0.621	121.06	0.652	sp3
HOMPOP	1.016	105.66	3.552	2.980	2.792	2.196	4.246	3.617	3.442	2.486	0.650	127.33	0.572	sp2
DPDHPS01	1.247	98.51	3.327	2.983	2.944	1.550	3.651	3.645	3.579	0.721	0.635	94.87	0.344	P
XAFVEH	1.236	98.77	2.840	2.986	2.813	0.391	3.625	3.674	3.575	0.600	0.762	120.36	-0.146	sp
BIBDOG	1.226	97.96	3.887	2.987	2.089	3.278	4.465	3.740	3.187	3.127	1.098	110.41	0.900	sp2
JURJAH	0.990	83.80	3.730	2.994	2.802	2.462	3.990	3.366	2.991	2.641	0.189	98.00	0.736	Ru
ICOPUN	1.158	97.37	3.018	3.002	2.928	0.732	3.846	3.895	3.806	0.553	0.878	128.34	0.016	sp2
MEDCOO	1.185	99.89	3.887	3.002	2.214	3.195	4.285	3.642	3.230	2.816	1.016	101.59	0.885	sp2
AZIKEZ	1.181	97.05	3.644	3.013	2.625	2.527	4.452	3.676	3.257	3.035	0.632	126.70	0.631	sp2
YISBOS	1.227	96.43	3.856	3.015	2.164	3.192	4.115	3.492	3.174	2.619	1.010	93.40	0.841	sp2
HUDFIW	1.200	109.48	3.449	3.016	2.847	1.947	4.592	4.110	3.923	2.387	1.076	159.55	0.433	sp2
MELFAK	1.200	109.49	3.197	3.019	2.988	1.137	3.324	3.590	3.323	0.082	0.335	85.40	0.178	sp2
KECXIA	1.221	94.28	3.802	3.028	2.712	2.665	3.716	3.378	3.343	1.623	0.631	76.70	0.774	sp2
DIXFAS	1.180	99.20	3.733	3.040	2.843	2.419	3.930	3.626	3.620	1.530	0.777	90.78	0.693	sp3
COLDEN	1.327	83.11	3.500	3.055	2.948	1.887	4.579	3.929	3.391	3.077	0.443	138.45	0.445	sp2
DUCPIB	1.200	109.46	3.139	3.058	3.011	0.887	3.486	3.666	3.478	0.236	0.467	96.56	0.081	sp2
COLDEN	1.356	100.51	3.975	3.066	2.569	3.033	4.470	3.645	2.761	3.515	0.192	102.54	0.909	sp2

Table C5 continued

GUJNOP	1.246	111.77	3.568	3.074	3.026	1.890	4.161	3.508	3.187	2.675	0.161	109.90	0.494	Mo
WAKHOF02	1.257	94.04	4.158	3.074	1.394	3.917	4.481	3.297	1.725	4.136	0.331	96.67	1.084	Si
QUSSUS	1.200	109.50	3.886	3.084	2.541	2.940	4.221	3.572	3.356	2.560	0.815	97.88	0.802	Zn
XUGPEU	0.905	106.19	2.890	3.108	2.881	0.228	3.731	3.762	3.674	0.650	0.793	155.26	-0.218	sp2
SIZBAE	1.396	97.94	3.827	3.110	2.901	2.496	4.781	4.213	3.927	2.727	1.026	125.92	0.717	sp3
LAPWAB	1.263	101.43	4.176	3.121	1.245	3.986	3.957	2.758	1.521	3.653	0.276	71.35	1.055	sp3
WESQER	1.334	90.12	2.950	3.122	2.930	0.343	3.873	3.760	3.727	1.053	0.797	125.03	-0.172	sp3
MEDDOP	1.227	95.04	3.955	3.125	2.278	3.233	4.080	3.427	3.143	2.602	0.865	87.01	0.830	sp2
WUJWAZ	1.314	93.55	3.848	3.139	2.774	2.667	4.154	3.612	3.464	2.293	0.690	94.10	0.709	sp2
DPDHPS01	1.247	98.51	4.263	3.146	1.222	4.084	4.311	3.145	1.612	3.998	0.390	83.87	1.117	P
LUVKOD	1.286	95.53	3.450	3.147	3.058	1.597	4.527	3.926	3.594	2.753	0.536	141.45	0.303	Si
MEDCII	1.235	99.89	4.059	3.147	2.611	3.108	4.718	3.712	2.701	3.868	0.090	115.14	0.912	sp2
YIQHOW	1.396	91.90	4.011	3.153	2.278	3.301	3.998	3.345	3.102	2.522	0.824	79.44	0.858	sp2
XODNOU	1.250	101.94	4.175	3.166	2.293	3.489	3.897	3.066	2.463	3.020	0.170	68.60	1.009	sp2
DPDHPS10	1.272	101.22	4.277	3.174	1.305	4.073	4.305	3.139	1.617	3.990	0.312	82.74	1.103	P
MOCGIW	1.156	102.82	3.784	3.179	3.023	2.276	4.755	3.924	3.336	3.388	0.313	142.66	0.605	Ti
HUCMIC	1.200	109.47	4.046	3.196	2.729	2.987	3.905	3.396	3.337	2.028	0.608	74.71	0.850	sp2
GUJNOP	1.246	111.77	3.621	3.200	3.179	1.734	4.121	4.072	4.037	0.828	0.858	104.92	0.421	Mo
WIXTII	1.200	109.46	3.240	3.213	3.129	0.841	4.253	3.911	3.896	1.706	0.767	142.27	0.027	sp2
VECXIL01	1.185	99.28	3.773	3.216	2.897	2.417	4.622	3.793	2.914	3.588	0.017	129.81	0.557	sp2
WAWQOC	0.897	108.76	3.567	3.219	3.205	1.566	3.924	3.583	3.518	1.738	0.313	106.97	0.348	sp2
VECXIL	1.255	96.43	3.761	3.221	2.931	2.357	4.581	3.742	2.834	3.599	-0.097	123.91	0.540	sp2
GIHVAV	1.295	94.81	4.368	3.222	1.813	3.974	4.736	3.790	2.264	4.160	0.451	98.49	1.146	P
GIQDER	1.259	96.21	3.846	3.251	3.127	2.239	4.371	3.755	3.345	2.814	0.218	106.41	0.595	sp2
VECXIL	1.255	96.43	4.403	3.256	1.230	4.228	4.655	3.689	2.307	4.043	1.077	93.69	1.147	sp2
LEFKEO	1.246	91.39	4.351	3.274	1.677	4.015	4.510	3.674	2.747	3.577	1.070	89.23	1.077	sp2

Table C5 continued

WIXTII	1.200	109.46	3.543	3.275	3.190	1.542	4.245	4.190	4.151	0.888	0.961	118.26	0.268	sp2
NAMROI	1.200	109.49	4.280	3.276	2.253	3.639	4.410	3.573	3.068	3.168	0.815	88.29	1.004	Zn
VECXIL01	1.185	99.28	4.425	3.276	1.320	4.224	4.707	3.744	2.302	4.106	0.982	96.44	1.149	sp2
SOSPAT	1.203	92.20	3.043	3.277	3.038	0.174	3.740	3.639	3.614	0.963	0.576	116.62	-0.234	Al
MEDBUT	1.120	99.64	4.009	3.284	2.909	2.759	4.370	3.950	3.824	2.115	0.915	101.37	0.725	ssp2
HIRBEQ	0.998	116.89	4.379	3.302	2.179	3.798	4.396	3.574	2.870	3.330	0.691	84.44	1.077	sp2
XUDQOE	1.200	109.50	3.517	3.305	3.250	1.344	4.434	4.407	4.349	0.864	1.099	133.94	0.212	sp2
VECXIL	1.251	97.16	4.016	3.309	2.757	2.920	4.653	3.662	2.659	3.818	-0.098	113.17	0.707	sp2
VECXIL01	1.184	100.93	4.016	3.310	2.768	2.910	4.698	3.710	2.723	3.828	-0.045	118.52	0.706	sp2
GAQQAR	1.226	97.35	4.386	3.315	1.885	3.960	4.602	3.721	2.751	3.689	0.866	92.34	1.071	sp2
FULGIE	1.214	95.37	3.417	3.329	3.283	0.948	4.132	3.638	3.556	2.104	0.273	118.20	0.088	sp2
DEBYUG	1.251	102.78	4.156	3.337	2.643	3.207	4.265	3.501	3.115	2.913	0.472	86.42	0.819	sp2
DIXFAS	1.161	90.04	4.363	3.348	2.039	3.857	4.594	3.832	3.041	3.443	1.002	94.08	1.015	sp3
HUCMIC	1.201	109.51	3.681	3.351	3.336	1.556	4.595	4.135	4.013	2.238	0.677	133.86	0.330	sp2
LAWKIE	0.959	110.63	3.181	3.354	3.170	0.264	3.725	3.609	3.562	1.090	0.392	117.65	-0.173	sp3
BOZTOB	1.339	107.26	3.622	3.358	3.296	1.502	4.711	4.208	4.119	2.286	0.823	138.69	0.264	sp3
PEDXIF	0.998	99.92	3.228	3.368	3.202	0.409	4.027	3.948	3.900	1.003	0.698	138.21	-0.140	Ir
COLDAJ	1.243	96.30	3.928	3.396	3.283	2.157	4.637	3.855	3.214	3.342	-0.069	117.63	0.532	sp2
RONVEV	1.099	93.57	4.773	3.415	0.027	4.773	3.928	2.696	0.082	3.927	0.055	35.29	1.358	sp2
YIQHOW	1.213	94.33	4.321	3.426	2.467	3.548	4.223	3.487	3.113	2.854	0.646	77.27	0.895	sp2
LEHKIT	1.127	110.46	4.369	3.452	2.796	3.357	4.089	3.506	3.320	2.387	0.524	68.36	0.917	sp2
TASBIY	1.200	84.40	4.584	3.470	2.210	4.016	4.598	3.729	3.149	3.350	0.939	83.17	1.114	Mg
COLDEN	1.322	103.47	4.175	3.474	3.172	2.715	4.452	3.700	3.226	3.068	0.054	93.36	0.701	sp2
COLDEN	1.327	83.11	4.284	3.474	3.052	3.006	4.656	4.055	3.605	2.947	0.553	97.93	0.810	sp2
GAQQAR	1.078	100.93	4.549	3.486	1.757	4.196	4.615	3.750	2.723	3.726	0.966	86.75	1.063	sp2
XIGPUA	1.158	97.46	4.172	3.491	3.289	2.567	3.906	3.619	3.613	1.484	0.324	68.79	0.681	sp2

Table C5 continued

HUSNEP	1.317	95.72	4.178	3.496	3.268	2.603	4.703	3.986	3.376	3.274	0.108	105.43	0.682	sp2
MPTMCH01	1.200	109.46	4.551	3.497	2.183	3.993	4.543	3.552	2.255	3.944	0.072	82.02	1.054	sp2
BEJTAM	0.956	109.17	4.160	3.508	3.334	2.488	4.502	4.096	4.061	1.943	0.727	104.91	0.652	sp3
ILOQAB	1.007	98.37	4.204	3.523	3.141	2.794	4.623	3.883	3.558	2.952	0.417	108.49	0.681	sp2
FATQIC	1.200	109.50	4.170	3.536	3.215	2.656	4.127	3.550	3.430	2.295	0.215	79.65	0.634	sp2
BEJTAM	0.956	109.17	3.953	3.542	3.510	1.818	4.063	3.969	3.946	0.968	0.436	89.75	0.411	sp3
MEDDOP	1.227	95.04	4.425	3.591	2.676	3.524	4.414	3.827	3.492	2.700	0.816	81.53	0.834	sp2
YIRGEM	1.196	97.55	4.214	3.595	3.163	2.784	3.725	3.424	3.355	1.619	0.192	58.17	0.619	sp2
HUDFIW	1.200	109.48	4.531	3.615	2.997	3.398	4.798	4.090	3.608	3.163	0.611	95.54	0.916	sp2
MURMER	1.220	99.59	4.919	3.621	1.286	4.748	3.852	2.677	1.335	3.613	0.049	25.37	1.298	sp3
RONVAR01	1.403	93.42	3.943	3.628	3.593	1.624	4.738	4.063	3.647	3.025	0.054	116.47	0.315	sp2
MOCGIW	1.131	101.07	4.504	3.653	2.975	3.382	4.636	4.028	3.643	2.867	0.668	89.57	0.851	Ti
HOMPOP	1.020	94.95	4.536	3.655	2.741	3.614	4.691	3.898	3.246	3.387	0.505	92.39	0.881	sp2
ICOPUN	1.186	97.77	3.689	3.660	3.600	0.805	3.531	3.662	3.510	0.385	-0.090	73.06	0.029	sp2
COLDEN	1.356	100.51	4.389	3.681	3.128	3.079	4.226	3.733	3.620	2.181	0.492	74.19	0.708	sp2
RONVAR01	1.403	93.42	5.081	3.711	0.242	5.075	4.021	2.770	0.188	4.017	-0.054	35.40	1.370	sp2
GAQQAR	1.078	100.93	4.801	3.716	2.517	4.088	4.664	3.765	3.157	3.433	0.640	76.27	1.085	sp2
SIZBAE	1.312	100.26	5.050	3.742	0.657	5.007	3.939	2.742	1.061	3.793	0.404	27.92	1.308	sp3
YOKRIA	1.185	86.62	4.988	3.748	1.967	4.584	4.103	2.983	2.097	3.527	0.130	36.98	1.240	P
RONVAR01	1.403	93.42	4.393	3.759	3.559	2.575	4.525	3.932	3.504	2.863	-0.055	86.33	0.634	sp2
MURMER	1.220	99.59	4.432	3.779	3.489	2.733	4.792	3.944	3.441	3.335	-0.048	99.76	0.653	sp3
SIZBAE	1.396	97.94	5.124	3.786	1.078	5.009	3.970	2.755	1.001	3.842	-0.077	29.58	1.338	sp3
KEXNOR	1.139	101.53	4.875	3.793	2.608	4.119	4.728	3.745	2.846	3.775	0.238	75.85	1.082	sp3
RUFDAZ	1.200	109.51	4.415	3.806	3.508	2.681	4.025	3.712	3.660	1.675	0.152	63.48	0.609	sp3
KICROE	1.327	109.46	4.837	3.816	1.998	4.405	4.276	3.145	1.995	3.782	-0.003	57.63	1.021	Si
POHJON01	1.201	109.49	4.755	3.819	3.029	3.665	4.356	3.731	3.345	2.790	0.316	63.62	0.936	sp3

Table C5 continued

GIHVAV	1.295	94.81	4.819	3.870	3.096	3.693	4.714	4.094	3.864	2.700	0.768	77.58	0.949		P
LOKRUA	1.200	109.45	4.695	3.900	3.236	3.402	4.700	3.869	3.314	3.333	0.078	82.93	0.795		sp2
GAQQAR	1.226	97.35	4.709	3.930	3.361	3.298	4.361	3.792	3.541	2.546	0.180	66.19	0.779		sp2
HUDFIW	1.200	109.49	4.690	3.981	3.559	3.054	4.253	3.871	3.711	2.078	0.152	61.63	0.709		sp2
WIXTEE	1.151	110.13	4.745	3.986	3.503	3.201	4.736	4.124	3.927	2.647	0.424	82.59	0.759		sp2
WAKHOF01	1.180	96.20	5.385	4.078	1.575	5.150	4.426	3.232	1.695	4.089	0.120	31.78	1.307		Si
POHJON01	1.200	109.56	5.059	4.089	3.148	3.960	4.651	3.976	3.635	2.901	0.487	63.62	0.970		sp3
EXUVUO	1.131	94.93	5.338	4.091	2.104	4.906	4.412	3.284	2.117	3.871	0.013	31.42	1.247		Ge
LOKRUA	1.200	109.45	4.777	4.101	3.465	3.288	4.676	3.899	3.388	3.223	-0.077	77.95	0.676		sp2
WESZUQ	1.223	100.71	5.133	4.104	2.415	4.529	4.458	3.554	2.796	3.472	0.381	50.61	1.029		sp3
WAKHOF01	1.184	98.36	5.436	4.126	1.447	5.240	4.460	3.257	1.531	4.189	0.084	30.68	1.310		Si
WAKHOF02	1.275	96.17	5.468	4.153	1.543	5.246	4.413	3.219	1.730	4.060	0.187	30.18	1.315		Si
BODMOX	1.132	97.07	4.986	4.157	3.386	3.660	4.462	3.809	3.373	2.921	-0.013	56.49	0.829		sp2
KEXNOR	1.139	101.53	5.083	4.207	3.363	3.811	4.745	4.079	3.535	3.165	0.172	66.50	0.876		sp3
DIZGUP	1.200	109.49	5.128	4.268	3.231	3.982	4.580	3.886	3.418	3.049	0.187	56.69	0.860		sp2
COLDEN	1.350	109.48	5.021	4.283	3.507	3.593	4.689	4.062	3.806	2.739	0.299	68.13	0.738		sp2
RONVEV	1.099	93.57	5.012	4.306	3.618	3.468	4.662	3.995	3.674	2.870	0.056	65.40	0.706		sp2
ILOQEF	1.200	109.46	5.199	4.353	3.364	3.964	4.612	3.977	3.644	2.827	0.280	54.79	0.846		sp2
XUGPEU	0.905	106.19	5.286	4.354	3.508	3.954	4.713	3.971	3.563	3.085	0.055	46.83	0.932		sp2
BOZTOB	1.339	107.26	5.549	4.421	1.475	5.349	4.655	3.670	2.084	4.162	0.609	42.72	1.128		sp3
NAMRUO	1.200	109.45	5.679	4.561	2.857	4.908	4.660	3.658	2.777	3.742	-0.080	28.48	1.118		Zn
GAGTUF	1.375	92.92	5.698	4.568	2.611	5.065	4.648	3.595	2.563	3.877	-0.048	35.50	1.130		sp2
HELXUR01	0.961	109.12	3.170	2.484	2.209	2.274	3.820	3.266	3.027	2.330	0.818	126.41	0.686	A1	sp2
ZORQUU	0.950	109.52	2.889	2.553	2.474	1.492	3.724	3.123	2.927	2.302	0.453	147.31	0.336	A1	sp2
HELXUR01	0.963	109.20	3.290	2.566	2.178	2.466	3.901	3.263	2.979	2.519	0.801	123.15	0.724	A1	sp2
HELXUR01	0.959	109.15	3.002	2.634	2.508	1.650	3.827	3.192	2.926	2.467	0.418	145.04	0.368	A1	sp2

Table C5 continued

HELXUR01	0.959	109.31	3.220	2.750	2.596	1.905	4.022	3.340	2.992	2.688	0.396	142.39	0.470	A1	sp2
VEZLOC	1.281	97.94	2.821	2.780	2.696	0.830	4.056	3.994	3.947	0.934	1.251	161.44	0.041	A1	sp3
GOCGUB	1.327	102.12	3.477	2.786	2.394	2.522	4.662	4.101	3.667	2.879	1.273	148.70	0.691	A1	sp2
ZORQUU	1.180	117.01	3.408	3.026	2.946	1.713	4.060	3.348	2.973	2.765	0.027	115.63	0.382	A1	sp2
VEZLOC	1.281	97.94	3.300	3.040	2.981	1.415	3.640	3.771	3.623	0.351	0.642	94.86	0.260	A1	sp3
YOMWAY01	1.136	107.93	2.948	3.182	2.941	0.203	3.783	3.645	3.620	1.098	0.679	130.31	-0.234	A1	sp2
ZORQUU	0.956	100.72	4.066	3.296	2.620	3.109	3.719	3.147	2.927	2.294	0.307	62.30	0.770	A1	sp2
KASJIZ	1.388	109.86	3.570	3.533	3.470	0.839	3.504	3.517	3.434	0.697	-0.036	76.02	0.037	A1	sp2
HELXUR01	0.963	109.20	4.181	3.710	3.654	2.032	4.117	3.761	3.654	1.897	0.000	79.56	0.471	A1	sp2
GOCGUB	1.319	98.27	4.591	3.789	2.976	3.496	4.541	3.822	3.380	3.033	0.404	79.54	0.802	A1	sp2
GOCGUB	1.327	102.12	4.917	4.163	3.408	3.544	4.569	3.856	3.431	3.017	0.023	67.18	0.754	A1	sp2
VEZLOC	1.281	97.94	5.377	4.166	1.098	5.264	4.363	3.301	1.466	4.109	0.368	33.27	1.211	A1	sp3
VEZLOC	1.281	97.94	5.428	4.232	1.102	5.315	4.333	3.257	1.399	4.101	0.297	27.55	1.196	A1	sp3
KASJIZ	1.353	108.58	5.440	4.553	3.226	4.380	4.693	3.918	3.302	3.335	0.076	50.33	0.887	A1	sp2
CEDYEQ	1.129	113.94	3.638	2.802	2.076	2.988	4.707	3.878	3.001	3.626	0.925	158.45	0.836	A	sp3
IKEGOV	1.200	109.50	3.917	2.819	1.560	3.593	4.634	3.747	2.677	3.783	1.117	119.96	1.098	A	sp3
VOPBEH	1.102	106.39	3.214	2.927	2.827	1.529	4.192	3.638	3.420	2.424	0.593	148.44	0.287	A	sp3
XORTOP	0.979	99.94	3.485	2.963	2.834	2.028	4.054	3.786	3.703	1.650	0.869	119.23	0.522	A	sp2
VOPBEH	1.199	96.12	2.753	2.979	2.742	0.246	3.466	3.402	3.356	0.866	0.614	117.04	-0.226	A	sp3
XORTOP	0.980	102.96	3.506	2.981	2.742	2.185	4.064	3.732	3.633	1.821	0.891	118.38	0.525	A	sp2
KAPNIZ	0.986	104.86	3.488	3.008	2.809	2.068	3.837	3.525	3.510	1.550	0.701	103.28	0.480	A	sp2
XORTIJ	1.205	89.96	3.538	3.020	2.966	1.929	4.597	4.089	3.988	2.287	1.022	147.15	0.518	A	sp2
XORTOP	0.979	99.94	4.067	3.028	2.325	3.337	4.305	3.440	2.893	3.188	0.568	97.46	1.039	A	sp2
XORTIJ	1.197	98.15	3.514	3.051	3.001	1.828	4.506	4.093	4.035	2.006	1.034	140.86	0.463	A	sp2
BAJNUW	1.247	95.97	2.953	3.059	2.917	0.460	3.679	3.630	3.591	0.800	0.674	116.24	-0.106	A	Se
BAPBOK	1.265	91.32	3.461	3.059	3.033	1.667	3.377	3.488	3.344	0.471	0.311	75.60	0.402	A	sp2

Table C5 continued

WOGRIT01	1.137	97.56	3.679	3.091	3.000	2.130	4.709	4.184	4.042	2.416	1.042	151.49	0.588	A	sp3
XORTOP	0.981	91.71	3.992	3.242	2.671	2.967	4.582	3.789	3.284	3.195	0.613	121.54	0.750	A	sp2
XORTOP	0.980	102.96	3.252	3.244	3.178	0.690	3.620	3.372	3.356	1.357	0.178	104.24	0.008	A	sp2
XORTOP	0.980	102.96	3.671	3.244	3.190	1.817	4.130	3.525	3.351	2.414	0.161	111.28	0.427	A	sp2
XORTOP	0.981	91.71	3.978	3.262	2.740	2.884	4.149	3.529	3.326	2.480	0.586	93.14	0.716	A	sp2
VOPBEH	1.102	106.39	4.314	3.390	2.840	3.247	4.757	3.987	3.469	3.255	0.629	107.16	0.924	A	sp3
IKEGOV	1.200	109.50	3.496	3.402	3.370	0.930	4.345	4.403	4.311	0.542	0.941	128.42	0.094	A	sp3
XORTIJ	1.197	98.15	4.125	3.483	3.042	2.786	4.699	3.792	3.096	3.535	0.054	111.63	0.642	A	sp2
WOGRIT01	1.137	97.56	4.517	3.507	1.794	4.145	4.701	3.832	2.836	3.749	1.042	92.22	1.010	A	sp3
XORTOP	0.980	105.91	4.304	3.512	3.207	2.870	3.895	3.407	3.333	2.015	0.126	59.23	0.792	A	sp2
XORTIJ	1.205	89.96	4.958	3.529	0.272	4.951	4.173	2.839	0.230	4.167	-0.042	43.86	1.429	A	sp2
XORTIJ	1.202	94.02	5.015	3.583	0.140	5.013	4.183	2.840	0.163	4.180	0.023	41.00	1.432	A	sp2
XORTIJ	1.203	93.94	5.011	3.585	0.279	5.003	4.165	2.833	0.234	4.158	-0.045	40.18	1.426	A	sp2
XORTIJ	1.197	98.15	5.055	3.632	0.252	5.049	4.166	2.829	0.198	4.161	-0.054	37.22	1.423	A	sp2
VOPBEH	1.199	96.12	4.980	3.941	2.755	4.149	4.722	3.970	3.336	3.342	0.581	70.75	1.039	A	sp3
KECXIA	1.221	94.28	3.417	2.544	2.228	2.591	3.971	2.769	1.598	3.635	-0.630	108.18	0.873	C	sp2
KOYJIT	1.240	98.99	3.622	2.699	2.123	2.935	4.023	2.794	1.273	3.816	-0.850	99.75	0.923	C	sp3
KOYJIT	1.257	96.83	3.604	2.704	2.289	2.784	3.992	2.787	1.456	3.717	-0.833	98.69	0.900	C	sp3
TAPZIU	1.337	96.76	3.661	2.733	2.059	3.027	3.987	2.779	1.012	3.856	-1.047	94.12	0.928	C	sp3
SIZBAE	1.396	97.94	3.652	2.745	2.300	2.837	3.945	2.745	1.274	3.734	-1.026	91.58	0.907	C	sp3
TERTEP	1.242	94.71	3.722	2.761	2.187	3.012	3.998	2.790	1.690	3.623	-0.497	93.61	0.961	C	sp3
XAFVEH	1.236	98.77	3.831	2.873	2.402	2.984	3.987	2.789	1.639	3.635	-0.763	88.13	0.958	C	sp3
MURMER	1.220	99.59	4.169	3.073	0.961	4.057	4.101	2.841	0.110	4.100	-0.851	78.37	1.096	C	sp3
SIZBAE	1.312	100.26	4.149	3.102	2.298	3.454	3.976	2.782	1.173	3.799	-1.125	73.34	1.047	C	sp3
WAKHOF01	1.180	96.20	4.143	3.159	2.413	3.368	4.439	3.257	1.570	4.152	-0.843	96.72	0.984	C	Si
WAKHOF01	1.184	98.36	4.159	3.201	2.513	3.314	4.461	3.281	1.676	4.134	-0.837	97.04	0.958	C	Si

Table C5 continued

ADELOL	1.075	87.31	4.064	3.238	2.837	2.910	3.743	2.915	2.488	2.796	-0.349	65.25	0.826	C	sp2
WAKHOF02	1.275	96.17	4.236	3.245	2.481	3.433	4.457	3.271	1.523	4.189	-0.958	91.55	0.991	C	Si
JURJAH	1.180	99.19	3.990	3.388	3.173	2.419	4.531	3.727	2.897	3.484	-0.276	110.05	0.602	C	Ru
NAMROI	1.200	109.49	4.090	3.443	3.253	2.479	4.481	3.504	2.747	3.540	-0.506	101.25	0.647	C	Zn
KICROE	1.326	109.47	4.381	3.467	2.987	3.205	4.368	3.223	2.047	3.859	-0.940	80.72	0.914	C	Si
GUJNOP	0.972	102.79	3.851	3.483	3.461	1.689	3.750	3.276	3.127	2.070	-0.334	76.77	0.368	C	Mo
XEYHUF	1.117	90.99	4.393	3.550	2.931	3.272	4.691	3.733	2.366	4.051	-0.565	98.52	0.843	C	sp2
GADNIJ	1.199	109.49	3.902	3.619	3.521	1.682	3.574	3.435	3.408	1.077	-0.113	65.43	0.283	C	sp2
QUSSUS	1.200	109.50	4.243	3.630	3.481	2.426	4.534	3.581	2.907	3.479	-0.574	96.30	0.613	C	Zn
NUZLAW	1.200	109.44	4.446	3.665	3.259	3.024	4.048	3.339	2.857	2.868	-0.402	63.16	0.781	C	Pd
TAPZIU	1.337	96.76	4.979	3.680	1.288	4.810	3.870	2.700	1.103	3.709	-0.185	29.40	1.299	C	sp3
HUCGES	1.200	109.48	4.223	3.701	3.581	2.238	4.385	3.696	3.153	3.047	-0.428	89.73	0.522	C	sp2
RUFDAZ	1.200	109.51	4.935	3.704	1.954	4.532	3.802	2.676	1.802	3.348	-0.152	16.81	1.231	C	sp3
JIVNAD	1.171	105.99	4.482	3.738	3.415	2.903	4.650	3.709	2.873	3.656	-0.542	90.89	0.744	C	sp2
ICOPUN	1.158	97.37	3.948	3.748	3.739	1.268	4.014	3.528	3.354	2.205	-0.385	84.91	0.200	C	sp2
ELAQIR	1.200	109.47	4.036	3.753	3.705	1.601	3.886	3.647	3.572	1.530	-0.133	74.28	0.283	C	sp2
HUCMIC	1.199	109.52	4.686	3.843	3.087	3.525	4.746	3.629	2.267	4.170	-0.820	85.55	0.843	C	sp2
WIXTAA	1.200	109.45	4.878	3.876	3.027	3.825	4.694	3.802	2.917	3.678	-0.110	74.14	1.002	C	sp2
XIGPUA	1.204	96.77	3.999	3.886	3.851	1.078	3.719	3.536	3.500	1.257	-0.351	68.02	0.113	C	sp2
GAQQAR	1.226	97.35	4.499	3.912	3.518	2.804	4.414	3.603	3.176	3.065	-0.342	78.19	0.587	C	sp2
WIXSUT	1.151	112.22	4.982	3.995	3.128	3.878	4.643	3.765	2.888	3.636	-0.240	66.43	0.987	C	sp2
YAFTUW	1.200	109.45	4.477	4.021	3.928	2.148	3.792	3.641	3.592	1.215	-0.336	48.62	0.456	C	sp2
YIRGEM	1.196	97.55	4.947	4.048	3.176	3.793	3.758	2.936	2.462	2.839	-0.714	5.20	0.899	C	sp2
HOMPOP	1.016	105.66	5.010	4.069	3.301	3.769	3.995	3.097	2.650	2.990	-0.651	2.83	0.941	C	sp2
FULGIE	1.214	95.37	5.017	4.109	3.294	3.784	3.823	3.006	2.563	2.837	-0.731	9.10	0.908	C	sp2
KEXNOR	1.139	101.53	4.116	4.113	4.053	0.717	3.837	3.933	3.816	0.401	-0.237	67.99	0.003	C	sp3

Table C5 continued

HIRBEQ	0.998	116.89	5.110	4.170	3.305	3.897	4.197	3.256	2.614	3.284	-0.691	21.47	0.940	C	sp2
MELFAK	1.200	109.49	4.766	4.171	3.724	2.974	4.602	3.739	3.143	3.362	-0.581	74.91	0.595	C	sp2
KOYJIT	1.240	98.99	5.212	4.172	2.766	4.417	4.537	3.645	2.444	3.822	-0.322	51.11	1.040	C	sp3
NUMWUP	1.236	100.29	5.120	4.207	3.468	3.767	3.887	3.045	2.691	2.805	-0.777	3.70	0.913	C	sp2
GAGTUF	1.304	96.13	4.862	4.217	3.957	2.825	4.134	3.725	3.661	1.920	-0.296	49.45	0.645	C	sp2
ILOQAB	1.007	98.37	4.930	4.237	3.892	3.026	4.419	3.830	3.475	2.730	-0.417	54.32	0.693	C	sp2
FATQIC	1.200	109.50	4.942	4.250	3.926	3.002	4.202	3.783	3.711	1.971	-0.215	46.20	0.692	C	sp2
KUKGAZ	1.201	97.21	4.914	4.253	3.984	2.877	4.416	3.934	3.697	2.415	-0.287	58.98	0.661	C	sp3
JURJAH	1.180	99.19	5.163	4.255	2.888	4.280	4.467	3.394	2.017	3.986	-0.871	48.38	0.908	C	Ru
HUCGES	1.199	109.45	4.906	4.262	3.989	2.856	4.012	3.643	3.597	1.777	-0.392	36.88	0.644	C	sp2
WUSBAN	1.207	90.94	4.999	4.266	3.831	3.211	4.576	3.776	3.036	3.424	-0.795	62.82	0.733	C	Al
PESPIN	1.231	97.82	5.257	4.268	3.249	4.133	4.682	3.757	3.139	3.474	-0.110	56.01	0.989	C	sp3
TASPUY	1.099	109.29	5.285	4.298	3.285	4.140	4.191	3.260	2.677	3.225	-0.608	4.88	0.987	C	sp3
WONHAK	1.288	96.52	5.088	4.316	3.724	3.467	4.270	3.505	2.911	3.124	-0.813	44.73	0.772	C	sp2
ELAQIR	1.200	109.47	4.935	4.324	4.087	2.766	4.095	3.616	3.545	2.050	-0.542	40.39	0.611	C	sp2
NAMRUO	1.200	109.45	5.018	4.335	3.993	3.039	4.498	3.567	2.881	3.454	-1.112	57.98	0.683	C	Zn
XEYHUF	1.117	90.99	4.259	4.364	4.243	0.369	3.698	3.510	3.500	1.194	-0.743	53.09	-0.105	C	sp2
WUJWAZ	1.314	93.55	5.441	4.365	1.941	5.083	4.198	3.226	1.788	3.798	-0.153	16.50	1.076	C	sp2
DIXFAS	1.180	99.20	4.957	4.379	4.215	2.609	4.551	3.837	3.468	2.947	-0.747	63.33	0.578	C	sp3
LEFKEO	1.246	91.39	5.125	4.385	3.959	3.255	4.765	3.792	2.890	3.789	-1.069	66.39	0.740	C	sp2
MEDBUT	1.120	99.64	5.172	4.405	3.827	3.479	4.359	3.809	3.468	2.641	-0.359	39.03	0.767	C	sp2
TAZFOR	1.372	100.02	4.726	4.423	4.362	1.819	4.492	3.966	3.865	2.289	-0.497	71.87	0.303	C	sp2
ODOSAD	1.319	108.54	4.648	4.468	4.465	1.291	3.590	3.705	3.566	0.414	-0.899	31.51	0.180	C	sp2
LUVKOD	1.286	95.53	5.368	4.469	3.676	3.912	4.137	3.324	2.954	2.896	-0.722	14.62	0.899	C	Si
KOYJIT	1.240	98.99	5.265	4.478	3.719	3.727	4.576	3.656	2.515	3.823	-1.204	50.45	0.787	C	sp3
WONHAK	1.288	96.52	4.912	4.489	4.383	2.217	3.785	3.229	3.151	2.097	-1.232	25.08	0.423	C	sp2

Table C5 continued

ODOSAD	1.313	105.71	4.424	4.495	4.395	0.506	4.025	3.903	3.883	1.060	-0.512	64.01	-0.071	C	sp2
WIXSON	1.200	109.54	5.425	4.497	3.641	4.022	4.745	3.851	2.956	3.712	-0.685	50.04	0.928	C	sp2
GIQDER	1.259	96.21	5.058	4.500	4.340	2.598	4.260	3.491	3.191	2.822	-1.149	44.94	0.558	C	sp2
GAQQAR	1.078	100.93	5.176	4.513	4.120	3.133	4.247	3.735	3.480	2.434	-0.640	27.30	0.663	C	sp2
KOYJIT	1.257	96.83	5.447	4.523	3.524	4.153	4.684	3.696	2.440	3.998	-1.084	47.14	0.924	C	sp3
COLDAJ	1.243	96.30	5.160	4.554	4.038	3.212	4.174	3.423	3.040	2.860	-0.998	33.08	0.606	C	sp2
KOYJIT	1.257	96.83	5.403	4.578	3.841	3.800	4.565	3.640	2.650	3.717	-1.191	43.02	0.825	C	sp3
YUCYUR	1.200	109.45	4.795	4.585	4.545	1.528	4.046	3.644	3.545	1.950	-1.000	45.57	0.210	C	sp2
BODMOX	1.162	94.79	5.132	4.591	4.377	2.679	4.568	3.974	3.553	2.871	-0.824	55.06	0.541	C	sp2
COLDEN	1.327	83.11	5.350	4.594	4.001	3.552	4.064	3.315	3.009	2.732	-0.992	12.45	0.756	C	sp2
GAQQAR	1.226	97.35	5.442	4.598	3.693	3.997	4.358	3.461	2.587	3.507	-1.106	24.72	0.844	C	sp2
WIVSEA	1.007	103.42	5.368	4.603	4.163	3.389	4.727	3.806	3.188	3.490	-0.975	46.21	0.765	C	sp2
LUVKOD	1.286	95.53	4.934	4.604	4.539	1.934	4.326	3.720	3.448	2.613	-1.091	54.98	0.330	C	Si
KILCUE	1.200	109.41	5.344	4.633	4.198	3.307	4.223	3.611	3.428	2.466	-0.770	18.50	0.711	C	sp2
MEDBUT	1.120	99.64	5.286	4.642	4.309	3.062	4.386	3.909	3.841	2.117	-0.468	32.78	0.644	C	sp2
DEBYUG	1.251	102.78	4.780	4.655	4.645	1.128	3.708	3.848	3.694	0.322	-0.951	26.91	0.125	C	sp2
HUSNEP	1.317	95.72	4.924	4.658	4.642	1.642	4.761	4.180	3.910	2.716	-0.732	75.19	0.266	C	sp2
TASBIY	1.200	84.40	5.503	4.698	4.082	3.691	4.433	3.588	3.061	3.207	-1.021	23.98	0.805	C	Mg
KOYJIT	1.240	98.99	5.657	4.703	2.916	4.848	4.452	3.526	2.438	3.725	-0.478	12.10	0.954	C	sp3
YUKDIR	1.280	116.38	5.511	4.707	3.769	4.021	4.550	3.683	2.592	3.740	-1.177	36.69	0.804	C	W
COLDAJ	1.238	103.89	5.545	4.708	3.613	4.206	4.526	3.659	2.725	3.614	-0.888	30.84	0.837	C	sp2
TAZFOR	1.372	100.02	5.434	4.745	4.349	3.258	4.711	4.234	4.146	2.237	-0.203	51.78	0.689	C	sp2
QUSSUS	1.200	109.50	5.737	4.753	3.659	4.419	4.555	3.597	2.912	3.503	-0.747	8.75	0.984	C	Zn
GIHVAV	1.295	94.81	5.228	4.788	4.600	2.484	3.975	3.617	3.573	1.742	-1.027	12.70	0.440	C	P
MELFAK	1.200	109.49	5.586	4.836	4.225	3.654	4.621	3.735	3.062	3.461	-1.163	32.61	0.750	C	sp2
KUKGAZ	1.105	95.19	5.305	4.864	4.635	2.581	4.254	3.764	3.560	2.329	-1.075	16.20	0.441	C	sp3

Table C5 continued

COLDEN	1.322	103.47	5.555	4.864	4.290	3.529	4.524	3.671	3.134	3.263	-1.156	34.30	0.691	C	sp2
WESQER	1.334	90.12	5.865	4.865	3.561	4.660	4.622	3.567	2.574	3.839	-0.987	18.76	1.000	C	sp3
COLDEN	1.344	87.89	5.601	4.872	4.195	3.711	4.549	3.698	3.190	3.243	-1.005	33.96	0.729	C	sp2
DPDHPS01	1.247	98.51	5.072	4.908	4.897	1.321	4.038	3.864	3.847	1.227	-1.050	29.87	0.164	C	P
DPDHPS10	1.272	101.22	5.122	4.958	4.947	1.327	4.089	3.923	3.908	1.203	-1.039	31.31	0.164	C	P
HUCGES	1.200	109.51	5.619	4.965	4.312	3.603	4.469	3.842	3.465	2.822	-0.847	14.77	0.654	C	sp2
BECVUC	1.236	93.10	5.239	4.979	4.953	1.707	4.232	3.868	3.846	1.766	-1.107	31.39	0.260	C	sp3
XAFVEH	1.236	98.77	5.296	5.056	5.003	1.737	4.104	3.838	3.821	1.498	-1.182	13.55	0.240	C	sp3
HEBPOT	1.329	98.07	5.997	5.094	3.949	4.513	4.686	3.788	2.958	3.634	-0.991	8.18	0.903	C	sp2
DUCNUL	1.200	109.46	5.606	5.098	4.797	2.901	4.723	4.130	3.680	2.960	-1.117	38.25	0.508	C	sp2
GAYKUO	1.059	102.96	3.152	2.472	2.305	2.150	3.816	3.336	3.140	2.168	0.835	121.73	0.680	D	Ru, sp3
WORNAS	1.200	110.39	3.161	2.583	2.502	1.932	3.712	3.274	3.193	1.893	0.691	108.00	0.578	D	Ru, H_2S
WORNEW	1.242	100.71	2.917	2.585	2.548	1.420	3.906	3.328	3.014	2.484	0.466	135.86	0.332	D	Ru, H_2S
SULTOI	1.187	115.03	3.244	2.653	2.544	2.013	4.177	3.274	2.728	3.163	0.184	135.68	0.591	D	Ru, sp3
DIRVAB	1.225	104.52	3.072	2.661	2.673	1.514	4.031	3.545	3.445	2.093	0.772	134.87	0.411	D	Ru, Ru
LEGPOC	1.289	111.72	3.643	2.663	2.052	3.010	4.553	3.553	2.092	4.044	0.040	128.08	0.980	D	Ru+, sp3
DIRVAB	1.170	109.79	3.057	2.671	2.581	1.638	4.114	3.565	3.232	2.545	0.651	150.32	0.386	D	Ru, Ru
JIHHUD	1.349	108.14	3.389	2.718	2.526	2.259	4.163	3.382	2.895	2.992	0.369	116.09	0.671	D	Ru, sp3
CEMROB	1.037	104.36	3.561	2.760	2.429	2.604	4.042	3.331	2.811	2.904	0.382	110.42	0.801	D	Rh, Rh
WORNAS	1.200	110.39	2.951	2.803	2.734	1.111	3.671	3.185	3.145	1.893	0.411	117.96	0.148	D	Ru, H_2S
SULTOI	1.187	115.03	3.361	2.819	2.576	2.159	4.258	3.429	2.651	3.332	0.075	132.84	0.542	D	Ru, sp3
GAYKUO	1.059	102.96	3.197	2.846	2.829	1.489	3.734	3.127	2.990	2.237	0.161	112.58	0.351	D	Ru, sp3
RODBIV	1.149	113.16	3.534	2.942	2.848	2.092	4.257	3.682	3.464	2.474	0.616	122.09	0.592	D	Mn, Mn
PUMDUX	1.201	109.45	3.753	2.943	2.438	2.853	3.957	3.319	3.103	2.455	0.665	90.83	0.810	D	Re
SIZCAF	1.197	107.39	3.949	3.138	2.807	2.778	3.904	3.302	2.936	2.573	0.129	79.10	0.811	D	Ru, H_2S
VETKAH	1.303	102.18	4.048	3.177	2.741	2.979	3.719	3.274	3.120	2.024	0.379	66.24	0.871	D	Rh, sp2

Table C5 continued

RODBIV	1.149	113.16	3.816	3.194	3.037	2.311	4.556	4.079	3.984	2.210	0.947	123.74	0.622	D	Mn, Mn
WORNEW	1.242	100.71	4.310	3.236	1.231	4.130	4.562	3.644	2.433	3.859	1.202	93.72	1.074	D	Ru, H_2S
COCKAG	1.354	97.02	4.240	3.293	2.672	3.292	4.065	3.410	3.013	2.729	0.341	73.34	0.947	D	Ir, sp3
PUDWER	1.258	100.36	4.372	3.397	2.600	3.515	4.239	3.317	2.805	3.178	0.205	75.62	0.975	D	Sb, sp3
ZOZBIB	1.207	101.61	3.980	3.955	3.896	0.813	4.648	4.287	4.207	1.976	0.311	116.63	0.025	D	Ni, Ni
XUHZOQ	1.000	113.93	4.697	4.041	3.683	2.915	4.486	3.826	3.583	2.699	-0.100	71.75	0.656	D	Cu, Cu
DIRVAB	1.170	109.79	5.572	4.615	2.587	4.935	4.696	3.820	2.744	3.811	0.157	37.33	0.957	D	Ru, Ru
YUKDIR01	1.280	116.38	3.254	2.767	2.653	1.884	4.315	3.607	3.313	2.765	0.660	140.07	0.487	E	W
WEKYAO	1.314	100.24	3.813	3.040	2.675	2.717	3.670	3.226	3.175	1.841	0.500	73.79	0.773	F	Ru^+
VOPBEH	1.199	96.12	3.647	2.705	2.278	2.848	3.985	2.787	1.625	3.639	-0.653	97.50	0.942	A, C	sp3
VOPBEH	1.102	106.39	4.139	3.114	2.420	3.358	3.931	2.769	1.706	3.542	-0.714	71.52	1.025	A, C	sp3
XORTOP	0.981	91.71	4.825	3.383	0.239	4.819	4.195	2.822	0.125	4.193	-0.114	45.44	1.442	A, C	sp2
MKMPIN	1.294	94.29	4.506	3.433	2.283	3.885	4.244	3.351	2.179	3.642	-0.104	70.15	1.073	A, C	sp2
BAPBOK	1.265	91.32	4.864	3.489	0.358	4.851	3.962	2.700	0.065	3.961	-0.293	39.02	1.375	A, C	sp2
HUKJUT	0.948	99.12	3.989	3.800	3.782	1.268	3.758	3.509	3.488	1.399	-0.294	69.14	0.189	A, C	sp3
KAPNIZ	0.986	104.86	4.806	3.847	3.070	3.698	4.668	3.695	2.753	3.770	-0.317	76.12	0.959	A, C	sp2
JOLBIX	1.200	109.47	4.959	4.072	3.149	3.831	4.296	3.338	2.201	3.689	-0.948	50.46	0.887	A, C	Re
BAJNUW	1.247	95.97	4.409	4.105	4.051	1.740	3.619	3.592	3.537	0.766	-0.514	44.10	0.304	A, C	Se
HUKJUT	0.948	99.12	4.964	4.161	3.636	3.379	4.296	3.627	3.404	2.621	-0.232	41.16	0.803	A, C	sp3
BAPBOK	1.265	91.32	4.958	4.210	3.797	3.188	4.169	3.675	3.503	2.260	-0.294	45.43	0.748	A, C	sp2
VAGLAR	0.986	100.71	5.532	4.545	2.728	4.813	4.581	3.668	2.409	3.896	-0.319	13.99	0.987	A, C	Mo
XORTOP	0.979	99.94	5.604	4.629	3.306	4.525	4.800	3.728	2.437	4.135	-0.869	31.85	0.975	A, C	sp2
WOGRIT	1.239	95.18	5.471	4.640	3.840	3.897	4.646	3.939	3.609	2.926	-0.231	43.23	0.831	A, C	sp3
WOGRIT01	1.137	97.56	5.281	4.808	4.597	2.599	4.632	4.026	3.680	2.813	-0.917	49.98	0.473	A, C	sp3
CEDYEQ	1.129	113.94	5.766	4.877	3.845	4.297	4.743	3.784	2.924	3.734	-0.921	22.51	0.889	A, C	sp3
IKEGOV	1.200	109.50	5.861	4.912	3.478	4.718	4.683	3.816	2.927	3.656	-0.551	9.94	0.949	A, C	sp3

Table C5 continued

ZORQUU	1.180	117.01	3.607	3.254	3.194	1.676	3.947	3.316	2.900	2.677	-0.294	97.96	0.353	A1, C	sp2
HELXUR01	0.959	109.31	3.921	3.554	3.438	1.885	3.924	3.292	2.897	2.647	-0.541	83.19	0.367	A1, C	sp2
ZORQUU	0.950	109.52	3.862	3.590	3.536	1.553	3.813	3.257	3.009	2.342	-0.527	80.00	0.272	A1, C	sp2
HELXUR01	0.959	109.15	3.863	3.591	3.526	1.578	3.848	3.273	3.019	2.386	-0.507	81.97	0.272	A1, C	sp2
KASJIZ	1.352	98.41	3.826	3.747	3.711	0.931	3.709	3.414	3.416	1.445	-0.295	74.79	0.079	A1, C	sp2
ZORQUU	0.956	100.72	4.674	3.993	3.489	3.110	3.798	3.278	3.031	2.289	-0.458	21.23	0.681	A1, C	sp2
HELXUR01	0.961	109.12	4.473	4.034	3.892	2.205	3.746	3.151	2.943	2.318	-0.949	36.63	0.439	A1, C	sp2
HELXUR01	0.963	109.20	4.619	4.131	3.864	2.531	3.882	3.249	2.912	2.567	-0.952	36.09	0.488	A1, C	sp2
KASJIZ	1.216	102.95	4.741	4.360	4.216	2.169	4.714	3.990	3.535	3.119	-0.681	81.38	0.381	A1, C	sp2
YOMWAY01	1.136	107.93	5.720	4.899	4.184	3.900	4.584	3.788	3.335	3.145	-0.849	1.48	0.821	A1, C	sp2
WUDQUI	1.317	94.66	3.863	2.861	1.897	3.365	4.025	2.810	1.673	3.661	-0.224	87.42	1.002	C, D	sp3, sp3
CEMROB	1.037	104.36	3.489	3.129	3.120	1.562	3.655	3.184	3.014	2.068	-0.106	90.89	0.360	C, D	Rh, Rh
XUHZOQ	1.000	113.93	4.883	3.879	2.985	3.864	4.799	3.769	2.592	4.039	-0.393	79.29	1.004	C, D	Cu, Cu
XEHDOF	1.255	98.82	4.481	3.977	3.853	2.288	4.522	3.811	3.477	2.891	-0.376	83.85	0.504	C, D	sp3
RODBIV	1.149	113.16	4.669	4.216	4.057	2.311	4.290	3.725	3.554	2.403	-0.503	63.93	0.453	C, D	Mn, Mn
WORNEW	1.242	100.71	4.564	4.249	4.134	1.934	3.634	3.232	3.218	1.688	-0.916	36.04	0.315	C, D	Ru, H_2S
GAYKUO	1.059	102.96	5.432	4.342	2.610	4.764	4.783	3.834	2.347	4.168	-0.263	47.64	1.090	C, D	Ru, sp3
VETKAH	1.303	102.18	4.819	4.378	4.227	2.314	3.599	3.125	2.960	2.047	-1.267	17.65	0.441	C, D	Rh, sp2
GAYKUO	1.059	102.96	5.008	4.447	4.179	2.760	4.767	3.953	3.344	3.397	-0.835	70.88	0.561	C, D	Ru, sp3
LEGPOC	1.289	111.72	5.281	4.560	4.011	3.435	4.177	3.550	3.109	2.790	-0.902	27.23	0.721	C, D	Ru+, sp3
SULTOI	1.187	115.03	5.682	4.702	2.936	4.865	4.633	3.565	2.030	4.165	-0.906	24.90	0.980	C, D	Ru, sp3
WORNAS	1.200	110.39	5.023	4.708	4.685	1.811	4.667	4.377	4.274	1.875	-0.411	66.09	0.315	C, D	Ru, H_2S
COCKAG	1.354	97.02	5.433	4.774	4.353	3.251	4.442	3.711	3.411	2.845	-0.942	37.84	0.659	C, D	Ir, sp3
WORNAS	1.200	110.39	5.046	4.848	4.825	1.477	4.795	4.278	4.134	2.429	-0.691	71.16	0.198	C, D	Ru, H_2S
JIHHUD	1.349	108.14	5.800	4.982	4.153	4.049	4.493	3.689	3.224	3.129	-0.929	12.72	0.818	C, D	Ru, sp3
ZOZBIB	1.207	101.61	5.593	5.002	4.452	3.385	4.513	4.082	3.863	2.333	-0.589	23.68	0.591	C, D	Ni, Ni
YUKDIR01	1.280	116.38	5.511	4.707	3.769	4.021	4.550	3.683	2.592	3.740	-1.177	36.69	0.804	C, E	W
SIZCAF10	1.197	107.39	3.949	3.138	2.807	2.778	3.904	3.302	2.936	2.573	0.129	79.10	0.811	D, E	Ru, H_2S

Table C6. Structure entries for thiol-aromatic interactions, with the cylinder restriction on the hydrogen atom after normalizing the thiol S–H bond.

All entries in analysis of thiol-aromatic interactions with normalized S–H bond lengths (1.338 Å[119]), following initial search parameters of the CSD and cylinder restriction on the normalized hydrogen atom. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the thiol-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55. Z refers to the atom or hybridized carbon that is bound to the sulfur atom.
Annotations: (A) polycyclic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (C) thiol groups were "pointed away" from the plane of the aromatic ring (as defined by ($d_{S–plane} - d_{H–plane}$) ≥ –0.10 Å); (D) where thiol groups where the sulfur atom is apparently tri-coordinate; (E) redundant structures or entries; (F) unusual structures with questionable features.

Refcode	Angle θ, Z–S–H bond angle, degrees	$d_{H–centroid}$, Å	$d_{H–Cmin}$, Å	H d_{plane}, Å	H r, Å	$d_{S–centroid}$, Å	$d_{S–Cmin}$, Å	S d_{plane}, Å	S r, Å	Δ S d_{plane} – H d_{plane}, Å	Angle S–H–centroid, degrees	$\Delta d_{centroid–Cmin}$, Å	Parsing Rationale	Z
WANCIX	107.05	2.537	2.145	2.135	1.370	3.729	3.152	3.099	2.074	0.964	146.797	0.392		sp2
FULGIE	95.37	2.774	2.272	2.251	1.621	3.932	3.087	2.684	2.873	0.433	143.612	0.502		sp2
HOMPOP	94.95	2.851	2.330	2.290	1.698	3.992	3.138	2.762	2.882	0.472	142.075	0.521		sp2
NUMWUP	100.29	2.743	2.355	2.360	1.398	3.875	3.089	2.837	2.640	0.477	141.01	0.388		sp2
DUCPIB	109.49	3.301	2.432	2.044	2.592	4.604	3.704	3.025	3.471	0.981	164.37	0.869		sp2

Table C6 continued

ODOSAD	110.13	2.621	2.446	2.404	1.044	3.746	3.644	3.612	0.993	1.208	139.973	0.175	sp2
KOYJIT	98.99	3.016	2.476	2.438	1.776	4.127	3.767	3.737	1.751	1.299	139.614	0.540	sp3
COLDEN	103.47	2.887	2.477	2.451	1.526	4.204	3.668	3.568	2.223	1.117	167.826	0.410	sp2
DEBYUG	102.78	3.167	2.485	2.235	2.244	3.812	3.220	3.072	2.257	0.837	108.655	0.682	sp2
VECXIL	96.64	2.677	2.499	2.430	1.123	3.593	3.621	3.539	0.621	1.109	123.512	0.178	sp2
YULZUA	113.44	2.934	2.502	2.484	1.561	4.026	3.764	3.690	1.610	1.206	137.809	0.432	sp2
YUCYUR	109.45	3.131	2.517	2.282	2.144	4.443	3.838	3.397	2.864	1.115	166.305	0.614	sp2
KOYJIT	96.83	2.916	2.535	2.515	1.476	4.169	3.839	3.749	1.824	1.234	155.364	0.381	sp3
COLDEN	100.51	3.248	2.540	2.414	2.173	4.273	3.737	3.537	2.398	1.123	132.911	0.708	sp2
YULZUA	104.69	2.753	2.541	2.504	1.144	4.010	3.782	3.725	1.485	1.221	155.591	0.212	sp2
EXUVUO	94.93	2.886	2.553	2.550	1.351	4.061	3.767	3.723	1.622	1.173	145.646	0.333	Ge
YULZOU	99.53	2.552	2.554	2.456	0.693	3.684	3.649	3.598	0.791	1.142	140.503	-0.002	sp2
COLDEN	99.16	2.694	2.554	2.499	1.006	4.014	3.801	3.796	1.305	1.297	168.354	0.140	sp2
VECXIL01	96.13	2.781	2.556	2.493	1.232	3.624	3.639	3.562	0.667	1.069	119.018	0.225	sp2
GAGTUF	92.92	2.852	2.568	2.535	1.307	3.721	3.722	3.652	0.713	1.117	120.911	0.284	sp2
HIYSAK	100.54	2.412	2.570	2.374	0.426	3.750	3.784	3.697	0.628	1.323	178.389	-0.158	sp3
NAMQEY	102.78	3.460	2.579	2.248	2.630	4.573	3.737	3.177	3.289	0.929	140.557	0.881	sp2
VECXIL01	100.93	2.630	2.595	2.515	0.769	3.599	3.607	3.537	0.665	1.022	127.126	0.035	sp2
VECXIL	97.16	2.595	2.600	2.494	0.717	3.551	3.581	3.497	0.617	1.003	126.038	-0.005	sp2
COLDAJ	96.30	2.712	2.626	2.555	0.909	3.779	3.693	3.629	1.054	1.074	135.022	0.086	sp2
MIBSAR	104.36	2.637	2.635	2.519	0.780	3.960	3.862	3.828	1.014	1.309	169.372	0.002	sp2
KILCUE	109.41	2.591	2.650	2.527	0.572	3.715	3.558	3.506	1.228	0.979	139.749	-0.059	sp2
ODOSAD	100.68	2.715	2.661	2.573	0.867	3.805	3.933	3.791	0.326	1.218	137.081	0.054	sp2
PESPIN	97.82	3.110	2.678	2.527	1.813	3.927	3.139	2.832	2.720	0.305	118.359	0.432	sp3
XUDQOE	109.50	3.200	2.678	2.619	1.839	4.159	3.919	3.892	1.466	1.273	128.009	0.522	sp2
ICOPUN	97.77	2.460	2.684	2.444	0.280	3.573	3.667	3.538	0.499	1.094	138.39	-0.224	sp2

Table C6 continued

QAXPEM01	102.81	3.108	2.687	2.553	1.773	4.076	3.315	3.018	2.740	0.465	128.389	0.421	Si
LIGDOU	106.57	3.294	2.699	2.504	2.140	4.445	3.690	2.891	3.376	0.387	143.825	0.595	Ru
COLDEN	87.89	2.643	2.700	2.579	0.578	3.878	3.625	3.619	1.393	1.040	152.273	-0.057	sp2
VECXIL	95.88	3.072	2.708	2.592	1.649	4.229	3.778	3.720	2.011	1.128	144.067	0.364	sp2
GUJNOP	111.77	3.073	2.711	2.632	1.586	3.903	3.413	3.351	2.001	0.719	119.086	0.362	Mo
GIQDER	96.21	3.111	2.720	2.630	1.662	4.407	4.021	3.851	2.143	1.221	162.8	0.391	sp2
WAKHOF02	94.04	3.469	2.728	2.395	2.510	4.764	4.035	3.362	3.375	0.967	163.018	0.741	Si
YUKDIR	116.38	3.210	2.737	2.623	1.850	4.315	3.607	3.313	2.765	0.690	139.406	0.473	W
ESABAB	94.47	3.373	2.739	2.613	2.133	4.573	4.045	3.830	2.499	1.217	149.155	0.634	sp3
SOSPAT	92.96	2.796	2.742	2.682	0.790	3.559	3.737	3.552	0.223	0.870	114.132	0.054	Al
HUDFIW	109.51	3.641	2.742	2.363	2.770	4.166	3.583	3.315	2.523	0.952	103.705	0.899	sp2
FAMZIC01	104.99	3.414	2.748	2.531	2.291	4.688	3.893	3.196	3.430	0.665	158.987	0.666	S
SIZBAE	100.26	2.599	2.761	2.570	0.387	3.733	3.873	3.717	0.345	1.147	140.746	-0.162	sp3
MEDCII	99.89	3.644	2.764	2.070	2.999	4.195	3.568	3.221	2.688	1.151	105.032	0.880	sp2
ZEPVOF	95.44	3.120	2.765	2.682	1.594	4.195	3.570	3.393	2.467	0.711	136.635	0.355	Mo
SULTOI	115.03	3.260	2.768	2.566	2.011	4.258	3.429	2.651	3.332	0.085	130.892	0.492	Al
KUKGAZ	96.11	3.469	2.777	2.551	2.351	4.620	3.650	2.837	3.646	0.286	144.116	0.692	sp3
VECXIL01	98.86	3.158	2.778	2.645	1.725	4.259	3.823	3.763	1.995	1.118	138.985	0.380	sp2
FAMZIC01	104.99	3.537	2.798	2.588	2.411	4.114	3.682	3.518	2.133	0.930	106.125	0.739	S
PESPIN	97.82	2.783	2.809	2.698	0.683	3.489	3.565	3.445	0.552	0.747	110.791	-0.026	Ir
TAXMUA	96.41	2.559	2.810	2.550	0.214	3.725	3.810	3.689	0.517	1.139	144.016	-0.251	sp2
HOMPOP	105.66	3.366	2.825	2.586	2.155	4.246	3.617	3.442	2.486	0.856	122.979	0.541	sp2
WAQFEZ	104.07	2.989	2.830	2.786	1.083	3.912	3.816	3.757	1.090	0.971	124.927	0.159	sp3
XUGPEU	106.19	2.503	2.831	2.502	0.071	3.731	3.762	3.674	0.650	1.172	151.113	-0.328	sp2
KECXIA	94.28	3.593	2.833	2.539	2.542	4.798	4.125	3.602	3.170	1.063	149.966	0.760	sp2
GAGTUF	86.42	3.559	2.836	2.578	2.454	4.260	3.857	3.803	1.920	1.225	112.768	0.723	sp2

Table C6 continued

ZZZLWW01	95.39	3.636	2.838	2.563	2.579	4.337	3.871	3.704	2.256	1.141	113.028	0.798	sp2
HUCJOF	109.45	3.375	2.847	2.795	1.892	3.630	3.577	3.539	0.808	0.744	89.956	0.528	sp2
ICOPUN	97.37	2.910	2.852	2.791	0.824	3.846	3.895	3.806	0.553	1.015	125.567	0.058	sp2
FAMZIC01	104.99	3.400	2.872	2.800	1.929	3.922	3.264	3.016	2.507	0.216	102.892	0.528	S
HUDFIW	109.48	3.320	2.879	2.723	1.899	4.592	4.110	3.923	2.387	1.200	158.717	0.441	sp2
POHJON01	109.37	3.561	2.891	2.740	2.274	4.372	3.765	3.433	2.707	0.693	119.157	0.670	sp3
PUMDUX	109.45	3.754	2.899	2.362	2.918	3.957	3.319	3.103	2.455	0.741	88.734	0.855	sp3
XAFVEH	98.77	2.790	2.901	2.750	0.471	3.625	3.674	3.575	0.600	0.825	118.556	-0.111	sp3
WONHAK	96.52	3.781	2.910	2.564	2.779	4.403	3.622	2.892	3.320	0.328	109.012	0.871	sp2
AZIKEZ	97.05	3.553	2.924	2.541	2.483	4.452	3.676	3.257	3.035	0.716	124.677	0.629	sp2
KUKGAZ	97.21	3.369	2.926	2.907	1.703	4.366	3.557	3.227	2.941	0.320	131.057	0.443	sp3
DPDHPS10	101.22	3.281	2.928	2.890	1.553	3.702	3.707	3.638	0.685	0.748	97.562	0.353	P
JURJAH	83.80	3.698	2.932	2.736	2.488	3.990	3.366	2.991	2.641	0.255	92.654	0.766	Ru
DPDHPS01	98.51	3.321	2.939	2.898	1.622	3.651	3.645	3.579	0.721	0.681	93.301	0.382	P
MELFAK	109.49	3.211	2.973	2.949	1.270	3.324	3.590	3.323	0.082	0.374	82.943	0.238	sp2
DUCPIB	109.46	3.126	2.987	2.957	1.014	3.486	3.666	3.478	0.236	0.521	94.043	0.139	sp2
DIXFAS	99.20	3.734	2.988	2.739	2.538	3.930	3.626	3.620	1.530	0.881	88.354	0.746	sp3
KECXIA	94.28	3.831	3.018	2.652	2.765	3.716	3.378	3.343	1.623	0.691	74.999	0.813	sp2
COLDEN	83.11	3.492	3.049	2.945	1.876	4.579	3.929	3.391	3.077	0.446	138.338	0.443	sp2
QUSSUS	109.50	3.870	3.053	2.447	2.998	4.221	3.572	3.356	2.560	0.909	95.846	0.817	Zn
GUJNOP	111.77	3.538	3.060	3.014	1.853	4.161	3.508	3.187	2.675	0.173	108.495	0.478	Mo
RODBIV	113.16	3.715	3.065	2.882	2.344	4.556	4.079	3.984	2.210	1.102	121.315	0.650	sp3
MOCGIW	102.82	3.641	3.086	2.974	2.101	4.755	3.924	3.336	3.388	0.362	140.922	0.555	Ti
WAWQOC	108.76	3.464	3.119	3.051	1.640	3.924	3.583	3.518	1.738	0.467	99.981	0.345	sp2
WESQER	90.12	2.948	3.120	2.928	0.343	3.873	3.760	3.727	1.053	0.799	124.964	-0.172	sp3
LUVKOD	95.53	3.409	3.122	3.036	1.550	4.527	3.926	3.594	2.753	0.558	140.903	0.287	Si

Table C6 continued

WUJWAZ	93.55	3.846	3.130	2.762	2.676	4.154	3.612	3.464	2.293	0.702	93.741	0.716	sp2
PEDXIF	99.92	2.983	3.131	2.965	0.327	4.027	3.948	3.900	1.003	0.935	133.849	-0.148	sp2
GUJNOP	111.77	3.598	3.145	3.115	1.801	4.121	4.072	4.037	0.828	0.922	103.504	0.453	Mo
SIZCAF	107.39	3.978	3.147	2.792	2.834	3.904	3.302	2.936	2.573	0.144	77.105	0.831	sp3
WIXTII	109.46	3.132	3.149	3.041	0.749	4.253	3.911	3.896	1.706	0.855	140.729	-0.017	sp2
SIZBAE	97.94	3.861	3.152	2.944	2.498	4.781	4.213	3.927	2.727	0.983	126.62	0.709	sp3
VECXIL01	99.28	3.677	3.154	2.894	2.268	4.622	3.793	2.914	3.588	0.020	127.981	0.523	sp2
XUDQOE	109.50	3.422	3.182	3.124	1.397	4.434	4.407	4.349	0.864	1.225	132.268	0.240	sp2
WIXTII	109.46	3.480	3.182	3.079	1.622	4.245	4.190	4.151	0.888	1.072	116.257	0.298	sp2
MEDBUT	99.64	3.971	3.185	2.732	2.882	4.370	3.950	3.824	2.115	1.092	98.295	0.786	sp2
GIQDER	96.21	3.824	3.230	3.113	2.221	4.371	3.755	3.345	2.814	0.232	105.27	0.594	sp2
LAWKIE	110.63	3.024	3.239	3.015	0.233	3.725	3.609	3.562	1.090	0.547	111.284	-0.215	sp3
VECXIL01	100.93	3.945	3.271	2.774	2.805	4.698	3.710	2.723	3.828	-0.051	116.56	0.674	sp2
HUCMIC	109.51	3.587	3.277	3.259	1.498	4.595	4.135	4.013	2.238	0.754	132.281	0.310	sp2
FULGIE	95.37	3.361	3.321	3.255	0.837	4.132	3.638	3.556	2.104	0.301	116.336	0.040	sp2
BEJTAM	109.17	4.079	3.321	3.043	2.716	4.502	4.096	4.061	1.943	1.018	99.711	0.758	sp3
BOZTOB	107.26	3.623	3.358	3.297	1.502	4.711	4.208	4.119	2.286	0.822	138.697	0.265	sp3
COLDAJ	96.30	3.885	3.378	3.288	2.069	4.637	3.855	3.214	3.342	-0.074	116.386	0.507	sp2
BEJTAM	109.17	3.974	3.431	3.335	2.161	4.063	3.969	3.946	0.968	0.611	84.235	0.543	sp3
ILOQAB	98.37	4.111	3.461	3.004	2.806	4.623	3.883	3.558	2.952	0.554	104.114	0.650	sp2
COLDEN	103.47	4.174	3.472	3.171	2.714	4.452	3.700	3.226	3.068	0.055	93.138	0.702	sp2
HUSNEP	95.72	4.173	3.492	3.266	2.598	4.703	3.986	3.376	3.274	0.110	105.152	0.681	sp2
XIGPUA	97.46	4.241	3.505	3.239	2.738	3.906	3.619	3.613	1.484	0.374	66.524	0.736	sp2
FATQIC	109.50	4.197	3.526	3.190	2.727	4.127	3.550	3.430	2.295	0.240	77.803	0.671	sp2
YIRGEM	97.55	4.291	3.641	3.140	2.925	3.725	3.424	3.355	1.619	0.215	56.552	0.650	sp2
RONVAR01	93.42	4.389	3.757	3.556	2.573	4.525	3.932	3.504	2.863	-0.052	87.171	0.632	sp2

Table C6 continued

MURMER	99.59	4.413	3.784	3.494	2.696	4.792	3.944	3.441	3.335	-0.053	98.249	0.629		sp3
RUFDAZ	109.51	4.478	3.830	3.490	2.806	4.025	3.712	3.660	1.675	0.170	61.9	0.648		sp2
HELXUR01	109.12	2.962	2.167	1.889	2.281	3.820	3.266	3.027	2.330	1.138	120.536	0.795	A1	sp2
HELXUR01	109.20	3.101	2.266	1.867	2.476	3.901	3.263	2.979	2.519	1.112	117.343	0.835	A1	sp2
ZORQUU	109.52	2.571	2.371	2.289	1.171	3.724	3.123	2.927	2.302	0.638	142.629	0.200	A1	sp2
HELXUR01	109.15	2.699	2.403	2.342	1.342	3.827	3.192	2.926	2.467	0.584	140.417	0.296	A1	sp2
HELXUR01	109.31	2.928	2.503	2.440	1.619	4.022	3.340	2.992	2.688	0.552	137.849	0.425	A1	sp2
VEZLOC	97.94	2.767	2.725	2.641	0.825	4.056	3.994	3.947	0.934	1.306	161.064	0.042	A1	sp3
GOCGUB	102.12	3.468	2.775	2.384	2.519	4.662	4.101	3.667	2.879	1.283	148.602	0.693	A1	sp2
ZORQUU	117.01	3.343	2.989	2.942	1.588	4.060	3.348	2.973	2.765	0.031	113.183	0.354	A1	sp2
VEZLOC	97.94	3.296	3.014	2.952	1.466	3.640	3.771	3.623	0.351	0.671	93.879	0.282	A1	sp3
YOMWAY01	107.93	2.821	3.130	2.821	0.000	3.783	3.645	3.620	1.098	0.799	127.182	-0.309	A1	sp2
KASJIZ	109.86	3.558	3.541	3.469	0.791	3.504	3.517	3.434	0.697	-0.035	76.802	0.017	A1	sp2
CEDYEQ	113.94	3.445	2.607	1.905	2.870	4.707	3.878	3.001	3.626	1.096	157.169	0.838	A	sp3
XORTOP	99.94	3.324	2.688	2.515	2.173	4.054	3.786	3.703	1.650	1.188	113.821	0.636	A	sp2
XORTOP	102.96	3.351	2.719	2.418	2.320	4.064	3.732	3.633	1.821	1.215	112.993	0.632	A	sp2
QAXPEM	94.08	3.315	2.752	2.606	2.049	4.186	3.367	2.939	2.981	0.333	122.313	0.563	A	Re
VOPBEH	106.39	3.015	2.777	2.700	1.342	4.192	3.638	3.420	2.424	0.720	146.09	0.238	A	sp3
KAPNIZ	104.86	3.425	2.830	2.558	2.278	3.837	3.525	3.510	1.550	0.952	97.535	0.595	A	sp2
VOPBEH	96.12	2.692	2.866	2.671	0.336	3.466	3.402	3.356	0.866	0.685	114.401	-0.174	A	sp3
WOGRIT01	97.56	3.504	2.902	2.816	2.085	4.709	4.184	4.042	2.416	1.226	149.924	0.602	A	sp3
XORTIJ	89.96	3.427	2.909	2.854	1.897	4.597	4.089	3.988	2.287	1.134	145.949	0.518	A	sp2
XORTIJ	98.15	3.405	2.936	2.879	1.818	4.506	4.093	4.035	2.006	1.156	139.362	0.469	A	sp2
XORTOP	91.71	3.817	2.985	2.448	2.929	4.582	3.789	3.284	3.195	0.836	116.966	0.832	A	sp2
BAJNUW	95.97	2.914	2.994	2.868	0.516	3.679	3.630	3.591	0.800	0.723	114.641	-0.080	A	Se
BAPBOK	91.32	3.480	3.049	3.015	1.738	3.377	3.488	3.344	0.471	0.329	74.441	0.431	A	sp2

Table C6 continued

XORTOP	102.96	3.183	3.210	3.114	0.659	3.620	3.372	3.356	1.357	0.242	97.987	-0.027	A	sp2
XORTOP	102.96	3.557	3.210	3.132	1.686	4.130	3.525	3.351	2.414	0.219	105.909	0.347	A	sp2
IKEGOV	109.50	3.412	3.284	3.262	1.001	4.345	4.403	4.311	0.542	1.049	126.611	0.128	A	sp3
XORTIJ	98.15	4.075	3.429	3.036	2.718	4.699	3.792	3.096	3.535	0.060	109.788	0.646	A	sp2
KECXIA	94.28	3.382	2.552	2.289	2.490	3.971	2.769	1.598	3.635	-0.691	106.301	0.830	C	sp2
ZEPVOF	95.44	3.620	2.714	2.198	2.876	4.643	3.506	1.986	4.197	-0.212	133.46	0.906	C	Mo
KOYJIT	98.99	3.607	2.716	2.190	2.866	4.023	2.794	1.273	3.816	-0.917	98.219	0.891	C	sp3
KOYJIT	96.83	3.592	2.718	2.343	2.723	3.992	2.787	1.456	3.717	-0.887	97.406	0.874	C	sp3
SIZBAE	97.94	3.654	2.731	2.258	2.873	3.945	2.745	1.274	3.734	-0.984	92.488	0.923	C	sp3
TERTEP	94.71	3.718	2.781	2.225	2.979	3.998	2.790	1.690	3.623	-0.535	92.129	0.937	C	sp3
XAFVEH	98.77	3.836	2.903	2.465	2.939	3.987	2.789	1.639	3.635	-0.826	86.607	0.933	C	sp3
VECXIL	96.43	3.715	3.196	2.938	2.274	4.581	3.742	2.834	3.599	-0.104	122.845	0.519	C	sp2
VECXIL	97.16	3.983	3.275	2.763	2.869	4.653	3.662	2.659	3.818	-0.104	112.016	0.708	C	sp2
JURJAH	99.19	3.938	3.371	3.210	2.281	4.531	3.727	2.897	3.484	-0.313	107.881	0.567	C	Ru
NAMROI	109.49	4.065	3.463	3.311	2.358	4.481	3.504	2.747	3.540	-0.564	99.342	0.602	C	Zn
GUJNOP	102.79	3.951	3.625	3.587	1.656	3.750	3.276	3.127	2.070	-0.460	71.6	0.326	C	Mo
GADNIJ	109.49	3.962	3.632	3.534	1.791	3.574	3.435	3.408	1.077	-0.126	63.6	0.330	C	sp2
ICOPUN	97.77	3.736	3.653	3.612	0.955	3.531	3.662	3.510	0.385	-0.102	70.836	0.083	C	sp2
QUSSUS	109.50	4.230	3.661	3.547	2.305	4.534	3.581	2.907	3.479	-0.640	94.436	0.569	C	Si
HUCGES	109.48	4.226	3.723	3.630	2.164	4.385	3.696	3.153	3.047	-0.477	87.858	0.503	C	sp2
JIVNAD	105.99	4.483	3.772	3.492	2.811	4.650	3.709	2.873	3.656	-0.619	88.754	0.711	C	sp2
GAQQAR	97.35	4.523	3.921	3.549	2.804	4.414	3.603	3.176	3.065	-0.373	76.806	0.602	C	sp2
GAYKUO	102.96	3.014	2.254	2.085	2.176	3.816	3.336	3.140	2.168	1.055	117.208	0.760	D	Ru, sp3
WORNAS	110.39	3.121	2.503	2.423	1.967	3.712	3.274	3.193	1.893	0.770	105.595	0.618	D	Ru, H_2S
WORNEW	100.71	2.849	2.544	2.512	1.344	3.906	3.328	3.014	2.484	0.502	134.521	0.305	D	Ru, H_2S
DIRVAB	109.79	2.912	2.554	2.487	1.515	4.114	3.565	3.232	2.545	0.745	148.685	0.358	D	Ru, Ru

Table C6 continued

DIRVAB	104.52	2.993	2.573	2.602	1.479	4.031	3.545	3.445	2.093	0.843	133.337	0.420	D	Ru, Ru
SULTOI	115.03	3.138	2.602	2.520	1.870	4.177	3.274	2.728	3.163	0.208	133.75	0.536	D	Ru, sp3
CEMROB	104.36	3.468	2.627	2.318	2.580	4.042	3.331	2.811	2.904	0.493	105.757	0.841	D	Rh, Rh
LEGPOC	111.72	3.613	2.636	2.051	2.974	4.553	3.553	2.092	4.044	0.041	127.469	0.977	D	Ru^{+}, sp3
JIHHUD	108.14	3.394	2.724	2.529	2.263	4.163	3.382	2.895	2.992	0.366	116.252	0.670	D	Ru, sp3
WORNAS	110.39	2.889	2.765	2.686	1.064	3.671	3.185	3.145	1.893	0.459	115.542	0.124	D	Ru, H_2S
GAYKUO	102.96	3.101	2.834	2.786	1.362	3.734	3.127	2.990	2.237	0.204	107.812	0.267	D	Ru, sp3
RODBIV	113.16	3.437	2.846	2.746	2.067	4.257	3.682	3.464	2.474	0.718	119.42	0.591	D	Mn, Mn
SIZCAF10	107.39	3.978	3.147	2.792	2.834	3.904	3.302	2.936	2.573	0.144	77.105	0.831	D	Ru, H_2S
RONVAR01	93.42	3.972	3.637	3.595	1.689	4.738	4.063	3.647	3.025	0.052	117.307	0.335	D	Mn, Mn
YUKDIR01	116.38	3.210	2.737	2.623	1.850	4.315	3.607	3.313	2.765	0.690	139.406	0.473	E	W
WEKYAO	100.24	3.820	3.042	2.665	2.737	3.670	3.226	3.175	1.841	0.510	73.44	0.778	F	Ru^{+}
VOPBEH	96.12	3.632	2.730	2.354	2.766	3.985	2.787	1.625	3.639	-0.729	95.322	0.902	A, C	sp3
ADELOL	87.31	4.181	3.365	2.922	2.990	3.743	2.915	2.488	2.796	-0.434	61.982	0.816	A, C	sp2
ZORQUU	117.01	3.588	3.270	3.233	1.556	3.947	3.316	2.900	2.677	-0.333	95.458	0.318	A1, C	sp2
CEMROB	104.36	3.497	3.177	3.150	1.519	3.655	3.184	3.014	2.068	-0.136	85.954	0.320	C, D	Rh, Rh
SOSPAT	92.20	2.985	3.191	2.973	0.267	3.740	3.639	3.614	0.963	0.641	114.297	-0.206	D, E	Ru, H_2S

Table C7. Structure entries for thiol-aromatic interactions, with the cylinder restriction on the sulfur atom after normalizing the thiol S–H bond.

All entries in analysis of thiol-aromatic interactions with normalized S–H bond lengths (1.338 Å[119]), following initial search parameters of the CSD and cylinder restriction on the sulfur atom. The entries were examined and annotated for specific criteria (described above). Entries with annotations A, B, C, D, E, and F were excluded from study, since these structures contained elements that would compete with the thiol-π interaction, and effect the geometry. Definitions of distances are shown in Figure 2.55. Z refers to the atom or hybridized carbon that is bound to the sulfur atom.
Annotations: (A) polycyclic rings; (A1) complex macrocyclic rings; (B) carbanions within or adjacent to the interacting aromatic rings; (C) thiol groups were "pointed away" from the plane of the aromatic ring (as defined by ($d_{S–plane} - d_{H–plane}$) ≥ –0.10 Å); (D) where thiol groups where the sulfur atom is apparently tri-coordinate; (E) redundant structures or entries; (F) unusual structures with questionable features.

Refcode	Angle θ, Z–S–H bond angle, degrees	$d_{H–centroid}$, Å	$d_{H–Cmin}$, Å	H d_{plane}, Å	H r, Å	$d_{S–centroid}$, Å	$d_{S–Cmin}$, Å	S d_{plane}, Å	S r, Å	Δ S d_{plane} – H d_{plane}, Å	Angle S–H–centroid, degrees	$\Delta d_{centroid–Cmin}$, Å	Parsing Rationale	Z
WANCIX	107.05	2.537	2.145	2.135	1.370	3.729	3.152	3.099	2.074	0.964	146.797	0.392		sp2
FULGIE	95.37	2.774	2.272	2.251	1.621	3.932	3.087	2.684	2.873	0.433	143.612	0.502		sp2
HOMPOP	94.95	2.851	2.330	2.290	1.698	3.992	3.138	2.762	2.882	0.472	142.075	0.521		sp2
NUZLAW	109.44	3.506	2.425	1.784	3.018	3.955	3.215	2.777	2.816	0.993	99.592	1.081		Pd
DUCPIB	109.49	3.301	2.432	2.044	2.592	4.604	3.704	3.025	3.471	0.981	164.37	0.869		sp2
ODOSAD	110.13	2.621	2.446	2.404	1.044	3.746	3.644	3.612	0.993	1.208	139.973	0.175		sp2

Table C7 continued

COLDEN	103.47	2.887	2.477	2.451	1.526	4.204	3.668	3.568	2.223	1.117	167.826	0.410	sp2
DEBYUG	102.78	3.167	2.485	2.235	2.244	3.812	3.220	3.072	2.257	0.837	108.655	0.682	sp2
VECXIL	96.64	2.677	2.499	2.430	1.123	3.593	3.621	3.539	0.621	1.109	123.512	0.178	sp2
YULZUA	113.44	2.934	2.502	2.484	1.561	4.026	3.764	3.690	1.610	1.206	137.809	0.432	sp2
TASBIY	84.40	3.653	2.505	1.334	3.401	4.466	3.510	2.652	3.593	1.318	119.51	1.148	Mg
YUCYUR	109.45	3.131	2.517	2.282	2.144	4.443	3.838	3.397	2.864	1.115	166.305	0.614	sp2
COLDEN	100.51	3.248	2.540	2.414	2.173	4.273	3.737	3.537	2.398	1.123	132.911	0.708	sp2
YULZUA	104.69	2.753	2.541	2.504	1.144	4.010	3.782	3.725	1.485	1.221	155.591	0.212	sp2
KECXIA	94.28	3.382	2.552	2.289	2.490	3.971	2.769	1.598	3.635	-0.691	106.301	0.830	sp2
EXUVUO	94.93	2.886	2.553	2.550	1.351	4.061	3.767	3.723	1.622	1.173	145.646	0.333	Ge
YULZOU	99.53	2.552	2.554	2.456	0.693	3.684	3.649	3.598	0.791	1.142	140.503	-0.002	sp2
VECXIL01	96.13	2.781	2.556	2.493	1.232	3.624	3.639	3.562	0.667	1.069	119.018	0.225	sp2
GAGTUF	92.92	2.852	2.568	2.535	1.307	3.721	3.722	3.652	0.713	1.117	120.911	0.284	sp2
HIYSAK	100.54	2.412	2.570	2.374	0.426	3.750	3.784	3.697	0.628	1.323	178.389	-0.158	sp3
NAMQEY	102.78	3.460	2.579	2.248	2.630	4.573	3.737	3.177	3.289	0.929	140.557	0.881	sp2
VECXIL01	100.93	2.630	2.595	2.515	0.769	3.599	3.607	3.537	0.665	1.022	127.126	0.035	sp2
VECXIL	97.16	2.595	2.600	2.494	0.717	3.551	3.581	3.497	0.617	1.003	126.038	-0.005	sp2
SIZBAE	100.26	3.637	2.609	1.564	3.284	3.917	2.754	1.605	3.573	0.041	91.919	1.028	sp3
MIBSAR	104.36	2.637	2.635	2.519	0.780	3.960	3.862	3.828	1.014	1.309	169.372	0.002	sp2
PESPIN	97.82	3.110	2.678	2.527	1.813	3.927	3.139	2.832	2.720	0.305	118.359	0.432	sp3
XUDQOE	109.50	3.200	2.678	2.619	1.839	4.159	3.919	3.892	1.466	1.273	128.009	0.522	sp2
ICOPUN	97.77	2.460	2.684	2.444	0.280	3.573	3.667	3.538	0.499	1.094	138.39	-0.224	sp2
QAXPEM01	102.81	3.108	2.687	2.553	1.773	4.076	3.315	3.018	2.740	0.465	128.389	0.421	Si
LIGDOU	106.57	3.294	2.699	2.504	2.140	4.445	3.690	2.891	3.376	0.387	143.825	0.595	Ru
COLDEN	87.89	2.643	2.700	2.579	0.578	3.878	3.625	3.619	1.393	1.040	152.273	-0.057	sp2
VECXIL	95.88	3.072	2.708	2.592	1.649	4.229	3.778	3.720	2.011	1.128	144.067	0.364	sp2

Table C7 continued

GUJNOP	111.77	3.073	2.711	2.632	1.586	3.903	3.413	3.351	2.001	0.719	119.086	0.362	Mo
ZEPVOF	95.44	3.620	2.714	2.198	2.876	4.643	3.506	1.986	4.197	-0.212	133.46	0.906	Mo
KOYJIT	98.99	3.607	2.716	2.190	2.866	4.023	2.794	1.273	3.816	-0.917	98.219	0.891	sp3
KOYJIT	96.83	3.592	2.718	2.343	2.723	3.992	2.787	1.456	3.717	-0.887	97.406	0.874	sp3
SIZBAE	97.94	3.654	2.731	2.258	2.873	3.945	2.745	1.274	3.734	-0.984	92.488	0.923	sp3
TAPZIU	96.76	3.661	2.733	2.060	3.026	3.987	2.779	1.012	3.856	-1.048	94.099	0.928	sp3
ESABAB	94.47	3.373	2.739	2.613	2.133	4.573	4.045	3.830	2.499	1.217	149.155	0.634	sp3
SOSPAT	92.96	2.796	2.742	2.682	0.790	3.559	3.737	3.552	0.223	0.870	114.132	0.054	Al
HUDFIW	109.51	3.641	2.742	2.363	2.770	4.166	3.583	3.315	2.523	0.952	103.705	0.899	sp2
FAMZIC01	104.99	3.414	2.748	2.531	2.291	4.688	3.893	3.196	3.430	0.665	158.987	0.666	S
QAXPEM	94.08	3.315	2.752	2.606	2.049	4.186	3.367	2.939	2.981	0.333	122.313	0.563	Si
MEDCII	99.89	3.644	2.764	2.070	2.999	4.195	3.568	3.221	2.688	1.151	105.032	0.880	sp2
KUKGAZ	96.11	3.469	2.777	2.551	2.351	4.620	3.650	2.837	3.646	0.286	144.116	0.692	sp3
VECXIL01	98.86	3.158	2.778	2.645	1.725	4.259	3.823	3.763	1.995	1.118	138.985	0.380	sp2
FAMZIC01	104.99	3.537	2.798	2.588	2.411	4.114	3.682	3.518	2.133	0.930	106.125	0.739	S
TAXMUA	96.41	2.559	2.810	2.550	0.214	3.725	3.810	3.689	0.517	1.139	144.016	-0.251	sp3
WAQFEZ	104.07	3.927	2.818	1.722	3.529	4.749	3.806	3.014	3.670	1.292	120.551	1.109	sp3
WAQFEZ	104.07	2.989	2.830	2.786	1.083	3.912	3.816	3.757	1.090	0.971	124.927	0.159	sp3
XUGPEU	106.19	2.503	2.831	2.502	0.071	3.731	3.762	3.674	0.650	1.172	151.113	-0.328	sp2
ZZZLWW01	95.39	3.636	2.838	2.563	2.579	4.337	3.871	3.704	2.256	1.141	113.028	0.798	sp2
HUCJOF	109.45	3.375	2.847	2.795	1.892	3.630	3.577	3.539	0.808	0.744	89.956	0.528	sp2
SUMLAN	89.40	3.939	2.862	1.291	3.721	4.333	3.148	1.643	4.009	0.352	98.034	1.077	P
FAMZIC01	104.99	3.400	2.872	2.800	1.929	3.922	3.264	3.016	2.507	0.216	102.892	0.528	S
HUDFIW	109.48	3.320	2.879	2.723	1.899	4.592	4.110	3.923	2.387	1.200	158.717	0.441	sp2
POHJON01	109.37	3.561	2.891	2.740	2.274	4.372	3.765	3.433	2.707	0.693	119.157	0.670	sp3
XAFVEH	98.77	3.836	2.903	2.465	2.939	3.987	2.789	1.639	3.635	-0.826	86.607	0.933	sp

Table C7 continued

WONHAK	96.52	3.781	2.910	2.564	2.779	4.403	3.622	2.892	3.320	0.328	109.012	0.871	sp2
SIZBAE	97.94	4.104	2.919	0.402	4.084	3.999	2.778	1.088	3.848	0.686	76.061	1.185	sp3
BIBDOG	97.96	3.849	2.920	1.988	3.296	4.465	3.740	3.187	3.127	1.199	108.84	0.929	sp2
AZIKEZ	97.05	3.553	2.924	2.541	2.483	4.452	3.676	3.257	3.035	0.716	124.677	0.629	sp2
KUKGAZ	97.21	3.369	2.926	2.907	1.703	4.366	3.557	3.227	2.941	0.320	131.057	0.443	sp3
DPDHPS10	101.22	3.281	2.928	2.890	1.553	3.702	3.707	3.638	0.685	0.748	97.562	0.353	P
MEDCOO	99.89	3.859	2.931	2.083	3.249	4.285	3.642	3.230	2.816	1.147	99.369	0.928	sp2
DPDHPS01	98.51	3.321	2.939	2.898	1.622	3.651	3.645	3.579	0.721	0.681	93.301	0.382	P
YISBOS	96.43	3.851	2.979	2.073	3.245	4.115	3.492	3.174	2.619	1.101	91.751	0.872	sp2
DUCPIB	109.46	3.126	2.987	2.957	1.014	3.486	3.666	3.478	0.236	0.521	94.043	0.139	sp2
KECXIA	94.28	3.831	3.018	2.652	2.765	3.716	3.378	3.343	1.623	0.691	74.999	0.813	sp2
WAKHOF02	94.04	4.149	3.047	1.373	3.915	4.481	3.297	1.725	4.136	0.352	95.556	1.102	Si
COLDEN	83.11	3.492	3.049	2.945	1.876	4.579	3.929	3.391	3.077	0.446	138.338	0.443	sp2
GUJNOP	111.77	3.538	3.060	3.014	1.853	4.161	3.508	3.187	2.675	0.173	108.495	0.478	Mo
COLDEN	100.51	3.979	3.072	2.572	3.036	4.470	3.645	2.761	3.515	0.189	102.79	0.907	sp2
MOCGIW	102.82	3.641	3.086	2.974	2.101	4.755	3.924	3.336	3.388	0.362	140.922	0.555	Ti
MURMER	99.59	4.194	3.107	1.064	4.057	4.101	2.841	0.110	4.100	-0.954	76.797	1.087	sp3
SIZBAE	100.26	4.156	3.113	2.320	3.448	3.976	2.782	1.173	3.799	-1.147	72.995	1.043	sp3
MEDDOP	95.04	3.963	3.117	2.200	3.296	4.080	3.427	3.143	2.602	0.943	85.411	0.846	sp2
MEDCII	99.89	4.016	3.117	2.603	3.058	4.718	3.712	2.701	3.868	0.098	113.809	0.899	sp2
WAWQOC	108.76	3.464	3.119	3.051	1.640	3.924	3.583	3.518	1.738	0.467	99.981	0.345	sp2
PEDXIF	99.92	2.983	3.131	2.965	0.327	4.027	3.948	3.900	1.003	0.935	133.849	-0.148	Ir
DPDHPS01	98.51	4.273	3.137	1.193	4.103	4.311	3.145	1.612	3.998	0.419	82.655	1.136	P
WIXTII	109.46	3.132	3.149	3.041	0.749	4.253	3.911	3.896	1.706	0.855	140.729	-0.017	sp2
YIQHOW	91.90	4.001	3.149	2.312	3.265	3.998	3.345	3.102	2.522	0.790	80.249	0.852	sp2
LAPWAB	101.43	4.200	3.150	1.228	4.016	3.957	2.758	1.521	3.653	0.293	70.385	1.050	sp3

Table C7 continued

VECXIL01	99.28	3.677	3.154	2.894	2.268	4.622	3.793	2.914	3.588	0.020	127.981	0.523	sp2
DPDHPS10	101.22	4.286	3.169	1.289	4.088	4.305	3.139	1.617	3.990	0.328	81.86	1.117	P
WAKHOF01	96.20	4.128	3.179	2.526	3.265	4.439	3.257	1.570	4.152	-0.956	94.538	0.949	Si
XODNOU	101.94	4.208	3.181	2.281	3.536	3.897	3.066	2.463	3.020	0.182	67.486	1.027	sp2
XUDQOE	109.50	3.422	3.182	3.124	1.397	4.434	4.407	4.349	0.864	1.225	132.268	0.240	sp2
WIXTII	109.46	3.480	3.182	3.079	1.622	4.245	4.190	4.151	0.888	1.072	116.257	0.298	sp2
SOSPAT	92.20	2.985	3.191	2.973	0.267	3.740	3.639	3.614	0.963	0.641	114.297	-0.206	AL
VECXIL	96.43	3.715	3.196	2.938	2.274	4.581	3.742	2.834	3.599	-0.104	122.845	0.519	sp2
HUCMIC	109.47	4.084	3.201	2.659	3.100	3.905	3.396	3.337	2.028	0.678	72.842	0.883	sp2
GIHVAV	94.81	4.362	3.209	1.798	3.974	4.736	3.790	2.264	4.160	0.466	97.928	1.153	P
GIQDER	96.21	3.824	3.230	3.113	2.221	4.371	3.755	3.345	2.814	0.232	105.27	0.594	sp2
VECXIL01	99.28	4.410	3.234	1.194	4.245	4.707	3.744	2.302	4.106	1.108	94.461	1.176	sp2
VECXIL	96.43	4.398	3.237	1.158	4.243	4.655	3.689	2.307	4.043	1.149	92.601	1.161	sp2
LAWKIE	110.63	3.024	3.239	3.015	0.233	3.725	3.609	3.562	1.090	0.547	111.284	-0.215	sp3
DAHCOE	108.92	4.133	3.242	2.529	3.269	4.524	3.722	3.329	3.063	0.800	98.281	0.891	sp2
WAKHOF02	96.17	4.235	3.257	2.529	3.397	4.457	3.271	1.523	4.189	-1.006	90.695	0.978	Si
LEFKEO	91.39	4.353	3.258	1.598	4.049	4.510	3.674	2.747	3.577	1.149	88.016	1.095	sp2
VECXIL01	100.93	3.945	3.271	2.774	2.805	4.698	3.710	2.723	3.828	-0.051	116.56	0.674	sp2
VECXIL	97.16	3.983	3.275	2.763	2.869	4.653	3.662	2.659	3.818	-0.104	112.016	0.708	sp2
GAQQAR	97.35	4.383	3.286	1.806	3.994	4.602	3.721	2.751	3.689	0.945	90.876	1.097	sp2
DIXFAS	90.04	4.354	3.300	1.887	3.924	4.594	3.832	3.041	3.443	1.154	91.756	1.054	sp3
DEBYUG	102.78	4.163	3.319	2.610	3.243	4.265	3.501	3.115	2.913	0.505	85.224	0.844	sp2
BEJTAM	109.17	4.079	3.321	3.043	2.716	4.502	4.096	4.061	1.943	1.018	99.711	0.758	sp3
BOZTOB	107.26	3.623	3.358	3.297	1.502	4.711	4.208	4.119	2.286	0.822	138.697	0.265	sp3
YAJHEW	97.41	4.511	3.392	1.991	4.048	4.572	3.777	2.870	3.559	0.879	84.133	1.119	sp3
BEJTAM	109.17	3.974	3.431	3.335	2.161	4.063	3.969	3.946	0.968	0.611	84.235	0.543	sp3

Table C7 continued

YIQHOW	94.33	4.350	3.444	2.401	3.627	4.223	3.487	3.113	2.854	0.712	75.669	0.906	sp2
NAMROI	109.49	4.065	3.463	3.311	2.358	4.481	3.504	2.747	3.540	-0.564	99.342	0.602	Zn
GAQQAR	100.93	4.571	3.469	1.524	4.309	4.615	3.750	2.723	3.726	1.199	83.502	1.102	sp2
COLDEN	83.11	4.282	3.471	3.047	3.009	4.656	4.055	3.605	2.947	0.558	97.785	0.811	sp2
COLDEN	103.47	4.174	3.472	3.171	2.714	4.452	3.700	3.226	3.068	0.055	93.138	0.702	sp2
LEHKIT	110.46	4.451	3.483	2.699	3.539	4.089	3.506	3.320	2.387	0.621	65.842	0.968	sp2
HUSNEP	95.72	4.173	3.492	3.266	2.598	4.703	3.986	3.376	3.274	0.110	105.152	0.681	sp2
MPTMCH01	109.46	4.572	3.497	2.174	4.022	4.543	3.552	2.255	3.944	0.081	80.308	1.075	sp2
FATQIC	109.50	4.197	3.526	3.190	2.727	4.127	3.550	3.430	2.295	0.240	77.803	0.671	sp2
HOMPOP	94.95	4.534	3.578	2.584	3.726	4.691	3.898	3.246	3.387	0.662	88.366	0.956	sp2
MEDDOP	95.04	4.443	3.581	2.602	3.601	4.414	3.827	3.492	2.700	0.890	80.11	0.862	sp2
HUDFIW	109.48	4.520	3.582	2.927	3.444	4.798	4.090	3.608	3.163	0.681	93.792	0.938	sp2
RONVEV	93.57	4.970	3.596	0.015	4.970	3.928	2.696	0.082	3.927	0.067	33.695	1.374	sp2
MOCGIW	101.07	4.511	3.604	2.853	3.494	4.636	4.028	3.643	2.867	0.790	86.936	0.907	Ti
GUJNOP	102.79	3.951	3.625	3.587	1.656	3.750	3.276	3.127	2.070	-0.460	71.6	0.326	Mo
RONVAR01	93.42	3.972	3.637	3.595	1.689	4.738	4.063	3.647	3.025	0.052	117.307	0.335	sp2
ICOPUN	97.77	3.736	3.653	3.612	0.955	3.531	3.662	3.510	0.385	-0.102	70.836	0.083	sp2
RONVAR	93.21	5.019	3.657	0.283	5.011	4.008	2.767	0.206	4.003	-0.077	35.629	1.362	sp2
QUSSUS	109.50	4.230	3.661	3.547	2.305	4.534	3.581	2.907	3.479	-0.640	94.436	0.569	Zn
RONVAR01	93.42	5.028	3.661	0.240	5.022	4.021	2.770	0.188	4.017	-0.052	35.83	1.367	sp2
MURMER	99.59	5.025	3.721	1.282	4.859	3.852	2.677	1.335	3.613	0.053	24.793	1.304	sp3
SIZBAE	97.94	5.073	3.738	1.075	4.958	3.970	2.755	1.001	3.842	-0.074	29.906	1.335	sp3
GAQQAR	100.93	4.869	3.751	2.363	4.257	4.664	3.765	3.157	3.433	0.794	73.304	1.118	sp2
RONVAR01	93.42	4.389	3.757	3.556	2.573	4.525	3.932	3.504	2.863	-0.052	87.171	0.632	sp2
SIZBAE	100.26	5.073	3.764	0.649	5.031	3.939	2.742	1.061	3.793	0.412	27.788	1.309	sp3
ICOPUN	97.37	3.968	3.813	3.799	1.146	4.014	3.528	3.354	2.205	-0.445	82.321	0.155	sp2

Table C7 continued

KICROE	109.46	4.843	3.823	1.998	4.412	4.276	3.145	1.995	3.782	-0.003	57.516	1.020	Si
RUFDAZ	109.51	5.067	3.828	1.972	4.668	3.802	2.676	1.802	3.348	-0.170	16.355	1.239	sp3
KEXNOR	101.53	4.928	3.836	2.566	4.207	4.728	3.745	2.846	3.775	0.280	73.602	1.092	sp3
RONVAR	93.21	4.453	3.849	3.660	2.536	4.518	3.971	3.584	2.751	-0.076	84.2	0.604	sp2
POHJON01	109.49	4.817	3.853	2.993	3.774	4.356	3.731	3.345	2.790	0.352	62.157	0.964	sp3
YOKRIA	86.62	5.111	3.863	1.951	4.724	4.103	2.983	2.097	3.527	0.146	35.95	1.248	P
HUCMIC	109.52	4.699	3.891	3.182	3.458	4.746	3.629	2.267	4.170	-0.915	83.865	0.808	sp2
LOKRUA	109.45	4.714	3.892	3.227	3.436	4.700	3.869	3.314	3.333	0.087	81.261	0.822	sp2
XIGPUA	96.77	4.050	3.935	3.889	1.131	3.719	3.536	3.500	1.257	-0.389	66.264	0.115	sp2
WIXTEE	110.13	4.773	3.995	3.434	3.315	4.736	4.124	3.927	2.647	0.493	80.357	0.778	sp2
HUDFIW	109.49	4.757	4.015	3.542	3.175	4.253	3.871	3.711	2.078	0.169	60.166	0.742	sp2
LOKRUA	109.45	4.808	4.108	3.474	3.324	4.676	3.899	3.388	3.223	-0.086	76.338	0.700	sp2
POHJON01	109.56	5.122	4.124	3.092	4.083	4.651	3.976	3.635	2.901	0.543	62.233	0.998	sp3
WESZUQ	100.71	5.207	4.170	2.379	4.632	4.458	3.554	2.796	3.472	0.417	49.631	1.037	sp3
YIRGEM	97.55	5.089	4.183	3.261	3.907	3.758	2.936	2.462	2.839	-0.799	5.051	0.906	sp2
WAKHOF02	96.17	5.523	4.204	1.534	5.306	4.413	3.219	1.730	4.060	0.196	29.854	1.319	Si
FULGIE	95.37	5.140	4.225	3.369	3.882	3.823	3.006	2.563	2.837	-0.806	8.876	0.915	sp2
DAHCOE	108.92	5.216	4.226	3.260	4.072	4.500	3.806	3.549	2.767	0.289	51.206	0.990	sp2
GAGTUF	96.13	4.884	4.234	3.965	2.852	4.134	3.725	3.661	1.920	-0.304	49.152	0.650	sp2
BODMOX	97.07	5.103	4.235	3.389	3.815	4.462	3.809	3.373	2.921	-0.016	54.561	0.868	sp2
KEXNOR	101.53	5.166	4.251	3.332	3.948	4.745	4.079	3.535	3.165	0.203	64.472	0.915	sp3
WAKHOF01	98.36	5.570	4.251	1.436	5.382	4.460	3.257	1.531	4.189	0.095	29.867	1.319	Si
EXUVUO	94.93	5.515	4.254	2.101	5.099	4.412	3.284	2.117	3.871	0.016	30.298	1.261	Ge
NUMWUP	100.29	5.222	4.305	3.532	3.846	3.887	3.045	2.691	2.805	-0.841	3.623	0.917	sp2
DIZGUP	109.49	5.205	4.317	3.209	4.098	4.580	3.886	3.418	3.049	0.209	55.426	0.888	sp2
PESPIN	97.82	5.318	4.327	3.258	4.203	4.682	3.757	3.139	3.474	-0.119	55.054	0.991	sp3

Table C7 continued

RONVEV	93.57	5.116	4.353	3.606	3.629	4.662	3.995	3.674	2.870	0.068	62.963	0.763	sp2
JURJAH	99.19	5.269	4.356	3.005	4.328	4.467	3.394	2.017	3.986	-0.988	47.087	0.913	Ru
HOMPOP	105.66	5.331	4.381	3.507	4.015	3.995	3.097	2.650	2.990	-0.857	2.658	0.950	sp2
WUJWAZ	93.55	5.464	4.386	1.944	5.106	4.198	3.226	1.788	3.798	-0.156	16.425	1.078	sp2
ILOQAB	98.37	5.130	4.413	4.029	3.176	4.419	3.830	3.475	2.730	-0.554	51.315	0.717	sp2
ILOQEF	109.46	5.279	4.416	3.332	4.095	4.612	3.977	3.644	2.827	0.312	53.57	0.863	sp2
BOZTOB	107.26	5.549	4.420	1.476	5.349	4.655	3.670	2.084	4.162	0.608	42.727	1.129	sp3
DIXFAS	99.20	5.030	4.470	4.316	2.583	4.551	3.837	3.468	2.947	-0.848	61.717	0.560	sp3
ODOSAD	108.54	4.664	4.480	4.478	1.304	3.590	3.705	3.566	0.414	-0.912	31.391	0.184	sp2
HIRBEQ	116.89	5.428	4.489	3.541	4.114	4.197	3.256	2.614	3.284	-0.927	20.153	0.939	sp2
LUVKOD	95.53	5.419	4.517	3.705	3.955	4.137	3.324	2.954	2.896	-0.751	14.484	0.902	Si
GAGTUF	92.92	5.668	4.539	2.610	5.031	4.648	3.595	2.563	3.877	-0.047	35.718	1.129	sp2
MEDBUT	99.64	5.342	4.544	3.897	3.654	4.359	3.809	3.468	2.641	-0.429	37.562	0.798	ssp2
GIQDER	96.21	5.114	4.568	4.412	2.586	4.260	3.491	3.191	2.822	-1.221	44.315	0.546	sp2
XUGPEU	106.19	5.591	4.589	3.482	4.374	4.713	3.971	3.563	3.085	0.081	43.596	1.002	sp2
COLDEN	83.11	5.360	4.604	4.009	3.558	4.064	3.315	3.009	2.732	-1.000	12.429	0.756	sp2
ZEPVOF	95.44	5.475	4.623	3.429	4.268	4.234	3.297	2.392	3.494	-1.037	19.173	0.852	Mo
COLDAJ	96.30	5.240	4.644	4.114	3.245	4.174	3.423	3.040	2.860	-1.074	32.506	0.596	sp2
NAMRUO	109.45	5.801	4.674	2.866	5.044	4.660	3.658	2.777	3.742	-0.089	27.831	1.127	Zn
GAQQAR	97.35	5.544	4.704	3.794	4.042	4.358	3.461	2.587	3.507	-1.207	24.232	0.840	sp2
KILCUE	109.41	5.475	4.756	4.286	3.407	4.223	3.611	3.428	2.466	-0.858	18.045	0.719	sp2
YUKDIR	116.38	5.558	4.756	3.822	4.035	4.550	3.683	2.592	3.740	-1.230	36.33	0.802	W
COLDAJ	103.89	5.630	4.789	3.684	4.257	4.526	3.659	2.725	3.614	-0.959	30.316	0.841	sp2
GIHVAV	94.81	5.270	4.828	4.634	2.510	3.975	3.617	3.573	1.742	-1.061	12.594	0.442	P
TASBIY	84.40	5.629	4.828	4.199	3.749	4.433	3.588	3.061	3.207	-1.138	23.41	0.801	Mg
COLDEN	87.89	5.596	4.867	4.190	3.709	4.549	3.698	3.190	3.243	-1.000	33.996	0.729	sp2

Table C7 continued

WESQER	90.12	5.869	4.869	3.564	4.663	4.622	3.567	2.574	3.839	-0.990	18.743	1.000		sp3
COLDEN	103.47	5.568	4.879	4.303	3.534	4.524	3.671	3.134	3.263	-1.169	34.206	0.689		sp2
MELFAK	109.49	5.703	4.965	4.359	3.677	4.621	3.735	3.062	3.461	-1.297	31.863	0.738		sp2
HELXUR01	109.20	3.101	2.266	1.867	2.476	3.901	3.263	2.979	2.519	1.112	117.343	0.835	A1	sp2
ZORQUU	109.52	2.571	2.371	2.289	1.171	3.724	3.123	2.927	2.302	0.638	142.629	0.200	A1	sp2
HELXUR01	109.15	2.699	2.403	2.342	1.342	3.827	3.192	2.926	2.467	0.584	140.417	0.296	A1	sp2
VEZLOC	97.94	2.767	2.725	2.641	0.825	4.056	3.994	3.947	0.934	1.306	161.064	0.042	A1	sp3
GOCGUB	102.12	3.468	2.775	2.384	2.519	4.662	4.101	3.667	2.879	1.283	148.602	0.693	A1	sp2
VEZLOC	97.94	3.296	3.014	2.952	1.466	3.640	3.771	3.623	0.351	0.671	93.879	0.282	A1	sp3
ZORQUU	117.01	3.588	3.270	3.233	1.556	3.947	3.316	2.900	2.677	-0.333	95.458	0.318	A1	sp2
ZORQUU	100.72	4.257	3.386	2.498	3.447	3.719	3.147	2.927	2.294	0.429	57.747	0.871	A1	sp2
HELXUR01	109.31	3.983	3.706	3.652	1.590	3.924	3.292	2.897	2.647	-0.755	77.764	0.277	A1	sp2
KASJIZ	98.41	3.822	3.744	3.708	0.927	3.709	3.414	3.416	1.445	-0.292	75.001	0.078	A1	sp2
GOCGUB	98.27	4.595	3.791	2.970	3.506	4.541	3.822	3.380	3.033	0.410	79.313	0.804	A1	sp2
GOCGUB	102.12	4.921	4.167	3.408	3.550	4.569	3.856	3.431	3.017	0.023	67.064	0.754	A1	sp2
VEZLOC	97.94	5.424	4.210	1.081	5.315	4.363	3.301	1.466	4.109	0.385	32.945	1.214	A1	sp3
VEZLOC	97.94	5.478	4.279	1.089	5.369	4.333	3.257	1.399	4.101	0.310	27.279	1.199	A1	sp3
HELXUR01	109.12	4.780	4.390	4.264	2.160	3.746	3.151	2.943	2.318	-1.321	33.932	0.390	A1	sp2
HELXUR01	109.20	4.927	4.476	4.235	2.518	3.882	3.249	2.912	2.567	-1.323	33.521	0.451	A1	sp2
KASJIZ	108.58	5.430	4.546	3.227	4.367	4.693	3.918	3.302	3.335	0.075	50.449	0.884	A1	sp2
KASJIZ	105.14	5.557	4.616	2.863	4.763	4.760	3.846	2.949	3.736	0.086	47.695	0.941	A1	sp2
YOMWAY01	107.93	5.922	5.098	4.335	4.035	4.584	3.788	3.335	3.145	-1.000	1.433	0.824	A1	sp2
IKEGOV	109.50	3.850	2.726	1.432	3.574	4.634	3.747	2.677	3.783	1.245	118.188	1.124	A	sp3
VOPBEH	96.12	3.632	2.730	2.354	2.766	3.985	2.787	1.625	3.639	-0.729	95.322	0.902	A	sp3
KAPNIZ	104.86	3.425	2.830	2.558	2.278	3.837	3.525	3.510	1.550	0.952	97.535	0.595	A	sp2
VOPBEH	96.12	2.692	2.866	2.671	0.336	3.466	3.402	3.356	0.866	0.685	114.401	-0.174	A	sp3

Table C7 continued

XORTIJ	89.96	3.427	2.909	2.854	1.897	4.597	4.089	3.988	2.287	1.134	145.949	0.518	A	sp2
XORTIJ	98.15	3.405	2.936	2.879	1.818	4.506	4.093	4.035	2.006	1.156	139.362	0.469	A	sp2
XORTOP	99.94	4.037	2.945	2.117	3.437	4.305	3.440	2.893	3.188	0.776	92.4	1.092	A	sp2
XORTOP	91.71	3.817	2.985	2.448	2.929	4.582	3.789	3.284	3.195	0.836	116.966	0.832	A	sp2
XORTOP	91.71	3.974	3.161	2.527	3.067	4.149	3.529	3.326	2.480	0.799	87.992	0.813	A	sp2
XORTOP	102.96	3.183	3.210	3.114	0.659	3.620	3.372	3.356	1.357	0.242	97.987	-0.027	A	sp2
XORTOP	102.96	3.557	3.210	3.132	1.686	4.130	3.525	3.351	2.414	0.219	105.909	0.347	A	sp2
VOPBEH	106.39	4.220	3.232	2.574	3.344	3.931	2.769	1.706	3.542	-0.868	68.472	0.988	A	sp3
XORTIJ	98.15	4.075	3.429	3.036	2.718	4.699	3.792	3.096	3.535	0.060	109.788	0.646	A	sp2
WOGRIT01	97.56	4.514	3.486	1.610	4.217	4.701	3.832	2.836	3.749	1.226	89.671	1.028	A	sp3
BAPBOK	91.32	4.920	3.543	0.375	4.906	3.962	2.700	0.065	3.961	-0.310	38.486	1.377	A	sp2
XORTOP	105.91	4.497	3.616	3.161	3.199	3.895	3.407	3.333	2.015	0.172	55.308	0.881	A	sp2
XORTIJ	89.96	5.054	3.621	0.277	5.046	4.173	2.839	0.230	4.167	-0.047	42.815	1.433	A	sp2
XORTOP	91.71	5.082	3.631	0.280	5.074	4.195	2.822	0.125	4.193	-0.155	42.569	1.451	A	sp2
XORTIJ	94.02	5.118	3.682	0.138	5.116	4.183	2.840	0.163	4.180	0.025	40.005	1.436	A	sp2
XORTIJ	93.94	5.115	3.684	0.284	5.107	4.165	2.833	0.234	4.158	-0.050	39.21	1.431	A	sp2
XORTIJ	98.15	5.168	3.741	0.258	5.162	4.166	2.829	0.198	4.161	-0.060	36.272	1.427	A	sp2
KAPNIZ	104.86	4.902	3.959	3.183	3.728	4.668	3.695	2.753	3.770	-0.430	72.114	0.943	A	sp2
VOPBEH	96.12	5.028	3.961	2.688	4.249	4.722	3.970	3.336	3.342	0.648	69.248	1.067	A	sp3
BAJNUW	95.97	4.474	4.155	4.089	1.816	3.619	3.592	3.537	0.766	-0.552	43.286	0.319	A	Se
WOGRIT01	97.56	5.412	4.947	4.759	2.577	4.632	4.026	3.680	2.813	-1.079	48.355	0.465	A	sp3
XORTOP	99.94	5.912	4.967	3.625	4.670	4.800	3.728	2.437	4.135	-1.188	30.012	0.945	A	sp2
IKEGOV	109.50	5.997	5.040	3.541	4.840	4.683	3.816	2.927	3.656	-0.614	9.716	0.957	A	sp3
CEDYEQ	113.94	5.960	5.082	4.016	4.404	4.743	3.784	2.924	3.734	-1.092	21.739	0.878	A	sp3
NUMWUP	100.29	2.743	2.355	2.360	1.398	3.875	3.089	2.837	2.640	0.477	141.01	0.388	C	sp2
KOYJIT	98.99	3.016	2.476	2.438	1.776	4.127	3.767	3.737	1.751	1.299	139.614	0.540	C	sp3

Table C7 continued

KOYJIT	96.83	2.916	2.535	2.515	1.476	4.169	3.839	3.749	1.824	1.234	155.364	0.381	C	sp3
COLDEN	99.16	2.694	2.554	2.499	1.006	4.014	3.801	3.796	1.305	1.297	168.354	0.140	C	sp2
COLDAJ	96.30	2.712	2.626	2.555	0.909	3.779	3.693	3.629	1.054	1.074	135.022	0.086	C	sp2
KILCUE	109.41	2.591	2.650	2.527	0.572	3.715	3.558	3.506	1.228	0.979	139.749	-0.059	C	sp2
ODOSAD	100.68	2.715	2.661	2.573	0.867	3.805	3.933	3.791	0.326	1.218	137.081	0.054	C	sp2
GIQDER	96.21	3.111	2.720	2.630	1.662	4.407	4.021	3.851	2.143	1.221	162.8	0.391	C	sp2
WAKHOF02	94.04	3.469	2.728	2.395	2.510	4.764	4.035	3.362	3.375	0.967	163.018	0.741	C	Si
YUKDIR	116.38	3.210	2.737	2.623	1.850	4.315	3.607	3.313	2.765	0.690	139.406	0.473	C	W
SIZBAE	100.26	2.599	2.761	2.570	0.387	3.733	3.873	3.717	0.345	1.147	140.746	-0.162	C	sp3
ZEPVOF	95.44	3.120	2.765	2.682	1.594	4.195	3.570	3.393	2.467	0.711	136.635	0.355	C	Mo
TERTEP	94.71	3.718	2.781	2.225	2.979	3.998	2.790	1.690	3.623	-0.535	92.129	0.937	C	sp3
PESPIN	97.82	2.783	2.809	2.698	0.683	3.489	3.565	3.445	0.552	0.747	110.791	-0.026	C	sp3
HOMPOP	105.66	3.366	2.825	2.586	2.155	4.246	3.617	3.442	2.486	0.856	122.979	0.541	C	sp2
KECXIA	94.28	3.593	2.833	2.539	2.542	4.798	4.125	3.602	3.170	1.063	149.966	0.760	C	sp2
GAGTUF	86.42	3.559	2.836	2.578	2.454	4.260	3.857	3.803	1.920	1.225	112.768	0.723	C	sp2
ICOPUN	97.37	2.910	2.852	2.791	0.824	3.846	3.895	3.806	0.553	1.015	125.567	0.058	C	sp2
XAFVEH	98.77	2.790	2.901	2.750	0.471	3.625	3.674	3.575	0.600	0.825	118.556	-0.111	C	sp3
JURJAH	83.80	3.698	2.932	2.736	2.488	3.990	3.366	2.991	2.641	0.255	92.654	0.766	C	Ru
TAPZIU	96.76	4.095	2.956	0.809	4.014	3.957	2.769	1.391	3.704	0.582	74.65	1.139	C	sp3
MELFAK	109.49	3.211	2.973	2.949	1.270	3.324	3.590	3.323	0.082	0.374	82.943	0.238	C	sp2
DIXFAS	99.20	3.734	2.988	2.739	2.538	3.930	3.626	3.620	1.530	0.881	88.354	0.746	C	sp3
QUSSUS	109.50	3.870	3.053	2.447	2.998	4.221	3.572	3.356	2.560	0.909	95.846	0.817	C	Zn
WESQER	90.12	2.948	3.120	2.928	0.343	3.873	3.760	3.727	1.053	0.799	124.964	-0.172	C	sp3
LUVKOD	95.53	3.409	3.122	3.036	1.550	4.527	3.926	3.594	2.753	0.558	140.903	0.287	C	Si
WUJWAZ	93.55	3.846	3.130	2.762	2.676	4.154	3.612	3.464	2.293	0.702	93.741	0.716	C	sp2
GUJNOP	111.77	3.598	3.145	3.115	1.801	4.121	4.072	4.037	0.828	0.922	103.504	0.453	C	Mo

Table C7 continued

SIZBAE	97.94	3.861	3.152	2.944	2.498	4.781	4.213	3.927	2.727	0.983	126.62	0.709	C	sp3
MEDBUT	99.64	3.971	3.185	2.732	2.882	4.370	3.950	3.824	2.115	1.092	98.295	0.786	C	sp2
WAKHOF01	98.36	4.143	3.222	2.622	3.208	4.461	3.281	1.676	4.134	-0.946	94.917	0.921	C	Si
NAMROI	109.49	4.286	3.269	2.160	3.702	4.410	3.573	3.068	3.168	0.908	86.448	1.017	C	Zn
HIRBEQ	116.89	4.425	3.274	1.943	3.976	4.396	3.574	2.870	3.330	0.927	80.052	1.151	C	sp2
HUCMIC	109.51	3.587	3.277	3.259	1.498	4.595	4.135	4.013	2.238	0.754	132.281	0.310	C	sp2
FULGIE	95.37	3.361	3.321	3.255	0.837	4.132	3.638	3.556	2.104	0.301	116.336	0.040	C	sp2
ADELOL	87.31	4.181	3.365	2.922	2.990	3.743	2.915	2.488	2.796	-0.434	61.982	0.816	C	sp2
JURJAH	99.19	3.938	3.371	3.210	2.281	4.531	3.727	2.897	3.484	-0.313	107.881	0.567	C	Ru
COLDAJ	96.30	3.885	3.378	3.288	2.069	4.637	3.855	3.214	3.342	-0.074	116.386	0.507	C	sp2
ILOQAB	98.37	4.111	3.461	3.004	2.806	4.623	3.883	3.558	2.952	0.554	104.114	0.650	C	sp2
TASBIY	84.40	4.603	3.466	2.103	4.095	4.598	3.729	3.149	3.350	1.046	81.463	1.137	C	Mg
KICROE	109.47	4.383	3.472	2.995	3.200	4.368	3.223	2.047	3.859	-0.948	80.567	0.911	C	Si
XIGPUA	97.46	4.241	3.505	3.239	2.738	3.906	3.619	3.613	1.484	0.374	66.524	0.736	C	sp2
XEYHUF	90.99	4.366	3.554	3.042	3.132	4.691	3.733	2.366	4.051	-0.676	95.658	0.812	C	sp2
GADNIJ	109.49	3.962	3.632	3.534	1.791	3.574	3.435	3.408	1.077	-0.126	63.6	0.330	C	sp2
YIRGEM	97.55	4.291	3.641	3.140	2.925	3.725	3.424	3.355	1.619	0.215	56.552	0.650	C	sp2
TAPZIU	96.76	4.980	3.680	1.288	4.811	3.870	2.700	1.103	3.709	-0.185	29.394	1.300	C	sp3
COLDEN	100.51	4.384	3.681	3.134	3.066	4.226	3.733	3.620	2.181	0.486	74.418	0.703	C	sp2
NUZLAW	109.44	4.510	3.716	3.305	3.069	4.048	3.339	2.857	2.868	-0.448	61.596	0.794	C	Pd
HUCGES	109.48	4.226	3.723	3.630	2.164	4.385	3.696	3.153	3.047	-0.477	87.858	0.503	C	sp2
JIVNAD	105.99	4.483	3.772	3.492	2.811	4.650	3.709	2.873	3.656	-0.619	88.754	0.711	C	sp2
MURMER	99.59	4.413	3.784	3.494	2.696	4.792	3.944	3.441	3.335	-0.053	98.249	0.629	C	sp3
ELAQIR	109.47	4.075	3.790	3.720	1.663	3.886	3.647	3.572	1.530	-0.148	72.408	0.285	C	sp2
RUFDAZ	109.51	4.478	3.830	3.490	2.806	4.025	3.712	3.660	1.675	0.170	61.9	0.648	C	sp3
GIHVAV	94.81	4.829	3.870	3.070	3.728	4.714	4.094	3.864	2.700	0.794	77.081	0.959	C	P

Table C7 continued

WIXTAA	109.45	4.917	3.908	3.040	3.865	4.694	3.802	2.917	3.678	-0.123	72.597	1.009	C	sp2
GAQQAR	97.35	4.523	3.921	3.549	2.804	4.414	3.603	3.176	3.065	-0.373	76.806	0.602	C	sp2
GAQQAR	97.35	4.755	3.950	3.344	3.380	4.361	3.792	3.541	2.546	0.197	64.955	0.805	C	sp2
WIXSUT	112.22	5.060	4.062	3.168	3.946	4.643	3.765	2.888	3.636	-0.280	64.489	0.998	C	sp2
YAFTUW	109.45	4.569	4.080	3.967	2.267	3.792	3.641	3.592	1.215	-0.375	47.316	0.489	C	sp2
KEXNOR	101.53	4.194	4.128	4.095	0.906	3.837	3.933	3.816	0.401	-0.279	65.473	0.066	C	sp3
WAKHOF01	96.20	5.520	4.204	1.559	5.295	4.426	3.232	1.695	4.089	0.136	30.91	1.316	C	Si
KOYJIT	98.99	5.274	4.225	2.791	4.475	4.537	3.645	2.444	3.822	-0.347	50.279	1.049	C	sp3
MELFAK	109.49	4.804	4.240	3.791	2.951	4.602	3.739	3.143	3.362	-0.648	73.316	0.564	C	sp2
COLDEN	109.48	5.017	4.281	3.510	3.585	4.689	4.062	3.806	2.739	0.296	68.252	0.736	C	sp2
KUKGAZ	97.21	4.986	4.306	4.017	2.954	4.416	3.934	3.697	2.415	-0.320	57.626	0.680	C	sp3
FATQIC	109.50	5.039	4.322	3.951	3.127	4.202	3.783	3.711	1.971	-0.240	45.071	0.717	C	sp2
WUSBAN	90.94	5.061	4.336	3.917	3.205	4.576	3.776	3.036	3.424	-0.881	61.499	0.725	C	Al
HUCGES	109.45	5.018	4.350	4.034	2.984	4.012	3.643	3.597	1.777	-0.437	35.933	0.668	C	sp2
WONHAK	96.52	5.124	4.350	3.756	3.485	4.270	3.505	2.911	3.124	-0.845	44.331	0.774	C	sp2
ZEPVOF	95.44	5.539	4.386	2.665	4.856	4.479	3.511	2.725	3.555	0.060	33.146	1.153	C	Mo
TAZFOR	100.02	4.715	4.415	4.350	1.819	4.492	3.966	3.865	2.289	-0.485	72.26	0.300	C	sp2
ELAQIR	109.47	5.041	4.419	4.149	2.863	4.095	3.616	3.545	2.050	-0.604	39.374	0.622	C	sp2
NAMRUO	109.45	5.093	4.435	4.121	2.993	4.498	3.567	2.881	3.454	-1.240	56.657	0.658	C	Zn
LEFKEO	91.39	5.163	4.440	4.039	3.216	4.765	3.792	2.890	3.789	-1.149	65.451	0.723	C	sp2
ODOSAD	105.71	4.435	4.501	4.405	0.515	4.025	3.903	3.883	1.060	-0.522	63.712	-0.066	C	sp2
TASPUY	109.29	5.523	4.527	3.417	4.339	4.191	3.260	2.677	3.225	-0.740	4.672	0.996	C	sp3
WONHAK	96.52	4.958	4.539	4.431	2.224	3.785	3.229	3.151	2.097	-1.280	24.829	0.419	C	sp2
XEYHUF	90.99	4.395	4.546	4.390	0.210	3.698	3.510	3.500	1.194	-0.890	50.791	-0.151	C	sp2
KOYJIT	98.99	5.328	4.551	3.814	3.720	4.576	3.656	2.515	3.823	-1.299	49.639	0.777	C	sp3
KOYJIT	96.83	5.503	4.583	3.594	4.167	4.684	3.696	2.440	3.998	-1.154	46.515	0.920	C	sp3

Table C7 continued

WIXSON	109.54	5.515	4.585	3.720	4.071	4.745	3.851	2.956	3.712	-0.764	48.943	0.930	C	sp2
LUVKOD	95.53	4.964	4.644	4.584	1.905	4.326	3.720	3.448	2.613	-1.136	54.483	0.320	C	Si
KOYJIT	96.83	5.463	4.644	3.918	3.807	4.565	3.640	2.650	3.717	-1.268	42.44	0.819	C	sp3
HUSNEP	95.72	4.930	4.668	4.653	1.629	4.761	4.180	3.910	2.716	-0.743	74.958	0.262	C	sp2
YUCYUR	109.45	4.892	4.691	4.660	1.489	4.046	3.644	3.545	1.950	-1.115	44.414	0.201	C	sp2
BODMOX	94.79	5.235	4.699	4.502	2.672	4.568	3.974	3.553	2.871	-0.949	53.481	0.536	C	sp2
GAQQAR	100.93	5.408	4.718	4.274	3.314	4.247	3.735	3.480	2.434	-0.794	26.037	0.690	C	sp2
DEBYUG	102.78	4.857	4.719	4.711	1.182	3.708	3.848	3.694	0.322	-1.017	26.446	0.138	C	sp2
TAZFOR	100.02	5.414	4.729	4.344	3.231	4.711	4.234	4.146	2.237	-0.198	52.06	0.685	C	sp2
KOYJIT	98.99	5.753	4.797	2.954	4.937	4.452	3.526	2.438	3.725	-0.516	11.897	0.956	C	sp3
MEDBUT	99.64	5.470	4.801	4.400	3.250	4.386	3.909	3.841	2.117	-0.559	31.541	0.669	C	sp2
WIVSEA	103.42	5.602	4.882	4.482	3.361	4.727	3.806	3.188	3.490	-1.294	43.764	0.720	C	sp2
QUSSUS	109.50	5.874	4.888	3.745	4.525	4.555	3.597	2.912	3.503	-0.833	8.541	0.986	C	Zn
DPDHPS01	98.51	5.151	4.982	4.974	1.339	4.038	3.864	3.847	1.227	-1.127	29.363	0.169	C	P
DPDHPS10	101.22	5.179	5.010	5.001	1.346	4.089	3.923	3.908	1.203	-1.093	30.93	0.169	C	P
BECVUC	93.10	5.326	5.073	5.044	1.710	4.232	3.868	3.846	1.766	-1.198	30.821	0.253	C	sp3
HUCGES	109.51	5.753	5.088	4.409	3.696	4.469	3.842	3.465	2.822	-0.944	14.417	0.665	C	sp2
KUKGAZ	95.19	5.528	5.096	4.861	2.632	4.254	3.764	3.560	2.329	-1.301	15.526	0.432	C	sp3
HEBPOT	98.07	6.006	5.103	3.956	4.519	4.686	3.788	2.958	3.634	-0.998	8.171	0.903	C	sp2
XAFVEH	98.77	5.395	5.156	5.100	1.760	4.104	3.838	3.821	1.498	-1.279	13.293	0.239	C	sp3
DUCNUL	109.46	5.715	5.214	4.925	2.899	4.723	4.130	3.680	2.960	-1.245	37.397	0.501	C	sp2
GAYKUO	102.96	3.014	2.254	2.085	2.176	3.816	3.336	3.140	2.168	1.055	117.208	0.760	D	Ru, sp3
WORNEW	100.71	2.849	2.544	2.512	1.344	3.906	3.328	3.014	2.484	0.502	134.521	0.305	D	Ru, H_2S
DIRVAB	109.79	2.912	2.554	2.487	1.515	4.114	3.565	3.232	2.545	0.745	148.685	0.358	D	Ru, Ru
DIRVAB	104.52	2.993	2.573	2.602	1.479	4.031	3.545	3.445	2.093	0.843	133.337	0.420	D	Ru, Ru
SULTOI	115.03	3.138	2.602	2.520	1.870	4.177	3.274	2.728	3.163	0.208	133.75	0.536	D	Ru, sp3

Table C7 continued

SULTOI	115.03	3.260	2.768	2.566	2.011	4.258	3.429	2.651	3.332	0.085	130.892	0.492	D	Ru, sp3
PUMDUX	109.45	3.754	2.899	2.362	2.918	3.957	3.319	3.103	2.455	0.741	88.734	0.855	D	Re
RODBIV	113.16	3.715	3.065	2.882	2.344	4.556	4.079	3.984	2.210	1.102	121.315	0.650	D	Mn, Mn
SIZCAF	107.39	3.978	3.147	2.792	2.834	3.904	3.302	2.936	2.573	0.144	77.105	0.831	D	Ru, H_2S
CEMROB	104.36	3.497	3.177	3.150	1.519	3.655	3.184	3.014	2.068	-0.136	85.954	0.320	D	Rh, Rh
VETKAH	102.18	4.062	3.182	2.731	3.007	3.719	3.274	3.120	2.024	0.389	65.786	0.880	D	Rh, sp2
PUDWER	100.36	4.393	3.418	2.587	3.550	4.239	3.317	2.805	3.178	0.218	74.612	0.975	D	Sb, sp3
ZOZBIB	101.61	3.923	3.939	3.863	0.683	4.648	4.287	4.207	1.976	0.344	114.921	-0.016	D	Ni, Ni
XUHZOQ	113.93	4.814	4.165	3.716	3.060	4.486	3.826	3.583	2.699	-0.133	67.926	0.649	D	Cu, Cu
RODBIV	113.16	4.755	4.283	4.140	2.339	4.290	3.725	3.554	2.403	-0.586	61.882	0.472	D	Mn, Mn
WORNEW	100.71	4.642	4.324	4.205	1.966	3.634	3.232	3.218	1.688	-0.987	35.349	0.318	D	Ru, H_2S
LEGPOC	111.72	5.325	4.600	4.045	3.463	4.177	3.550	3.109	2.790	-0.936	26.989	0.725	D	Ru+, sp3
GAYKUO	102.96	5.106	4.610	4.399	2.592	4.767	3.953	3.344	3.397	-1.055	67.921	0.496	D	Ru, sp3
DIRVAB	109.79	5.707	4.731	2.564	5.099	4.696	3.820	2.744	3.811	0.180	36.302	0.976	D	Ru, Ru
COCKAG	97.02	5.420	4.761	4.342	3.244	4.442	3.711	3.411	2.845	-0.931	37.941	0.659	D	Ir, sp3
WORNAS	110.39	5.081	4.764	4.732	1.851	4.667	4.377	4.274	1.875	-0.458	64.664	0.317	D	Ru, H_2S
WORNAS	110.39	5.092	4.928	4.904	1.371	4.795	4.278	4.134	2.429	-0.770	69.69	0.164	D	Ru, H_2S
JIHHUD	108.14	5.789	4.971	4.145	4.041	4.493	3.689	3.224	3.129	-0.921	12.743	0.818	D	Ru, sp3
YUKDIR01	116.38	3.210	2.737	2.623	1.850	4.315	3.607	3.313	2.765	0.690	139.406	0.473	E	W
WEKYAO	100.24	3.820	3.042	2.665	2.737	3.670	3.226	3.175	1.841	0.510	73.44	0.778	F	Ru^+
CEDYEQ	113.94	3.445	2.607	1.905	2.870	4.707	3.878	3.001	3.626	1.096	157.169	0.838	A, C	sp3
XORTOP	99.94	3.324	2.688	2.515	2.173	4.054	3.786	3.703	1.650	1.188	113.821	0.636	A, C	sp2
XORTOP	102.96	3.351	2.719	2.418	2.320	4.064	3.732	3.633	1.821	1.215	112.993	0.632	A, C	sp2
VOPBEH	106.39	3.015	2.777	2.700	1.342	4.192	3.638	3.420	2.424	0.720	146.09	0.238	A, C	sp3
WOGRIT01	97.56	3.504	2.902	2.816	2.085	4.709	4.184	4.042	2.416	1.226	149.924	0.602	A, C	sp3
BAJNUW	95.97	2.914	2.994	2.868	0.516	3.679	3.630	3.591	0.800	0.723	114.641	-0.080	A, C	Se

Table C7 continued

BAPBOK	91.32	3.480	3.049	3.015	1.738	3.377	3.488	3.344	0.471	0.329	74.441	0.431	A, C	sp2
IKEGOV	109.50	3.412	3.284	3.262	1.001	4.345	4.403	4.311	0.542	1.049	126.611	0.128	A, C	sp3
VOPBEH	106.39	4.250	3.296	2.705	3.278	4.757	3.987	3.469	3.255	0.764	104.109	0.954	A, C	sp3
MKMPIN	94.29	4.521	3.444	2.287	3.900	4.244	3.351	2.179	3.642	-0.108	69.62	1.077	A, C	sp2
HUKJUT	99.12	4.144	3.955	3.904	1.390	3.758	3.509	3.488	1.399	-0.416	64.095	0.189	A, C	sp3
JOLBIX	109.47	5.048	4.162	3.258	3.856	4.296	3.338	2.201	3.689	-1.057	49.254	0.886	A, C	Re
BAPBOK	91.32	5.009	4.251	3.813	3.248	4.169	3.675	3.503	2.260	-0.310	44.84	0.758	A, C	sp2
HUKJUT	99.12	5.264	4.422	3.732	3.712	4.296	3.627	3.404	2.621	-0.328	38.359	0.842	A, C	sp3
WOGRIT	95.18	5.543	4.705	3.858	3.980	4.646	3.939	3.609	2.926	-0.249	42.528	0.838	A, C	sp3
VAGLAR	100.71	5.874	4.864	2.842	5.141	4.581	3.668	2.409	3.896	-0.433	13.155	1.010	A, C	Mo
PIHMOK	93.83	5.321	5.007	4.934	1.992	4.105	3.748	3.621	1.934	-1.313	21.394	0.314	A, C	sp2
HELXUR01	109.12	2.962	2.167	1.889	2.281	3.820	3.266	3.027	2.330	1.138	120.536	0.795	A1, C	sp2
HELXUR01	109.31	2.928	2.503	2.440	1.619	4.022	3.340	2.992	2.688	0.552	137.849	0.425	A1, C	sp2
ZORQUU	117.01	3.343	2.989	2.942	1.588	4.060	3.348	2.973	2.765	0.031	113.183	0.354	A1, C	sp2
YOMWAY01	107.93	2.821	3.130	2.821	0.000	3.783	3.645	3.620	1.098	0.799	127.182	-0.309	A1, C	sp2
KASJIZ	109.86	3.558	3.541	3.469	0.791	3.504	3.517	3.434	0.697	-0.035	76.802	0.017	A1, C	sp2
HELXUR01	109.20	4.265	3.757	3.654	2.200	4.117	3.761	3.654	1.897	0.000	74.601	0.508	A1, C	sp2
HELXUR01	109.15	3.934	3.777	3.727	1.259	3.848	3.273	3.019	2.386	-0.708	76.495	0.157	A1, C	sp2
ZORQUU	109.52	3.948	3.787	3.751	1.232	3.813	3.257	3.009	2.342	-0.742	74.442	0.161	A1, C	sp2
ZORQUU	100.72	5.031	4.305	3.672	3.439	3.798	3.278	3.031	2.289	-0.641	19.658	0.726	A1, C	sp2
KASJIZ	102.95	4.761	4.414	4.285	2.075	4.714	3.990	3.535	3.119	-0.750	79.921	0.347	A1, C	sp2
WORNAS	110.39	3.121	2.503	2.423	1.967	3.712	3.274	3.193	1.893	0.770	105.595	0.618	C, D	Ru, H_2S
CEMROB	104.36	3.468	2.627	2.318	2.580	4.042	3.331	2.811	2.904	0.493	105.757	0.841	C, D	Rh, Rh
LEGPOC	111.72	3.613	2.636	2.051	2.974	4.553	3.553	2.092	4.044	0.041	127.469	0.977	C, D	Ru+, sp3
JIHHUD	108.14	3.394	2.724	2.529	2.263	4.163	3.382	2.895	2.992	0.366	116.252	0.670	C, D	Ru, sp3
WORNAS	110.39	2.889	2.765	2.686	1.064	3.671	3.185	3.145	1.893	0.459	115.542	0.124	C, D	Ru, H_2S

Table C7 continued

GAYKUO	102.96	3.101	2.834	2.786	1.362	3.734	3.127	2.990	2.237	0.204	107.812	0.267	C, D	Ru, sp3
RODBIV	113.16	3.437	2.846	2.746	2.067	4.257	3.682	3.464	2.474	0.718	119.42	0.591	C, D	Mn, Mn
WUDQUI	94.66	3.864	2.867	1.901	3.364	4.025	2.810	1.673	3.661	-0.228	87.105	0.997	C, D	sp3, sp3
WORNEW	100.71	4.304	3.219	1.138	4.151	4.562	3.644	2.433	3.859	1.295	92.451	1.085	C, D	Ru, H_2S
COCKAG	97.02	4.236	3.291	2.676	3.284	4.065	3.410	3.013	2.729	0.337	73.551	0.945	C, D	Ir, sp3
XUHZOQ	113.93	4.957	3.973	3.117	3.854	4.799	3.769	2.592	4.039	-0.525	75.453	0.984	C, D	Cu, Cu
XEHDOF	98.82	4.491	4.002	3.878	2.265	4.522	3.811	3.477	2.891	-0.401	82.791	0.489	C, D	sp3
VETKAH	102.18	4.853	4.412	4.261	2.323	3.599	3.125	2.960	2.047	-1.301	17.527	0.441	C, D	Rh, sp2
GAYKUO	102.96	5.624	4.508	2.680	4.944	4.783	3.834	2.347	4.168	-0.333	45.536	1.116	C, D	Ru, sp3
SULTOI	115.03	5.820	4.848	3.052	4.956	4.633	3.565	2.030	4.165	-1.022	24.274	0.972	C, D	Ru, sp3
ZOZBIB	101.61	5.713	5.108	4.516	3.499	4.513	4.082	3.863	2.333	-0.653	23.153	0.605	C, D	Ni, Ni
YUKDIR01	116.38	5.558	4.756	3.822	4.035	4.550	3.683	2.592	3.740	-1.230	36.33	0.802	C, E	W
SIZCAF10	107.39	3.978	3.147	2.792	2.834	3.904	3.302	2.936	2.573	0.144	77.105	0.831	D, E	Ru, H_2S

Appendix D

NMR SPECTRA FOR NOVEL COMPOUNDS

Compounds were described and utilized in Chapter 3: Electronic Control of Proline Cis-Trans Isomerism via a C–H/π Aromatic Interaction: Insights into the Nature of C–H/π Interactions.

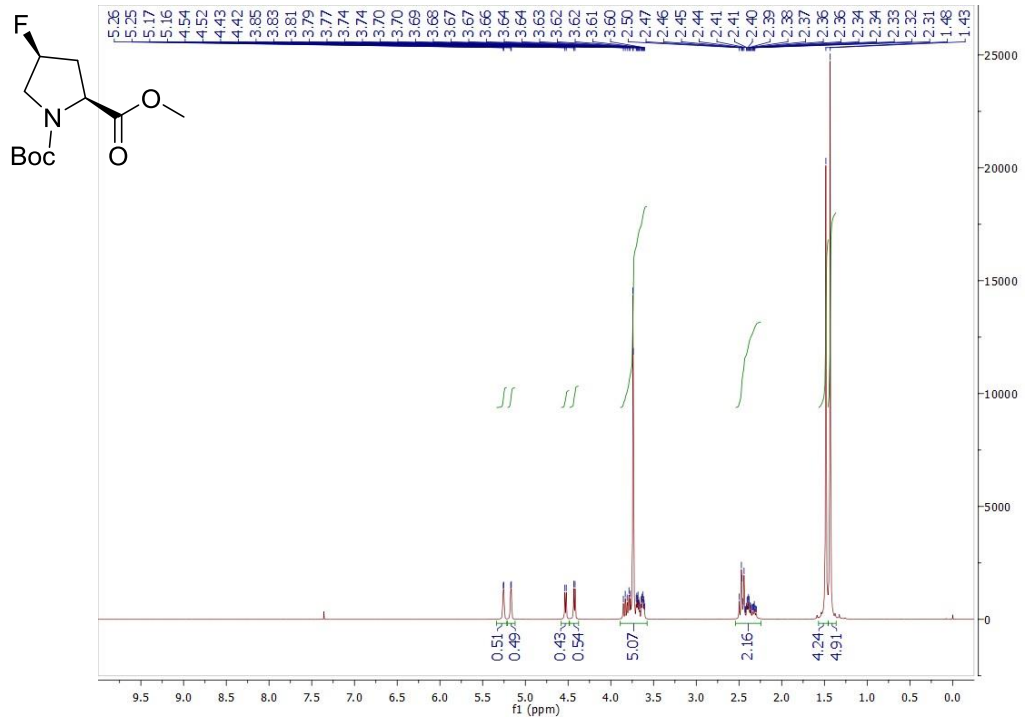


Figure D1 **^{1}H NMR spectrum of Boc-4*S*-fluoroproline-methyl ester in $CDCl_3$**

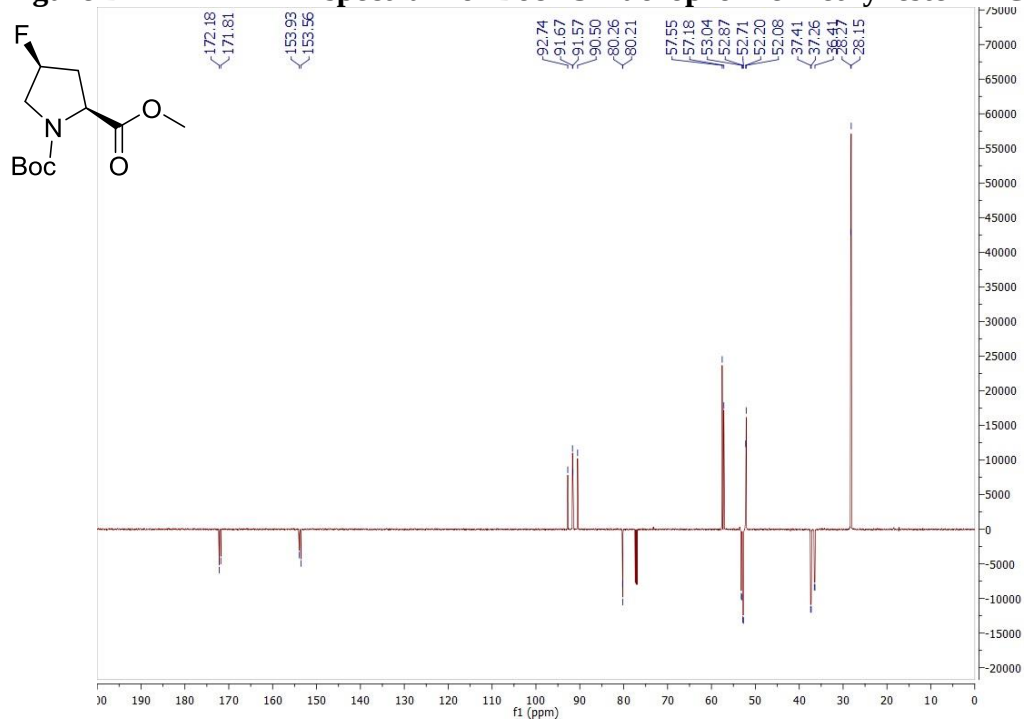


Figure D2 **^{13}C NMR spectrum of Boc-4*S*-fluoroproline-methyl ester in $CDCl_3$**

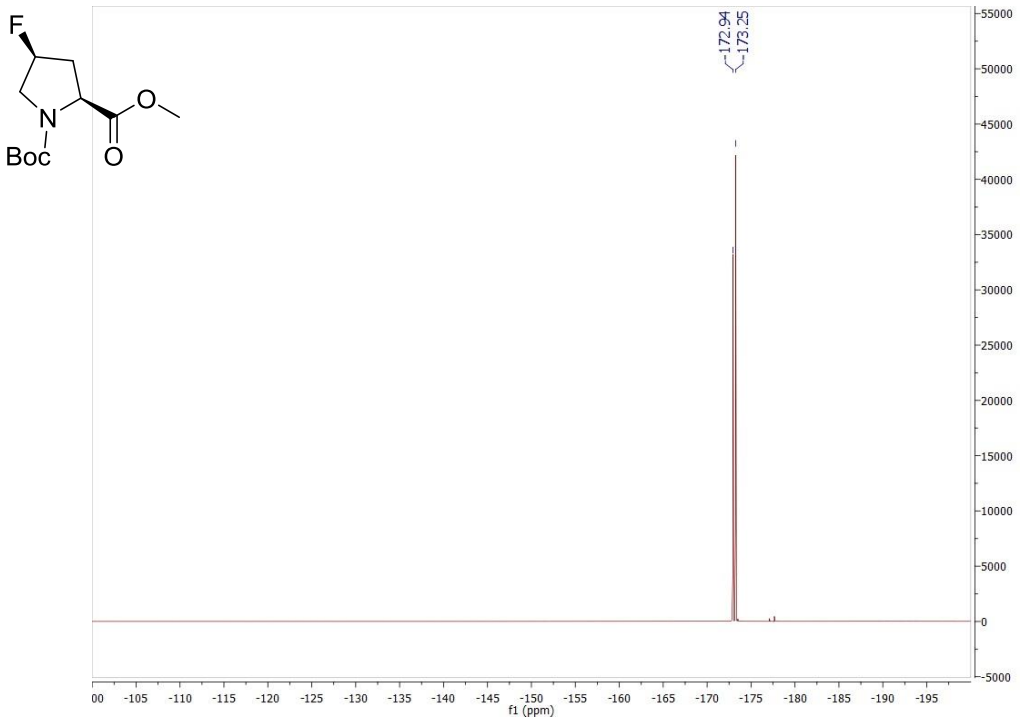


Figure D3 ^{19}F NMR spectrum of Boc-4*S*-fluoroproline-methyl ester in $CDCl_3$.

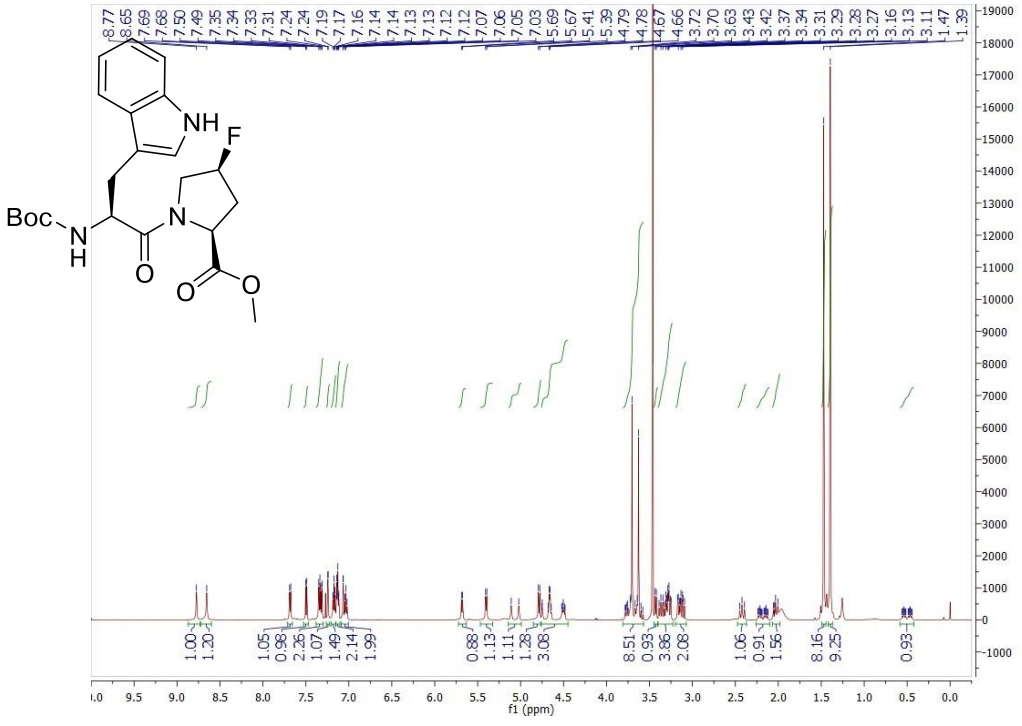


Figure D4 ^{1}H NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

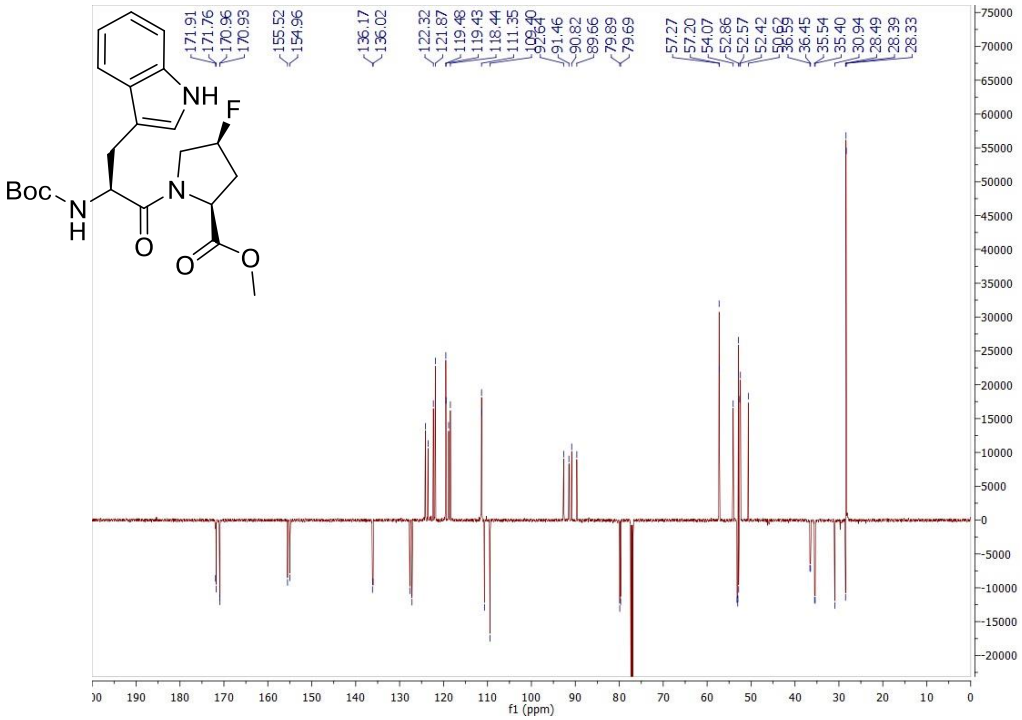


Figure D5 ^{13}C NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

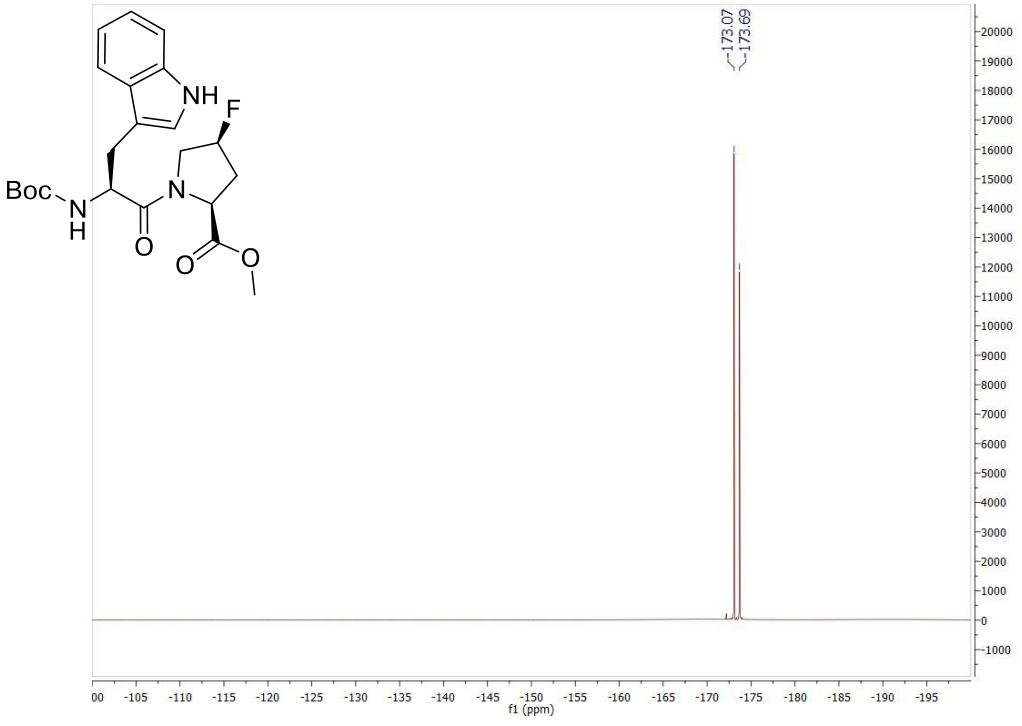


Figure D6 ^{19}F NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

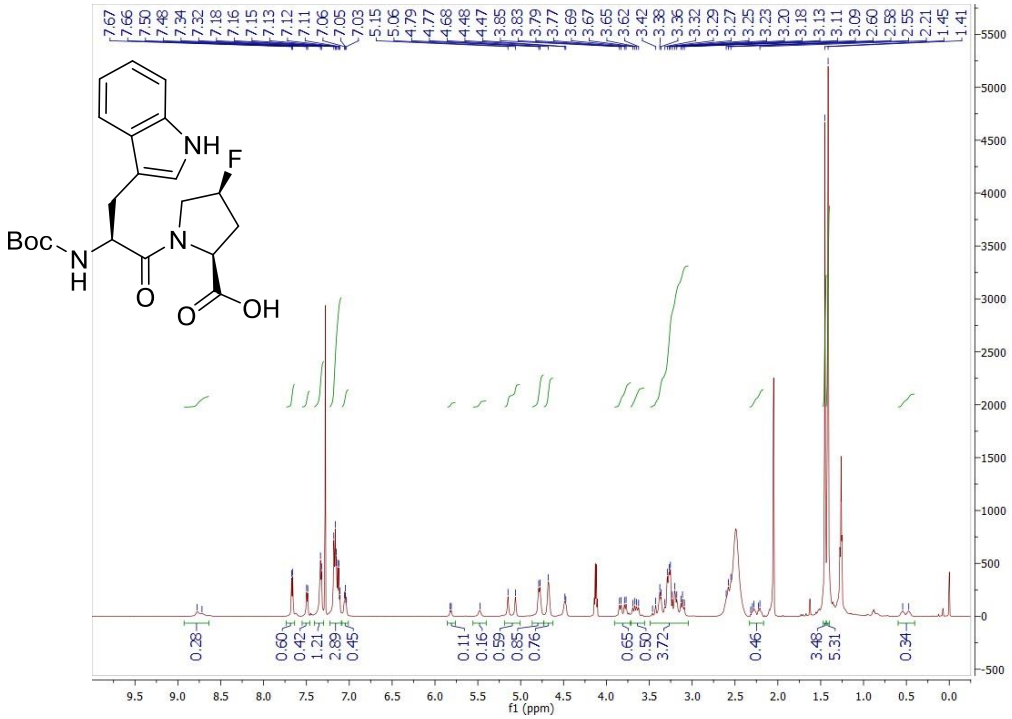


Figure D7 ^{1}H NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline) in $CDCl_3$ with ethyl acetate as a cosolvent.

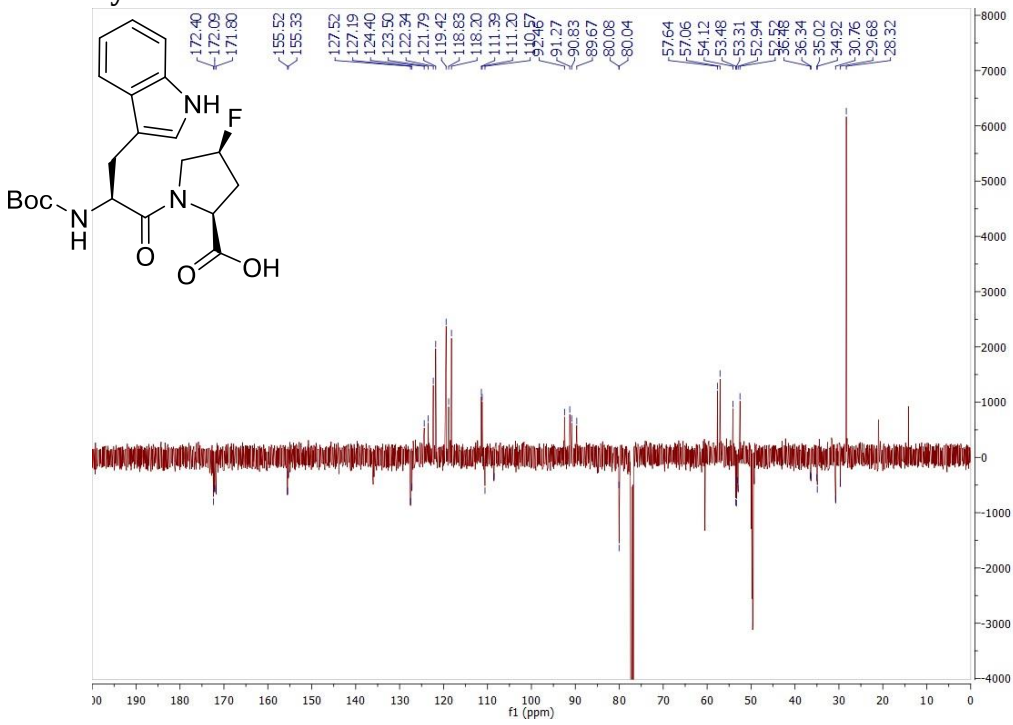


Figure D8 ^{13}C NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline) in $CDCl_3$ with ethyl acetate as a cosolvent.

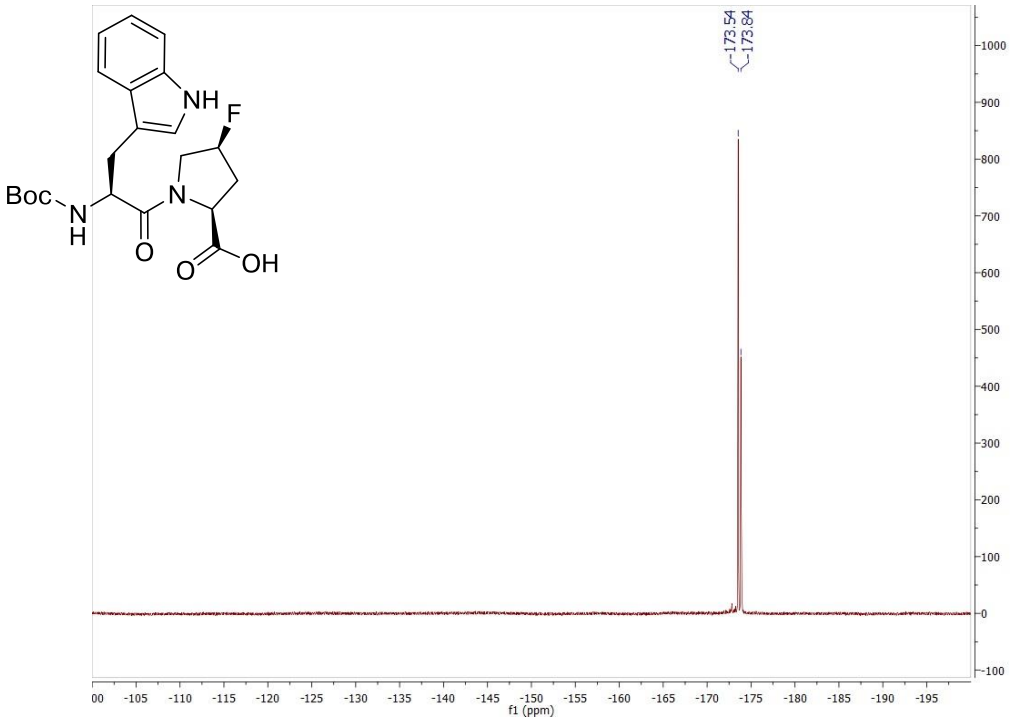


Figure D9 ^{19}F NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline) in $CDCl_3$ with ethyl acetate as a cosolvent.

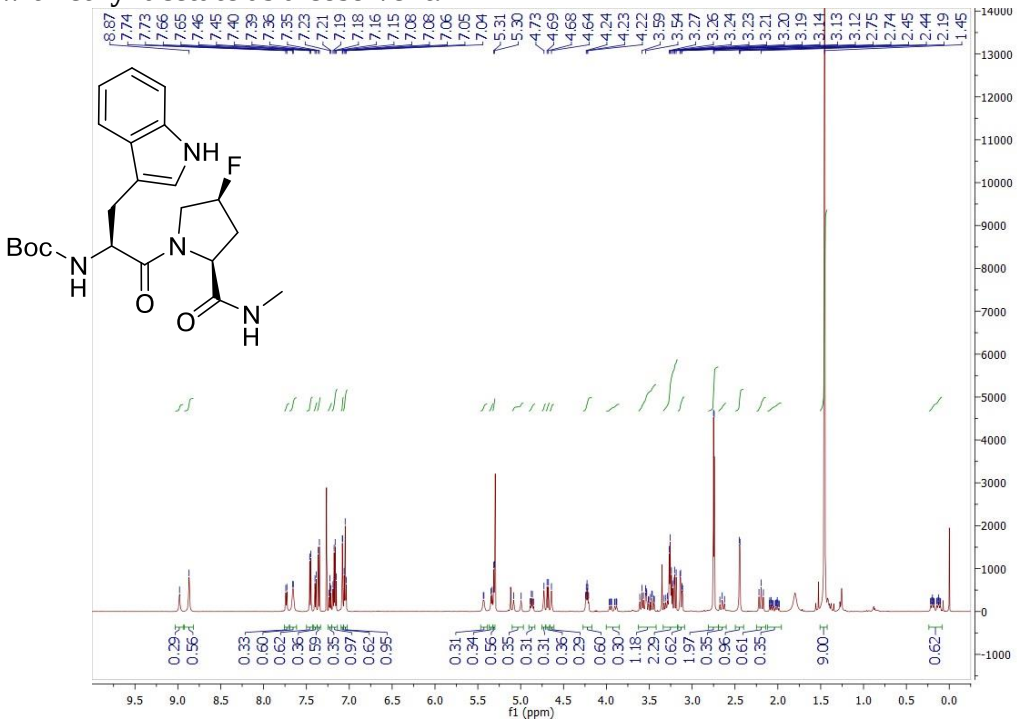


Figure D10 ^{1}H NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl amide in $CDCl_3$.

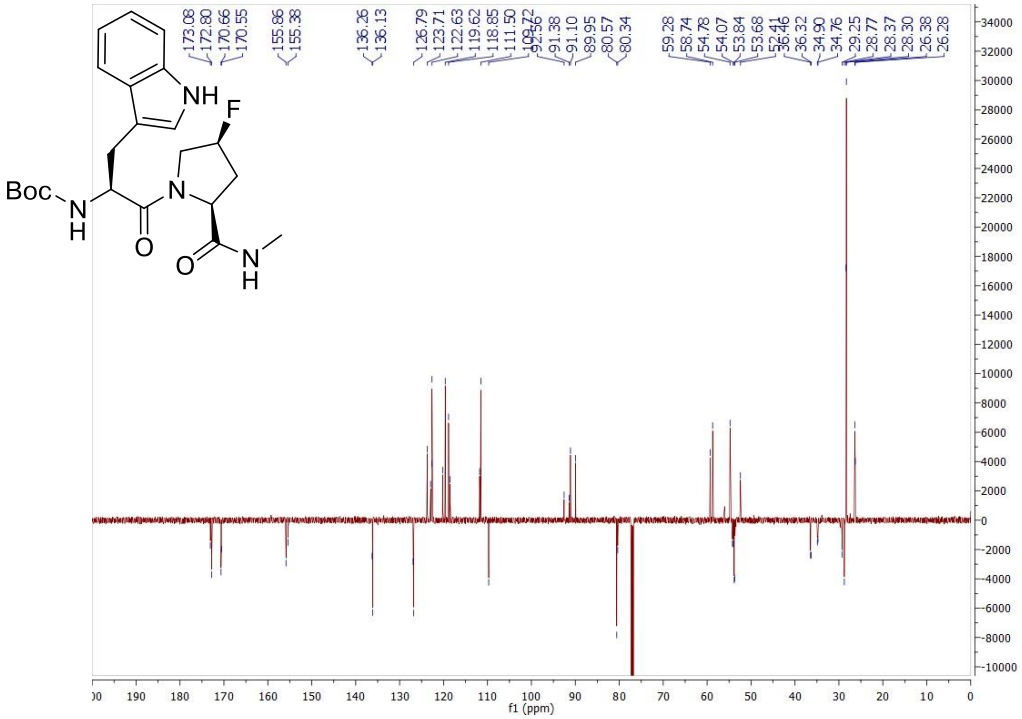


Figure D11 ^{13}C NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl amide in $CDCl_3$.

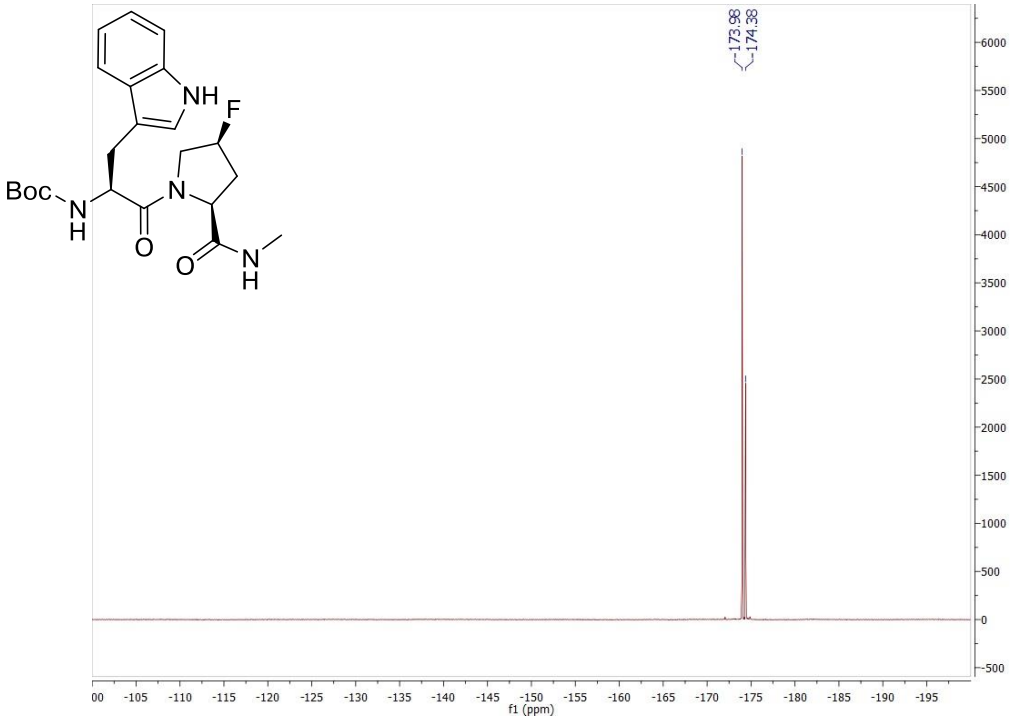


Figure D12 ^{19}F NMR spectrum of Boc-tryptophan-(4*S*-fluoroproline)-methyl amide in $CDCl_3$.

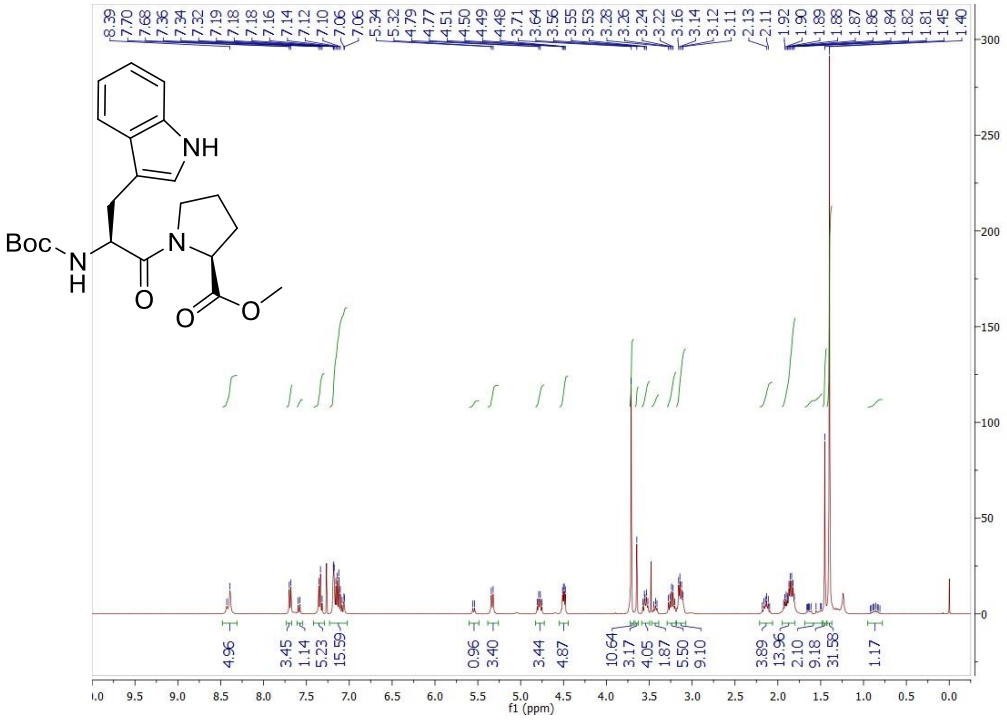


Figure D13 ^{1}H NMR spectrum of Boc-tryptophan-proline-methyl ester in $CDCl_3$.

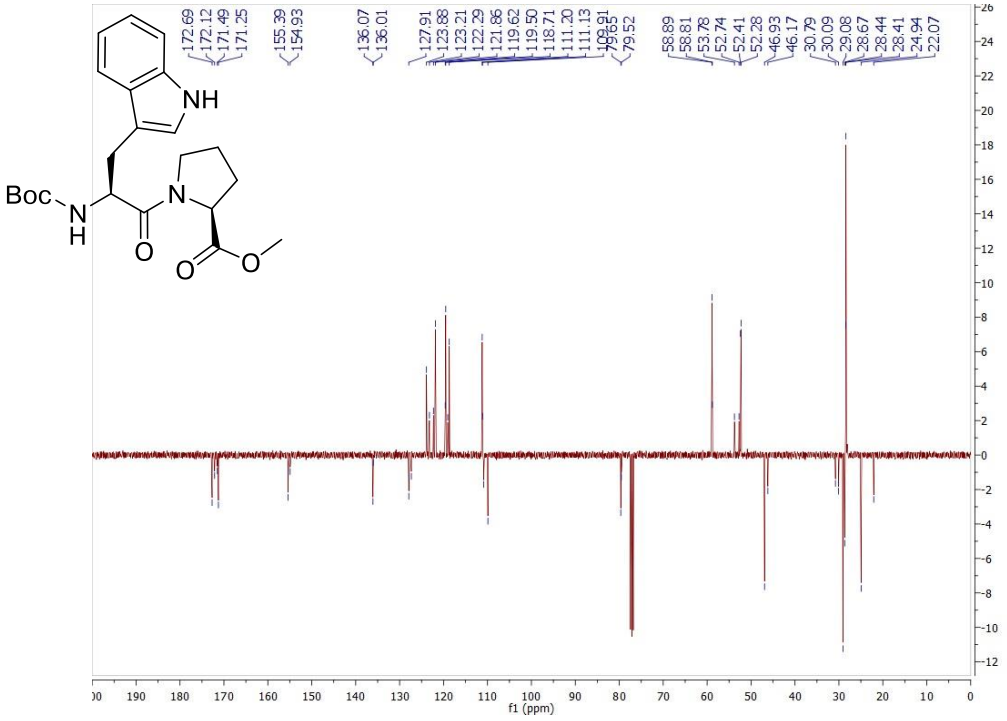


Figure D14 ^{13}C NMR spectrum of Boc-tryptophan-proline-methyl ester in $CDCl_3$.

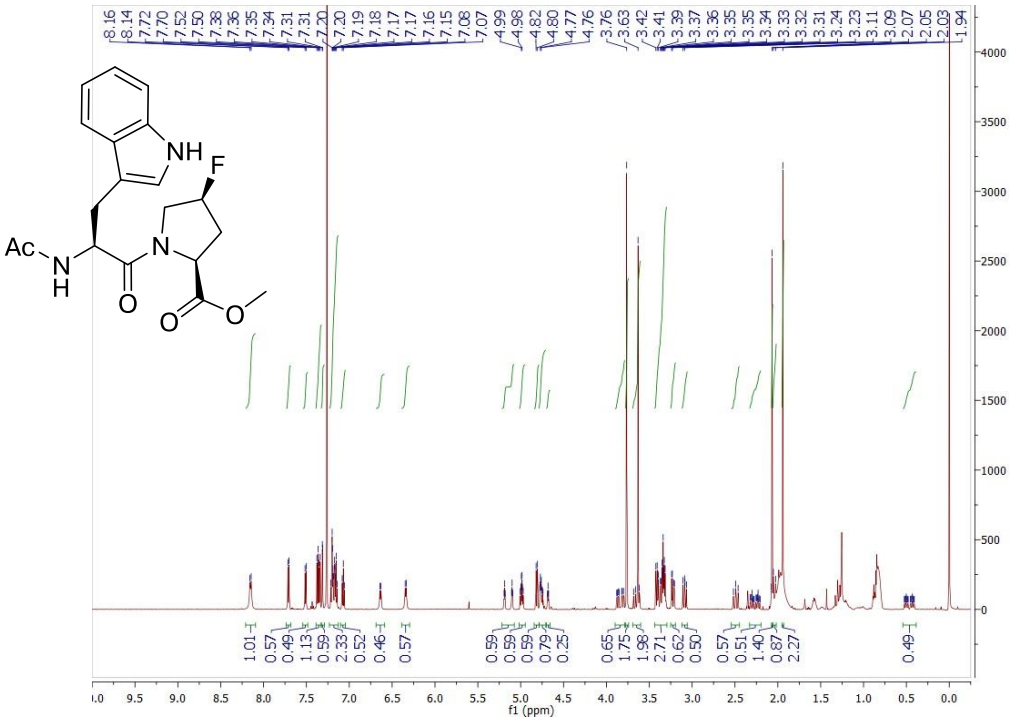


Figure D15 ^{1}H NMR spectrum of Ac-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

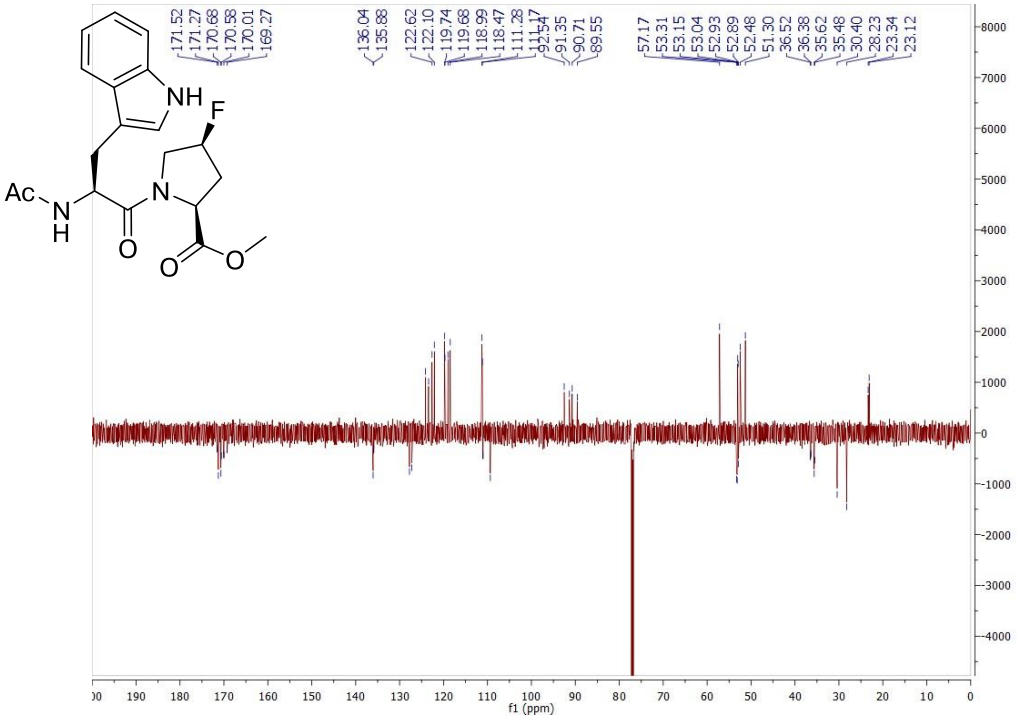


Figure D16 ^{13}C NMR spectrum of Ac-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

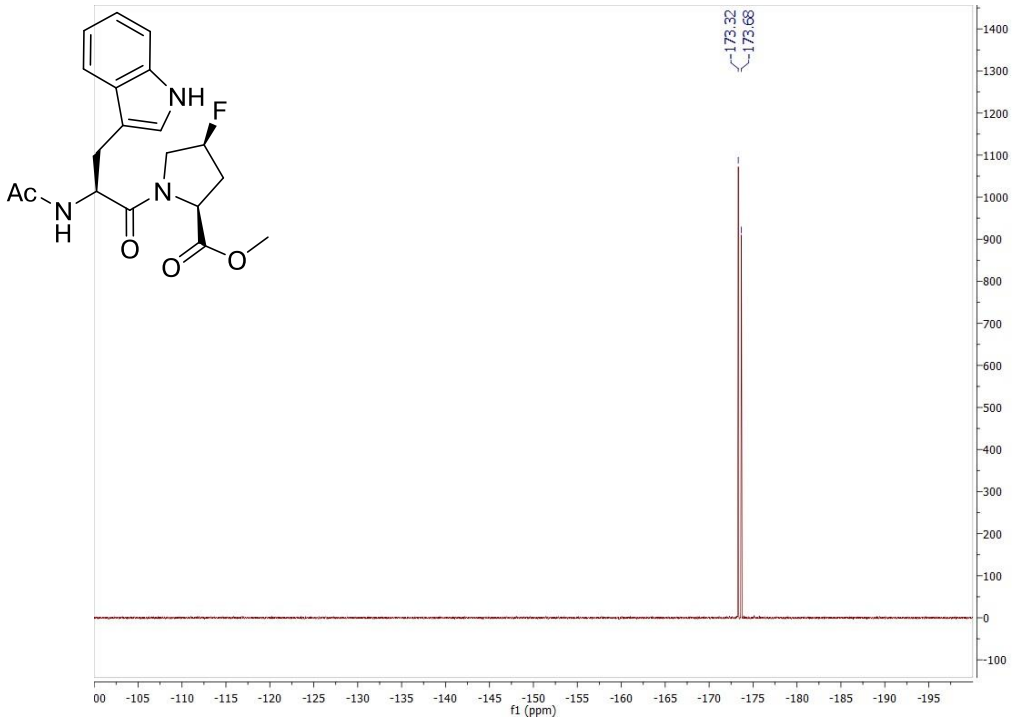


Figure D17 ^{19}F NMR spectrum of Ac-tryptophan-(4*S*-fluoroproline)-methyl ester in $CDCl_3$.

Appendix E

X-RAY CRYSTALLOGRAPHIC INFORMATION FOR DIPEPTIDES

Compounds were described and utilized in Chapter 3: Electronic Control of Proline Cis-Trans Isomerism via a C–H/π Aromatic Interaction: Insights into the Nature of C–H/π Interactions.

Crystal Structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2

Crystals were mounted using viscous oil onto a plastic mesh and cooled to the data collection temperature. Data were collected on a Bruker-AXS APEX II DUO CCD diffractometer with Cu-Kα radiation (λ = 1.54178 Å) micro focused with Goebel mirrors. Unit cell parameters were obtained from 36 data frames, 0.5º ω, from three different sections of the Ewald sphere. The systematic absences in the diffraction data are uniquely consistent for $P2_12_12_1$. The data-set was treated with multi-scan absorption corrections (Apex3 software suite, Madison, WI, 2015). The structure was solved using direct methods and refined with full-matrix, least-squares procedures on F^2 (Sheldrick, G.M. 2008. Acta Cryst. A64, 112-122). Refinement of the absolute structure parameter yielded nil indicating the true hand of the data has been determined. All non-hydrogen atoms were refined with anisotropic displacement parameters with rigid bond U_{eq} restraints. Phenyl groups were constrained to be idealized flat hexagonal rigid groups. H-atoms were placed in calculated positions with U_{iso} equal to 1.2 (1.5 for methyl H) U_{eq} of the attached atom. Atomic scattering factors are contained in the SHELXTL program library (Sheldrick, G., *op. cit*.). The compound consistently crystallizes with a thin plate habit and the data presented represent the best of several trials. The CIFs have been deposited with the Cambridge Crystallographic Database under CCDC 1442050.

Crystal Structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2

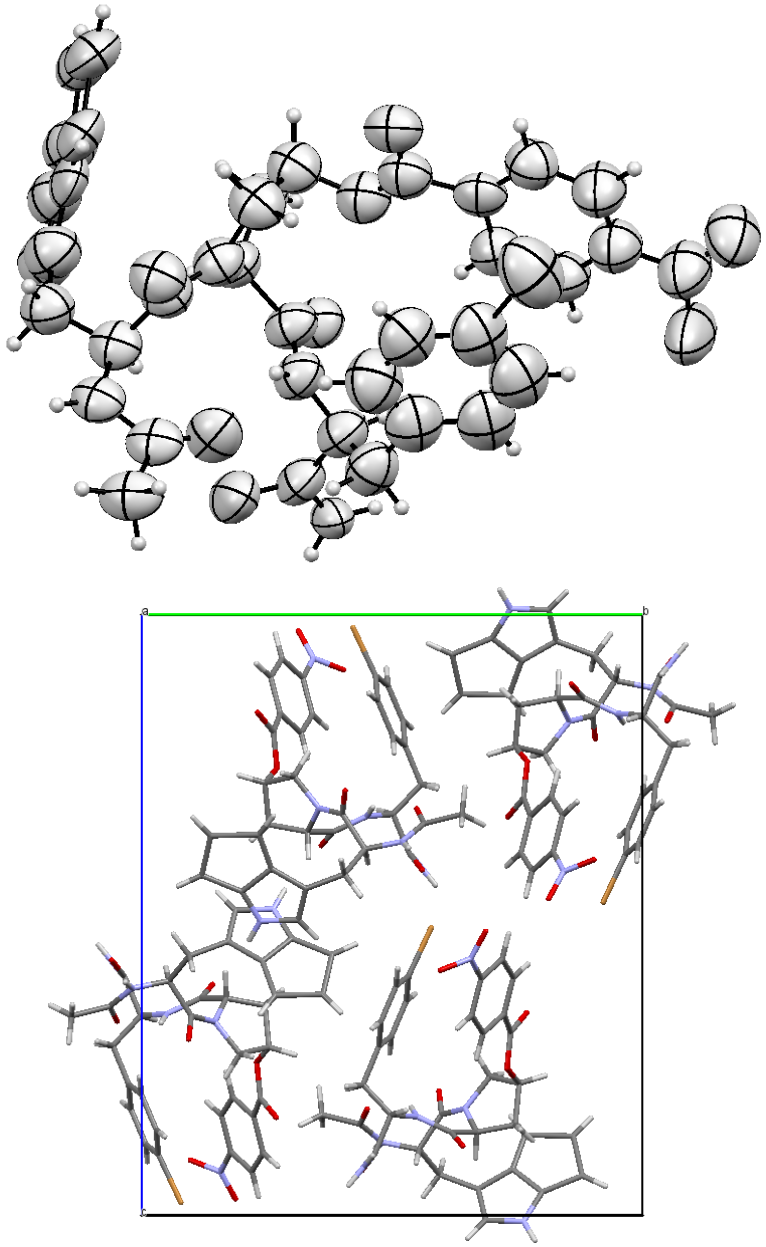


Figure E1 Crystal structure of Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.
Top: ORTEP diagram of the Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2 monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature from a solution of 60% methanol in chloroform.

Table E1. Crystallographic data and refinement details for Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

empirical formula	$C_{34}H_{33}BrN_6O_8$	
formula weight	733.57	
T (K)	200(2)	
wavelength (Å)	1.54178	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 8.8418(5)	α = 90
	b = 17.9487(9)	β = 90
	c = 21.6358(11)	γ = 90
Volume ($Å^3$)	3433.6(3)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.419	
absorption coefficient (mm^{-1})	2.138	
F(000)	1512	
crystal size (mm)	0.163 x 0.112 x 0.038	
θ range for data collection	4.086 to 44.552°	
Index ranges	$-6 \le h \le 8$, $-16 \le k \le 16$, $-15 \le l \le 19$	
Reflections collected/ unique	8109/2667 [R(int) = 0.0810]	
Coverage of independent reflections	99.9%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7488 and 0.6327	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	2667 / 218 / 398	
Goodness-of-fit on F^2	1.192	
Final *R* indices	1105 data; I>2σ(I)	R1 = 0.1325, wR2 = 0.3283
	all data	R1 = 0.2191, wR2 = 0.3919
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.2000P)^2]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.521 and -0.582 $eÅ^{-3}$	
R.M.S. deviation from mean	0.112 $eÅ^{-3}$	

Table E2. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
Br(1)	4689(8)	4217(4)	216(3)	179(3)
C(1)	6460(30)	5229(17)	1775(18)	131(15)
C(2)	6240(40)	4927(18)	1191(15)	152(17)
C(3)	4820(50)	4665(16)	1022(11)	132(14)
C(4)	3620(30)	4705(17)	1438(17)	125(14)
C(5)	3850(30)	5008(17)	2022(15)	136(15)
C(6)	5270(40)	5270(15)	2191(12)	123(13)
C(7)	5320(50)	5580(20)	2843(18)	110(12)
C(8)	5480(50)	5020(20)	3342(19)	106(11)
C(9)	5580(60)	5350(20)	3990(20)	103(12)
C(10)	4460(60)	3800(20)	3650(20)	93(11)
C(11)	1680(50)	3040(20)	2671(18)	113(12)
C(12)	2870(50)	2450(20)	2810(20)	109(12)
C(13)	3250(50)	2530(20)	3502(18)	110(12)
C(14)	2890(50)	3340(20)	3618(18)	94(11)
C(15)	4370(60)	2620(20)	1950(20)	99(12)
C(16)	6080(30)	2573(14)	1046(13)	115(13)
C(17)	7400(40)	2791(16)	747(9)	128(14)
C(18)	8470(30)	3216(16)	1059(14)	115(12)
C(19)	8230(30)	3422(14)	1670(13)	97(12)
C(20)	6910(40)	3204(15)	1969(9)	100(12)
C(21)	5830(30)	2780(15)	1657(13)	89(11)
C(22)	460(60)	4020(30)	3310(20)	102(12)
C(23)	210(40)	4460(20)	3897(17)	96(11)
C(24)	640(60)	5660(30)	3460(20)	98(13)
C(25)	-220(50)	6400(20)	3314(19)	127(16)
C(26)	-900(40)	4050(20)	4320(17)	106(12)
N(1)	6810(40)	5563(14)	4186(14)	88(10)
N(2)	4260(40)	4487(18)	3354(14)	96(9)
N(3)	1680(40)	3552(19)	3209(16)	97(9)
N(4)	9710(50)	3490(30)	730(20)	131(13)

Table E2 continued

N(5)	-360(40)	5197(19)	3766(14)	96(9)
N(6)	640(20)	2367(16)	5126(9)	128(12)
C(27)	400(30)	3119(15)	5063(11)	117(13)
C(28)	-390(30)	3271(11)	4539(12)	111(12)
C(29)	-662(19)	2574(10)	4240(8)	122(13)
C(30)	-1360(30)	2361(15)	3689(9)	131(15)
C(31)	-1360(30)	1620(16)	3528(11)	165(17)
C(32)	-690(30)	1086(12)	3902(14)	162(18)
C(33)	-20(30)	1274(10)	4452(13)	141(16)
C(34)	-2(17)	2020(11)	4619(9)	130(14)
O(1)	4230(30)	5600(15)	4250(16)	139(11)
O(2)	5620(30)	3586(14)	3844(12)	102(9)
O(3)	4320(30)	2695(14)	2555(13)	109(8)
O(4)	3390(40)	2303(15)	1628(14)	116(10)
O(5)	10320(40)	4060(20)	917(16)	201(18)
O(6)	10210(40)	3140(20)	286(16)	156(12)
O(7)	-470(30)	4047(15)	2920(13)	109(9)
O(8)	1790(40)	5484(16)	3245(14)	123(12)

Table E3. Bond lengths [Å] and angles [°] for Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

Bond lengths

Br(1)-C(3)	1.92(2)	C(18)-C(19)	1.3900
C(1)-C(2)	1.3900	C(18)-N(4)	1.40(5)
C(1)-C(6)	1.3900	C(19)-C(20)	1.3900
C(1)-H(1)	0.9500	C(19)-H(19)	0.9500
C(2)-C(3)	1.3900	C(20)-C(21)	1.3900
C(2)-H(2)	0.9500	C(20)-H(20)	0.9500
C(3)-C(4)	1.3900	C(22)-O(7)	1.18(4)
C(4)-C(5)	1.3900	C(22)-N(3)	1.38(5)
C(4)-H(4)	0.9500	C(22)-C(23)	1.51(5)
C(5)-C(6)	1.3900	C(23)-N(5)	1.45(4)
C(5)-H(5)	0.9500	C(23)-C(26)	1.53(4)
C(6)-C(7)	1.52(4)	C(23)-H(23)	1.0000
C(7)-C(8)	1.49(5)	C(24)-O(8)	1.16(5)
C(7)-H(7A)	0.9900	C(24)-N(5)	1.38(5)
C(7)-H(7B)	0.9900	C(24)-C(25)	1.57(5)
C(8)-N(2)	1.44(4)	C(25)-H(25A)	0.9800
C(8)-C(9)	1.52(5)	C(25)-H(25B)	0.9800
C(8)-H(8)	1.0000	C(25)-H(25C)	0.9800
C(9)-N(1)	1.23(5)	C(26)-C(28)	1.54(4)
C(9)-O(1)	1.39(5)	C(26)-H(26A)	0.9900
C(10)-O(2)	1.18(5)	C(26)-H(26B)	0.9900
C(10)-N(2)	1.40(5)	N(1)-H(1A)	0.8800
C(10)-C(14)	1.62(6)	N(1)-H(1B)	0.8800
C(11)-N(3)	1.48(4)	N(2)-H(2A)	0.8800
C(11)-C(12)	1.53(5)	N(4)-O(6)	1.22(4)
C(11)-H(11A)	0.9900	N(4)-O(5)	1.22(4)
C(11)-H(11B)	0.9900	N(5)-H(5A)	0.8800
C(12)-O(3)	1.47(5)	N(6)-C(27)	1.3755
C(12)-C(13)	1.53(5)	N(6)-C(34)	1.3848
C(12)-H(12)	1.0000	N(6)-H(6)	0.8800
C(13)-C(14)	1.51(5)	C(27)-C(28)	1.3593
C(13)-H(13A)	0.9900	C(27)-H(27)	0.9500

Table E3 continued

C(13)-H(13B)	0.9900	C(28)-C(29)	1.4292
C(14)-N(3)	1.44(4)	C(29)-C(30)	1.3961
C(14)-H(14)	1.0000	C(29)-C(34)	1.4143
C(15)-O(4)	1.25(5)	C(30)-C(31)	1.3746
C(15)-O(3)	1.32(5)	C(30)-H(30)	0.9500
C(15)-C(21)	1.47(5)	C(31)-C(32)	1.3872
C(16)-C(17)	1.3900	C(31)-H(31)	0.9500
C(16)-C(21)	1.3900	C(32)-C(33)	1.3733
C(16)-H(16)	0.9500	C(32)-H(32)	0.9500
C(17)-C(18)	1.3900	C(33)-C(34)	1.3870
C(17)-H(17)	0.9500	C(33)-H(33)	0.9500

Table E3 continued

Bond angles

C(2)-C(1)-C(6)	120.0	C(20)-C(19)-H(19)	120.0
C(2)-C(1)-H(1)	120.0	C(19)-C(20)-C(21)	120.0
C(6)-C(1)-H(1)	120.0	C(19)-C(20)-H(20)	120.0
C(1)-C(2)-C(3)	120.0	C(21)-C(20)-H(20)	120.0
C(1)-C(2)-H(2)	120.0	C(20)-C(21)-C(16)	120.0
C(3)-C(2)-H(2)	120.0	C(20)-C(21)-C(15)	120(3)
C(4)-C(3)-C(2)	120.0	C(16)-C(21)-C(15)	120(3)
C(4)-C(3)-Br(1)	124(3)	O(7)-C(22)-N(3)	118(4)
C(2)-C(3)-Br(1)	116(3)	O(7)-C(22)-C(23)	118(5)
C(3)-C(4)-C(5)	120.0	N(3)-C(22)-C(23)	124(5)
C(3)-C(4)-H(4)	120.0	N(5)-C(23)-C(26)	109(3)
C(5)-C(4)-H(4)	120.0	N(5)-C(23)-C(22)	111(3)
C(4)-C(5)-C(6)	120.0	C(26)-C(23)-C(22)	110(3)
C(4)-C(5)-H(5)	120.0	N(5)-C(23)-H(23)	108.7
C(6)-C(5)-H(5)	120.0	C(26)-C(23)-H(23)	108.7
C(5)-C(6)-C(1)	120.0	C(22)-C(23)-H(23)	108.7
C(5)-C(6)-C(7)	113(3)	O(8)-C(24)-N(5)	127(5)
C(1)-C(6)-C(7)	127(3)	O(8)-C(24)-C(25)	124(5)
C(8)-C(7)-C(6)	115(3)	N(5)-C(24)-C(25)	107(4)
C(8)-C(7)-H(7A)	108.4	C(24)-C(25)-H(25A)	109.5
C(6)-C(7)-H(7A)	108.4	C(24)-C(25)-H(25B)	109.5
C(8)-C(7)-H(7B)	108.4	H(25A)-C(25)-H(25B)	109.5
C(6)-C(7)-H(7B)	108.4	C(24)-C(25)-H(25C)	109.5
H(7A)-C(7)-H(7B)	107.5	H(25A)-C(25)-H(25C)	109.5
N(2)-C(8)-C(7)	113(4)	H(25B)-C(25)-H(25C)	109.5
N(2)-C(8)-C(9)	107(4)	C(23)-C(26)-C(28)	115(3)
C(7)-C(8)-C(9)	114(3)	C(23)-C(26)-H(26A)	108.4
N(2)-C(8)-H(8)	107.6	C(28)-C(26)-H(26A)	108.4
C(7)-C(8)-H(8)	107.6	C(23)-C(26)-H(26B)	108.4
C(9)-C(8)-H(8)	107.6	C(28)-C(26)-H(26B)	108.4
N(1)-C(9)-O(1)	121(4)	H(26A)-C(26)-H(26B)	107.5
N(1)-C(9)-C(8)	120(5)	C(9)-N(1)-H(1A)	120.0
O(1)-C(9)-C(8)	117(5)	C(9)-N(1)-H(1B)	120.0
O(2)-C(10)-N(2)	124(4)	H(1A)-N(1)-H(1B)	120.0
O(2)-C(10)-C(14)	127(4)	C(10)-N(2)-C(8)	120(3)
N(2)-C(10)-C(14)	109(4)	C(10)-N(2)-H(2A)	120.1
N(3)-C(11)-C(12)	106(3)	C(8)-N(2)-H(2A)	120.1

Table E3 continued

N(3)-C(11)-H(11A)	110.6	C(22)-N(3)-C(14)	130(4)
C(12)-C(11)-H(11A)	110.6	C(22)-N(3)-C(11)	120(4)
N(3)-C(11)-H(11B)	110.6	C(14)-N(3)-C(11)	109(3)
C(12)-C(11)-H(11B)	110.6	O(6)-N(4)-O(5)	122(5)
H(11A)-C(11)-H(11B)	108.7	O(6)-N(4)-C(18)	120(4)
O(3)-C(12)-C(11)	108(4)	O(5)-N(4)-C(18)	118(4)
O(3)-C(12)-C(13)	99(3)	C(24)-N(5)-C(23)	115(4)
C(11)-C(12)-C(13)	107(4)	C(24)-N(5)-H(5A)	122.7
O(3)-C(12)-H(12)	113.9	C(23)-N(5)-H(5A)	122.7
C(11)-C(12)-H(12)	113.9	C(27)-N(6)-C(34)	107.3
C(13)-C(12)-H(12)	113.9	C(27)-N(6)-H(6)	126.4
C(14)-C(13)-C(12)	102(3)	C(34)-N(6)-H(6)	126.4
C(14)-C(13)-H(13A)	111.4	C(28)-C(27)-N(6)	111.1
C(12)-C(13)-H(13A)	111.4	C(28)-C(27)-H(27)	124.4
C(14)-C(13)-H(13B)	111.4	N(6)-C(27)-H(27)	124.4
C(12)-C(13)-H(13B)	111.4	C(27)-C(28)-C(29)	106.7
H(13A)-C(13)-H(13B)	109.3	C(27)-C(28)-C(26)	126(2)
N(3)-C(14)-C(13)	108(3)	C(29)-C(28)-C(26)	127(2)
N(3)-C(14)-C(10)	122(4)	C(30)-C(29)-C(34)	119.0
C(13)-C(14)-C(10)	109(4)	C(30)-C(29)-C(28)	134.5
N(3)-C(14)-H(14)	105.8	C(34)-C(29)-C(28)	106.5
C(13)-C(14)-H(14)	105.8	C(31)-C(30)-C(29)	118.8
C(10)-C(14)-H(14)	105.8	C(31)-C(30)-H(30)	120.6
O(4)-C(15)-O(3)	125(5)	C(29)-C(30)-H(30)	120.6
O(4)-C(15)-C(21)	118(4)	C(30)-C(31)-C(32)	121.3
O(3)-C(15)-C(21)	116(4)	C(30)-C(31)-H(31)	119.3
C(17)-C(16)-C(21)	120.0	C(32)-C(31)-H(31)	119.3
C(17)-C(16)-H(16)	120.0	C(33)-C(32)-C(31)	121.4
C(21)-C(16)-H(16)	120.0	C(33)-C(32)-H(32)	119.3
C(16)-C(17)-C(18)	120.0	C(31)-C(32)-H(32)	119.3
C(16)-C(17)-H(17)	120.0	C(32)-C(33)-C(34)	117.8
C(18)-C(17)-H(17)	120.0	C(32)-C(33)-H(33)	121.1
C(19)-C(18)-C(17)	120.0	C(34)-C(33)-H(33)	121.1
C(19)-C(18)-N(4)	121(3)	N(6)-C(34)-C(33)	130.1
C(17)-C(18)-N(4)	119(3)	N(6)-C(34)-C(29)	108.3
C(18)-C(19)-C(20)	120.0	C(33)-C(34)-C(29)	121.6
C(18)-C(19)-H(19)	120.0	C(15)-O(3)-C(12)	112(4)

Table E4. Anisotropic atomic displacement parameters (Å2) for Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Br(1)	225(7)	189(6)	125(4)	12(4)	11(5)	37(6)
C(1)	150(30)	110(30)	130(30)	60(30)	0(30)	10(30)
C(2)	170(40)	160(30)	130(30)	40(30)	0(30)	0(30)
C(3)	160(40)	170(30)	70(20)	20(20)	-10(30)	0(30)
C(4)	140(30)	130(30)	110(30)	50(30)	-30(30)	-20(30)
C(5)	150(30)	170(30)	90(30)	10(30)	-30(30)	30(30)
C(6)	120(30)	130(30)	120(30)	40(20)	-20(30)	20(30)
C(7)	100(30)	90(20)	140(30)	10(20)	-20(30)	0(20)
C(8)	110(30)	100(30)	110(30)	30(20)	-10(30)	-10(20)
C(9)	90(30)	110(30)	110(30)	20(20)	0(30)	-10(30)
C(10)	90(30)	60(20)	130(30)	0(20)	0(30)	-30(20)
C(11)	100(30)	140(30)	100(30)	0(30)	-20(30)	0(30)
C(12)	90(30)	120(30)	120(30)	10(20)	-10(30)	-10(20)
C(13)	110(30)	100(30)	120(30)	20(30)	-20(30)	-10(20)
C(14)	100(30)	90(20)	90(20)	-10(20)	0(20)	-20(20)
C(15)	110(30)	80(20)	110(30)	-10(30)	0(30)	0(20)
C(16)	120(30)	110(30)	110(30)	-10(20)	20(30)	10(20)
C(17)	110(30)	140(30)	140(30)	-30(30)	20(30)	-10(30)
C(18)	110(30)	110(30)	120(30)	10(30)	20(30)	-20(20)
C(19)	110(30)	90(20)	90(30)	-10(20)	10(20)	-10(20)
C(20)	90(30)	100(30)	120(30)	10(20)	0(20)	0(20)
C(21)	120(30)	70(20)	80(20)	10(20)	30(20)	-10(20)
C(22)	80(30)	150(30)	80(30)	-20(30)	20(20)	0(30)
C(23)	80(20)	120(30)	80(20)	0(20)	-30(20)	10(20)
C(24)	90(30)	80(30)	120(30)	-10(20)	-60(30)	-10(30)
C(25)	130(40)	80(30)	170(40)	20(30)	-70(30)	0(30)
C(26)	100(30)	130(30)	90(20)	10(20)	0(20)	20(20)
N(1)	60(20)	60(20)	140(30)	-53(19)	0(20)	-9(16)
N(2)	90(20)	80(20)	120(20)	-20(20)	3(19)	-16(18)

Table E4 continued

N(3)	80(20)	120(20)	90(20)	-1(19)	-20(20)	-15(19)
N(4)	110(30)	160(30)	130(30)	-30(30)	10(30)	-40(30)
N(5)	70(20)	90(20)	130(20)	-20(20)	-10(20)	-10(19)
N(6)	100(20)	130(30)	150(30)	10(20)	-10(20)	0(20)
C(27)	100(30)	160(30)	90(30)	0(30)	10(20)	0(30)
C(28)	90(30)	140(30)	110(30)	20(20)	-20(20)	10(30)
C(29)	70(20)	130(30)	170(30)	10(30)	10(20)	-20(20)
C(30)	110(30)	90(30)	190(40)	0(30)	10(30)	-50(20)
C(31)	160(30)	100(30)	240(40)	-40(30)	-30(30)	-50(30)
C(32)	140(40)	70(30)	270(50)	-10(30)	30(30)	-10(30)
C(33)	120(30)	60(20)	250(40)	30(30)	30(30)	10(20)
C(34)	120(30)	90(30)	190(30)	20(30)	0(30)	-10(20)
O(1)	90(20)	110(20)	220(30)	0(20)	0(20)	-1(19)
O(2)	70(19)	120(20)	120(20)	5(16)	-19(18)	24(16)
O(3)	110(20)	112(18)	109(19)	10(18)	-20(20)	-10(17)
O(4)	130(20)	90(20)	130(20)	-22(18)	-30(20)	-3(18)
O(5)	180(30)	260(40)	160(30)	-110(30)	80(30)	-120(30)
O(6)	120(30)	180(30)	170(30)	-30(30)	20(30)	-20(20)
O(7)	100(20)	110(20)	120(20)	-2(18)	-18(18)	11(18)
O(8)	100(20)	100(20)	160(30)	20(20)	20(20)	40(20)

Table E5. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Ac-Trp(2*S*,4*S*-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

	x/a	y/b	z/c	U(eq)
H(1)	7433	5409	1890	158
H(2)	7054	4899	907	183
H(4)	2649	4526	1323	150
H(5)	3028	5036	2306	163
H(7A)	6183	5931	2871	132
H(7B)	4385	5869	2917	132
H(8)	6436	4735	3264	127
H(11A)	1947	3317	2289	136
H(11B)	672	2813	2615	136
H(12)	2563	1936	2689	130
H(13A)	2607	2198	3759	131
H(13B)	4325	2418	3583	131
H(14)	2454	3364	4045	112
H(16)	5351	2283	833	138
H(17)	7572	2650	329	153
H(19)	8957	3712	1883	116
H(20)	6736	3345	2386	119
H(23)	1204	4505	4116	115
H(25A)	411	6827	3433	190
H(25B)	-1172	6416	3545	190
H(25C)	-432	6425	2870	190
H(26A)	-1874	3998	4100	127
H(26B)	-1081	4364	4689	127
H(1A)	6852	5875	4500	106
H(1B)	7654	5401	4015	106
H(2A)	3391	4595	3176	115
H(5A)	-1279	5340	3871	115
H(6)	1122	2146	5432	154
H(27)	734	3485	5349	141
H(30)	-1831	2721	3431	157
H(31)	-1835	1471	3153	198
H(32)	-700	580	3775	194
H(33)	422	905	4710	169

Table E6. Torsion angles (°) for Ac-Trp(2*S*,*4*S-*p*-nitrobenzoate-hyp)(4-Br-Phe)-NH_2.

C(1)-C(2)-C(3)-C(4)	0.0	C(14)-C(10)-N(2)-C(8)	177(3)
C(1)-C(2)-C(3)-Br(1)	-177.1(18)	C(14)-C(10)-N(2)-C(8)	160(3)
C(2)-C(3)-C(4)-C(5)	0.0	C(9)-C(8)-N(2)-C(10)	-75(5)
Br(1)-C(3)-C(4)-C(5)	177(2)	O(7)-C(22)-N(3)-C(14)	-173(4)
C(3)-C(4)-C(5)-C(6)	0.0	C(23)-C(22)-N(3)-C(14)	4(7)
C(4)-C(5)-C(6)-C(1)	0.0	O(7)-C(22)-N(3)-C(11)	-8(6)
C(4)-C(5)-C(6)-C(7)	-180(3)	C(23)-C(22)-N(3)-C(11)	169(4)
C(2)-C(1)-C(6)-C(5)	0.0	C(13)-C(14)-N(3)-C(22)	142(4)
C(2)-C(1)-C(6)-C(7)	180(3)	C(10)-C(14)-N(3)-C(22)	-91(5)
C(5)-C(6)-C(7)-C(8)	81(4)	C(13)-C(14)-N(3)-C(11)	-24(4)
C(1)-C(6)-C(7)-C(8)	-99(4)	C(10)-C(14)-N(3)-C(11)	103(4)
C(6)-C(7)-C(8)-N(2)	-60(5)	C(12)-C(11)-N(3)-C(22)	-161(4)
C(6)-C(7)-C(8)-C(9)	178(4)	C(12)-C(11)-N(3)-C(14)	6(5)
N(2)-C(8)-C(9)-N(1)	150(4)	C(19)-C(18)-N(4)-O(6)	156(3)
C(7)-C(8)-C(9)-N(1)	-85(5)	C(17)-C(18)-N(4)-O(6)	-31(5)
N(2)-C(8)-C(9)-O(1)	-48(5)	C(19)-C(18)-N(4)-O(5)	-19(6)
C(7)-C(8)-C(9)-O(1)	77(5)	C(17)-C(18)-N(4)-O(5)	154(4)
N(3)-C(11)-C(12)-O(3)	-92(4)	O(8)-C(24)-N(5)-C(23)	10(6)
N(3)-C(11)-C(12)-C(13)	13(5)	C(25)-C(24)-N(5)-C(23)	175(3)
O(3)-C(12)-C(13)-C(14)	86(4)	C(26)-C(23)-N(5)-C(24)	171(3)
C(11)-C(12)-C(13)-C(14)	-27(4)	C(22)-C(23)-N(5)-C(24)	-67(4)
C(12)-C(13)-C(14)-N(3)	31(4)	C(34)-N(6)-C(27)-C(28)	0.5
C(12)-C(13)-C(14)-C(10)	-103(4)	N(6)-C(27)-C(28)-C(29)	-0.6
O(2)-C(10)-C(14)-N(3)	-160(4)	N(6)-C(27)-C(28)-C(26)	179(3)
N(2)-C(10)-C(14)-N(3)	15(5)	C(23)-C(26)-C(28)-C(27)	91(3)
O(2)-C(10)-C(14)-C(13)	-34(6)	C(23)-C(26)-C(28)-C(29)	-90(3)
N(2)-C(10)-C(14)-C(13)	141(3)	C(27)-C(28)-C(29)-C(30)	-178.2
C(21)-C(16)-C(17)-C(18)	0.0	C(26)-C(28)-C(29)-C(30)	2(3)
C(16)-C(17)-C(18)-C(19)	0.0	C(27)-C(28)-C(29)-C(34)	0.4
C(16)-C(17)-C(18)-N(4)	-173(3)	C(26)-C(28)-C(29)-C(34)	-179(3)
C(17)-C(18)-C(19)-C(20)	0.0	C(34)-C(29)-C(30)-C(31)	-1.2

Table E6 continued

N(4)-C(18)-C(19)-C(20)	173(3)	C(28)-C(29)-C(30)-C(31)	177.2
C(18)-C(19)-C(20)-C(21)	0.0	C(29)-C(30)-C(31)-C(32)	0.3
C(19)-C(20)-C(21)-C(16)	0.0	C(30)-C(31)-C(32)-C(33)	0.9
C(19)-C(20)-C(21)-C(15)	-173(3)	C(31)-C(32)-C(33)-C(34)	-1.2
C(17)-C(16)-C(21)-C(20)	0.0	C(27)-N(6)-C(34)-C(33)	177.4
C(17)-C(16)-C(21)-C(15)	173(3)	C(27)-N(6)-C(34)-C(29)	-0.3
O(4)-C(15)-C(21)-C(20)	173(3)	C(32)-C(33)-C(34)-N(6)	-177.1
O(3)-C(15)-C(21)-C(20)	-20(4)	C(32)-C(33)-C(34)-C(29)	0.3
O(4)-C(15)-C(21)-C(16)	0(5)	C(30)-C(29)-C(34)-N(6)	178.7
O(3)-C(15)-C(21)-C(16)	167(3)	C(28)-C(29)-C(34)-N(6)	-0.1
O(7)-C(22)-C(23)-N(5)	-43(6)	C(30)-C(29)-C(34)-C(33)	0.8
N(3)-C(22)-C(23)-N(5)	140(4)	C(28)-C(29)-C(34)-C(33)	-178.0
O(7)-C(22)-C(23)-C(26)	78(5)	O(4)-C(15)-O(3)-C(12)	-8(6)
N(3)-C(22)-C(23)-C(26)	-98(5)	C(21)-C(15)-O(3)-C(12)	-174(3)
N(5)-C(23)-C(26)-C(28)	-176(3)	C(11)-C(12)-O(3)-C(15)	-75(4)
C(22)-C(23)-C(26)-C(28)	61(5)	C(13)-C(12)-O(3)-C(15)	174(3)

Crystal Structure of Ac-Trpflp-OMe

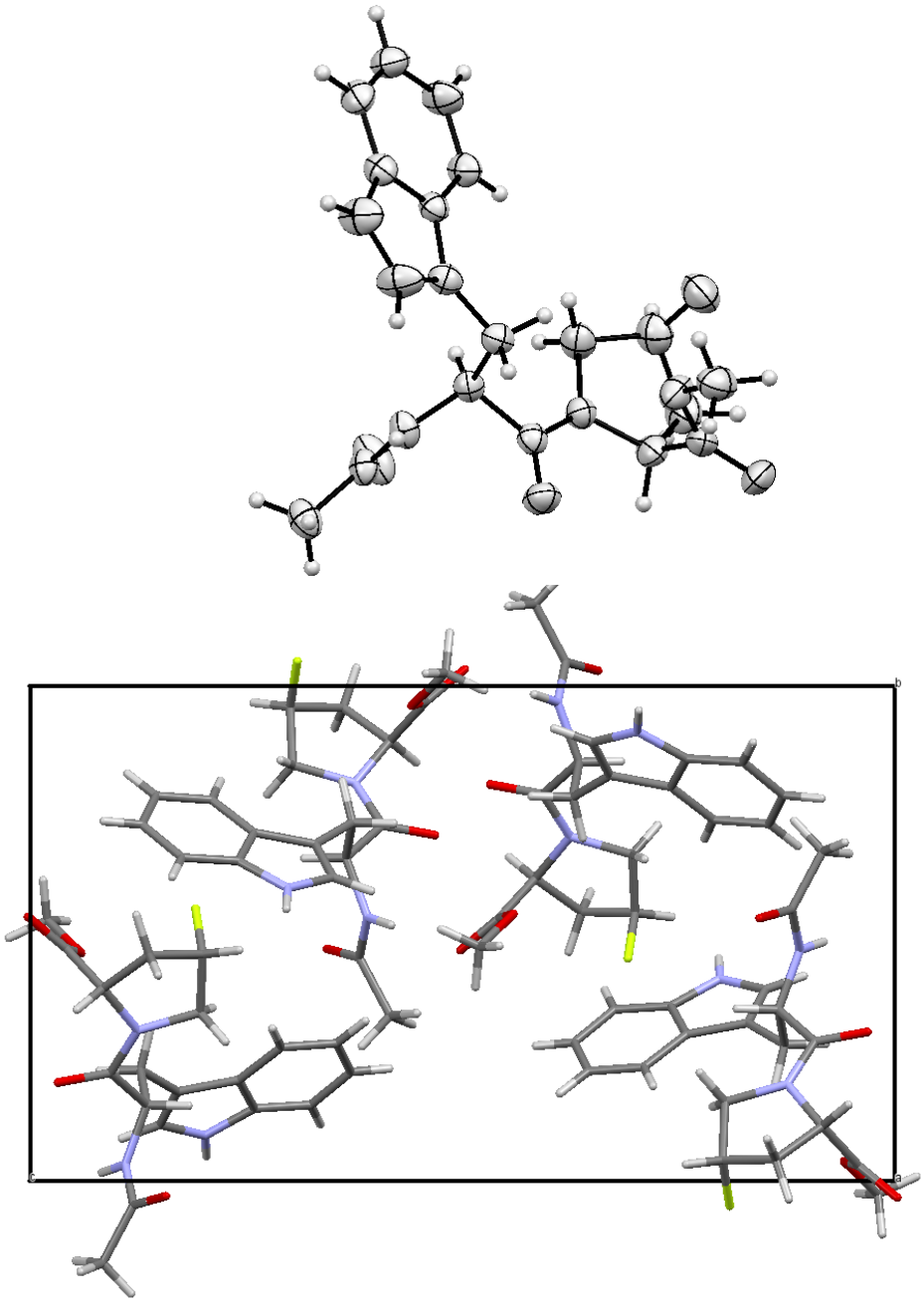


Figure E2 Crystal structure of Ac-Trpflp-OMe.
Top: ORTEP diagram of the Ac-Trpflp-OMe monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing.

Table E7. Crystallographic data and refinement details for Ac-Trpflp-OMe.

empirical formula	$C_{19}H_{22}FN_3O_4$	
formula weight	375.39	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 8.3447(11)	α = 90
	b = 11.2429(15)	β = 90
	c = 19.688(3)	γ = 90
Volume ($Å^3$)	1847.1(4)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.350	
absorption coefficient (mm^{-1})	0.102	
F(000)	792	
crystal size (mm)	0.256 x 0.307 x 0.495	
θ range for data collection	2.069 to 27.543°	
Index ranges	$-10 \le h \le 9$, $-14 \le k \le 14$, $-24 \le l \le 25$	
Reflections collected/ unique	14175/4262 [R(int) = 0.0466]	
Coverage of independent reflections	99.9%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7456 and 0.6569	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	4262 / 0 / 252	
Goodness-of-fit on F^2	1.014	
Final *R* indices	3160 data; $I>2\sigma(I)$	R1 = 0.0486, wR2 = 0.1025
	all data	R1 = 0.0714, wR2 = 0.1127
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0495P)^2+0.1409P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.170 and -0.185 $eÅ^{-3}$	
R.M.S. deviation from mean	0.040 $eÅ^{-3}$	

Table E8. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Ac-Trpflp-OMe.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
C1	0.0321(4)	0.6135(4)	0.64990(18)	0.0532(10)
C2	0.9057(4)	0.6806(3)	0.67027(15)	0.0356(7)
C3	0.9335(4)	0.7039(2)	0.74101(15)	0.0321(6)
C4	0.8530(4)	0.7701(3)	0.79033(16)	0.0399(7)
C5	0.9171(4)	0.7781(3)	0.85448(17)	0.0492(9)
C6	0.0598(4)	0.7201(3)	0.87088(17)	0.0459(8)
C7	0.1412(4)	0.6535(3)	0.82433(17)	0.0432(8)
C8	0.0769(4)	0.6465(3)	0.75920(16)	0.0362(7)
C9	0.7706(4)	0.7246(3)	0.62737(15)	0.0356(7)
C10	0.6140(4)	0.6538(2)	0.63490(15)	0.0324(6)
C11	0.5208(4)	0.4512(3)	0.61949(16)	0.0359(7)
C12	0.5408(4)	0.3341(3)	0.58357(18)	0.0454(8)
C13	0.4865(4)	0.7184(3)	0.59276(16)	0.0334(7)
C14	0.4124(4)	0.8403(3)	0.69658(15)	0.0430(8)
C15	0.3233(4)	0.9580(3)	0.69943(18)	0.0482(8)
C16	0.2150(4)	0.9563(3)	0.63790(18)	0.0520(9)
C17	0.3110(4)	0.8890(3)	0.58407(17)	0.0415(8)
C18	0.4204(4)	0.9692(3)	0.54202(15)	0.0413(8)
C19	0.6885(5)	0.0266(3)	0.51961(18)	0.0527(9)
F1	0.4353(3)	0.05043(17)	0.69118(10)	0.0588(6)
N1	0.1346(4)	0.5917(3)	0.70213(16)	0.0552(8)
N2	0.6334(3)	0.5332(2)	0.60991(12)	0.0352(6)
N3	0.4028(3)	0.8043(2)	0.62447(12)	0.0346(6)
O1	0.4059(3)	0.4692(2)	0.65667(13)	0.0582(7)
O2	0.4710(3)	0.6989(2)	0.53199(11)	0.0456(6)
O3	0.3713(3)	0.0336(2)	0.49754(12)	0.0576(7)
O4	0.5731(3)	0.9602(2)	0.55887(11)	0.0438(5)

Table E9. Bond lengths [Å] and angles [°] for Ac-Trpflp-OMe.

Bond lengths

C1-C2	1.358(4)	C1-N1	1.360(5)
C1-H1	0.93	C2-C3	1.436(4)
C2-C9	1.492(4)	C3-C4	1.395(4)
C3-C8	1.407(4)	C4-C5	1.375(5)
C4-H4	0.93	C5-C6	1.395(5)
C5-H5	0.93	C6-C7	1.365(5)
C6-H6	0.93	C7-C8	1.392(4)
C7-H7	0.93	C8-N1	1.369(4)
C9-C10	1.537(4)	C9-H9A	0.97
C9-H9B	0.97	C10-N2	1.451(4)
C10-C13	1.533(4)	C10-H10	0.98
C11-O1	1.224(4)	C11-N2	1.329(4)
C11-C12	1.504(4)	C12-H12A	0.96
C12-H12B	0.96	C12-H12C	0.96
C13-O2	1.223(3)	C13-N3	1.346(4)
C14-N3	1.478(4)	C14-C15	1.519(5)
C14-H14A	0.97	C14-H14B	0.97
C15-F1	1.407(4)	C15-C16	1.511(5)
C15-H15	0.98	C16-C17	1.529(4)
C16-H16A	0.97	C16-H16B	0.97
C17-N3	1.458(4)	C17-C18	1.528(5)
C17-H17	0.98	C18-O3	1.208(4)
C18-O4	1.321(4)	C19-O4	1.443(4)
C19-H19A	0.96	C19-H19B	0.96
C19-H19C	0.96	N1-H1N	0.86(4)
N2-H2N	0.82(3)		

Table E9 continued

Bond angles

C2-C1-N1	111.4(3)	C2-C1-H1	124.3
N1-C1-H1	124.3	C1-C2-C3	105.2(3)
C1-C2-C9	127.1(3)	C3-C2-C9	127.6(3)
C4-C3-C8	118.5(3)	C4-C3-C2	134.0(3)
C8-C3-C2	107.5(3)	C5-C4-C3	119.1(3)
C5-C4-H4	120.4	C3-C4-H4	120.4
C4-C5-C6	120.9(3)	C4-C5-H5	119.5
C6-C5-H5	119.5	C7-C6-C5	121.7(3)
C7-C6-H6	119.1	C5-C6-H6	119.1
C6-C7-C8	117.2(3)	C6-C7-H7	121.4
C8-C7-H7	121.4	N1-C8-C7	130.3(3)
N1-C8-C3	107.3(3)	C7-C8-C3	122.4(3)
C2-C9-C10	114.6(2)	C2-C9-H9A	108.6
C10-C9-H9A	108.6	C2-C9-H9B	108.6
C10-C9-H9B	108.6	H9A-C9-H9B	107.6
N2-C10-C13	109.7(2)	N2-C10-C9	110.9(2)
C13-C10-C9	107.0(2)	N2-C10-H10	109.7
C13-C10-H10	109.7	C9-C10-H10	109.7
O1-C11-N2	121.6(3)	O1-C11-C12	120.8(3)
N2-C11-C12	117.5(3)	C11-C12-H12A	109.5
C11-C12-H12B	109.5	H12A-C12-H12B	109.5
C11-C12-H12C	109.5	H12A-C12-H12C	109.5
H12B-C12-H12C	109.5	O2-C13-N3	121.8(3)
O2-C13-C10	121.2(3)	N3-C13-C10	116.7(3)
N3-C14-C15	104.3(3)	N3-C14-H14A	110.9
C15-C14-H14A	110.9	N3-C14-H14B	110.9
C15-C14-H14B	110.9	H14A-C14-H14B	108.9
F1-C15-C16	108.3(3)	F1-C15-C14	108.3(3)
C16-C15-C14	104.6(3)	F1-C15-H15	111.8
C16-C15-H15	111.8	C14-C15-H15	111.8
C15-C16-C17	104.4(3)	C15-C16-H16A	110.9
C17-C16-H16A	110.9	C15-C16-H16B	110.9
C17-C16-H16B	110.9	H16A-C16-H16B	108.9
N3-C17-C18	111.5(3)	N3-C17-C16	102.7(3)
C18-C17-C16	113.4(3)	N3-C17-H17	109.7
C18-C17-H17	109.7	C16-C17-H17	109.7

Table E9 continued

O3-C18-O4	123.8(3)	O3-C18-C17	122.9(3)
O4-C18-C17	113.3(3)	O4-C19-H19A	109.5
O4-C19-H19B	109.5	H19A-C19-H19B	109.5
O4-C19-H19C	109.5	H19A-C19-H19C	109.5
H19B-C19-H19C	109.5	C1-N1-C8	108.6(3)
C1-N1-H1N	131.(3)	C8-N1-H1N	121.(3)
C11-N2-C10	121.4(3)	C11-N2-H2N	117.(2)
C10-N2-H2N	121.(2)	C13-N3-C17	119.2(2)
C13-N3-C14	127.9(3)	C17-N3-C14	112.0(2)
C18-O4-C19	118.0(3)		

Table E10. Anisotropic atomic displacement parameters (Å2) for Ac-Trpflp-OMe.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C1	0.042(2)	0.075(3)	0.0421(19)	-0.0218(18)	0.0001(17)	0.0132(18)
C2	0.0343(17)	0.0380(16)	0.0344(16)	-0.0040(13)	0.0024(14)	-0.0023(13)
C3	0.0302(16)	0.0297(13)	0.0363(15)	-0.0025(13)	-0.0011(14)	-0.0031(12)
C4	0.0344(18)	0.0406(16)	0.0448(17)	-0.0093(15)	-0.0023(15)	0.0096(14)
C5	0.049(2)	0.056(2)	0.0433(19)	-0.0167(16)	0.0008(17)	0.0095(18)
C6	0.044(2)	0.055(2)	0.0393(17)	-0.0043(16)	-0.0094(16)	0.0012(16)
C7	0.0315(17)	0.0429(16)	0.055(2)	0.0001(16)	-0.0081(16)	0.0038(15)
C8	0.0281(16)	0.0372(15)	0.0432(17)	-0.0054(14)	0.0007(14)	0.0012(13)
C9	0.0363(18)	0.0366(16)	0.0339(16)	-0.0015(14)	0.0009(14)	-0.0032(13)
C10	0.0353(16)	0.0320(13)	0.0300(15)	-0.0001(12)	0.0042(13)	0.0010(13)
C11	0.0280(16)	0.0340(15)	0.0457(18)	0.0028(14)	0.0028(14)	0.0054(13)
C12	0.0367(19)	0.0365(17)	0.063(2)	-0.0051(16)	0.0048(17)	0.0013(14)
C13	0.0320(17)	0.0309(15)	0.0372(18)	0.0009(14)	0.0001(13)	-0.0030(13)
C14	0.052(2)	0.0426(16)	0.0350(17)	-0.0028(14)	0.0067(16)	0.0058(15)
C15	0.048(2)	0.0470(18)	0.050(2)	-0.0062(17)	0.0075(17)	0.0084(16)
C16	0.042(2)	0.0497(19)	0.064(2)	-0.0039(18)	-0.0012(18)	0.0115(17)
C17	0.0363(18)	0.0393(16)	0.049(2)	-0.0023(15)	-0.0122(16)	0.0066(14)
C18	0.054(2)	0.0327(15)	0.0374(17)	-0.0057(14)	-0.0152(16)	0.0027(15)

Table E10 continued

C19	0.059(2)	0.053(2)	0.0464(19)	0.0008(18)	-0.0062(18)	-0.0151(18)
F1	0.0654(14)	0.0437(10)	0.0673(13)	-0.0099(10)	-0.0027(12)	-0.0014(10)
N1	0.0390(17)	0.070(2)	0.0567(18)	-0.0207(16)	-0.0038(16)	0.0208(15)
N2	0.0353(15)	0.0310(12)	0.0391(14)	-0.0013(11)	0.0124(12)	0.0033(12)
N3	0.0359(15)	0.0334(12)	0.0346(13)	-0.0013(11)	-0.0009(11)	0.0034(11)
O1	0.0398(14)	0.0453(12)	0.0894(19)	-0.0112(14)	0.0296(13)	-0.0039(12)
O2	0.0516(14)	0.0512(13)	0.0341(12)	-0.0073(10)	-0.0068(11)	0.0082(11)
O3	0.0685(18)	0.0497(13)	0.0545(14)	0.0132(12)	-0.0278(13)	-0.0025(14)
O4	0.0395(13)	0.0473(12)	0.0445(12)	0.0103(11)	-0.0058(11)	-0.0035(11)

Table E11. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Ac-Trpflp-OMe.

	x/a	y/b	z/c	U(eq)
H1	1.0467	0.5859	0.6058	0.064
H4	0.7572	0.8082	0.7799	0.048
H5	0.8646	0.8228	0.8874	0.059
H6	1.1004	0.7270	0.9147	0.055
H7	1.2357	0.6144	0.8357	0.052
H9A	0.8036	0.7222	0.5801	0.043
H9B	0.7497	0.8070	0.6389	0.043
H10	0.5816	0.6520	0.6827	0.039
H12A	0.5594	0.2725	0.6164	0.068
H12B	0.6304	0.3386	0.5531	0.068
H12C	0.4454	0.3164	0.5583	0.068
H14A	0.5230	0.8500	0.7106	0.052
H14B	0.3616	0.7816	0.7256	0.052
H15	0.2623	0.9665	0.7417	0.058
H16A	0.1912	1.0365	0.6229	0.062
H16B	0.1153	0.9156	0.6479	0.062
H17	0.2376	0.8460	0.5539	0.05
H19A	0.6992	0.9913	0.4754	0.079
H19B	0.7903	1.0253	0.5423	0.079
H19C	0.6528	1.1073	0.5149	0.079
H1N	1.221(5)	0.550(3)	0.7037(19)	0.066
H2N	0.705(4)	0.517(3)	0.5829(17)	0.042

Table E12. Torsion angles (°) for Ac-Trpflp-OMe.

N1-C1-C2-C3	0.9(4)	N1-C1-C2-C9	178.7(3)
C1-C2-C3-C4	177.3(4)	C9-C2-C3-C4	-0.4(5)
C1-C2-C3-C8	-0.9(3)	C9-C2-C3-C8	-178.6(3)
C8-C3-C4-C5	0.8(5)	C2-C3-C4-C5	-177.3(3)
C3-C4-C5-C6	-0.8(5)	C4-C5-C6-C7	0.0(5)
C5-C6-C7-C8	0.6(5)	C6-C7-C8-N1	176.7(3)
C6-C7-C8-C3	-0.6(5)	C4-C3-C8-N1	-178.0(3)
C2-C3-C8-N1	0.5(3)	C4-C3-C8-C7	-0.1(4)
C2-C3-C8-C7	178.4(3)	C1-C2-C9-C10	101.9(4)
C3-C2-C9-C10	-80.8(4)	C2-C9-C10-N2	-65.7(3)
C2-C9-C10-C13	174.7(2)	N2-C10-C13-O2	-34.7(4)
C9-C10-C13-O2	85.6(3)	N2-C10-C13-N3	150.8(2)
C9-C10-C13-N3	-88.9(3)	N3-C14-C15-F1	-93.1(3)
N3-C14-C15-C16	22.2(3)	F1-C15-C16-C17	81.3(3)
C14-C15-C16-C17	-34.1(4)	C15-C16-C17-N3	32.3(3)
C15-C16-C17-C18	-88.1(3)	N3-C17-C18-O3	168.7(3)
C16-C17-C18-O3	-75.9(4)	N3-C17-C18-O4	-10.8(4)
C16-C17-C18-O4	104.6(3)	C2-C1-N1-C8	-0.6(5)
C7-C8-N1-C1	-177.6(3)	C3-C8-N1-C1	0.0(4)
O1-C11-N2-C10	-9.2(5)	C12-C11-N2-C10	171.6(3)
C13-C10-N2-C11	-71.1(3)	C9-C10-N2-C11	171.0(3)
O2-C13-N3-C17	-10.2(4)	C10-C13-N3-C17	164.3(3)
O2-C13-N3-C14	-178.0(3)	C10-C13-N3-C14	-3.6(4)
C18-C17-N3-C13	-67.0(3)	C16-C17-N3-C13	171.3(3)
C18-C17-N3-C14	102.7(3)	C16-C17-N3-C14	-19.0(3)
C15-C14-N3-C13	166.8(3)	C15-C14-N3-C17	-1.8(3)
O3-C18-O4-C19	-3.3(4)	C17-C18-O4-C19	176.3(3)
N1-C1-C2-C3	0.9(4)	N1-C1-C2-C9	178.7(3)
C1-C2-C3-C4	177.3(4)	C9-C2-C3-C4	-0.4(5)
C1-C2-C3-C8	-0.9(3)	C9-C2-C3-C8	-178.6(3)

Crystal Structure of Boc-Trpflp-OMe

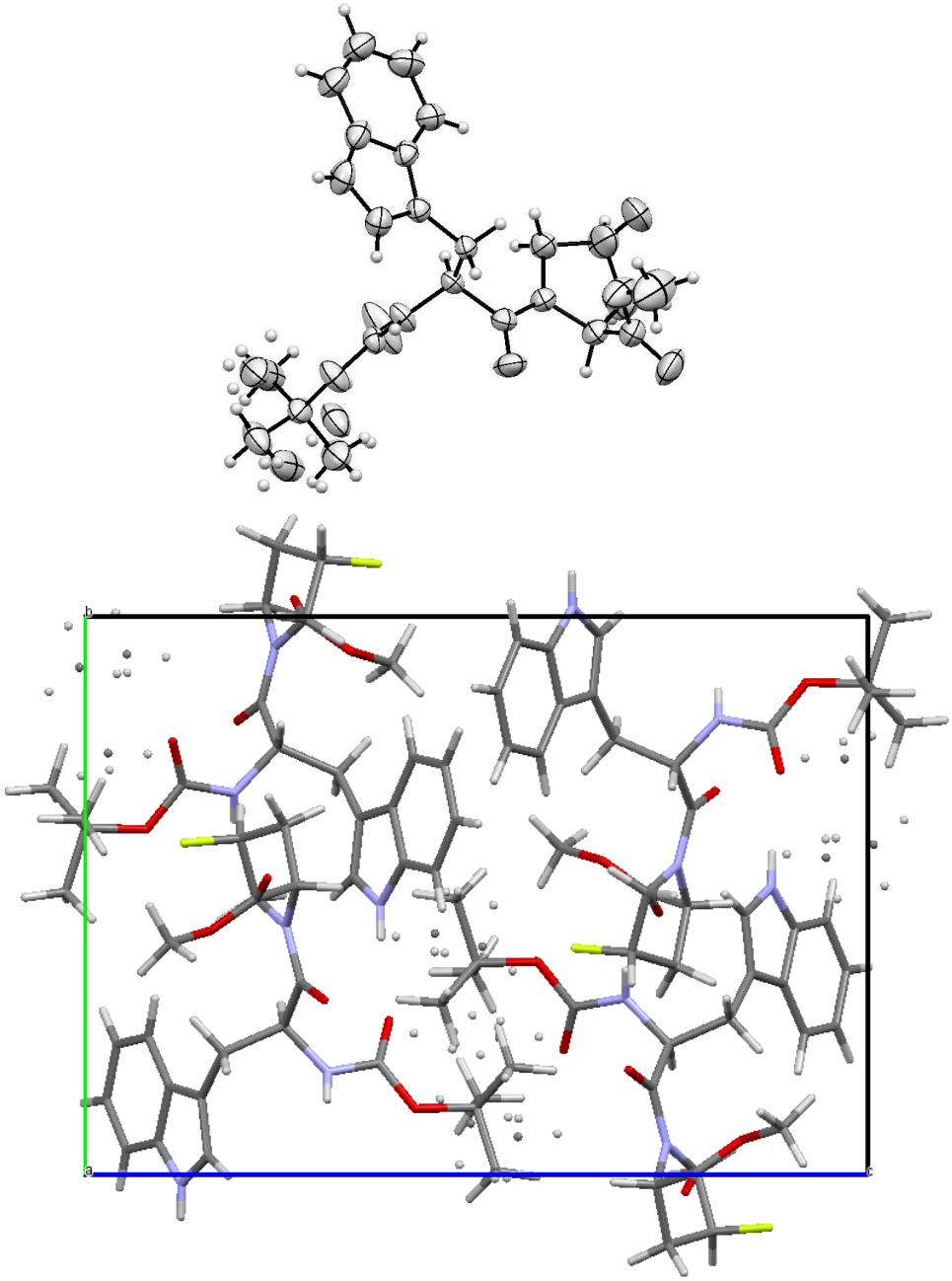


Figure E3 Crystal structure of Boc-Trpflp-OMe.
Top: ORTEP diagram of the Boc-Trpflp-OMe monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in methanol.

Table E13. Crystallographic data and refinement details for Boc-Trpflp-OMe.

empirical formula	$C_{22}H_{28}FN_3O_5$	
formula weight	443.47	
T (K)	200(2)	
wavelength (Å)	1.54178	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 10.2717(3)	α = 90
	b = 12.7320(3)	β = 90
	c = 17.7592(4)	γ = 90
Volume (Å^3)	2322.54(10)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.240	
absorption coefficient (mm^{-1})	0.781	
F(000)	920	
crystal size (mm)	0.177 x 0.236 x 0.448	
θ range for data collection	4.974 to 75.125°	
Index ranges	$-12 \le h \le 12$, $-15 \le k \le 13$, $-21 \le l \le 22$	
Reflections collected/ unique	21502/4738 [R(int) = 0.0549]	
Coverage of independent reflections	99.6%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7539 and 0.4275	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	4738 / 76 / 297	
Goodness-of-fit on F^2	1.069	
Final *R* indices	4414 data; I>2σ(I)	R1 = 0.0686, wR2 = 0.1894
	all data	R1 = 0.0718, wR2 = 0.1953
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.1341P)^2+0.5206P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.540 and -0.417 eÅ^{-3}	
R.M.S. deviation from mean	0.066 eÅ^{-3}	

Table E14. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-Trpflp-OMe.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
F1	0.2211(4)	0.5942(3)	0.1262(2)	0.1065(14)
N1	0.2219(3)	0.4460(2)	0.25698(15)	0.0401(6)
N2	0.3227(3)	0.1837(2)	0.30970(16)	0.0423(6)
N3	0.4681(4)	0.9841(2)	0.1230(2)	0.0555(8)
O1	0.8913(3)	0.5306(3)	0.2329(2)	0.0779(10)
O2	0.0261(3)	0.4373(3)	0.16053(18)	0.0726(10)
O3	0.1056(3)	0.3127(2)	0.30544(17)	0.0575(7)
O4	0.4465(5)	0.2761(3)	0.3884(2)	0.0953(16)
O5	0.3664(3)	0.1183(2)	0.42177(14)	0.0541(7)
C1	0.9216(7)	0.4155(8)	0.1088(4)	0.109(3)
C2	0.9982(4)	0.4955(3)	0.2193(2)	0.0520(9)
C3	0.1143(4)	0.5183(3)	0.2699(2)	0.0472(8)
C4	0.1761(6)	0.6257(3)	0.2550(3)	0.0711(13)
C5	0.2808(6)	0.6021(4)	0.1978(4)	0.0810(16)
C6	0.3326(4)	0.4950(3)	0.2183(2)	0.0494(8)
C7	0.2058(3)	0.3443(3)	0.27545(17)	0.0398(7)
C8	0.3138(3)	0.2669(2)	0.25365(16)	0.0352(6)
C9	0.2785(3)	0.2192(2)	0.17646(16)	0.0371(6)
C10	0.3648(4)	0.0331(3)	0.1569(2)	0.0484(8)
C11	0.3744(3)	0.1398(3)	0.14864(17)	0.0391(7)
C12	0.4921(3)	0.1581(3)	0.10767(18)	0.0406(7)
C13	0.5551(4)	0.2468(3)	0.0805(2)	0.0514(8)
C14	0.6712(5)	0.2335(4)	0.0406(3)	0.0682(12)
C15	0.7212(5)	0.1343(5)	0.0268(3)	0.0733(13)
C16	0.6625(5)	0.0448(4)	0.0527(3)	0.0641(11)
C17	0.5472(4)	0.0583(3)	0.0926(2)	0.0505(8)
C18	0.3831(3)	0.2000(2)	0.37518(18)	0.0371(6)
C19	0.4149(4)	0.1208(3)	0.49974(19)	0.0485(8)
C20	0.3321(7)	0.1926(5)	0.5443(3)	0.0731(18)

Table E14 continued

C21	0.3871(8)	0.0064(4)	0.5253(3)	0.0691(16)
C22	0.5563(5)	0.1364(8)	0.5034(4)	0.083(2)
C20'	0.5529(19)	0.0916(19)	0.4924(19)	0.0731(18)
C21'	0.424(3)	0.2448(12)	0.5294(11)	0.0691(16)
C22'	0.3342(18)	0.068(2)	0.5531(12)	0.083(2)

Table E15. Bond lengths [Å] and angles [°] for Boc-Trpflp-OMe.

Bond lengths

F1-C5	1.415(8)	N1-C7	1.345(4)
N1-C3	1.456(4)	N1-C6	1.468(5)
N2-C18	1.334(4)	N2-C8	1.457(4)
N2-H2N	0.74(4)	N3-C17	1.358(6)
N3-C10	1.370(5)	N3-H3N	0.76(4)
O1-C2	1.211(5)	O2-C2	1.311(5)
O2-C1	1.440(6)	O3-C7	1.226(4)
O4-C18	1.191(5)	O5-C18	1.340(4)
O5-C19	1.472(4)	C1-H1A	0.98
C1-H1B	0.98	C1-H1C	0.98
C2-C3	1.521(6)	C3-C4	1.530(6)
C3-H3	1.0	C4-C5	1.509(7)
C4-H4A	0.99	C4-H4B	0.99
C5-C6	1.508(5)	C5-H5	1.0
C6-H6A	0.99	C6-H6B	0.99
C7-C8	1.534(4)	C8-C9	1.543(4)
C8-H8	1.0	C9-C11	1.495(4)
C9-H9A	0.99	C9-H9B	0.99
C10-C11	1.370(5)	C10-H10	0.95
C11-C12	1.430(5)	C12-C13	1.388(5)
C12-C17	1.416(5)	C13-C14	1.398(6)
C13-H13	0.95	C14-C15	1.385(8)
C14-H14	0.95	C15-C16	1.369(8)
C15-H15	0.95	C16-C17	1.391(6)
C16-H16	0.95	C19-C22'	1.425(15)
C19-C22	1.468(6)	C19-C20'	1.472(16)
C19-C20	1.479(6)	C19-C21	1.551(7)

Table E15 continued

C19-C21'	1.667(15)	C20-H20A	0.98
C20-H20B	0.98	C20-H20C	0.98
C21-H21A	0.98	C21-H21B	0.98
C21-H21C	0.98	C22-H22A	0.98
C22-H22B	0.98	C22-H22C	0.98
C20'-H20D	0.98	C20'-H20E	0.98
C20'-H20F	0.98	C21'-H21D	0.98
C21'-H21E	0.98	C21'-H21F	0.98
C22'-H22D	0.98	C22'-H22E	0.98
C22'-H22F	0.98		

Bond angles

C7-N1-C3	118.5(3)	C7-N1-C6	128.2(3)
C3-N1-C6	113.1(3)	C18-N2-C8	120.8(3)
C18-N2-H2N	113.(4)	C8-N2-H2N	125.(4)
C17-N3-C10	108.7(3)	C17-N3-H3N	127.(4)
C10-N3-H3N	124.(4)	C2-O2-C1	117.0(4)
C18-O5-C19	121.4(3)	O2-C1-H1A	109.5
O2-C1-H1B	109.5	H1A-C1-H1B	109.5
O2-C1-H1C	109.5	H1A-C1-H1C	109.5
H1B-C1-H1C	109.5	O1-C2-O2	124.4(4)
O1-C2-C3	121.6(4)	O2-C2-C3	114.0(3)
N1-C3-C2	112.4(3)	N1-C3-C4	102.9(3)
C2-C3-C4	113.1(4)	N1-C3-H3	109.4
C2-C3-H3	109.4	C4-C3-H3	109.4
C5-C4-C3	103.5(3)	C5-C4-H4A	111.1
C3-C4-H4A	111.1	C5-C4-H4B	111.1
C3-C4-H4B	111.1	H4A-C4-H4B	109.0
F1-C5-C6	107.8(5)	F1-C5-C4	108.0(5)
C6-C5-C4	105.6(4)	F1-C5-H5	111.7
C6-C5-H5	111.7	C4-C5-H5	111.7
N1-C6-C5	102.9(3)	N1-C6-H6A	111.2
C5-C6-H6A	111.2	N1-C6-H6B	111.2
C5-C6-H6B	111.2	H6A-C6-H6B	109.1
O3-C7-N1	121.7(3)	O3-C7-C8	120.4(3)
N1-C7-C8	117.9(3)	N2-C8-C7	109.9(2)
N2-C8-C9	109.6(3)	C7-C8-C9	107.9(2)

Table E15 continued

N2-C8-H8	109.8	C7-C8-H8	109.8
C9-C8-H8	109.8	C11-C9-C8	113.9(3)
C11-C9-H9A	108.8	C8-C9-H9A	108.8
C11-C9-H9B	108.8	C8-C9-H9B	108.8
H9A-C9-H9B	107.7	C11-C10-N3	110.4(3)
C11-C10-H10	124.8	N3-C10-H10	124.8
C10-C11-C12	106.0(3)	C10-C11-C9	126.0(3)
C12-C11-C9	128.0(3)	C13-C12-C17	118.6(3)
C13-C12-C11	134.6(3)	C17-C12-C11	106.8(3)
C12-C13-C14	118.4(4)	C12-C13-H13	120.8
C14-C13-H13	120.8	C15-C14-C13	121.1(5)
C15-C14-H14	119.5	C13-C14-H14	119.5
C16-C15-C14	122.4(4)	C16-C15-H15	118.8
C14-C15-H15	118.8	C15-C16-C17	116.4(4)
C15-C16-H16	121.8	C17-C16-H16	121.8
N3-C17-C16	128.8(4)	N3-C17-C12	108.1(3)
C16-C17-C12	123.1(4)	O4-C18-N2	123.5(3)
O4-C18-O5	125.4(3)	N2-C18-O5	111.0(3)
C22'-C19-O5	114.7(10)	C22-C19-O5	112.3(4)
C22'-C19-C20'	120.0(13)	O5-C19-C20'	103.7(14)
C22-C19-C20	117.5(6)	O5-C19-C20	108.8(4)
C22-C19-C21	107.2(5)	O5-C19-C21	101.1(3)
C20-C19-C21	108.6(4)	C22'-C19-C21'	105.5(11)
O5-C19-C21'	109.6(8)	C20'-C19-C21'	102.4(11)
C19-C20-H20A	109.5	C19-C20-H20B	109.5
H20A-C20-H20B	109.5	C19-C20-H20C	109.5
H20A-C20-H20C	109.5	H20B-C20-H20C	109.5
C19-C21-H21A	109.5	C19-C21-H21B	109.5
H21A-C21-H21B	109.5	C19-C21-H21C	109.5
H21A-C21-H21C	109.5	H21B-C21-H21C	109.5
C19-C22-H22A	109.5	C19-C22-H22B	109.5
H22A-C22-H22B	109.5	C19-C22-H22C	109.5
H22A-C22-H22C	109.5	H22B-C22-H22C	109.5
C19-C20'-H20D	109.5	C19-C20'-H20E	109.5
H20D-C20'-H20E	109.5	C19-C20'-H20F	109.5
H20D-C20'-H20F	109.5	H20E-C20'-H20F	109.5
C19-C21'-H21D	109.5	C19-C21'-H21E	109.5

Table E15 continued

H21D-C21'-H21E	109.5	C19-C21'-H21F	109.5
H21D-C21'-H21F	109.5	H21E-C21'-H21F	109.5
C19-C22'-H22D	109.5	C19-C22'-H22E	109.5
H22D-C22'-H22E	109.5	C19-C22'-H22F	109.5
H22D-C22'-H22F	109.5	H22E-C22'-H22F	109.5

Table E16. Anisotropic atomic displacement parameters (Å2) for Boc-Trpflp-OMe.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U11	U22	U33	U23	U13	U12
F1	0.123(3)	0.109(3)	0.088(2)	0.059(2)	0.032(2)	0.048(2)
N1	0.0411(13)	0.0423(14)	0.0371(12)	0.0003(10)	0.0053(10)	0.0070(11)
N2	0.0507(15)	0.0391(13)	0.0372(13)	0.0116(12)	-0.0094(11)	-0.0119(12)
N3	0.0566(18)	0.0350(14)	0.075(2)	-0.0041(15)	-0.0014(16)	0.0107(13)
O1	0.0546(17)	0.091(2)	0.089(2)	-0.006(2)	0.0009(16)	0.0350(17)
O2	0.0554(17)	0.112(3)	0.0508(15)	-0.0196(17)	-0.0062(13)	0.0182(18)
O3	0.0552(15)	0.0578(15)	0.0594(15)	0.0022(13)	0.0263(13)	0.0009(12)
O4	0.156(4)	0.0514(17)	0.079(2)	0.0226(16)	-0.064(3)	-0.049(2)
O5	0.0663(16)	0.0551(15)	0.0408(13)	0.0124(11)	-0.0150(12)	-0.0105(13)
C1	0.072(3)	0.181(8)	0.073(3)	-0.037(4)	-0.023(3)	0.017(4)
C2	0.0505(19)	0.055(2)	0.0501(18)	0.0050(16)	0.0060(15)	0.0166(16)
C3	0.0531(18)	0.0453(17)	0.0433(16)	-0.0035(14)	0.0068(15)	0.0131(15)
C4	0.079(3)	0.042(2)	0.093(3)	-0.006(2)	0.014(3)	0.011(2)
C5	0.087(3)	0.044(2)	0.111(4)	0.027(2)	0.032(3)	0.012(2)
C6	0.0521(18)	0.0398(17)	0.056(2)	0.0082(14)	0.0098(16)	0.0010(14)
C7	0.0454(16)	0.0420(15)	0.0319(13)	0.0018(12)	0.0050(12)	0.0025(13)
C8	0.0388(14)	0.0341(14)	0.0327(13)	0.0067(11)	0.0031(11)	0.0009(11)
C9	0.0400(14)	0.0401(15)	0.0312(13)	0.0044(11)	-0.0012(11)	0.0096(12)
C10	0.0452(17)	0.0436(18)	0.0563(19)	0.0013(15)	-0.0032(15)	0.0031(14)
C11	0.0402(15)	0.0416(16)	0.0354(14)	0.0006(12)	-0.0027(12)	0.0062(13)
C12	0.0432(16)	0.0422(16)	0.0364(14)	-0.0041(12)	-0.0013(12)	0.0080(13)
C13	0.052(2)	0.0511(19)	0.0508(19)	0.0021(16)	0.0089(16)	0.0031(16)
C14	0.064(3)	0.076(3)	0.065(3)	0.009(2)	0.018(2)	-0.007(2)
C15	0.057(2)	0.094(4)	0.069(3)	0.000(3)	0.015(2)	0.021(3)
C16	0.057(2)	0.066(3)	0.070(2)	-0.008(2)	0.011(2)	0.019(2)
C17	0.0491(19)	0.053(2)	0.0496(18)	-0.0071(16)	-0.0023(15)	0.0151(16)
C18	0.0443(15)	0.0298(13)	0.0370(13)	0.0013(11)	-0.0037(12)	0.0048(12)
C19	0.0488(18)	0.065(2)	0.0318(15)	0.0014(14)	-0.0022(13)	0.0163(16)
C20	0.101(4)	0.059(3)	0.060(3)	0.002(2)	0.019(3)	0.032(3)
C21	0.109(5)	0.058(3)	0.041(2)	0.0084(19)	-0.006(3)	0.015(3)

Table E16 continued

C22	0.045(2)	0.147(7)	0.056(3)	0.031(4)	-0.010(2)	-0.001(3)
C20'	0.101(4)	0.059(3)	0.060(3)	0.002(2)	0.019(3)	0.032(3)
C21'	0.109(5)	0.058(3)	0.041(2)	0.0084(19)	-0.006(3)	0.015(3)
C22'	0.045(2)	0.147(7)	0.056(3)	0.031(4)	-0.010(2)	-0.001(3)

Table E17. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-Trpflp-OMe.

	x/a	y/b	z/c	U(eq)
H2N	0.284(5)	0.135(3)	0.309(3)	0.051
H3N	0.477(5)	-0.075(3)	0.122(3)	0.067
H1A	-0.1537	0.3885	0.1366	0.163
H1B	-0.0498	0.3630	0.0720	0.163
H1C	-0.1029	0.4802	0.0825	0.163
H3	0.0866	0.5140	0.3238	0.057
H4A	0.1113	0.6756	0.2344	0.085
H4B	0.2139	0.6554	0.3016	0.085
H5	0.3509	0.6565	0.1983	0.097
H6A	0.3575	0.4548	0.1728	0.059
H6B	0.4090	0.5006	0.2521	0.059
H8	0.3989	0.3048	0.2501	0.042
H9A	0.1919	0.1856	0.1802	0.045
H9B	0.2719	0.2766	0.1390	0.045
H10	0.2962	-0.0019	0.1825	0.058
H13	0.5201	0.3148	0.0889	0.062
H14	0.7165	0.2936	0.0226	0.082
H15	0.7991	0.1282	-0.0017	0.088
H16	0.6986	-0.0228	0.0440	0.077
H20A	0.2403	0.1740	0.5370	0.11
H20B	0.3544	0.1864	0.5977	0.11
H20C	0.3465	0.2651	0.5276	0.11
H21A	0.2932	-0.0069	0.5236	0.104
H21B	0.4320	-0.0427	0.4916	0.104
H21C	0.4187	-0.0035	0.5769	0.104
H22A	0.5842	0.1375	0.5562	0.124
H22B	0.6002	0.0788	0.4770	0.124
H22C	0.5789	0.2032	0.4794	0.124
H20D	0.5939	0.1350	0.4536	0.11
H20E	0.5972	0.1028	0.5406	0.11
H20F	0.5596	0.0174	0.4783	0.11
H21D	0.4783	0.2854	0.4948	0.104

Table E17 continued

H21E	0.3361	0.2752	0.5309	0.104
H21F	0.4618	0.2466	0.5800	0.104
H22D	0.3733	0.0735	0.6033	0.124
H22E	0.2478	0.1008	0.5536	0.124
H22F	0.3261	-0.0060	0.5390	0.124

Crystal Structure of Boc-TrpPro-OMe

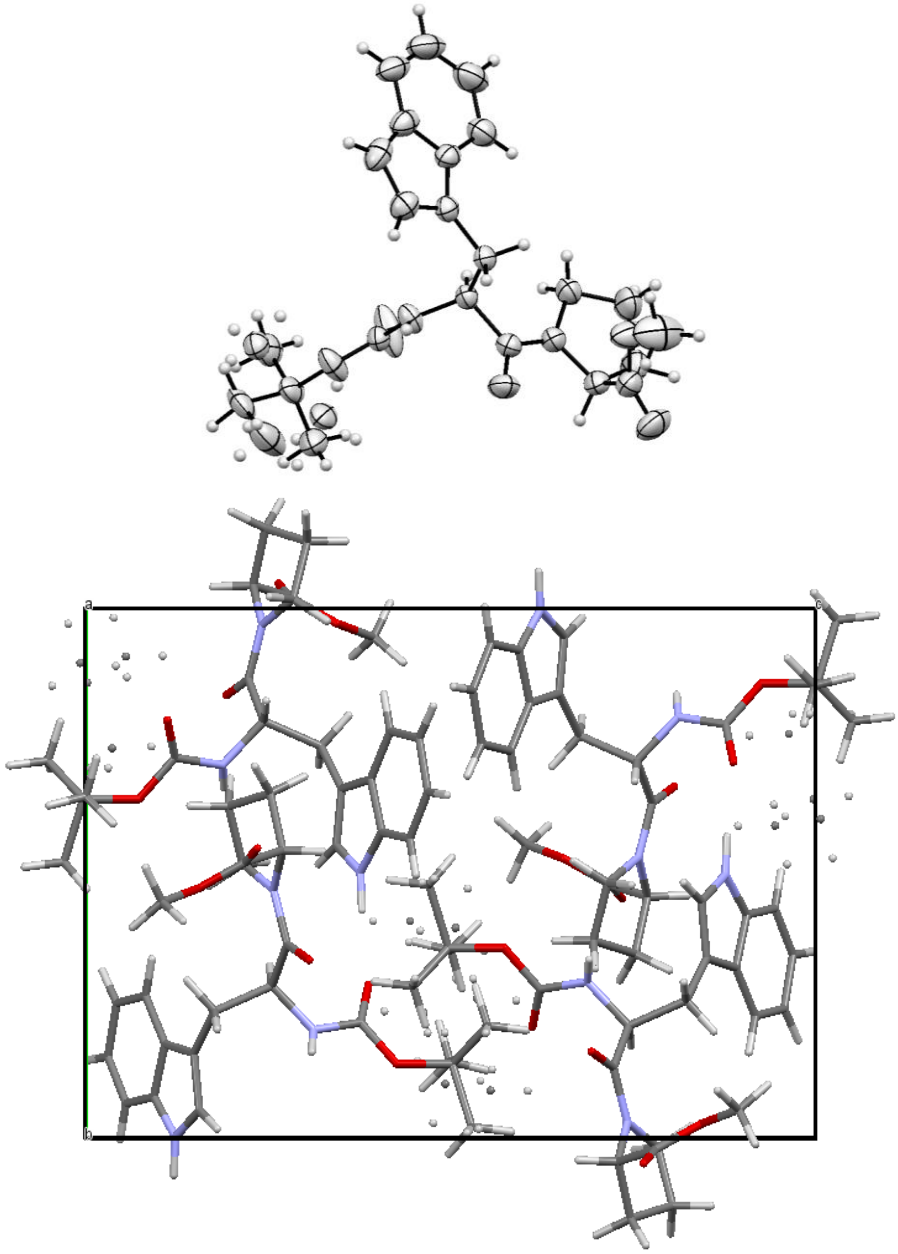


Figure E4 Crystal structure of Boc-TrpPro-OMe.
Top: ORTEP diagram of the Boc-TrpPro-OMe monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in methanol. Crystals grown from methanol/trifluoroethanol generated crystals with the same unit cell dimensions.

Table E18. Crystallographic data and refinement details for Boc-TrpPro-OMe.

empirical formula	$C_{22}H_{29}N_3O_5$	
formula weight	415.48	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 10.2770(16)	α = 90
	b = 12.609(2)	β = 90
	c = 17.573(3)	γ = 90
Volume (Å^3)	2277.2(6)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.212	
absorption coefficient (mm^{-1})	0.086	
F(000)	888	
crystal size (mm)	0.176 x 0.276 x 0.403	
θ range for data collection	2.296 to 27.596°	
Index ranges	$-12 \le h \le 13$, $-16 \le k \le 16$, $-22 \le l \le 22$	
Reflections collected/ unique	17412/5273 [R(int) = 0.0566]	
Coverage of independent reflections	99.7%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7456 and 0.6629	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	5273 / 66 / 312	
Goodness-of-fit on F^2	1.001	
Final *R* indices	3374 data; I>2σ(I)	R1 = 0.0599, wR2 = 0.1303
	all data	R1 = 0.1033, wR2 = 0.1524
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0688P)^2+0.2547P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.197 and -0.174 eÅ^{-3}	
R.M.S. deviation from mean	0.039 eÅ^{-3}	

Table E19. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-TrpPro-OMe.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
C1	0.0016(4)	0.5215(3)	0.7196(2)	0.0527(10)
C2	0.1173(4)	0.5439(3)	0.7703(2)	0.0471(10)
C3	0.1803(5)	0.6516(3)	0.7533(3)	0.0621(12)
C4	0.2842(5)	0.6267(4)	0.6957(3)	0.0673(13)
C5	0.3335(4)	0.5179(3)	0.7156(2)	0.0482(9)
C6	0.2063(4)	0.3665(3)	0.7766(2)	0.0400(9)
C7	0.3134(3)	0.2878(3)	0.75551(19)	0.0369(8)
C8	0.2791(4)	0.2399(3)	0.6770(2)	0.0394(9)
C9	0.3644(4)	0.0527(3)	0.6561(2)	0.0501(10)
C10	0.3755(3)	0.1602(3)	0.6488(2)	0.0402(8)
C11	0.4951(4)	0.1781(3)	0.6090(2)	0.0440(9)
C12	0.5597(4)	0.2690(4)	0.5823(2)	0.0588(11)
C13	0.6757(5)	0.2551(5)	0.5427(3)	0.0734(15)
C14	0.7289(5)	0.1541(6)	0.5307(3)	0.0796(17)
C15	0.6680(5)	0.0658(5)	0.5556(3)	0.0687(14)
C16	0.5505(4)	0.0781(4)	0.5937(2)	0.0529(11)
C17	0.3859(4)	0.2185(3)	0.8771(2)	0.0395(8)
C18	0.4251(4)	0.1402(4)	0.0026(2)	0.0531(11)
C19	0.3887(11)	0.0216(7)	0.0293(5)	0.063(3)
C20	0.5685(9)	0.1454(11)	0.9994(11)	0.062(3)
C21	0.3576(10)	0.2128(8)	0.0511(5)	0.068(3)
C18'	0.4251(4)	0.1402(4)	0.0026(2)	0.0531(11)
C19'	0.3383(11)	0.0903(14)	0.0552(7)	0.088(5)
C20'	0.5569(12)	0.1027(11)	0.9929(13)	0.049(4)
C21'	0.4408(14)	0.2632(8)	0.0364(6)	0.062(4)
C22	0.9287(5)	0.4458(6)	0.6052(3)	0.111(2)
N1	0.2241(3)	0.4696(2)	0.75767(17)	0.0413(7)
N2	0.3207(3)	0.2044(3)	0.81225(17)	0.0424(8)
N3	0.4688(4)	0.0023(3)	0.6232(2)	0.0597(10)

Table E19 continued

O1	0.0347(3)	0.4721(3)	0.65566(17)	0.0730(10)
O2	0.8925(3)	0.5472(3)	0.7355(2)	0.0747(10)
O3	0.1056(3)	0.3362(2)	0.80668(16)	0.0544(7)
O4	0.3632(3)	0.1405(2)	0.92642(15)	0.0583(8)
O5	0.4596(4)	0.2895(2)	0.88768(19)	0.0852(13)

Table E20. Bond lengths [Å] and angles [°] for Boc-TrpPro-OMe.

Bond lengths

C1-O2	1.200(5)	C1-O1	1.328(5)
C1-C2	1.512(6)	C2-N1	1.460(5)
C2-C3	1.533(6)	C2-H2A	1.0
C3-C4	1.504(7)	C3-H3A	0.99
C3-H3B	0.99	C4-C5	1.505(6)
C4-H4A	0.99	C4-H4B	0.99
C5-N1	1.477(5)	C5-H5A	0.99
C5-H5B	0.99	C6-O3	1.224(4)
C6-N1	1.354(5)	C6-C7	1.527(5)
C7-N2	1.451(5)	C7-C8	1.548(5)
C7-H7A	1.0	C8-C10	1.496(5)
C8-H8A	0.99	C8-H8B	0.99
C9-C10	1.366(6)	C9-N3	1.375(5)
C9-H9A	0.95	C10-C11	1.432(5)
C11-C12	1.405(6)	C11-C16	1.409(6)
C12-C13	1.391(7)	C12-H12A	0.95
C13-C14	1.401(8)	C13-H13A	0.95
C14-C15	1.351(8)	C14-H14A	0.95
C15-C16	1.390(6)	C15-H15A	0.95
C16-N3	1.373(6)	C17-O5	1.188(4)
C17-O4	1.332(4)	C17-N2	1.334(5)
C18-C21	1.431(8)	C18-C20	1.476(10)
C18-O4	1.482(4)	C18-C19	1.612(8)
C19-H19A	0.98	C19-H19B	0.98
C19-H19C	0.98	C20-H20A	0.98
C20-H20B	0.98	C20-H20C	0.98
C21-H21A	0.98	C21-H21B	0.98
C21-H21C	0.98	C18'-C19'	1.430(10)
C18'-C20'	1.444(12)	C18'-O4	1.482(4)
C18'-C21'	1.669(11)	C19'-H19D	0.98
C19'-H19E	0.98	C19'-H19F	0.98
C20'-H20D	0.98	C20'-H20E	0.98
C20'-H20F	0.98	C21'-H21D	0.98

Table E20 continued

C21'-H21E	0.98	C21'-H21F	0.98
C22-O1	1.443(6)	C22-H22A	0.98
C22-H22B	0.98	C22-H22C	0.98
N2-H2N	0.81(4)	N3-H3N	0.91(5)

Bond angles

O2-C1-O1	124.2(4)	O2-C1-C2	123.1(4)
O1-C1-C2	112.6(3)	N1-C2-C1	112.4(3)
N1-C2-C3	102.8(3)	C1-C2-C3	112.5(3)
N1-C2-H2A	109.6	C1-C2-H2A	109.6
C3-C2-H2A	109.6	C4-C3-C2	104.3(3)
C4-C3-H3A	110.9	C2-C3-H3A	110.9
C4-C3-H3B	110.9	C2-C3-H3B	110.9
H3A-C3-H3B	108.9	C5-C4-C3	105.8(4)
C5-C4-H4A	110.6	C3-C4-H4A	110.6
C5-C4-H4B	110.6	C3-C4-H4B	110.6
H4A-C4-H4B	108.7	N1-C5-C4	103.6(3)
N1-C5-H5A	111.0	C4-C5-H5A	111.0
N1-C5-H5B	111.0	C4-C5-H5B	111.0
H5A-C5-H5B	109.0	O3-C6-N1	121.4(3)
O3-C6-C7	120.7(3)	N1-C6-C7	117.8(3)
N2-C7-C6	110.0(3)	N2-C7-C8	110.0(3)
C6-C7-C8	107.8(3)	N2-C7-H7A	109.7
C6-C7-H7A	109.7	C8-C7-H7A	109.7
C10-C8-C7	114.0(3)	C10-C8-H8A	108.7
C7-C8-H8A	108.7	C10-C8-H8B	108.7
C7-C8-H8B	108.7	H8A-C8-H8B	107.6
C10-C9-N3	110.7(4)	C10-C9-H9A	124.6
N3-C9-H9A	124.6	C9-C10-C11	105.9(3)
C9-C10-C8	125.4(3)	C11-C10-C8	128.6(3)
C12-C11-C16	118.4(4)	C12-C11-C10	134.2(4)
C16-C11-C10	107.3(4)	C13-C12-C11	117.9(4)
C13-C12-H12A	121.0	C11-C12-H12A	121.0
C12-C13-C14	121.6(5)	C12-C13-H13A	119.2
C14-C13-H13A	119.2	C15-C14-C13	121.3(4)
C15-C14-H14A	119.3	C13-C14-H14A	119.3
C14-C15-C16	117.8(5)	C14-C15-H15A	121.1

Table E20 continued

C16-C15-H15A	121.1	N3-C16-C15	129.4(5)
N3-C16-C11	107.7(3)	C15-C16-C11	122.8(5)
O5-C17-O4	124.5(3)	O5-C17-N2	123.7(4)
O4-C17-N2	111.7(3)	C21-C18-C20	118.6(7)
C21-C18-O4	109.2(4)	C20-C18-O4	113.2(8)
C21-C18-C19	107.9(5)	C20-C18-C19	106.5(7)
O4-C18-C19	99.5(4)	C18-C19-H19A	109.5
C18-C19-H19B	109.5	H19A-C19-H19B	109.5
C18-C19-H19C	109.5	H19A-C19-H19C	109.5
H19B-C19-H19C	109.5	C18-C20-H20A	109.5
C18-C20-H20B	109.5	H20A-C20-H20B	109.5
C18-C20-H20C	109.5	H20A-C20-H20C	109.5
H20B-C20-H20C	109.5	C18-C21-H21A	109.5
C18-C21-H21B	109.5	H21A-C21-H21B	109.5
C18-C21-H21C	109.5	H21A-C21-H21C	109.5
H21B-C21-H21C	109.5	C19'-C18'-C20'	121.1(9)
C19'-C18'-O4	108.4(6)	C20'-C18'-O4	107.3(9)
C19'-C18'-C21'	103.9(7)	C20'-C18'-C21'	104.8(7)
O4-C18'-C21'	111.2(4)	C18'-C19'-H19D	109.5
C18'-C19'-H19E	109.5	H19D-C19'-H19E	109.5
C18'-C19'-H19F	109.5	H19D-C19'-H19F	109.5
H19E-C19'-H19F	109.5	C18'-C20'-H20D	109.5
C18'-C20'-H20E	109.5	H20D-C20'-H20E	109.5
C18'-C20'-H20F	109.5	H20D-C20'-H20F	109.5
H20E-C20'-H20F	109.5	C18'-C21'-H21D	109.5
C18'-C21'-H21E	109.5	H21D-C21'-H21E	109.5
C18'-C21'-H21F	109.5	H21D-C21'-H21F	109.5
H21E-C21'-H21F	109.5	O1-C22-H22A	109.5
O1-C22-H22B	109.5	H22A-C22-H22B	109.5
O1-C22-H22C	109.5	H22A-C22-H22C	109.5
H22B-C22-H22C	109.5	C6-N1-C2	118.5(3)
C6-N1-C5	128.4(3)	C2-N1-C5	112.6(3)
C17-N2-C7	121.1(3)	C17-N2-H2N	120.(3)
C7-N2-H2N	114.(3)	C9-N3-C16	108.3(3)
C9-N3-H3N	127.(3)	C16-N3-H3N	125.(3)
C1-O1-C22	115.7(4)	C17-O4-C18'	120.9(3)
C17-O4-C18	120.9(3)		

Table E21. Anisotropic atomic displacement parameters (Å2) for Boc-TrpPro-OMe.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C1	0.048(2)	0.064(3)	0.047(2)	0.003(2)	0.002(2)	0.017(2)
C2	0.051(2)	0.050(2)	0.040(2)	-0.0021(18)	0.0042(19)	0.0144(19)
C3	0.074(3)	0.047(2)	0.065(3)	0.003(2)	-0.001(2)	0.013(2)
C4	0.071(3)	0.053(3)	0.078(3)	0.020(2)	0.012(3)	0.008(2)
C5	0.048(2)	0.050(2)	0.047(2)	0.0115(19)	0.0035(19)	0.0019(19)
C6	0.042(2)	0.044(2)	0.0335(18)	0.0008(16)	0.0030(16)	0.0032(17)
C7	0.0362(18)	0.0420(19)	0.0324(17)	0.0050(16)	-0.0010(15)	0.0024(16)
C8	0.0393(19)	0.046(2)	0.0334(18)	0.0041(16)	-0.0025(16)	0.0067(17)
C9	0.043(2)	0.053(2)	0.055(2)	-0.004(2)	-0.0038(19)	0.0006(19)
C10	0.041(2)	0.045(2)	0.0348(18)	0.0016(17)	-0.0054(16)	0.0058(17)
C11	0.043(2)	0.056(2)	0.0334(19)	-0.0032(17)	-0.0013(17)	0.0113(19)
C12	0.056(2)	0.071(3)	0.050(2)	0.010(2)	0.005(2)	0.001(2)
C13	0.059(3)	0.103(4)	0.058(3)	0.016(3)	0.009(2)	-0.004(3)
C14	0.051(3)	0.133(5)	0.054(3)	-0.006(3)	0.011(2)	0.021(4)
C15	0.052(3)	0.103(4)	0.050(3)	-0.011(3)	0.003(2)	0.028(3)
C16	0.043(2)	0.073(3)	0.043(2)	-0.009(2)	-0.0073(19)	0.018(2)
C17	0.043(2)	0.041(2)	0.0343(18)	0.0031(17)	-0.0030(17)	0.0078(18)
C18	0.048(2)	0.082(3)	0.0300(18)	0.012(2)	-0.0037(17)	0.015(2)
C19	0.091(7)	0.066(5)	0.033(4)	0.018(4)	0.001(4)	-0.002(5)
C20	0.057(5)	0.079(9)	0.049(6)	0.009(8)	-0.010(4)	0.009(5)
C21	0.083(7)	0.074(6)	0.049(5)	0.004(4)	0.008(4)	0.033(6)
C18'	0.048(2)	0.082(3)	0.0300(18)	0.012(2)	-0.0037(17)	0.015(2)
C19'	0.061(7)	0.145(14)	0.058(7)	0.039(8)	-0.005(5)	-0.012(8)
C20'	0.056(6)	0.049(8)	0.041(7)	0.001(7)	-0.004(5)	0.009(5)
C21'	0.070(8)	0.086(7)	0.031(5)	0.002(5)	-0.001(5)	0.028(6)
C22	0.068(3)	0.195(7)	0.069(4)	-0.036(4)	-0.017(3)	0.002(4)
N1	0.0404(16)	0.0460(18)	0.0375(16)	0.0041(14)	0.0023(13)	0.0061(14)
N2	0.0449(18)	0.0446(18)	0.0377(16)	0.0091(15)	-0.0062(15)	-0.0112(15)
N3	0.063(2)	0.050(2)	0.066(2)	-0.012(2)	-0.0113(19)	0.0164(19)

Table E21 continued

O1	0.0516(17)	0.121(3)	0.0465(17)	-0.0186(18)	-0.0056(14)	0.0157(18)
O2	0.0493(18)	0.093(2)	0.081(2)	-0.0128(19)	-0.0013(16)	0.0277(18)
O3	0.0493(16)	0.0595(17)	0.0544(16)	0.0017(14)	0.0186(13)	0.0029(14)
O4	0.0541(17)	0.079(2)	0.0417(14)	0.0237(15)	-0.0130(13)	-0.0179(15)
O5	0.130(3)	0.0492(17)	0.076(2)	0.0227(16)	-0.058(2)	-0.040(2)

Table E22. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-TrpPro-OMe.

	x/a	y/b	z/c	U(eq)
H2A	0.0898	0.5412	0.8248	0.057
H3A	0.1157	0.7018	0.7322	0.075
H3B	0.2187	0.6826	0.7999	0.075
H4A	0.2477	0.6274	0.6436	0.081
H4B	0.3555	0.6794	0.6985	0.081
H5A	0.3541	0.4767	0.6692	0.058
H5B	0.4124	0.5222	0.7478	0.058
H7A	0.3988	0.3255	0.7524	0.044
H8A	0.1926	0.2056	0.6803	0.047
H8B	0.2726	0.2981	0.6393	0.047
H9A	0.2941	0.0176	0.6805	0.06
H12A	0.5253	0.3379	0.5910	0.071
H13A	0.7200	0.3155	0.5234	0.088
H14A	0.8094	0.1477	0.5045	0.095
H15A	0.7043	-0.0025	0.5473	0.082
H19A	0.4119	0.0124	1.0830	0.095
H19B	0.4371	-0.0297	0.9983	0.095
H19C	0.2951	0.0099	1.0227	0.095
H20A	0.6045	0.1284	1.0495	0.092
H20B	0.5954	0.2171	0.9846	0.092
H20C	0.6007	0.0943	0.9619	0.092
H21A	0.3879	0.2039	1.1036	0.103
H21B	0.2639	0.1986	1.0486	0.103
H21C	0.3746	0.2856	1.0344	0.103
H19D	0.3779	0.0895	1.1059	0.132
H19E	0.3213	0.0173	1.0387	0.132
H19F	0.2563	0.1298	1.0569	0.132
H20D	0.6010	0.1459	0.9544	0.073
H20E	0.5554	0.0285	0.9764	0.073
H20F	0.6035	0.1083	1.0414	0.073
H21D	0.5084	0.3007	1.0077	0.094
H21E	0.4653	0.2603	1.0903	0.094

Table E22 continued

H21F	0.3579	0.3010	1.0311	0.094
H22A	-0.0397	0.3987	0.5649	0.166
H22B	-0.1061	0.5108	0.5825	0.166
H22C	-0.1401	0.4099	0.6340	0.166
H2N	0.261(4)	0.163(3)	0.810(2)	0.051
H3N	0.482(4)	-0.069(4)	0.621(3)	0.072

Table E23. Torsion angles (°) for Boc-TrpPro-OMe.

O2-C1-C2-N1	155.2(4)	O1-C1-C2-N1	-25.8(5)
O2-C1-C2-C3	-89.4(5)	O1-C1-C2-C3	89.7(4)
N1-C2-C3-C4	29.4(4)	C1-C2-C3-C4	-91.8(4)
C2-C3-C4-C5	-33.5(5)	C3-C4-C5-N1	23.8(5)
O3-C6-C7-N2	-35.9(5)	N1-C6-C7-N2	147.1(3)
O3-C6-C7-C8	84.0(4)	N1-C6-C7-C8	-93.0(4)
N2-C7-C8-C10	-59.7(4)	C6-C7-C8-C10	-179.7(3)
N3-C9-C10-C11	-0.3(4)	N3-C9-C10-C8	179.2(3)
C7-C8-C10-C9	96.7(4)	C7-C8-C10-C11	-83.9(4)
C9-C10-C11-C12	177.9(4)	C8-C10-C11-C12	-1.7(7)
C9-C10-C11-C16	0.4(4)	C8-C10-C11-C16	-179.2(3)
C16-C11-C12-C13	-0.5(6)	C10-C11-C12-C13	-177.8(4)
C11-C12-C13-C14	-1.2(7)	C12-C13-C14-C15	1.5(8)
C13-C14-C15-C16	0.0(7)	C14-C15-C16-N3	178.6(4)
C14-C15-C16-C11	-1.8(6)	C12-C11-C16-N3	-178.3(4)
C10-C11-C16-N3	-0.3(4)	C12-C11-C16-C15	2.1(6)
C10-C11-C16-C15	-180.0(4)	O3-C6-N1-C2	-3.2(5)
C7-C6-N1-C2	173.8(3)	O3-C6-N1-C5	-174.3(3)
C7-C6-N1-C5	2.7(6)	C1-C2-N1-C6	-66.5(4)
C3-C2-N1-C6	172.3(3)	C1-C2-N1-C5	105.9(4)
C3-C2-N1-C5	-15.3(4)	C4-C5-N1-C6	166.6(4)
C4-C5-N1-C2	-4.9(4)	O5-C17-N2-C7	-13.6(6)
O4-C17-N2-C7	169.6(3)	C6-C7-N2-C17	-82.3(4)
C8-C7-N2-C17	159.1(3)	C10-C9-N3-C16	0.1(5)
C15-C16-N3-C9	179.8(4)	C11-C16-N3-C9	0.1(4)
O2-C1-O1-C22	-2.4(7)	C2-C1-O1-C22	178.6(5)
O5-C17-O4-C18'	3.5(6)	N2-C17-O4-C18'	-179.7(3)
O5-C17-O4-C18	3.5(6)	N2-C17-O4-C18	-179.7(3)
C19'-C18'-O4-C17	148.4(8)	C20'-C18'-O4-C17	-79.3(7)
C21'-C18'-O4-C17	34.8(7)	C21-C18-O4-C17	78.1(6)
C20-C18-O4-C17	-56.4(8)	C19-C18-O4-C17	-169.1(5)

Crystal Structure of Boc-Trpflp-OH

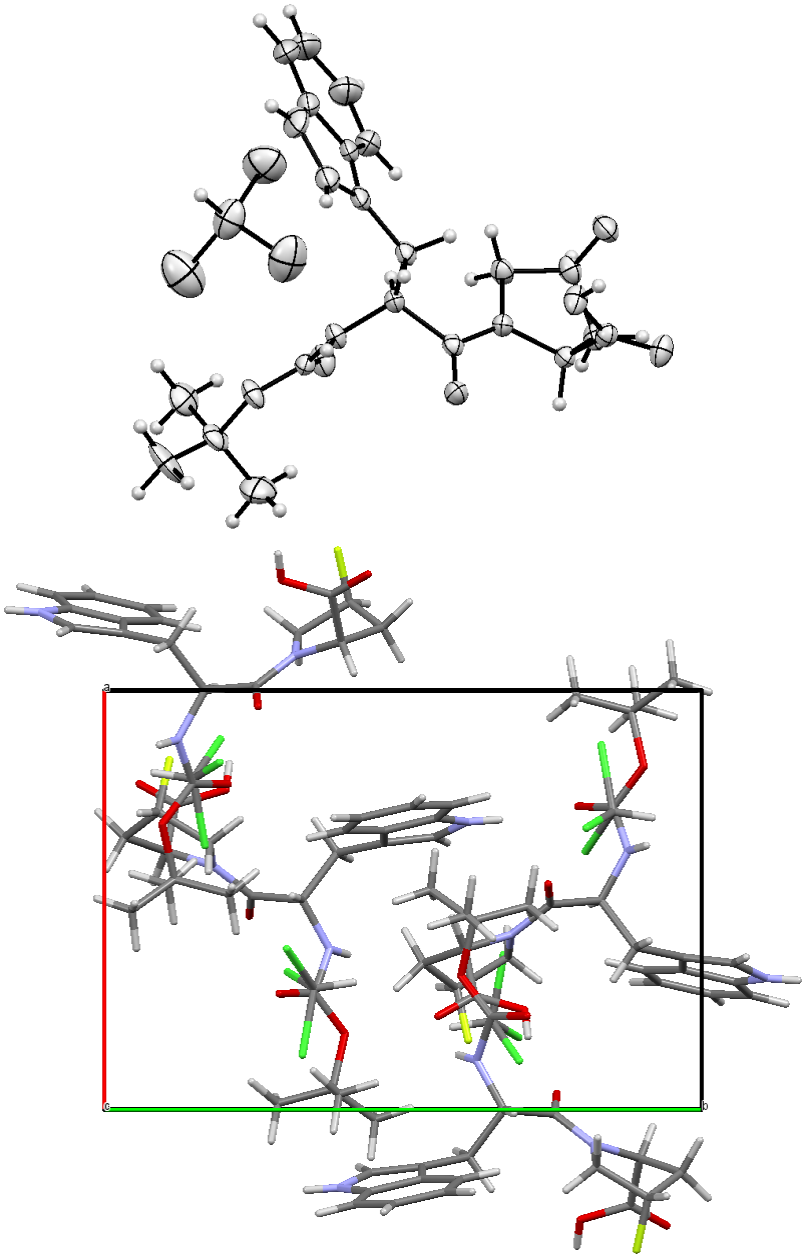


Figure E5 Crystal structure of Boc-Trpflp-OH.
Top: ORTEP diagram of the Boc-Trpflp-OH monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in $CDCl_3$.

Table E24. Crystallographic data and refinement details for Boc-Trpflp-OH.

empirical formula	$C_{22}H_{27}Cl_3FN_3O_5$	
formula weight	538.81	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 10.5930(6)	α = 90
	b = 15.0701(9)	β = 90
	c = 16.4238(11)	γ = 90
Volume (Å^3)	2621.9(3)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.365	
absorption coefficient (mm^{-1})	0.393	
F(000)	1120	
crystal size (mm)	0.579 x 0.501 x 0.445	
θ range for data collection	1.834 to 27.568°	
Index ranges	$-12 \le h \le 9$, $-19 \le k \le 12$, $-21 \le l \le 20$	
Reflections collected/ unique	10058/5727 [R(int) = 0.0250]	
Coverage of independent reflections	96.8%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7456 and 0.6726	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	5727 / 0 / 319	
Goodness-of-fit on F^2	1.031	
Final *R* indices	4952 data; I>2σ(I)	R1 = 0.0386, wR2 = 0.0982
	all data	R1 = 0.0457, wR2 = 0.1018
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0611P)^2]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.338 and -0.349 eÅ^{-3}	
R.M.S. deviation from mean	0.043 eÅ^{-3}	

Table E25. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-Trpflp-OH.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
Cl1	0.13017(13)	0.33157(10)	0.45148(7)	0.0882(4)
Cl2	0.32822(13)	0.30679(8)	0.56875(5)	0.0763(3)
Cl3	0.38384(12)	0.30554(8)	0.39642(6)	0.0748(3)
F1	0.83457(16)	0.10920(11)	0.79676(10)	0.0396(4)
N1	0.58270(19)	0.18116(13)	0.85582(11)	0.0231(4)
N2	0.3827(2)	0.37699(14)	0.84472(12)	0.0236(4)
N3	0.6867(3)	0.60755(18)	0.7657(2)	0.0483(7)
O1	0.7771(2)	0.05891(12)	0.99795(10)	0.0351(5)
O2	0.7627(2)	0.20321(13)	0.96661(12)	0.0333(4)
O3	0.46211(18)	0.24267(12)	0.95324(10)	0.0299(4)
O4	0.2765(2)	0.29271(13)	0.75290(11)	0.0353(5)
O5	0.18095(17)	0.40299(12)	0.82495(11)	0.0318(4)
C1	0.7287(3)	0.11962(16)	0.96177(14)	0.0251(5)
C2	0.6164(2)	0.10313(16)	0.90455(14)	0.0247(5)
C3	0.6503(3)	0.03425(16)	0.83914(16)	0.0295(6)
C4	0.7101(3)	0.08982(18)	0.77316(16)	0.0303(6)
C5	0.6360(3)	0.17594(18)	0.77288(14)	0.0290(6)
C6	0.5105(2)	0.24665(16)	0.88499(14)	0.0230(5)
C7	0.5008(2)	0.33103(16)	0.83278(14)	0.0228(5)
C8	0.6126(2)	0.38986(18)	0.85731(16)	0.0275(5)
C9	0.6504(3)	0.55392(19)	0.82960(19)	0.0386(7)
C10	0.6365(3)	0.46863(17)	0.80406(16)	0.0301(6)
C11	0.6658(3)	0.46807(18)	0.71787(17)	0.0322(6)
C12	0.6669(3)	0.4037(2)	0.65690(18)	0.0401(7)
C13	0.6959(4)	0.4283(3)	0.5774(2)	0.0530(9)
C14	0.7249(4)	0.5167(3)	0.5586(2)	0.0619(11)
C15	0.7266(3)	0.5810(2)	0.6167(2)	0.0544(10)
C16	0.6959(3)	0.5569(2)	0.6965(2)	0.0414(7)
C17	0.2779(3)	0.35186(15)	0.80369(14)	0.0233(5)

Table E25 continued

C18	0.0491(3)	0.3818(2)	0.79887(17)	0.0359(6)
C19	0.9760(3)	0.4536(3)	0.8431(3)	0.0638(11)
C20	0.0371(3)	0.3933(2)	0.7075(2)	0.0482(8)
C21	0.0148(4)	0.2905(3)	0.8283(2)	0.0601(10)
C22	0.2892(4)	0.3512(2)	0.4735(2)	0.0548(9)

Table E26. Bond lengths [Å] and angles [°] for Boc-Trpflp-OH.

Bond lengths

Cl1-C22	1.748(4)	Cl2-C22	1.750(3)
Cl3-C22	1.756(4)	F1-C4	1.405(3)
N1-C6	1.338(3)	N1-C2	1.466(3)
N1-C5	1.477(3)	N2-C17	1.353(3)
N2-C7	1.444(3)	N2-H2N	0.85(3)
N3-C16	1.372(5)	N3-C9	1.379(4)
N3-H3N	0.81(4)	O1-C1	1.206(3)
O2-C1	1.313(3)	O2-H2O	0.78(4)
O3-C6	1.234(3)	O4-C17	1.221(3)
O5-C17	1.330(3)	O5-C18	1.495(3)
C1-C2	1.536(4)	C2-C3	1.537(3)
C2-H2A	1.0	C3-C4	1.509(4)
C3-H3A	0.99	C3-H3B	0.99
C4-C5	1.517(4)	C4-H4	1.0
C5-H5A	0.99	C5-H5B	0.99
C6-C7	1.537(3)	C7-C8	1.533(4)
C7-H7	1.0	C8-C10	1.496(4)
C8-H8A	0.99	C8-H8B	0.99
C9-C10	1.360(4)	C9-H9	0.95
C10-C11	1.449(4)	C11-C12	1.394(4)
C11-C16	1.421(4)	C12-C13	1.391(4)
C12-H12	0.95	C13-C14	1.402(6)
C13-H13	0.95	C14-C15	1.360(6)
C14-H14	0.95	C15-C16	1.398(5)
C15-H15	0.95	C18-C21	1.503(5)
C18-C19	1.515(5)	C18-C20	1.516(4)
C19-H19A	0.98	C19-H19B	0.98
C19-H19C	0.98	C20-H20A	0.98
C20-H20B	0.98	C20-H20C	0.98
C21-H21A	0.98	C21-H21B	0.98
C21-H21C	0.98	C22-H22	1.0

Table E26 continued

Bond angles

C6-N1-C2	122.37(19)	C6-N1-C5	126.0(2)
C2-N1-C5	111.57(19)	C17-N2-C7	120.6(2)
C17-N2-H2N	116.(2)	C7-N2-H2N	120.(2)
C16-N3-C9	108.9(3)	C16-N3-H3N	127.(3)
C9-N3-H3N	124.(3)	C1-O2-H2O	111.(3)
C17-O5-C18	121.5(2)	O1-C1-O2	125.6(2)
O1-C1-C2	120.6(2)	O2-C1-C2	113.9(2)
N1-C2-C1	113.1(2)	N1-C2-C3	102.54(19)
C1-C2-C3	110.8(2)	N1-C2-H2A	110.0
C1-C2-H2A	110.0	C3-C2-H2A	110.0
C4-C3-C2	103.0(2)	C4-C3-H3A	111.2
C2-C3-H3A	111.2	C4-C3-H3B	111.2
C2-C3-H3B	111.2	H3A-C3-H3B	109.1
F1-C4-C3	108.1(2)	F1-C4-C5	108.0(2)
C3-C4-C5	105.1(2)	F1-C4-H4	111.8
C3-C4-H4	111.8	C5-C4-H4	111.8
N1-C5-C4	103.9(2)	N1-C5-H5A	111.0
C4-C5-H5A	111.0	N1-C5-H5B	111.0
C4-C5-H5B	111.0	H5A-C5-H5B	109.0
O3-C6-N1	121.8(2)	O3-C6-C7	121.3(2)
N1-C6-C7	116.7(2)	N2-C7-C8	110.9(2)
N2-C7-C6	112.2(2)	C8-C7-C6	106.30(19)
N2-C7-H7	109.1	C8-C7-H7	109.1
C6-C7-H7	109.1	C10-C8-C7	115.9(2)
C10-C8-H8A	108.3	C7-C8-H8A	108.3
C10-C8-H8B	108.3	C7-C8-H8B	108.3
H8A-C8-H8B	107.4	C10-C9-N3	110.4(3)
C10-C9-H9	124.8	N3-C9-H9	124.8
C9-C10-C11	106.5(2)	C9-C10-C8	126.0(3)
C11-C10-C8	127.1(2)	C12-C11-C16	118.5(3)
C12-C11-C10	135.0(2)	C16-C11-C10	106.5(3)
C13-C12-C11	119.3(3)	C13-C12-H12	120.3
C11-C12-H12	120.3	C12-C13-C14	120.6(3)
C12-C13-H13	119.7	C14-C13-H13	119.7
C15-C14-C13	121.7(3)	C15-C14-H14	119.1
C13-C14-H14	119.1	C14-C15-C16	118.0(3)

Table E26 continued

C14-C15-H15	121.0	C16-C15-H15	121.0
N3-C16-C15	130.4(3)	N3-C16-C11	107.7(3)
C15-C16-C11	121.9(3)	O4-C17-O5	126.4(2)
O4-C17-N2	123.7(2)	O5-C17-N2	109.90(19)
O5-C18-C21	109.2(2)	O5-C18-C19	100.8(2)
C21-C18-C19	112.1(3)	O5-C18-C20	109.8(2)
C21-C18-C20	113.7(3)	C19-C18-C20	110.5(3)
C18-C19-H19A	109.5	C18-C19-H19B	109.5
H19A-C19-H19B	109.5	C18-C19-H19C	109.5
H19A-C19-H19C	109.5	H19B-C19-H19C	109.5
C18-C20-H20A	109.5	C18-C20-H20B	109.5
H20A-C20-H20B	109.5	C18-C20-H20C	109.5
H20A-C20-H20C	109.5	H20B-C20-H20C	109.5
C18-C21-H21A	109.5	C18-C21-H21B	109.5
H21A-C21-H21B	109.5	C18-C21-H21C	109.5
H21A-C21-H21C	109.5	H21B-C21-H21C	109.5
Cl1-C22-Cl2	110.4(2)	Cl1-C22-Cl3	109.54(18)
Cl2-C22-Cl3	111.1(2)	Cl1-C22-H22	108.6
Cl2-C22-H22	108.6	Cl3-C22-H22	108.6

Table E27. Anisotropic atomic displacement parameters (Å2) for Boc-Trpflp-OH.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Cl1	0.0697(8)	0.1157(10)	0.0792(7)	-0.0256(7)	-0.0113(6)	0.0086(7)
Cl2	0.1086(9)	0.0768(7)	0.0434(4)	-0.0048(4)	-0.0117(5)	0.0145(7)
Cl3	0.0877(8)	0.0819(7)	0.0548(5)	-0.0071(5)	0.0149(5)	0.0032(7)
F1	0.0262(9)	0.0390(9)	0.0538(9)	0.0004(7)	0.0033(7)	0.0035(8)
N1	0.0221(11)	0.0230(10)	0.0241(9)	0.0025(8)	0.0001(8)	0.0016(9)
N2	0.0192(11)	0.0263(11)	0.0252(9)	-0.0041(8)	-0.0029(8)	0.0042(9)
N3	0.0432(17)	0.0226(12)	0.0791(19)	0.0112(13)	0.0002(14)	0.0004(13)
O1	0.0430(12)	0.0283(10)	0.0342(9)	0.0010(8)	-0.0126(8)	0.0031(9)
O2	0.0328(12)	0.0252(9)	0.0420(10)	0.0014(8)	-0.0145(8)	-0.0030(9)
O3	0.0304(10)	0.0294(10)	0.0299(8)	0.0036(7)	0.0050(7)	0.0052(8)
O4	0.0333(11)	0.0302(10)	0.0426(10)	-0.0098(8)	-0.0072(8)	0.0032(9)
O5	0.0170(10)	0.0341(10)	0.0443(10)	-0.0081(8)	-0.0061(8)	0.0028(8)
C1	0.0280(14)	0.0230(12)	0.0244(10)	0.0001(9)	0.0005(10)	0.0017(11)
C2	0.0224(13)	0.0196(12)	0.0322(12)	0.0022(9)	0.0006(10)	0.0002(10)
C3	0.0277(14)	0.0225(12)	0.0383(13)	-0.0041(10)	-0.0047(11)	-0.0002(11)
C4	0.0280(15)	0.0292(13)	0.0336(12)	-0.0066(10)	-0.0018(11)	0.0050(12)
C5	0.0294(15)	0.0328(14)	0.0249(11)	-0.0011(10)	0.0021(10)	0.0071(12)
C6	0.0179(13)	0.0239(12)	0.0272(11)	-0.0005(9)	-0.0034(9)	-0.0007(10)
C7	0.0196(13)	0.0235(12)	0.0253(10)	0.0012(9)	-0.0003(9)	0.0037(10)
C8	0.0182(13)	0.0272(13)	0.0370(13)	0.0035(10)	-0.0045(10)	0.0016(11)
C9	0.0271(16)	0.0320(15)	0.0566(17)	0.0048(13)	-0.0015(13)	0.0024(13)
C10	0.0160(13)	0.0276(13)	0.0465(14)	0.0057(11)	-0.0025(11)	0.0034(11)
C11	0.0162(13)	0.0314(14)	0.0490(15)	0.0142(11)	0.0019(11)	0.0064(11)
C12	0.0359(17)	0.0369(15)	0.0474(15)	0.0115(13)	0.0075(13)	0.0055(14)
C13	0.054(2)	0.057(2)	0.0479(17)	0.0104(15)	0.0154(15)	0.0067(18)
C14	0.051(2)	0.075(3)	0.059(2)	0.030(2)	0.0179(18)	0.010(2)
C15	0.0345(18)	0.050(2)	0.079(2)	0.0358(19)	0.0116(17)	0.0043(16)
C16	0.0243(16)	0.0330(15)	0.067(2)	0.0188(14)	0.0024(13)	0.0035(13)
C17	0.0211(13)	0.0206(11)	0.0282(11)	0.0041(9)	-0.0017(10)	0.0002(10)

Table E27 continued

C18	0.0153(14)	0.0442(16)	0.0482(15)	-0.0013(13)	-0.0057(11)	-0.0017(13)
C19	0.0191(17)	0.089(3)	0.083(3)	-0.026(2)	-0.0067(16)	0.0146(18)
C20	0.0310(18)	0.058(2)	0.0552(18)	0.0044(16)	-0.0170(14)	-0.0040(16)
C21	0.036(2)	0.066(2)	0.079(2)	0.015(2)	-0.0011(17)	-0.0174(18)
C22	0.079(3)	0.0394(17)	0.0457(16)	-0.0072(14)	-0.0074(17)	0.0061(18)

Table E28. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-Trpflp-OH.

	x/a	y/b	z/c	U(eq)
H2N	0.370(3)	0.4070(19)	0.8878(18)	0.028
H3N	0.692(4)	0.661(3)	0.769(2)	0.058
H2O	0.825(4)	0.209(2)	0.991(2)	0.05
H2A	0.5416	0.0825	0.9365	0.03
H3A	0.7105	-0.0103	0.8605	0.035
H3B	0.5740	0.0035	0.8188	0.035
H4	0.7070	0.0595	0.7190	0.036
H5A	0.6920	0.2271	0.7617	0.035
H5B	0.5682	0.1744	0.7314	0.035
H7	0.5093	0.3146	0.7740	0.027
H8A	0.5983	0.4112	0.9136	0.033
H8B	0.6898	0.3528	0.8580	0.033
H9	0.6369	0.5738	0.8838	0.046
H12	0.6481	0.3436	0.6695	0.048
H13	0.6960	0.3848	0.5356	0.064
H14	0.7440	0.5321	0.5039	0.074
H15	0.7479	0.6405	0.6034	0.065
H19A	-0.0082	0.4490	0.9017	0.096
H19B	-0.1144	0.4462	0.8325	0.096
H19C	0.0034	0.5120	0.8237	0.096
H20A	0.0627	0.4535	0.6923	0.072
H20B	-0.0508	0.3835	0.6912	0.072
H20C	0.0915	0.3502	0.6799	0.072
H21A	0.0661	0.2463	0.7996	0.09
H21B	-0.0748	0.2793	0.8174	0.09
H21C	0.0307	0.2863	0.8869	0.09
H22	0.3035	0.4167	0.4751	0.066

Table E29. Torsion angles (°) for Boc-Trpflp-OH.

C6-N1-C2-C1	-82.2(3)	C5-N1-C2-C1	98.9(2)
C6-N1-C2-C3	158.4(2)	C5-N1-C2-C3	-20.5(3)
O1-C1-C2-N1	-170.7(2)	O2-C1-C2-N1	9.8(3)
O1-C1-C2-C3	-56.1(3)	O2-C1-C2-C3	124.4(2)
N1-C2-C3-C4	34.5(3)	C1-C2-C3-C4	-86.5(2)
C2-C3-C4-F1	78.6(2)	C2-C3-C4-C5	-36.6(3)
C6-N1-C5-C4	179.4(2)	C2-N1-C5-C4	-1.8(3)
F1-C4-C5-N1	-91.2(2)	C3-C4-C5-N1	24.0(3)
C2-N1-C6-O3	-4.1(4)	C5-N1-C6-O3	174.6(2)
C2-N1-C6-C7	170.2(2)	C5-N1-C6-C7	-11.1(4)
C17-N2-C7-C8	157.9(2)	C17-N2-C7-C6	-83.3(3)
O3-C6-C7-N2	-33.9(3)	N1-C6-C7-N2	151.8(2)
O3-C6-C7-C8	87.5(3)	N1-C6-C7-C8	-86.9(3)
N2-C7-C8-C10	-68.4(3)	C6-C7-C8-C10	169.4(2)
C16-N3-C9-C10	0.6(4)	N3-C9-C10-C11	-0.2(3)
N3-C9-C10-C8	172.5(3)	C7-C8-C10-C9	128.6(3)
C7-C8-C10-C11	-60.1(4)	C9-C10-C11-C12	-178.5(3)
C8-C10-C11-C12	8.9(5)	C9-C10-C11-C16	-0.3(3)
C8-C10-C11-C16	-172.9(2)	C16-C11-C12-C13	-0.7(4)
C10-C11-C12-C13	177.3(3)	C11-C12-C13-C14	0.6(5)
C12-C13-C14-C15	0.4(6)	C13-C14-C15-C16	-1.2(6)
C9-N3-C16-C15	178.5(3)	C9-N3-C16-C11	-0.7(3)
C14-C15-C16-N3	-178.0(4)	C14-C15-C16-C11	1.1(5)
C12-C11-C16-N3	179.2(3)	C10-C11-C16-N3	0.6(3)
C12-C11-C16-C15	-0.1(4)	C10-C11-C16-C15	-178.7(3)
C18-O5-C17-O4	11.8(4)	C18-O5-C17-N2	-170.6(2)
C7-N2-C17-O4	-2.9(4)	C7-N2-C17-O5	179.5(2)
C17-O5-C18-C21	57.6(3)	C17-O5-C18-C19	175.8(3)
C17-O5-C18-C20	-67.6(3)		

Crystal Structure of Boc-Trpflp-NHMe

Crystals were mounted using viscous oil onto a plastic mesh and cooled to the data collection temperature. Data were collected on a Bruker-AXS APEX II DUO CCD diffractometer with Mo-Kα radiation (λ = 0.71073 Å) monochromated with graphite. Unit cell parameters were obtained from 36 data frames, 0.5º ω, from three different sections of the Ewald sphere. The systematic absences in the diffraction data are uniquely consistent for $P2_12_12_1$. The data-set was treated with multi-scan absorption corrections (Apex3 software suite, Madison, WI, 2015). The structure was solved using direct methods and refined with full-matrix, least-squares procedures on F^2 (Sheldrick, G.M. 2008. Acta Cryst. A64, 112-122). Refinement of the absolute structure parameter yielded nil indicating the true hand of the data has been determined. All non-hydrogen atoms were refined with anisotropic displacement parameters. H-atoms were placed in calculated positions with U_{iso} equal to 1.2 (1.5 for methyl H) U_{eq} of the attached atom. Atomic scattering factors are contained in the SHELXTL program library (Sheldrick, G., *op. cit*.). The CIFs have been deposited with the Cambridge Crystallographic Database under CCDC 1442051.

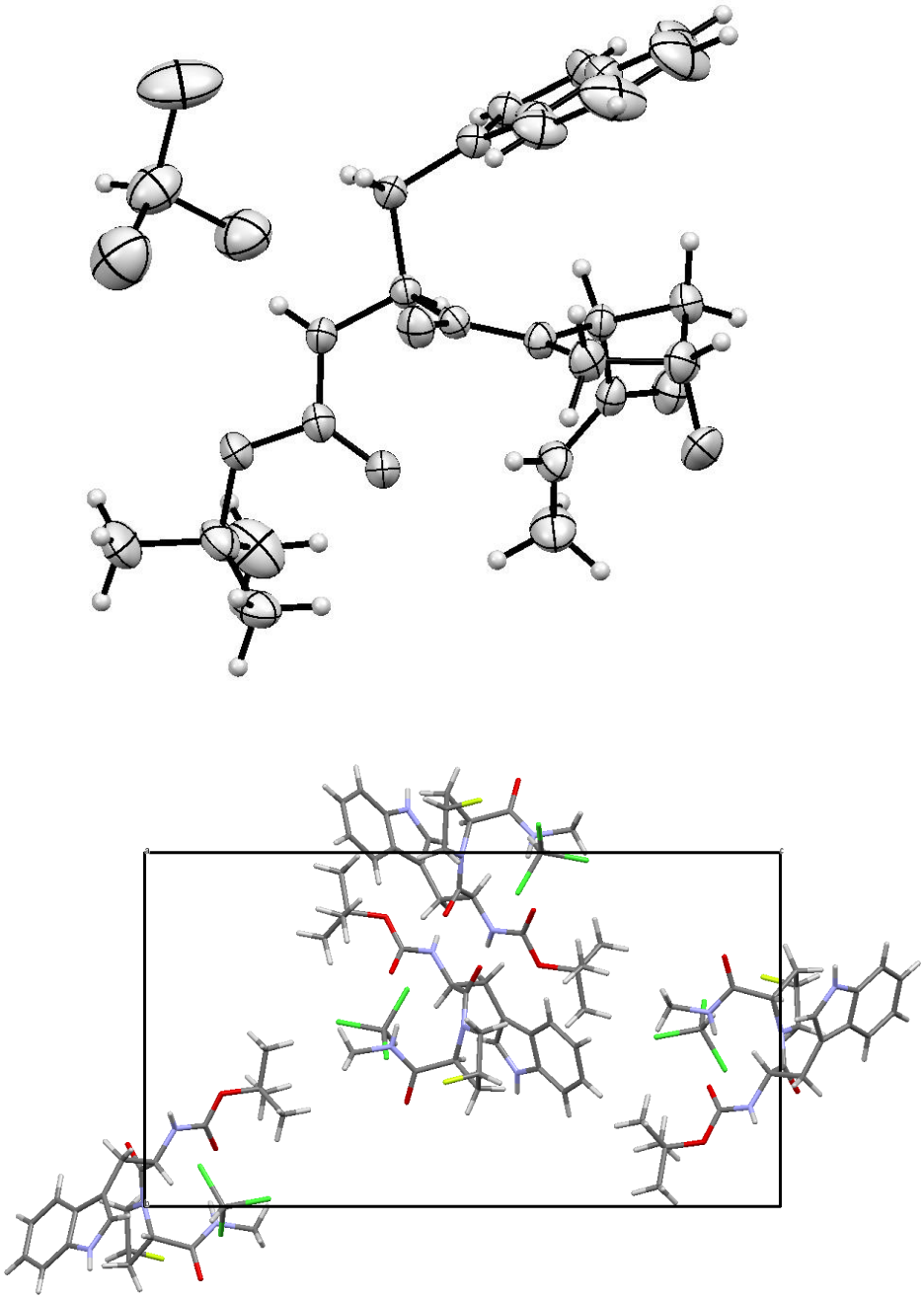


Figure E6 Crystal structure of Boc-Trpflp-NHMe.
Top: ORTEP diagram of the Boc-Trpflp-NHMe monomer with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in $CDCl_3$.

Table E30. Crystallographic data and refinement details for Boc-Trpflp-NHMe.

empirical formula	$C_{23}H_{30}Cl_3FN_4O_4$	
formula weight	551.86	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, P 2_1 2_1 2_1	
Unit cell dimensions (Å, °)	a = 9.1475(3)	α = 90
	b = 13.0425(5)	β = 90
	c = 23.5901(8)	γ = 90
Volume (Å^3)	2814.45(17)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.302	
absorption coefficient (mm^{-1})	0.366	
F(000)	1152	
crystal size (mm)	0.537 x 0.332 x 0.188	
θ range for data collection	1.726 to 28.445°	
Index ranges	$-12 \le h \le 12$, $-17 \le k \le 17$, $-29 \le l \le 31$	
Reflections collected/ unique	42018/7074 [R(int) = 0.0260]	
Coverage of independent reflections	99.9%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7457 and 0.6868	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	7074 / 0 / 329	
Goodness-of-fit on F^2	1.045	
Final *R* indices	6180 data; I>2σ(I)	R1 = 0.0453, wR2 = 0.1282
	all data	R1 = 0.0522, wR2 = 0.1345
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0861P)^2+0.3346P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.460 and -0.467 eÅ^{-3}	
R.M.S. deviation from mean	0.054 eÅ^{-3}	

Table E31. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-Trpflp-NHMe.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
C(1)	6894(5)	5659(3)	3273(1)	54(1)
C(2)	6601(3)	6182(2)	4256(1)	33(1)
C(3)	8780(3)	5092(2)	5283(1)	38(1)
C(4)	8963(3)	6245(2)	5254(1)	42(1)
C(5)	7423(3)	6658(2)	5250(1)	38(1)
C(6)	6593(3)	5880(2)	4882(1)	31(1)
C(7)	6775(2)	3954(2)	5054(1)	28(1)
C(8)	5217(2)	3759(2)	4840(1)	28(1)
C(9)	4173(3)	3673(2)	5350(1)	32(1)
C(10)	2879(3)	5371(2)	5531(1)	37(1)
C(11)	3952(3)	4682(2)	5647(1)	30(1)
C(12)	4832(3)	5128(2)	6083(1)	36(1)
C(13)	6068(3)	4812(3)	6385(1)	51(1)
C(14)	6647(4)	5482(4)	6784(2)	75(1)
C(15)	6036(5)	6432(4)	6890(2)	86(2)
C(16)	4813(4)	6761(3)	6603(2)	66(1)
C(17)	4224(3)	6096(2)	6205(1)	42(1)
C(18)	5986(3)	2717(2)	4051(1)	36(1)
C(19)	6179(4)	1670(2)	3193(1)	46(1)
C(20)	7811(4)	1787(3)	3116(1)	50(1)
C(21)	5336(5)	2407(4)	2823(2)	74(1)
C(22)	5744(6)	560(3)	3089(2)	73(1)
C(23)	6296(5)	-23(3)	6243(2)	67(1)
Cl(1)	6551(1)	1132(1)	5871(1)	76(1)
Cl(2)	5816(2)	212(2)	6948(1)	111(1)
Cl(3)	7875(2)	-791(1)	6196(1)	94(1)
F(1)	9637(2)	6485(2)	4733(1)	56(1)
N(1)	7014(3)	5482(2)	3880(1)	39(1)
N(2)	7339(2)	4912(2)	5018(1)	30(1)

Table E31 continued

N(3)	5194(2)	2804(2)	4524(1)	34(1)
N(4)	3029(3)	6225(2)	5864(1)	44(1)
O(1)	6181(2)	7050(1)	4122(1)	43(1)
O(2)	7477(2)	3253(1)	5268(1)	37(1)
O(3)	6842(2)	3367(1)	3885(1)	44(1)
O(4)	5741(2)	1816(2)	3792(1)	44(1)

Table E32. Bond lengths [Å] and angles [°] for Boc-Trpflp-NHMe.

Bond lengths

C(1)-N(1)	1.455(4)	C(12)-C(17)	1.409(4)
C(1)-H(1A)	0.9800	C(13)-C(14)	1.390(5)
C(1)-H(1B)	0.9800	C(13)-H(13)	0.9500
C(1)-H(1C)	0.9800	C(14)-C(15)	1.382(7)
C(2)-O(1)	1.237(3)	C(14)-H(14)	0.9500
C(2)-N(1)	1.329(4)	C(15)-C(16)	1.375(7)
C(2)-C(6)	1.528(3)	C(15)-H(15)	0.9500
C(3)-N(2)	1.478(3)	C(16)-C(17)	1.387(4)
C(3)-C(4)	1.515(4)	C(16)-H(16)	0.9500
C(3)-H(3A)	0.9900	C(17)-N(4)	1.366(4)
C(3)-H(3B)	0.9900	C(18)-O(3)	1.218(3)
C(4)-F(1)	1.411(4)	C(18)-N(3)	1.335(3)
C(4)-C(5)	1.508(4)	C(18)-O(4)	1.344(3)
C(4)-H(4)	1.0000	C(19)-O(4)	1.481(3)
C(5)-C(6)	1.535(3)	C(19)-C(21)	1.510(5)
C(5)-H(5A)	0.9900	C(19)-C(20)	1.511(5)
C(5)-H(5B)	0.9900	C(19)-C(22)	1.521(5)
C(6)-N(2)	1.471(3)	C(20)-H(20A)	0.9800
C(6)-H(6)	1.0000	C(20)-H(20B)	0.9800
C(7)-O(2)	1.225(3)	C(20)-H(20C)	0.9800
C(7)-N(2)	1.355(3)	C(21)-H(21A)	0.9800
C(7)-C(8)	1.533(3)	C(21)-H(21B)	0.9800
C(8)-N(3)	1.453(3)	C(21)-H(21C)	0.9800
C(8)-C(9)	1.539(3)	C(22)-H(22A)	0.9800
C(8)-H(8)	1.0000	C(22)-H(22B)	0.9800
C(9)-C(11)	1.505(3)	C(22)-H(22C)	0.9800
C(9)-H(9A)	0.9900	C(23)-Cl(2)	1.748(4)
C(9)-H(9B)	0.9900	C(23)-Cl(1)	1.758(4)
C(10)-C(11)	1.359(3)	C(23)-Cl(3)	1.761(5)
C(10)-N(4)	1.370(4)	C(23)-H(23)	1.0000
C(10)-H(10)	0.9500	N(1)-H(1N)	0.78(4)
C(11)-C(12)	1.430(3)	N(3)-H(3N)	0.81(4)
C(12)-C(13)	1.398(4)	N(4)-H(4N)	0.86(4)

Table E32 continued

Bond angles

N(1)-C(1)-H(1A)	109.5	C(12)-C(13)-H(13)	121.1
N(1)-C(1)-H(1B)	109.5	C(15)-C(14)-C(13)	122.2(4)
H(1A)-C(1)-H(1B)	109.5	C(15)-C(14)-H(14)	118.9
N(1)-C(1)-H(1C)	109.5	C(13)-C(14)-H(14)	118.9
H(1A)-C(1)-H(1C)	109.5	C(16)-C(15)-C(14)	121.3(3)
H(1B)-C(1)-H(1C)	109.5	C(16)-C(15)-H(15)	119.4
O(1)-C(2)-N(1)	123.1(2)	C(14)-C(15)-H(15)	119.4
O(1)-C(2)-C(6)	118.8(2)	C(15)-C(16)-C(17)	117.0(4)
N(1)-C(2)-C(6)	118.1(2)	C(15)-C(16)-H(16)	121.5
N(2)-C(3)-C(4)	103.7(2)	C(17)-C(16)-H(16)	121.5
N(2)-C(3)-H(3A)	111.0	N(4)-C(17)-C(16)	129.2(3)
C(4)-C(3)-H(3A)	111.0	N(4)-C(17)-C(12)	107.8(2)
N(2)-C(3)-H(3B)	111.0	C(16)-C(17)-C(12)	123.0(3)
C(4)-C(3)-H(3B)	111.0	O(3)-C(18)-N(3)	123.9(2)
H(3A)-C(3)-H(3B)	109.0	O(3)-C(18)-O(4)	124.7(2)
F(1)-C(4)-C(5)	108.8(3)	N(3)-C(18)-O(4)	111.3(2)
F(1)-C(4)-C(3)	107.9(2)	O(4)-C(19)-C(21)	109.3(3)
C(5)-C(4)-C(3)	104.6(2)	O(4)-C(19)-C(20)	111.6(2)
F(1)-C(4)-H(4)	111.7	C(21)-C(19)-C(20)	111.8(3)
C(5)-C(4)-H(4)	111.7	O(4)-C(19)-C(22)	101.8(3)
C(3)-C(4)-H(4)	111.7	C(21)-C(19)-C(22)	112.3(3)
C(4)-C(5)-C(6)	103.3(2)	C(20)-C(19)-C(22)	109.6(3)
C(4)-C(5)-H(5A)	111.1	C(19)-C(20)-H(20A)	109.5
C(6)-C(5)-H(5A)	111.1	C(19)-C(20)-H(20B)	109.5
C(4)-C(5)-H(5B)	111.1	H(20A)-C(20)-H(20B)	109.5
C(6)-C(5)-H(5B)	111.1	C(19)-C(20)-H(20C)	109.5
H(5A)-C(5)-H(5B)	109.1	H(20A)-C(20)-H(20C)	109.5
N(2)-C(6)-C(2)	115.37(19)	H(20B)-C(20)-H(20C)	109.5
N(2)-C(6)-C(5)	102.38(19)	C(19)-C(21)-H(21A)	109.5
C(2)-C(6)-C(5)	111.9(2)	C(19)-C(21)-H(21B)	109.5
N(2)-C(6)-H(6)	109.0	H(21A)-C(21)-H(21B)	109.5
C(2)-C(6)-H(6)	109.0	C(19)-C(21)-H(21C)	109.5
C(5)-C(6)-H(6)	109.0	H(21A)-C(21)-H(21C)	109.5
O(2)-C(7)-N(2)	121.0(2)	H(21B)-C(21)-H(21C)	109.5
O(2)-C(7)-C(8)	119.9(2)	C(19)-C(22)-H(22A)	109.5
N(2)-C(7)-C(8)	119.1(2)	C(19)-C(22)-H(22B)	109.5

Table E32 continued

N(3)-C(8)-C(7)	108.93(18)	H(22A)-C(22)-H(22B)	109.5
N(3)-C(8)-C(9)	109.23(19)	C(19)-C(22)-H(22C)	109.5
C(7)-C(8)-C(9)	109.41(19)	H(22A)-C(22)-H(22C)	109.5
N(3)-C(8)-H(8)	109.8	H(22B)-C(22)-H(22C)	109.5
C(7)-C(8)-H(8)	109.8	Cl(2)-C(23)-Cl(1)	111.0(3)
C(9)-C(8)-H(8)	109.8	Cl(2)-C(23)-Cl(3)	111.4(2)
C(11)-C(9)-C(8)	112.52(19)	Cl(1)-C(23)-Cl(3)	110.3(2)
C(11)-C(9)-H(9A)	109.1	Cl(2)-C(23)-H(23)	108
C(8)-C(9)-H(9A)	109.1	Cl(1)-C(23)-H(23)	108
C(11)-C(9)-H(9B)	109.1	Cl(3)-C(23)-H(23)	108
C(8)-C(9)-H(9B)	109.1	C(2)-N(1)-C(1)	121.9(2)
H(9A)-C(9)-H(9B)	107.8	C(2)-N(1)-H(1N)	122(3)
C(11)-C(10)-N(4)	110.5(2)	C(1)-N(1)-H(1N)	115(3)
C(11)-C(10)-H(10)	124.8	C(7)-N(2)-C(6)	128.98(19)
N(4)-C(10)-H(10)	124.8	C(7)-N(2)-C(3)	117.3(2)
C(10)-C(11)-C(12)	106.4(2)	C(6)-N(2)-C(3)	111.73(18)
C(10)-C(11)-C(9)	125.6(2)	C(18)-N(3)-C(8)	119.6(2)
C(12)-C(11)-C(9)	128.0(2)	C(18)-N(3)-H(3N)	121(2)
C(13)-C(12)-C(17)	118.7(3)	C(8)-N(3)-H(3N)	117(2)
C(13)-C(12)-C(11)	134.5(3)	C(17)-N(4)-C(10)	108.5(2)
C(17)-C(12)-C(11)	106.8(2)	C(17)-N(4)-H(4N)	125(3)
C(14)-C(13)-C(12)	117.8(3)	C(10)-N(4)-H(4N)	127(3)
C(14)-C(13)-H(13)	121.1	C(18)-O(4)-C(19)	120.1(2)

Table E33. Anisotropic atomic displacement parameters (Å2) for Boc-Trpflp-NHMe.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	76(2)	52(2)	35(1)	4(1)	-2(1)	-5(2)
C(2)	30(1)	27(1)	42(1)	1(1)	-2(1)	-6(1)
C(3)	26(1)	41(1)	47(1)	-5(1)	-7(1)	-3(1)
C(4)	36(1)	44(1)	48(1)	-11(1)	-2(1)	-12(1)
C(5)	42(1)	32(1)	40(1)	-6(1)	0(1)	-5(1)
C(6)	30(1)	25(1)	38(1)	-2(1)	1(1)	-3(1)
C(7)	28(1)	26(1)	30(1)	-3(1)	0(1)	1(1)
C(8)	27(1)	25(1)	33(1)	-3(1)	0(1)	-1(1)
C(9)	31(1)	27(1)	39(1)	-2(1)	6(1)	-2(1)
C(10)	32(1)	34(1)	44(1)	-2(1)	2(1)	1(1)
C(11)	30(1)	29(1)	32(1)	-1(1)	3(1)	0(1)
C(12)	36(1)	41(1)	30(1)	-4(1)	3(1)	-2(1)
C(13)	47(2)	74(2)	33(1)	-4(1)	-2(1)	8(2)
C(14)	52(2)	129(4)	44(2)	-23(2)	-10(2)	5(2)
C(15)	77(3)	122(4)	59(2)	-51(2)	-3(2)	-12(3)
C(16)	68(2)	65(2)	65(2)	-36(2)	9(2)	-10(2)
C(17)	45(1)	39(1)	43(1)	-10(1)	9(1)	-4(1)
C(18)	39(1)	29(1)	39(1)	-5(1)	0(1)	-3(1)
C(19)	63(2)	42(1)	33(1)	-10(1)	-1(1)	0(1)
C(20)	62(2)	52(2)	34(1)	1(1)	9(1)	7(1)
C(21)	81(3)	83(3)	59(2)	-9(2)	-25(2)	20(2)
C(22)	102(3)	60(2)	57(2)	-27(2)	14(2)	-23(2)
C(23)	64(2)	81(3)	57(2)	14(2)	-4(2)	-9(2)
Cl(1)	86(1)	76(1)	66(1)	13(1)	10(1)	4(1)
Cl(2)	99(1)	175(2)	57(1)	28(1)	14(1)	12(1)
Cl(3)	87(1)	86(1)	110(1)	19(1)	3(1)	8(1)
F(1)	46(1)	56(1)	68(1)	-6(1)	14(1)	-18(1)
N(1)	50(1)	32(1)	36(1)	3(1)	-1(1)	-2(1)
N(2)	25(1)	28(1)	35(1)	-2(1)	-3(1)	-2(1)
N(3)	33(1)	28(1)	40(1)	-6(1)	4(1)	-7(1)
N(4)	45(1)	30(1)	56(1)	-4(1)	5(1)	6(1)

Table E33 continued

O(1)	48(1)	29(1)	53(1)	6(1)	-3(1)	-1(1)
O(2)	34(1)	31(1)	47(1)	0(1)	-3(1)	7(1)
O(3)	51(1)	33(1)	47(1)	-8(1)	15(1)	-10(1)
O(4)	56(1)	35(1)	41(1)	-14(1)	12(1)	-12(1)

Table E34. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-Trpflp-NHMe.

	x/a	y/b	z/c	U(eq)
H(1A)	5957	5991	3189	81
H(1B)	6947	5002	3072	81
H(1C)	7696	6102	3147	81
H(3A)	9565	4739	5070	46
H(3B)	8790	4849	5681	46
H(4)	9536	6513	5584	51
H(5A)	7388	7352	5081	46
H(5B)	7014	6686	5638	46
H(6)	5558	5839	5017	37
H(8)	4902	4337	4590	34
H(9A)	3216	3413	5217	39
H(9B)	4573	3169	5623	39
H(10)	2131	5275	5257	44
H(13)	6499	4160	6319	61
H(14)	7491	5281	6990	90
H(15)	6468	6867	7166	103
H(16)	4390	7413	6675	79
H(20A)	8081	1592	2729	74
H(20B)	8322	1344	3387	74
H(20C)	8088	2503	3183	74
H(21A)	4285	2316	2888	112
H(21B)	5559	2271	2424	112
H(21C)	5615	3112	2917	112
H(22A)	4692	479	3156	110
H(22B)	6287	113	3348	110
H(22C)	5971	373	2696	110
H(23)	5472	-403	6059	81
H(1N)	7180(40)	4910(30)	3968(15)	47
H(3N)	4550(40)	2390(30)	4601(14)	41
H(4N)	2480(40)	6760(30)	5865(15)	52

Table E35. Torsion angles (°) for Boc-Trpflp-NHMe.

N(2)-C(3)-C(4)-F(1)	-88.9(2)	C(15)-C(16)-C(17)-C(12)	-0.7(5)
N(2)-C(3)-C(4)-C(5)	26.8(3)	C(13)-C(12)-C(17)-N(4)	-179.3(3)
F(1)-C(4)-C(5)-C(6)	77.2(2)	C(11)-C(12)-C(17)-N(4)	0.6(3)
C(3)-C(4)-C(5)-C(6)	-37.9(3)	C(13)-C(12)-C(17)-C(16)	1.2(4)
O(1)-C(2)-C(6)-N(2)	-169.8(2)	C(11)-C(12)-C(17)-C(16)	-178.9(3)
N(1)-C(2)-C(6)-N(2)	12.4(3)	O(1)-C(2)-N(1)-C(1)	-4.9(4)
O(1)-C(2)-C(6)-C(5)	-53.3(3)	C(6)-C(2)-N(1)-C(1)	172.7(3)
N(1)-C(2)-C(6)-C(5)	129.0(2)	O(2)-C(7)-N(2)-C(6)	-166.9(2)
C(4)-C(5)-C(6)-N(2)	33.6(3)	C(8)-C(7)-N(2)-C(6)	11.1(3)
C(4)-C(5)-C(6)-C(2)	-90.5(3)	O(2)-C(7)-N(2)-C(3)	-4.4(3)
O(2)-C(7)-C(8)-N(3)	-44.7(3)	C(8)-C(7)-N(2)-C(3)	173.6(2)
N(2)-C(7)-C(8)-N(3)	137.2(2)	C(2)-C(6)-N(2)-C(7)	-92.5(3)
O(2)-C(7)-C(8)-C(9)	74.6(3)	C(5)-C(6)-N(2)-C(7)	145.7(2)
N(2)-C(7)-C(8)-C(9)	-103.5(2)	C(2)-C(6)-N(2)-C(3)	104.2(2)
N(3)-C(8)-C(9)-C(11)	-171.0(2)	C(5)-C(6)-N(2)-C(3)	-17.7(3)
C(7)-C(8)-C(9)-C(11)	69.8(2)	C(4)-C(3)-N(2)-C(7)	-170.8(2)
N(4)-C(10)-C(11)-C(12)	0.2(3)	C(4)-C(3)-N(2)-C(6)	-5.4(3)
N(4)-C(10)-C(11)-C(9)	-177.6(2)	O(3)-C(18)-N(3)-C(8)	6.5(4)
C(8)-C(9)-C(11)-C(10)	91.4(3)	O(4)-C(18)-N(3)-C(8)	-175.0(2)
C(8)-C(9)-C(11)-C(12)	-85.8(3)	C(7)-C(8)-N(3)-C(18)	-62.8(3)
C(10)-C(11)-C(12)-C(13)	179.4(3)	C(9)-C(8)-N(3)-C(18)	177.7(2)
C(9)-C(11)-C(12)-C(13)	-2.9(5)	C(16)-C(17)-N(4)-C(10)	178.9(3)
C(10)-C(11)-C(12)-C(17)	-0.5(3)	C(12)-C(17)-N(4)-C(10)	-0.5(3)
C(9)-C(11)-C(12)-C(17)	177.2(2)	C(11)-C(10)-N(4)-C(17)	0.2(3)
C(17)-C(12)-C(13)-C(14)	-1.0(4)	O(3)-C(18)-O(4)-C(19)	-18.0(4)
C(11)-C(12)-C(13)-C(14)	179.1(3)	N(3)-C(18)-O(4)-C(19)	163.5(2)
C(12)-C(13)-C(14)-C(15)	0.4(6)	C(21)-C(19)-O(4)-C(18)	-62.7(4)
C(13)-C(14)-C(15)-C(16)	0.1(7)	C(20)-C(19)-O(4)-C(18)	61.5(3)
C(14)-C(15)-C(16)-C(17)	0.0(7)	C(22)-C(19)-O(4)-C(18)	178.4(3)
C(15)-C(16)-C(17)-N(4)	179.9(4)		

Appendix F

NMR SPECTRA OF SYNTHESIZED COMPOUNDS

Compounds were described and utilized in Chapter 4: Synthesis of 2-Thiophenylalanine and its Application into Native Chemical Ligation of Peptides and Proteins

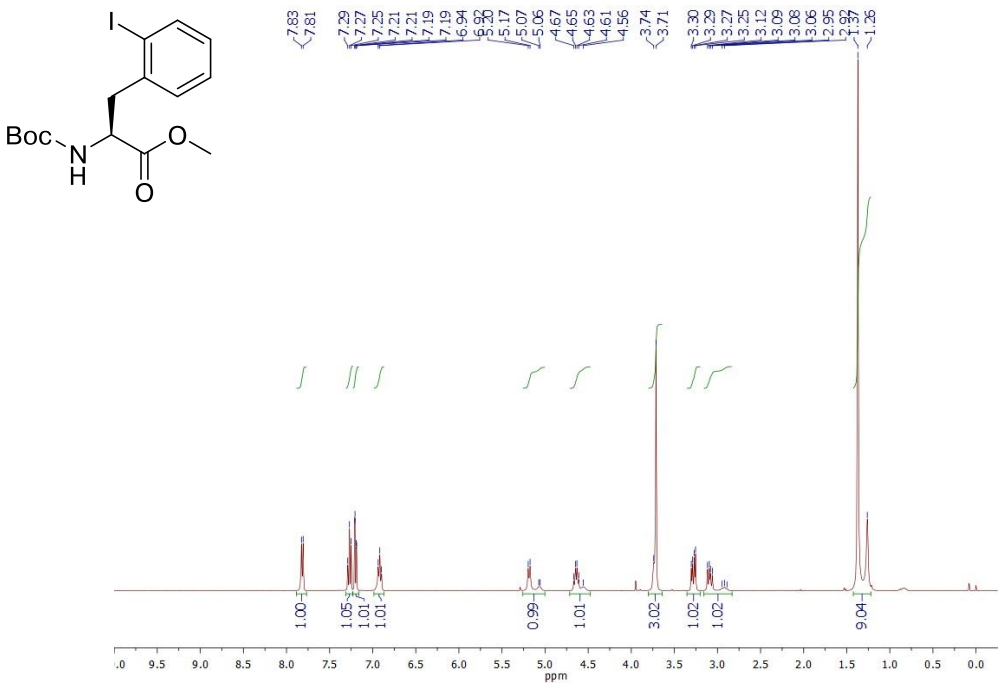


Figure F1 **^{1}H NMR spectrum of Boc-2-iodophenylalanine-methyl ester in $CDCl_3$.**

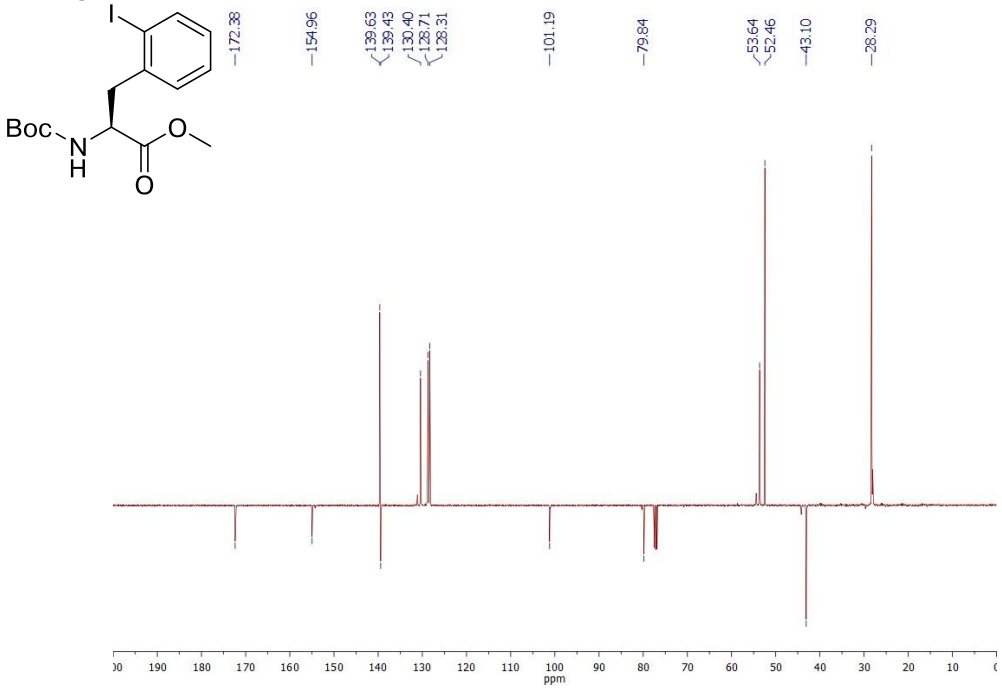


Figure F2 **^{13}C NMR spectrum of Boc-2-iodophenylalanine-methyl ester in $CDCl_3$.**

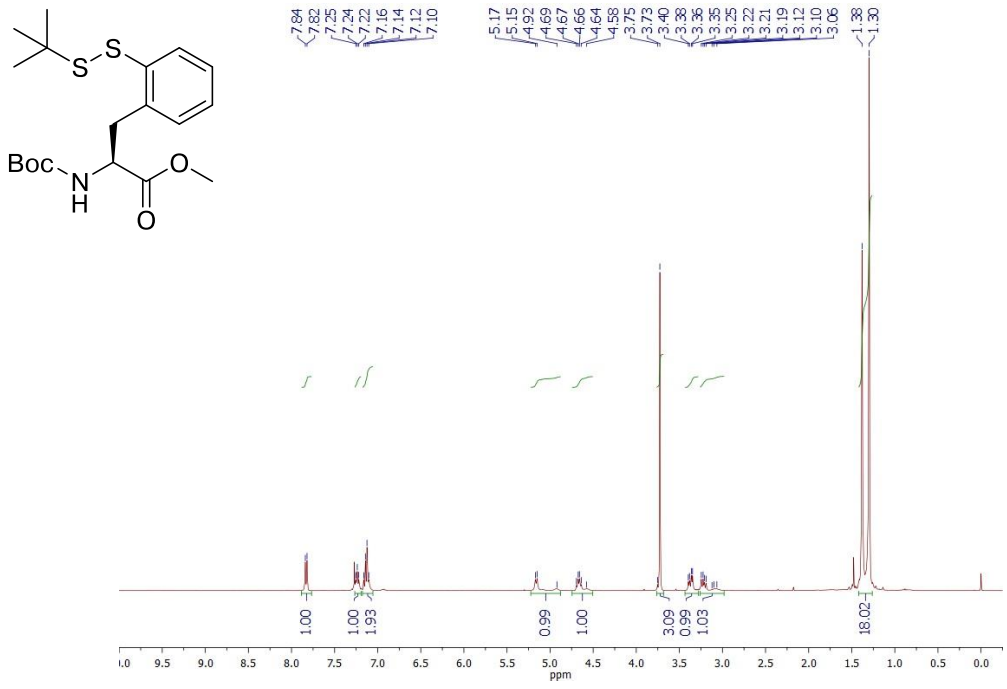


Figure F3 ^{1}H NMR spectrum of Boc-2-S(S-*tert*-butyl)thiophenylalanine-methyl ester in $CDCl_3$.

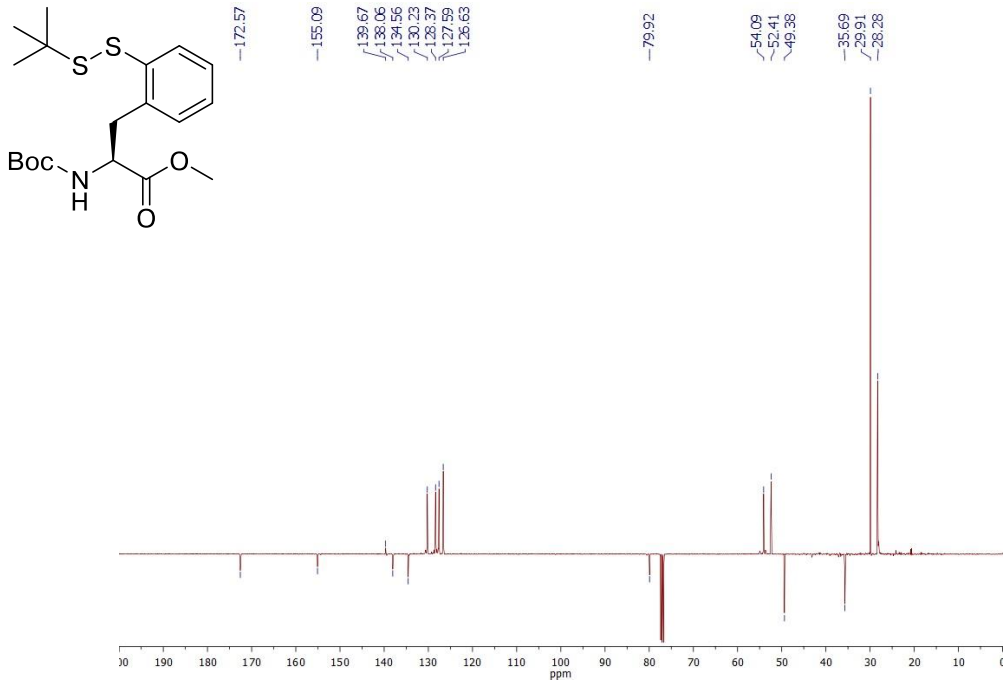


Figure F4 ^{13}C NMR spectrum of Boc-2-S(S-*tert*-butyl)thiophenylalanine-methyl ester in $CDCl_3$.

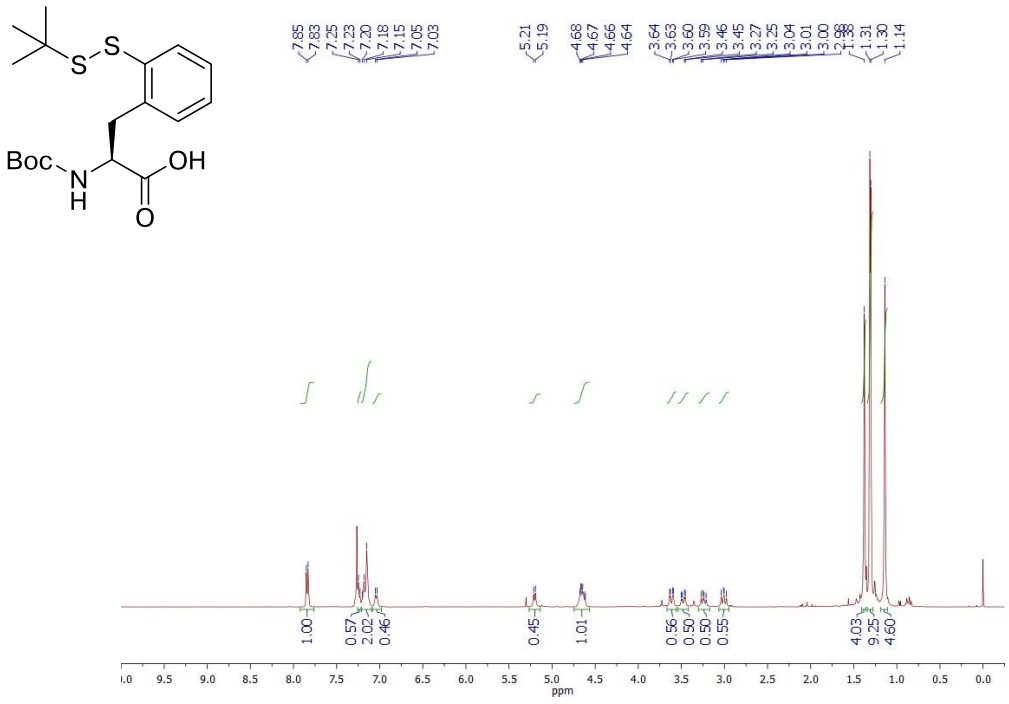


Figure F5 **^{1}H NMR spectrum of Boc-2-S(S-*tert*-butyl)-thiophenylalanine in $CDCl_3$.**

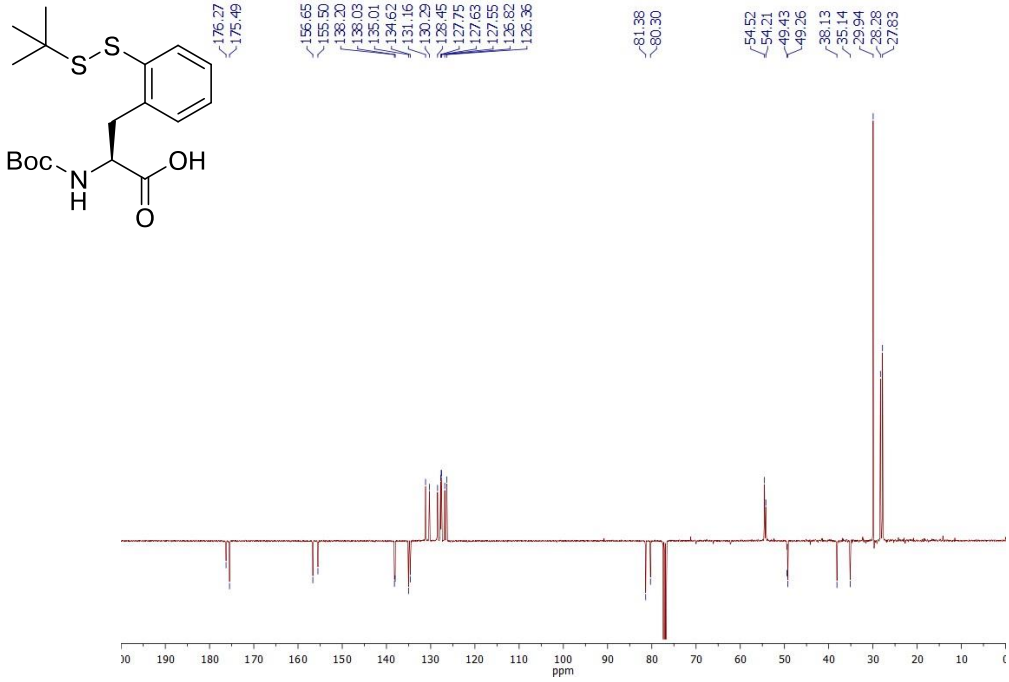


Figure F6 **^{13}C NMR spectrum of Boc-2-S(S-*tert*-butyl)-thiophenylalanine in $CDCl_3$.**

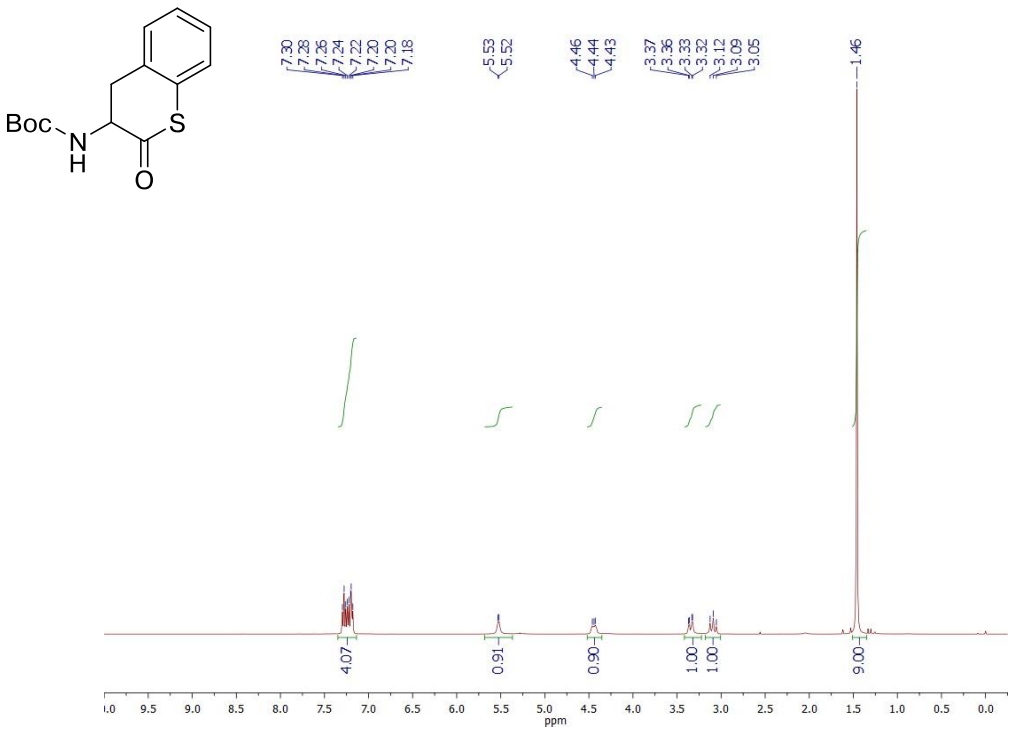


Figure F7 **^{1}H NMR spectrum of Boc-3-amino-δ-thiochromanone in $CDCl_3$.**

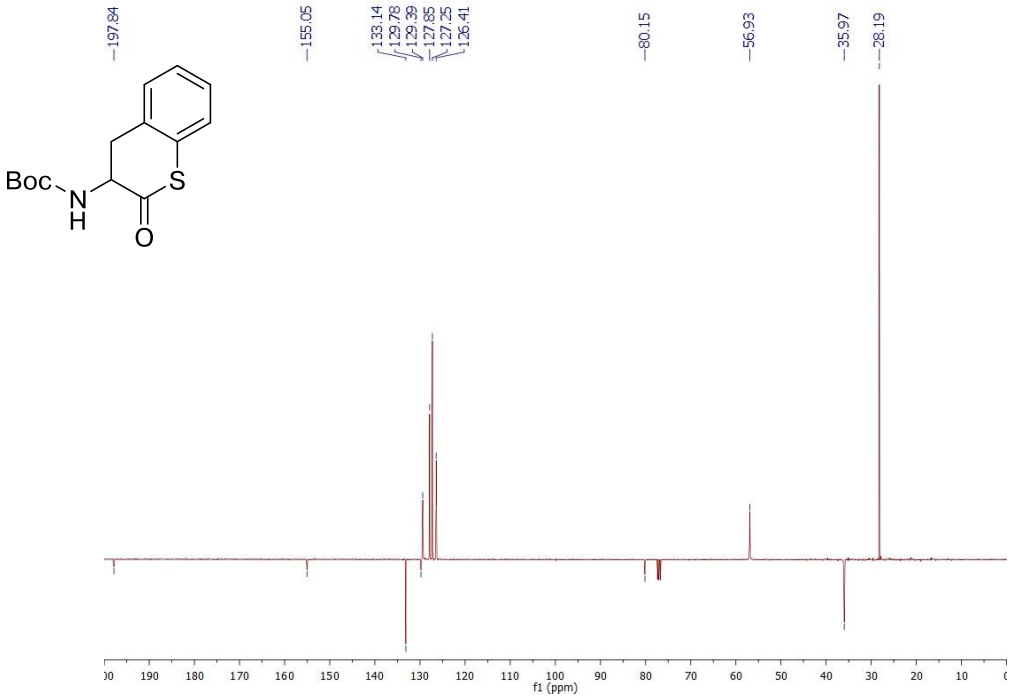


Figure F8 **^{13}C NMR spectrum of Boc-3-amino-δ-thiochromanone in $CDCl_3$.**

Appendix G

X-RAY CRYSTALLOGRAPHIC INFORMATION FOR SYNTHESIZED SMALL MOLECULES

Compounds were described and utilized in Chapter 4: Synthesis of 2-Thiophenylalanine and its Application into Native Chemical Ligation of Peptides and Proteins.

Crystal Structure of Boc-3-amino-δ-thiochromanone

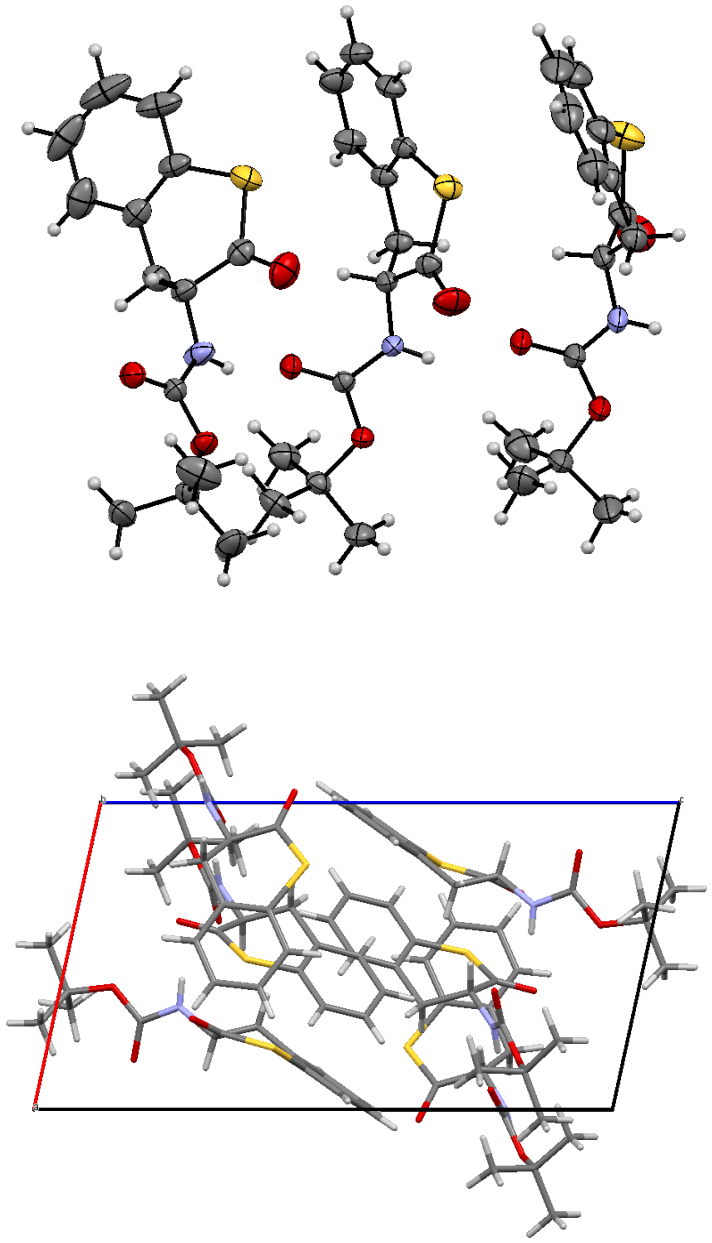


Figure G1 Crystal structure of Boc-3-amino-δ-thiochromanone.
Top: ORTEP diagram of the Boc-3-amino-δ-thiochromanone with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in $CDCl_3$.

Table G1. Crystallographic data and refinement details for Boc-3-amino-δ-thiochromanone.

empirical formula	$C_{14}H_{17}NO_3S$	
formula weight	279.35	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Triclinic, P $^{-1}$	
Unit cell dimensions (Å, °)	a = 9.9584(9)	α = 96.7990(10)
	b = 13.6429(12)	β = 99.3630(10)
	c = 17.4530(16)	γ = 108.3460(10)
Volume (Å^3)	2184.0(3)	
Z, Z′, calcd density (g/cm^3)	6, 0, 1.274	
absorption coefficient (mm^{-1})	0.225	
F(000)	888	
crystal size (mm)	0.398 x 0.422 x 0.439	
θ range for data collection	1.60 to 28.30°	
Index ranges	$-13 \le h \le 13$, $-18 \le k \le 18$, $-23 \le l \le 23$	
Reflections collected/ unique	59020/10808 [R(int) = 0.0641]	
Coverage of independent reflections	99.5%	
Absorption correction	multi-scan	
Max. and min. transmission	0.9160 and 0.9080	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	10808 / 0 / 532	
Goodness-of-fit on F^2	1.028	
Final *R* indices	7685 data; I>2σ(I)	R1 = 0.0474, wR2 = 0.1049
	all data	R1 = 0.0767, wR2 = 0.1230
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0437P)^2+1.1290P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.425 and -0.529 eÅ^{-3}	
R.M.S. deviation from mean	0.049 eÅ^{-3}	

Table G2. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-3-amino-δ-thiochromanone.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
C1	0.45159(18)	0.74420(13)	0.59281(10)	0.0289(4)
C2	0.3361(2)	0.74900(16)	0.53817(12)	0.0377(4)
C3	0.3025(2)	0.69278(17)	0.46216(12)	0.0454(5)
C4	0.3835(2)	0.63306(18)	0.44038(12)	0.0461(5)
C5	0.4987(2)	0.62950(16)	0.49480(11)	0.0389(4)
C6	0.53517(18)	0.68473(14)	0.57158(10)	0.0296(4)
C7	0.65885(19)	0.68066(15)	0.63171(10)	0.0330(4)
C8	0.61652(18)	0.66171(13)	0.71043(10)	0.0293(4)
C9	0.58102(19)	0.75334(15)	0.75001(11)	0.0332(4)
C10	0.72029(18)	0.54644(13)	0.77659(9)	0.0252(3)
C11	0.8764(2)	0.45324(14)	0.83329(10)	0.0319(4)
C12	0.0306(2)	0.49622(17)	0.88140(14)	0.0487(5)
C13	0.8694(2)	0.39136(17)	0.75361(12)	0.0442(5)
C14	0.7745(2)	0.38986(16)	0.87890(13)	0.0444(5)
C15	0.1343(2)	0.36511(17)	0.54725(11)	0.0401(4)
C16	0.0495(2)	0.3585(2)	0.47369(13)	0.0607(7)
C17	0.9975(3)	0.2670(3)	0.42068(15)	0.0786(10)
C18	0.0263(3)	0.1800(3)	0.43960(15)	0.0755(9)
C19	0.1119(3)	0.1855(2)	0.51258(14)	0.0608(7)
C20	0.1695(2)	0.27911(16)	0.56741(11)	0.0393(4)
C21	0.2702(2)	0.28971(16)	0.64447(11)	0.0392(4)
C22	0.23815(19)	0.34874(15)	0.71446(10)	0.0338(4)
C23	0.25024(19)	0.46057(15)	0.70584(11)	0.0361(4)
C24	0.28753(18)	0.31146(13)	0.84819(10)	0.0290(4)
C25	0.3823(2)	0.30490(14)	0.98627(10)	0.0320(4)
C26	0.3332(3)	0.18724(16)	0.97637(13)	0.0554(6)
C27	0.5354(3)	0.3519(2)	0.03555(13)	0.0615(7)
C28	0.2833(3)	0.3506(2)	0.02138(15)	0.0674(8)
C29	0.6501(2)	0.08578(16)	0.66915(14)	0.0467(5)

Table G2 continued

C30	0.5152(3)	0.0693(2)	0.62158(17)	0.0634(7)
C31	0.3974(3)	0.0619(2)	0.6563(2)	0.0728(9)
C32	0.4144(3)	0.0729(2)	0.7365(2)	0.0686(8)
C33	0.5480(2)	0.08879(18)	0.78294(17)	0.0558(6)
C34	0.6677(2)	0.09472(15)	0.75008(13)	0.0410(5)
C35	0.8109(2)	0.10461(16)	0.80023(13)	0.0426(5)
C36	0.8776(2)	0.03001(14)	0.76212(12)	0.0362(4)
C37	0.9204(2)	0.06093(16)	0.68685(14)	0.0452(5)
C38	0.02323(19)	0.93192(14)	0.81688(11)	0.0351(4)
C39	0.1909(2)	0.85828(15)	0.88709(14)	0.0460(5)
C40	0.0894(3)	0.80569(19)	0.93680(14)	0.0611(6)
C41	0.1928(3)	0.7837(2)	0.81624(16)	0.0596(6)
C42	0.3429(3)	0.9137(2)	0.9370(2)	0.0881(11)
N1	0.72918(17)	0.64262(12)	0.76346(10)	0.0354(4)
N2	0.33385(18)	0.35173(15)	0.78695(10)	0.0440(4)
N3	0.00033(18)	0.02423(13)	0.81522(11)	0.0445(4)
O1	0.61321(18)	0.78685(13)	0.81971(8)	0.0538(4)
O2	0.61406(13)	0.46786(9)	0.75117(7)	0.0324(3)
O3	0.84548(13)	0.55020(9)	0.82186(7)	0.0313(3)
O4	0.29532(17)	0.53352(12)	0.76032(9)	0.0563(4)
O5	0.16346(13)	0.25728(10)	0.84590(8)	0.0392(3)
O6	0.39957(13)	0.33828(10)	0.90989(7)	0.0379(3)
O7	0.03289(18)	0.06174(14)	0.66896(11)	0.0647(5)
O8	0.93647(14)	0.84781(10)	0.78236(9)	0.0439(3)
O9	0.15196(14)	0.94766(10)	0.86277(9)	0.0459(4)
S1	0.48897(6)	0.81739(4)	0.68887(3)	0.03753(12)
S2	0.20175(6)	0.48792(4)	0.61062(3)	0.04954(15)
S3	0.79745(7)	0.09823(6)	0.62134(4)	0.06043(18)

Table G3. Bond lengths [Å] and angles [°] for Boc-3-amino-δ-thiochromanone.

Bond lengths

C1-C2	1.393(2)	C1-C6	1.394(2)
C1-S1	1.7644(18)	C2-C3	1.383(3)
C3-C4	1.378(3)	C4-C5	1.383(3)
C5-C6	1.388(3)	C6-C7	1.503(2)
C7-C8	1.528(2)	C8-N1	1.444(2)
C8-C9	1.523(3)	C9-O1	1.200(2)
C9-S1	1.7746(18)	C10-O2	1.219(2)
C10-N1	1.337(2)	C10-O3	1.3456(19)
C11-O3	1.478(2)	C11-C14	1.512(3)
C11-C12	1.518(3)	C11-C13	1.518(3)
C15-C20	1.393(3)	C15-C16	1.393(3)
C15-S2	1.759(2)	C16-C17	1.359(4)
C17-C18	1.370(5)	C18-C19	1.392(4)
C19-C20	1.395(3)	C20-C21	1.503(3)
C21-C22	1.516(3)	C22-N2	1.443(2)
C22-C23	1.520(3)	C23-O4	1.205(2)
C23-S2	1.770(2)	C24-O5	1.215(2)
C24-N2	1.341(2)	C24-O6	1.341(2)
C25-O6	1.478(2)	C25-C28	1.492(3)
C25-C26	1.503(3)	C25-C27	1.518(3)
C29-C34	1.380(3)	C29-C30	1.396(3)
C29-S3	1.775(2)	C30-C31	1.388(4)
C31-C32	1.366(4)	C32-C33	1.379(4)
C33-C34	1.389(3)	C34-C35	1.504(3)
C35-C36	1.525(3)	C36-N3	1.438(2)
C36-C37	1.512(3)	C37-O7	1.208(3)
C37-S3	1.772(2)	C38-O8	1.212(2)
C38-O9	1.335(2)	C38-N3	1.351(2)
C39-O9	1.478(2)	C39-C40	1.505(3)
C39-C41	1.512(3)	C39-C42	1.521(3)

Table G3 continued

Bond angles

C2-C1-C6	120.88(17)	C2-C1-S1	117.54(14)
C6-C1-S1	121.57(13)	C3-C2-C1	119.63(17)
C4-C3-C2	120.21(18)	C3-C4-C5	119.77(19)
C4-C5-C6	121.49(18)	C5-C6-C1	118.02(16)
C5-C6-C7	122.17(16)	C1-C6-C7	119.81(16)
C6-C7-C8	112.20(14)	N1-C8-C9	109.71(14)
N1-C8-C7	111.42(15)	C9-C8-C7	111.59(14)
O1-C9-C8	124.68(17)	O1-C9-S1	117.38(15)
C8-C9-S1	117.93(13)	O2-C10-N1	124.59(15)
O2-C10-O3	125.72(15)	N1-C10-O3	109.69(14)
O3-C11-C14	110.68(14)	O3-C11-C12	102.17(14)
C14-C11-C12	110.98(16)	O3-C11-C13	109.39(14)
C14-C11-C13	112.50(17)	C12-C11-C13	110.63(16)
C20-C15-C16	120.9(2)	C20-C15-S2	121.94(15)
C16-C15-S2	117.05(19)	C17-C16-C15	120.3(3)
C16-C17-C18	120.2(2)	C17-C18-C19	120.2(3)
C18-C19-C20	120.8(3)	C15-C20-C19	117.5(2)
C15-C20-C21	120.30(18)	C19-C20-C21	122.2(2)
C20-C21-C22	113.83(16)	N2-C22-C21	111.07(16)
N2-C22-C23	108.67(16)	C21-C22-C23	112.75(15)
O4-C23-C22	123.63(18)	O4-C23-S2	117.27(16)
C22-C23-S2	119.06(14)	O5-C24-N2	124.18(16)
O5-C24-O6	126.57(16)	N2-C24-O6	109.23(15)
O6-C25-C28	110.19(16)	O6-C25-C26	110.36(15)
C28-C25-C26	113.0(2)	O6-C25-C27	102.46(15)
C28-C25-C27	110.7(2)	C26-C25-C27	109.60(18)
C34-C29-C30	120.9(2)	C34-C29-S3	121.92(16)
C30-C29-S3	117.1(2)	C31-C30-C29	119.4(3)
C32-C31-C30	120.1(2)	C31-C32-C33	120.0(3)
C32-C33-C34	121.4(3)	C29-C34-C33	118.1(2)
C29-C34-C35	120.45(18)	C33-C34-C35	121.3(2)
C34-C35-C36	111.24(17)	N3-C36-C37	109.94(17)
N3-C36-C35	111.97(17)	C37-C36-C35	111.34(16)
O7-C37-C36	124.0(2)	O7-C37-S3	118.40(18)
C36-C37-S3	117.60(14)	O8-C38-O9	126.07(17)
O8-C38-N3	123.79(17)	O9-C38-N3	110.12(15)

Table G3 continued

O9-C39-C40	109.10(18)	O9-C39-C41	111.09(18)
C40-C39-C41	112.97(19)	O9-C39-C42	101.54(16)
C40-C39-C42	110.7(2)	C41-C39-C42	110.9(2)
C10-N1-C8	122.35(15)	C24-N2-C22	123.49(16)
C38-N3-C36	120.74(16)	C10-O3-C11	121.21(13)
C24-O6-C25	121.83(14)	C38-O9-C39	120.67(14)
C1-S1-C9	104.89(8)	C15-S2-C23	105.13(9)
C37-S3-C29	103.72(10)		

Table G4. Anisotropic atomic displacement parameters (Å2) for Boc-3-amino-δ-thiochromanone.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U11	U22	U33	U23	U13	U12
C1	0.0299(9)	0.0301(9)	0.0300(9)	0.0110(7)	0.0088(7)	0.0115(7)
C2	0.0358(10)	0.0434(11)	0.0425(11)	0.0154(9)	0.0094(8)	0.0222(9)
C3	0.0432(11)	0.0567(13)	0.0384(11)	0.0151(10)	-0.0012(9)	0.0224(10)
C4	0.0555(13)	0.0564(13)	0.0284(10)	0.0066(9)	0.0041(9)	0.0244(11)
C5	0.0449(11)	0.0485(11)	0.0318(10)	0.0093(8)	0.0116(8)	0.0252(9)
C6	0.0285(9)	0.0355(9)	0.0299(9)	0.0122(7)	0.0093(7)	0.0139(7)
C7	0.0284(9)	0.0405(10)	0.0345(9)	0.0086(8)	0.0075(7)	0.0168(8)
C8	0.0251(8)	0.0270(8)	0.0328(9)	0.0091(7)	-0.0004(7)	0.0064(7)
C9	0.0331(9)	0.0368(10)	0.0311(9)	0.0097(8)	0.0082(7)	0.0119(8)
C10	0.0270(8)	0.0253(8)	0.0235(8)	0.0045(6)	0.0056(6)	0.0090(7)
C11	0.0369(10)	0.0295(9)	0.0334(9)	0.0081(7)	0.0056(8)	0.0172(8)
C12	0.0436(12)	0.0479(12)	0.0544(13)	0.0112(10)	-0.0056(10)	0.0224(10)
C13	0.0549(13)	0.0453(12)	0.0399(11)	0.0048(9)	0.0096(9)	0.0287(10)
C14	0.0547(13)	0.0426(11)	0.0480(12)	0.0216(9)	0.0184(10)	0.0247(10)
C15	0.0281(9)	0.0606(13)	0.0320(10)	0.0111(9)	0.0049(8)	0.0154(9)
C16	0.0421(12)	0.112(2)	0.0376(12)	0.0182(13)	0.0067(10)	0.0380(14)
C17	0.0371(13)	0.155(3)	0.0349(13)	-0.0013(17)	0.0033(10)	0.0287(17)
C18	0.0451(14)	0.104(2)	0.0444(14)	-0.0296(15)	0.0142(11)	-0.0064(15)
C19	0.0597(15)	0.0577(15)	0.0526(14)	-0.0086(11)	0.0197(12)	0.0059(12)
C20	0.0346(10)	0.0462(11)	0.0332(10)	0.0035(8)	0.0090(8)	0.0091(8)
C21	0.0409(11)	0.0390(10)	0.0418(11)	0.0116(8)	0.0089(9)	0.0175(9)
C22	0.0269(9)	0.0414(10)	0.0293(9)	0.0108(8)	0.0038(7)	0.0060(8)
C23	0.0285(9)	0.0415(11)	0.0387(10)	0.0061(8)	0.0096(8)	0.0120(8)
C24	0.0302(9)	0.0271(8)	0.0286(9)	0.0063(7)	0.0051(7)	0.0081(7)
C25	0.0413(10)	0.0287(9)	0.0251(8)	0.0079(7)	0.0073(7)	0.0093(8)
C26	0.0826(17)	0.0340(11)	0.0397(12)	0.0072(9)	-0.0062(11)	0.0155(11)
C27	0.0588(15)	0.0662(16)	0.0361(12)	0.0175(11)	-0.0061(10)	-0.0057(12)
C28	0.099(2)	0.0855(19)	0.0491(14)	0.0247(13)	0.0369(14)	0.0599(17)

Table G4 continued

C29	0.0462(12)	0.0380(11)	0.0635(14)	0.0142(10)	0.0091(10)	0.0242(9)
C30	0.0622(16)	0.0561(15)	0.0769(17)	0.0148(13)	-0.0021(13)	0.0345(13)
C31	0.0442(14)	0.0530(15)	0.120(3)	0.0117(16)	-0.0071(15)	0.0291(12)
C32	0.0422(13)	0.0503(14)	0.120(3)	0.0131(15)	0.0231(15)	0.0228(11)
C33	0.0475(13)	0.0458(13)	0.0792(17)	0.0087(12)	0.0224(12)	0.0196(10)
C34	0.0367(10)	0.0268(9)	0.0620(13)	0.0070(9)	0.0124(9)	0.0134(8)
C35	0.0376(11)	0.0372(11)	0.0509(12)	0.0053(9)	0.0087(9)	0.0112(9)
C36	0.0300(9)	0.0247(9)	0.0510(11)	0.0080(8)	0.0014(8)	0.0083(7)
C37	0.0449(12)	0.0375(11)	0.0599(13)	0.0112(10)	0.0136(10)	0.0212(9)
C38	0.0271(9)	0.0279(9)	0.0450(11)	0.0102(8)	0.0003(8)	0.0045(7)
C39	0.0378(11)	0.0299(10)	0.0631(14)	0.0143(9)	-0.0090(10)	0.0094(8)
C40	0.0759(17)	0.0499(14)	0.0483(13)	0.0141(11)	0.0002(12)	0.0134(12)
C41	0.0533(14)	0.0569(14)	0.0797(18)	0.0186(13)	0.0177(13)	0.0300(12)
C42	0.0513(15)	0.0477(15)	0.139(3)	0.0251(16)	-0.0417(17)	0.0081(12)
N1	0.0286(8)	0.0256(8)	0.0426(9)	0.0096(7)	-0.0087(7)	0.0030(6)
N2	0.0260(8)	0.0621(12)	0.0341(9)	0.0213(8)	0.0015(7)	-0.0002(8)
N3	0.0304(8)	0.0232(8)	0.0674(12)	0.0073(8)	-0.0091(8)	0.0021(6)
O1	0.0678(10)	0.0674(10)	0.0296(7)	0.0045(7)	0.0062(7)	0.0316(9)
O2	0.0271(6)	0.0257(6)	0.0403(7)	0.0062(5)	0.0036(5)	0.0052(5)
O3	0.0291(6)	0.0255(6)	0.0360(7)	0.0050(5)	-0.0029(5)	0.0099(5)
O4	0.0552(9)	0.0510(9)	0.0545(9)	-0.0102(8)	0.0094(8)	0.0153(8)
O5	0.0294(7)	0.0414(7)	0.0427(8)	0.0172(6)	0.0067(6)	0.0034(6)
O6	0.0329(7)	0.0462(8)	0.0270(6)	0.0128(6)	0.0028(5)	0.0023(6)
O7	0.0543(10)	0.0709(12)	0.0866(13)	0.0222(10)	0.0313(9)	0.0346(9)
O8	0.0335(7)	0.0247(7)	0.0615(9)	0.0062(6)	-0.0086(6)	0.0032(5)
O9	0.0313(7)	0.0263(7)	0.0683(10)	0.0110(6)	-0.0117(6)	0.0039(5)
S1	0.0459(3)	0.0366(3)	0.0361(2)	0.00461(19)	0.0076(2)	0.0236(2)
S2	0.0539(3)	0.0457(3)	0.0542(3)	0.0202(3)	0.0046(3)	0.0235(3)
S3	0.0691(4)	0.0767(4)	0.0586(4)	0.0320(3)	0.0229(3)	0.0449(4)

Table G5. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-3-amino-δ-thiochromanone.

	x/a	y/b	z/c	U(eq)
H2A	0.2808	0.7906	0.5530	0.045
H3A	0.2234	0.6954	0.4249	0.055
H4	0.3602	0.5945	0.3882	0.055
H5	0.5542	0.5883	0.4793	0.047
H7A	0.7405	0.7478	0.6405	0.04
H7B	0.6921	0.6237	0.6109	0.04
H8	0.5277	0.5977	0.6996	0.035
H12A	1.0942	0.5380	0.8511	0.073
H12B	1.0628	0.4378	0.8935	0.073
H12C	1.0344	0.5406	0.9307	0.073
H13A	0.7680	0.3573	0.7260	0.066
H13B	0.9143	0.3378	0.7614	0.066
H13C	0.9214	0.4390	0.7220	0.066
H14A	0.7760	0.4349	0.9274	0.067
H14B	0.8053	0.3315	0.8925	0.067
H14C	0.6760	0.3622	0.8464	0.067
H16	0.0279	0.4183	0.4606	0.073
H17	-0.0590	0.2634	0.3703	0.094
H18	-0.0122	0.1157	0.4028	0.091
H19	0.1313	0.1248	0.5252	0.073
H21A	0.3707	0.3266	0.6399	0.047
H21B	0.2641	0.2187	0.6545	0.047
H22	0.1367	0.3106	0.7187	0.041
H26A	0.2297	0.1576	0.9522	0.083
H26B	0.3503	0.1670	1.0281	0.083
H26C	0.3877	0.1604	0.9423	0.083
H27A	0.5995	0.3234	1.0102	0.092
H27B	0.5367	0.3341	1.0884	0.092
H27C	0.5688	0.4284	1.0398	0.092
H28A	0.3184	0.4269	1.0245	0.101
H28B	0.2811	0.3345	1.0745	0.101
H28C	0.1856	0.3204	0.9882	0.101

Table G5 continued

H30	0.5042	1.0633	0.5659	0.076
H31	0.3047	1.0493	0.6243	0.087
H32	0.3341	1.0695	0.7603	0.082
H33	0.5584	1.0959	0.8386	0.067
H35A	0.7970	1.0886	0.8528	0.051
H35B	0.8780	1.1777	0.8077	0.051
H36	0.8024	0.9583	0.7481	0.043
H40A	1.0885	0.8582	0.9800	0.092
H40B	1.1221	0.7522	0.9587	0.092
H40C	0.9916	0.7725	0.9041	0.092
H41A	1.0933	0.7443	0.7876	0.089
H41B	1.2398	0.7347	0.8339	0.089
H41C	1.2466	0.8238	0.7812	0.089
H42A	1.4066	0.9490	0.9039	0.132
H42B	1.3796	0.8621	0.9592	0.132
H42C	1.3402	0.9658	0.9799	0.132
H1	0.807(2)	0.6933(17)	0.7794(12)	0.042
H2	0.421(3)	0.3864(19)	0.7929(14)	0.053
H3	1.069(3)	1.0810(19)	0.8330(14)	0.053

Table G6. Torsion angles (°) for Boc-3-amino-δ-thiochromanone.

C6-C1-C2-C3	1.0(3)	S1-C1-C2-C3	-179.88(16)
C1-C2-C3-C4	-0.5(3)	C2-C3-C4-C5	0.0(3)
C3-C4-C5-C6	0.2(3)	C4-C5-C6-C1	0.2(3)
C4-C5-C6-C7	179.22(19)	C2-C1-C6-C5	-0.8(3)
S1-C1-C6-C5	-179.92(14)	C2-C1-C6-C7	-179.84(16)
S1-C1-C6-C7	1.1(2)	C5-C6-C7-C8	-132.61(18)
C1-C6-C7-C8	46.4(2)	C6-C7-C8-N1	171.51(15)
C6-C7-C8-C9	-65.47(19)	N1-C8-C9-O1	-17.3(3)
C7-C8-C9-O1	-141.33(19)	N1-C8-C9-S1	161.20(12)
C7-C8-C9-S1	37.21(19)	C20-C15-C16-C17	1.1(3)
S2-C15-C16-C17	177.91(19)	C15-C16-C17-C18	1.0(4)
C16-C17-C18-C19	-1.5(4)	C17-C18-C19-C20	0.0(4)
C16-C15-C20-C19	-2.5(3)	S2-C15-C20-C19	-179.17(16)
C16-C15-C20-C21	175.37(19)	S2-C15-C20-C21	-1.3(3)
C18-C19-C20-C15	2.0(3)	C18-C19-C20-C21	-175.9(2)
C15-C20-C21-C22	45.0(2)	C19-C20-C21-C22	-137.2(2)
C20-C21-C22-N2	177.20(16)	C20-C21-C22-C23	-60.6(2)
N2-C22-C23-O4	-20.8(3)	C21-C22-C23-O4	-144.38(19)
N2-C22-C23-S2	157.01(13)	C21-C22-C23-S2	33.4(2)
C34-C29-C30-C31	0.0(3)	S3-C29-C30-C31	-178.26(19)
C29-C30-C31-C32	1.4(4)	C30-C31-C32-C33	-1.5(4)
C31-C32-C33-C34	0.4(4)	C30-C29-C34-C33	-1.1(3)
S3-C29-C34-C33	177.05(16)	C30-C29-C34-C35	175.57(19)
S3-C29-C34-C35	-6.2(3)	C32-C33-C34-C29	1.0(3)
C32-C33-C34-C35	-175.7(2)	C29-C34-C35-C36	-42.0(3)
C33-C34-C35-C36	134.6(2)	C34-C35-C36-N3	-169.41(16)
C34-C35-C36-C37	67.1(2)	N3-C36-C37-O7	12.2(3)
C35-C36-C37-O7	136.9(2)	N3-C36-C37-S3	-167.70(14)
C35-C36-C37-S3	-43.0(2)	O2-C10-N1-C8	-6.3(3)
O3-C10-N1-C8	173.83(15)	C9-C8-N1-C10	132.72(17)

Crystal Structure of Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine

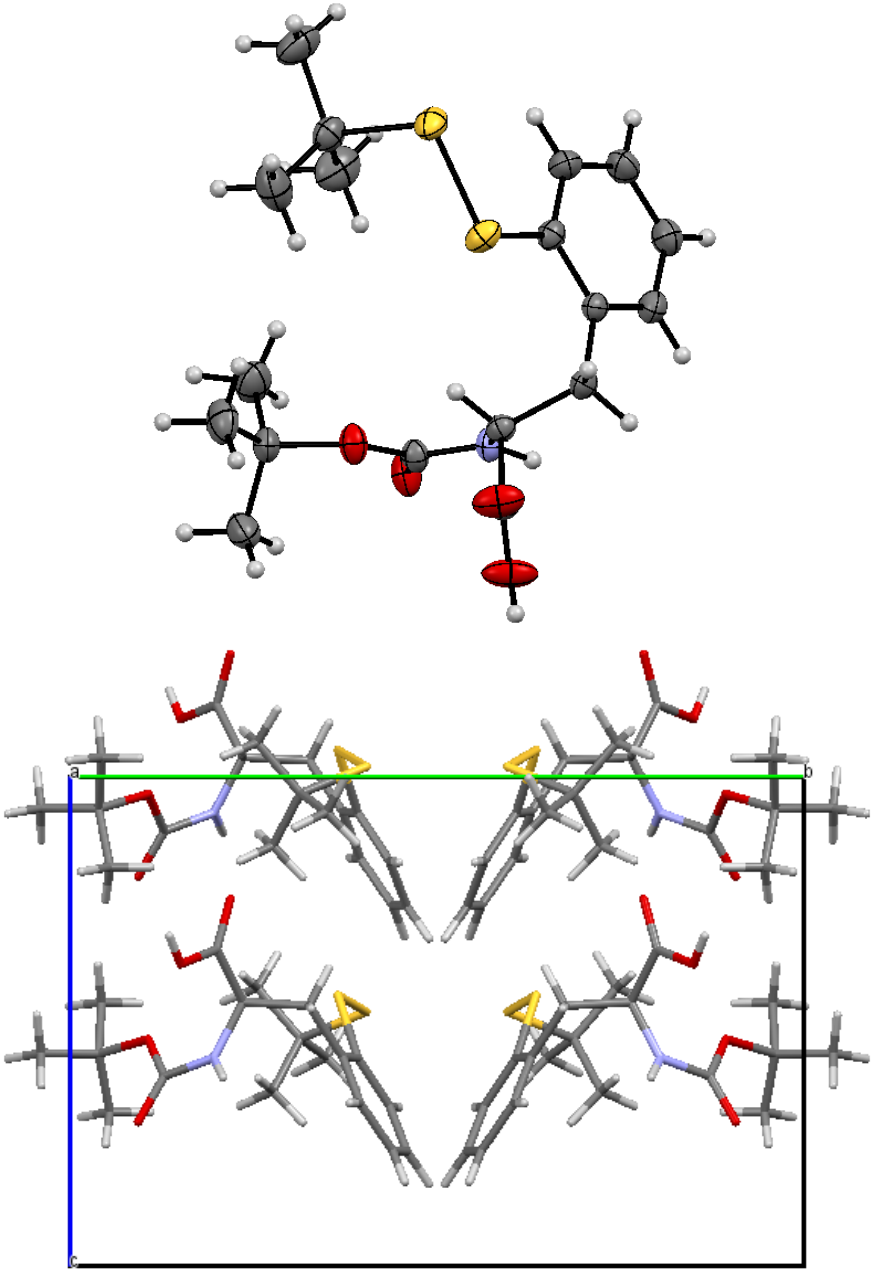


Figure G2 Crystal structure of Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.
Top: ORTEP diagram of the Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine with ellipsoids shown at 50% probability; bottom: overall crystal packing. Diffractable crystals were obtained via slow evaporation at room temperature in $CDCl_3$.

Table G7. Crystallographic data and refinement details for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

empirical formula	$C_{18}H_{27}NO_4S_2$	
formula weight	385.52	
T (K)	200(2)	
wavelength (Å)	0.71073	
crystal system, space group	Orthorhombic, P c a 2_1	
Unit cell dimensions (Å, °)	a = 11.1750(9)	α = 90
	b = 16.5912(14)	β = 90
	c = 11.0819(9)	γ = 90
Volume ($Å^3$)	2054.7(3)	
Z, Z′, calcd density (g/cm^3)	4, 0, 1.246	
absorption coefficient (mm^{-1})	0.280	
F(000)	824	
crystal size (mm)	0.300 x 0.438 x 0.473	
□ range for data collection	2.20 to 31.44°	
Index ranges	$-12 \le h \le 16$, $-24 \le k \le 24$, $-13 \le l \le 15$	
Reflections collected/ unique	29107/6362 [R(int) = 0.0248]	
Coverage of independent reflections	97.1%	
Absorption correction	multi-scan	
Max. and min. transmission	0.7462 and 0.6937	
Structure solution technique	direct methods	
Refinement method	Full-matrix least-squares on F^2	
Function minimized	$\Sigma\, w(F_o^2 - F_c^2)^2$	
Data / restraints / parameters	6362 / 3 / 238	
Goodness-of-fit on F^2	1.063	
Final *R* indices	5782 data; I>2σ(I)	R1 = 0.0338, wR2 = 0.0827
	all data	R1 = 0.0398, wR2 = 0.0869
Weighting scheme	$w=1/[\sigma^2(F_o^2)+(0.0528P)^2+0.1069P]$ where $P=(F_o^2+2F_c^2)/3$	
Largest diff. peak and hole	0.395 and -0.171 $eÅ^{-3}$	
R.M.S. deviation from mean	0.047 $eÅ^{-3}$	

Table G8. Atomic coordinates (× 10^4) and equivalent isotropic displacement parameters (A^2 × 10^3) for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

	x/a	y/b	z/c	U(eq)
S1	0.77898(4)	0.40358(3)	0.48020(6)	0.03407(13)
S2	0.61047(4)	0.36476(3)	0.44670(5)	0.03148(11)
N1	0.35854(13)	0.19531(8)	0.57423(15)	0.0246(3)
O1	0.34097(16)	0.21790(10)	0.25073(15)	0.0421(4)
O2	0.22983(17)	0.15014(11)	0.38318(15)	0.0454(4)
O3	0.38972(14)	0.09213(8)	0.70478(15)	0.0347(3)
O4	0.50270(13)	0.10584(8)	0.53626(14)	0.0312(3)
C1	0.86219(18)	0.31381(12)	0.5347(2)	0.0331(4)
C2	0.7991(2)	0.27589(17)	0.6417(3)	0.0509(6)
C3	0.8778(3)	0.25471(16)	0.4315(3)	0.0543(7)
C4	0.9828(2)	0.34940(17)	0.5732(3)	0.0526(7)
C5	0.52193(15)	0.39193(9)	0.57483(18)	0.0243(3)
C6	0.56771(17)	0.43516(11)	0.6725(2)	0.0320(4)
C7	0.49265(19)	0.45881(11)	0.7650(2)	0.0358(4)
C8	0.37236(19)	0.43922(12)	0.7630(2)	0.0353(4)
C9	0.32748(17)	0.39525(10)	0.6671(2)	0.0311(4)
C10	0.40002(15)	0.37098(9)	0.57143(19)	0.0238(3)
C11	0.34563(14)	0.32539(9)	0.46753(19)	0.0258(4)
C12	0.31655(17)	0.19921(10)	0.35311(19)	0.0277(4)
C13	0.38268(14)	0.23555(9)	0.46063(18)	0.0231(3)
C14	0.41523(16)	0.12795(10)	0.61163(18)	0.0254(4)
C15	0.58490(16)	0.03818(10)	0.5659(2)	0.0302(4)
C16	0.5157(2)	0.95958(11)	0.5646(3)	0.0433(6)
C17	0.6474(2)	0.05360(16)	0.6846(3)	0.0483(6)
C18	0.6719(2)	0.04227(13)	0.4612(3)	0.0447(6)

Table G9. Bond lengths [Å] and angles [°] for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

Bond lengths

S1-C1	1.857(2)	S1-S2	2.0245(6)
S2-C5	1.788(2)	N1-C14	1.350(2)
N1-C13	1.450(2)	N1-H1	0.889(3)
O1-C12	1.207(3)	O2-C12	1.309(2)
O2-H2	0.820(3)	O3-C14	1.225(2)
O4-C14	1.337(2)	O4-C15	1.487(2)
C1-C3	1.516(4)	C1-C2	1.516(4)
C1-C4	1.532(3)	C2-H2A	0.98
C2-H2B	0.98	C2-H2C	0.98
C3-H3A	0.98	C3-H3B	0.98
C3-H3C	0.98	C4-H4A	0.98
C4-H4B	0.98	C4-H4C	0.98
C5-C6	1.396(3)	C5-C10	1.406(2)
C6-C7	1.382(3)	C6-H6	0.95
C7-C8	1.383(3)	C7-H7	0.95
C8-C9	1.384(3)	C8-H8	0.95
C9-C10	1.394(3)	C9-H9	0.95
C10-C11	1.506(3)	C11-C13	1.549(2)
C11-H11A	0.99	C11-H11B	0.99
C12-C13	1.526(3)	C13-H13	1.0
C15-C17	1.511(3)	C15-C16	1.516(3)
C15-C18	1.516(3)	C16-H16A	0.98
C16-H16B	0.98	C16-H16C	0.98
C17-H17A	0.98	C17-H17B	0.98
C17-H17C	0.98	C18-H18A	0.98
C18-H18B	0.98	C18-H18C	0.98

Table G9 continued

Bond angles

C1-S1-S2	105.68(7)	C5-S2-S1	106.79(6)
C14-N1-C13	124.08(15)	C14-N1-H1	115.7(16)
C13-N1-H1	118.6(16)	C12-O2-H2	113.(2)
C14-O4-C15	121.37(15)	C3-C1-C2	112.0(2)
C3-C1-C4	111.0(2)	C2-C1-C4	110.6(2)
C3-C1-S1	109.35(17)	C2-C1-S1	110.74(16)
C4-C1-S1	102.82(15)	C1-C2-H2A	109.5
C1-C2-H2B	109.5	H2A-C2-H2B	109.5
C1-C2-H2C	109.5	H2A-C2-H2C	109.5
H2B-C2-H2C	109.5	C1-C3-H3A	109.5
C1-C3-H3B	109.5	H3A-C3-H3B	109.5
C1-C3-H3C	109.5	H3A-C3-H3C	109.5
H3B-C3-H3C	109.5	C1-C4-H4A	109.5
C1-C4-H4B	109.5	H4A-C4-H4B	109.5
C1-C4-H4C	109.5	H4A-C4-H4C	109.5
H4B-C4-H4C	109.5	C6-C5-C10	120.19(18)
C6-C5-S2	122.85(14)	C10-C5-S2	116.89(14)
C7-C6-C5	119.93(17)	C7-C6-H6	120.0
C5-C6-H6	120.0	C8-C7-C6	120.8(2)
C8-C7-H7	119.6	C6-C7-H7	119.6
C7-C8-C9	119.2(2)	C7-C8-H8	120.4
C9-C8-H8	120.4	C8-C9-C10	121.73(18)
C8-C9-H9	119.1	C10-C9-H9	119.1
C9-C10-C5	118.13(17)	C9-C10-C11	119.46(15)
C5-C10-C11	122.39(17)	C10-C11-C13	114.41(14)
C10-C11-H11A	108.7	C13-C11-H11A	108.7
C10-C11-H11B	108.7	C13-C11-H11B	108.7
H11A-C11-H11B	107.6	O1-C12-O2	124.50(19)
O1-C12-C13	121.52(16)	O2-C12-C13	113.92(17)
N1-C13-C12	113.94(14)	N1-C13-C11	110.52(15)
C12-C13-C11	106.82(14)	N1-C13-H13	108.5
C12-C13-H13	108.5	C11-C13-H13	108.5
O3-C14-O4	124.30(15)	O3-C14-N1	123.46(17)
O4-C14-N1	112.24(17)	O4-C15-C17	110.51(16)
O4-C15-C16	109.40(16)	C17-C15-C16	112.9(2)
O4-C15-C18	101.13(16)	C17-C15-C18	111.25(19)

Table G9 continued

C16-C15-C18	111.00(18)	C15-C16-H16A	109.5
C15-C16-H16B	109.5	H16A-C16-H16B	109.5
C15-C16-H16C	109.5	H16A-C16-H16C	109.5
H16B-C16-H16C	109.5	C15-C17-H17A	109.5
C15-C17-H17B	109.5	H17A-C17-H17B	109.5
C15-C17-H17C	109.5	H17A-C17-H17C	109.5
H17B-C17-H17C	109.5	C15-C18-H18A	109.5
C15-C18-H18B	109.5	H18A-C18-H18B	109.5
C15-C18-H18C	109.5	H18A-C18-H18C	109.5
H18B-C18-H18C	109.5		

Table G10. Anisotropic atomic displacement parameters (Å2) for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

The anisotropic atomic displacement factor exponent takes the form: -2π2[h2 a*2 U11 + ... + 2 h k a* b* U12]

	U11	U22	U33	U23	U13	U12
S1	0.02391(19)	0.03115(18)	0.0471(3)	0.0079(2)	0.00334(19)	-0.00246(15)
S2	0.02438(19)	0.0440(2)	0.0261(3)	0.0014(2)	0.00043(18)	-0.00158(16)
N1	0.0265(7)	0.0271(6)	0.0202(8)	0.0003(6)	0.0011(6)	0.0057(5)
O1	0.0463(9)	0.0595(9)	0.0204(9)	0.0033(7)	-0.0026(7)	-0.0211(8)
O2	0.0575(10)	0.0547(9)	0.0239(9)	0.0048(7)	-0.0075(7)	-0.0318(8)
O3	0.0448(8)	0.0331(6)	0.0261(8)	0.0063(5)	0.0105(6)	0.0129(5)
O4	0.0362(7)	0.0325(6)	0.0250(8)	0.0035(5)	0.0067(6)	0.0111(5)
C1	0.0305(9)	0.0368(8)	0.0319(12)	0.0039(8)	-0.0045(8)	0.0029(7)
C2	0.0514(14)	0.0587(14)	0.0427(16)	0.0218(12)	-0.0057(11)	-0.0034(11)
C3	0.0684(17)	0.0445(11)	0.0501(18)	-0.0036(11)	-0.0069(13)	0.0195(11)
C4	0.0325(10)	0.0647(14)	0.0605(19)	0.0115(13)	-0.0135(11)	-0.0033(10)
C5	0.0232(7)	0.0255(6)	0.0241(10)	0.0025(6)	-0.0010(7)	0.0017(5)
C6	0.0265(8)	0.0332(8)	0.0361(13)	-0.0045(8)	-0.0044(8)	-0.0013(7)
C7	0.0383(10)	0.0332(8)	0.0357(13)	-0.0101(8)	-0.0053(9)	0.0040(7)
C8	0.0358(10)	0.0338(8)	0.0362(13)	-0.0065(8)	0.0044(9)	0.0054(7)
C9	0.0243(8)	0.0297(7)	0.0394(13)	-0.0018(8)	0.0027(8)	0.0030(6)
C10	0.0231(7)	0.0218(6)	0.0265(10)	0.0018(6)	-0.0035(7)	0.0021(5)
C11	0.0234(7)	0.0260(6)	0.0282(11)	0.0012(7)	-0.0055(7)	0.0026(5)
C12	0.0314(8)	0.0277(7)	0.0239(10)	0.0005(7)	-0.0024(7)	-0.0038(6)
C13	0.0229(6)	0.0253(6)	0.0212(10)	-0.0005(6)	-0.0018(6)	-0.0008(5)
C14	0.0284(8)	0.0258(6)	0.0221(10)	-0.0024(6)	-0.0001(7)	0.0046(6)
C15	0.0266(8)	0.0277(7)	0.0361(12)	-0.0021(7)	0.0014(8)	0.0072(6)
C16	0.0394(10)	0.0296(8)	0.0609(17)	-0.0056(9)	0.0060(11)	0.0016(7)
C17	0.0431(12)	0.0537(12)	0.0482(17)	-0.0071(11)	-0.0134(11)	0.0118(10)
C18	0.0406(10)	0.0434(9)	0.0501(17)	-0.0008(10)	0.0167(11)	0.0102(8)

Table G11. Hydrogen atomic coordinates and isotropic atomic displacement parameters (Å2) for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

	x/a	y/b	z/c	U(eq)
H1	0.2920(12)	0.2078(14)	0.614(2)	0.03
H2	0.190(2)	0.1349(18)	0.3255(18)	0.054
H2A	0.8470	0.2309	0.6723	0.076
H2B	0.7204	0.2561	0.6164	0.076
H2C	0.7891	0.3163	0.7054	0.076
H3A	0.9270	0.2093	0.4584	0.082
H3B	0.9171	0.2818	0.3637	0.082
H3C	0.7992	0.2349	0.4058	0.082
H4A	0.9699	0.3899	0.6362	0.079
H4B	1.0214	0.3748	0.5034	0.079
H4C	1.0342	0.3064	0.6044	0.079
H6	0.6504	0.4483	0.6754	0.038
H7	0.5240	0.4889	0.8308	0.043
H8	0.3211	0.4558	0.8268	0.042
H9	0.2451	0.3812	0.6664	0.037
H11A	0.3690	0.3522	0.3913	0.031
H11B	0.2574	0.3285	0.4741	0.031
H13	0.4706	0.2327	0.4444	0.028
H16A	0.4567	-0.0402	0.6301	0.065
H16B	0.5712	-0.0855	0.5760	0.065
H16C	0.4745	-0.0463	0.4871	0.065
H17A	0.6801	0.1084	0.6850	0.073
H17B	0.7127	0.0147	0.6948	0.073
H17C	0.5901	0.0477	0.7509	0.073
H18A	0.6279	0.0371	0.3851	0.067
H18B	0.7300	-0.0017	0.4679	0.067
H18C	0.7140	0.0941	0.4627	0.067

Table G12. Torsion angles (°) for Boc-2-S(S-*tert*-butyl)-D,L-phenylalanine.

S2-S1-C1-C3	68.92(18)	S2-S1-C1-C2	-54.96(18)
S2-S1-C1-C4	-173.10(16)	S1-S2-C5-C6	2.68(16)
S1-S2-C5-C10	179.78(11)	C10-C5-C6-C7	-1.3(3)
S2-C5-C6-C7	175.75(16)	C5-C6-C7-C8	0.9(3)
C6-C7-C8-C9	0.2(3)	C7-C8-C9-C10	-0.8(3)
C8-C9-C10-C5	0.4(3)	C8-C9-C10-C11	-178.16(17)
C6-C5-C10-C9	0.6(2)	S2-C5-C10-C9	-176.58(13)
C6-C5-C10-C11	179.16(16)	S2-C5-C10-C11	2.0(2)
C9-C10-C11-C13	-109.73(18)	C5-C10-C11-C13	71.7(2)
C14-N1-C13-C12	83.4(2)	C14-N1-C13-C11	-156.37(16)
O1-C12-C13-N1	-168.61(18)	O2-C12-C13-N1	14.3(2)
O1-C12-C13-C11	69.1(2)	O2-C12-C13-C11	-108.03(19)
C10-C11-C13-N1	53.92(19)	C10-C11-C13-C12	178.37(15)
C15-O4-C14-O3	-5.5(3)	C15-O4-C14-N1	173.75(16)
C13-N1-C14-O3	-176.03(18)	C13-N1-C14-O4	4.7(2)
C14-O4-C15-C17	-57.1(2)	C14-O4-C15-C16	67.8(2)
C14-O4-C15-C18	-175.02(18)		

Appendix H

RIGHTS & PERMISSIONS

a) Data, figures, and analyses were reprinted (adapted) with permission from Forbes, C. R.; Zondlo, N. J. Synthesis of thiophenylalanine-containing peptides via Cu(I)-mediated cross-coupling. *Org. Lett.* **2012**, *14* (2), 464-467. Copyright 2012, American Chemical Society.

b) Data, figures, and analyses were reprinted (adapted) from the ACS AuthorChoice open access article Pandey, A. K.; Thomas, K. M.; Forbes, C. R.; Zondlo, N. J. Tunable control of polyproline helix (PPII) structure via aromatic electronic effects: An electronic switch of polyproline helix. *Biochemistry.* **2014**, *53* (32), 5307-5314.

c) Data, figures, and analyses were reprinted (adapted) from the following article under a Creative Commons Attribution-Noncommercial 3.0 Unported license published by the Royal Society of Chemistry: Forbes, C. R.; Pandey, A. K.; Ganguly, H. K.; Zondlo, N. J. 4*R*- and 4*S*-iodophenyl hydroxyproline, 4*R*-pentynoyl hydroxyproline, and S-propargyl-4-thiolphenylalanine: Conformationally biased and tunable amino acids for bioorthogonal reactions. *Org. & Biomol. Chem.* **2016**, *14* (7), 2327-2346.

d) Data, figures, and analyses were reprinted (adapted) with permission from Thomas, K. M.; Naduthambi, D.; Zondlo, N. J. Electronic control of amide cis−trans isomerism via the aromatic−prolyl interaction. *J. Am. Chem. Soc.* **2006**, *128* (7), 2216-2217. Copyright 2006, American Chemical Society.

e) Figure 2 of Pal, D.; Chakrabarti, P. Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations, *J. Mol. Biol*, **1999**, *294*, 271-288, was reprinted with permission from Elsevier.