SYNTHESIS AND APPLICATION OF NOVEL PERFLUORO-*TERT*-BUTYL

AMINO ACIDS FOR THE DETECTION OF PROTEIN ACTIVITY BY ^{19}F

NMR AND MECHANISTIC INSIGHTS INTO POLYPROLINE I AND

POLYPROLINE II FORMATION VIA 4,4-DIFLUOROPROLINE

by

Caitlin M Tressler

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

Winter 2017

SYNTHESIS AND APPLICATION OF NOVEL PERFLUORO-*TERT*-BUTYL AMINO ACIDS FOR THE DETECTION OF PROTEIN ACTIVITY BY ^{19}F NMR AND MECHANISTIC INSIGHTS INTO POLYPROLINE I AND POLYPROLINE II FORMATION VIA 4,4-DIFLUOROPROLINE

by

Caitlin M Tressler

Approved: _____
Murray V. Johnston, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____
George Watson, Ph.D.
Dean of 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: _____

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: _____

Sharon Rozovsky, 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: _____

Robert A. Sikes, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Neal Zondlo, for the opportunity and supported to complete this work. He not only pushed me to achieve my fullest potential but also encouraged my own natural curiosity and sense of adventure in science. I would also like to thank my committee, who were always willing to offer guidance and expertise to assist me throughout my graduate career.

I would also like to thank the Zondlo lab group, both past and present, for assisting me not only in a scientific capacity but also for their friendship and support. Specifically, I would like to thank Dr. Himal Ganguly, Drew Urmey, and Brice Ludwig for making the last few years some of the most enjoyable time I spent in the Zondlo lab. I would also like to thank Anil Pandey and Feng Guo whose research greatly influenced my graduate work. Finally, I would like to thank my amazing undergraduate students Gen Weist and Rachel Jung. My fondest memories of grad school will always be the opportunity to mentor two amazing students who I'm sure will go on to do great things.

I would also like to thank the many friends and family members who supported me during my graduate career. My parents and sister were always just a phone call away to offer support. I would like to thank my friends who offered advice, emotional support, and my own personal cheerleading section, Andy, Rebecca, Emily, Linn, Gabby, and Garrett. Finally, I would like to thank my husband, Justin, who stood by me through everything. After all the late nights and the long weekends, we can close this chapter and move on to the next adventure.

viii

LIST OF TABLES

ABSTRACT

We have developed a series of four novel amino acids containing perfluoro-*tert*-butyl ethers for use in detection of protein and cellular activities by ^{19}F magnetic resonance. All four amino acids were synthesized, as monomers, for introduction into peptides. These amino acids contain nine equivalent fluorine atoms which display as a single peak by ^{19}F NMR and are detectable in nanomolar concentrations in five minutes. Each amino acid was characterized to determine its structural preferences, in order to design potential applications for each designed amino acid.

Perfluoro-*tert*-butyl homoserine demonstrated the ability to form α-helical structures within a peptide context, and was designed into a series of Estrogen Receptor binding peptides to target the ER•co-activator protein-protein interaction. We were able to design a peptide containing perfluoro-*tert*-butyl homoserine which bound with similar affinity to the native co-activator ligand. We were also able to observe the ER•co-activator protein-protein interaction by ^{19}F NMR.

Perfluoro-*tert*-butyl tyrosine was designed into a series of peptides based on a high affinity p53 chimeric peptide due to the presence of aromatic residues within the target binding sequence. While these peptides did not bind with affinity near that of the native peptide, we were able to detection the perfluoro-*tert*-butyl tyrosine in 30 seconds by ^{19}F NMR in nanomolar quantities. This amino acid is also being explored as a potential expressible ^{19}F NMR-based sensor.

4R-Perfluoro-*tert*-butyl hydroxyproline and *4S*-perfluoro-*tert*-butyl hydroxyproline demonstrated opposite conformational preferences, leading to divergent applications in kinase-sensing activity. Both amino acids were introduced into kinase recognition motifs and phosphorylation was successfully detected by ^{19}F NMR. Furthermore, PKA and Akt kinases showed different stereochemical preferences between the two amino acids, indicating the potential of these amino acids to differentiate between similar protein kinases. PKA substrates containing 4*R*-perfluoro-*tert*-butyl hydroxyproline were also subjected to phosphorylation in HeLa cell lysates, which was detectable by ^{19}F NMR, indicating the potential for this technology to be used in live cell imaging.

Finally, 4,4-difluoroproline was synthesized and examined in proline oligomers to determine this amino acid's potential to be used as a probe for *cis-trans* isomerization within peptides. 4,4-Difluoroproline was introduced at the beginning, middle, and end of a nine residue proline oligomer and the interconversion between PPI and PPII helices was examined by circular dichroism and ^{19}F NMR. These experiments allowed us to develop a model of PPI helix stabilization as well as a mechanism of PPI and PPII helix interconversion based on favorable reverse n→π* and n→π* interactions respectively. These data are also indicative of the potential of 4,4-difluoroproline to be used as a ^{19}F probe of *cis-trans* isomerization within peptides and proteins.

Chapter 1

SYNTHESIS AND CHARACTERIZATION OF AMINO ACIDS CONTAINING PERFLUORO-*TERT*-BUTYL GROUPS

1.1 Introduction

Fluorine is a unique atom with distinct electronic properties, which is currently being used as a magnetic resonance probe.[1,2] Fluorine is the most electronegative element and is similar in size to a hydrogen atom.[1] [19]F is 100% naturally abundant and has favorable magnetic properties due to a spin-½ nucleus.[3] Fluorine also has a gyromagetic ratio similar to ^{1}H and much higher than other spin-½ nuclei. Fluorine's magnetic and electronic properties have led to an interest in development in drug molecules, for peptide and protein stability, and as magnetic resonance probes for NMR and MRI applications.[1-11]

1.1.1 Fluorine in Biological Systems, Pharmaceuticals, and Medical Imaging

With the exception of a few rare cases, biological systems do not utilize organofluorine. However, fluorine is not inherently toxic.[4] Fluorine is only present within biological systems as fluoride in tissues such as bone, which is not visible by ^{19}F MRI due to the lack of free rotation of the fluorine atoms. This immobilized fluorine has an extremely short spin-spin relaxation time which prevents detection by current technology, preventing these fluorine atoms from contributing to background ^{19}F signal.[2] However, fluorine has been used since the 1950's by the pharmaceutical industry to modulate properties of drug molecules.[12] Fluorine increases the

1

hydrophobicity of a molecule over hydrogen without adding significant steric bulk. This change can be beneficial in drug molecules that have to pass through a membrane and/or the blood-brain barrier.[13] The addition of C-F bonds in place of C-H bonds also leads to increased metabolic stability of drug molecules. Furthermore, introduction of a fluorine atom or multiple fluorine atoms can potentially greatly alter the electronic properties of a drug molecule.[10] For example, a common modification includes the replacement of a phenyl ring with a pentafluorophenyl ring or a single hydrogen with a fluorine.[4,10] In the case of the pentafluorophenyl group, the addition of mutliple fluorine atoms in place of hydrogen atoms leads to an extremely electron-deficient aromatic system. The substitution of a single hydrogen with fluorine is less dramatic, but still leads to substantial electronic changes within the system due to the electronegative nature of fluorine. Substitution to a fluorine also slightly increases bond length (C-F bond length ~1.35 Å vs C-H bond length ~ 1.09 Å). Despite the similar size to ^{1}H, the increased electronegativity means the electronics of a system can be altered without additional steric components. For example, it can increase a drug's binding affinity to its target through increased hydrophobicity as well as redistribute the electronic properties of the system making it more electron-rich in some parts of the molecule and less electron-rich in other parts of the molecule.[12] Altering these properties allows for tuning of drug molecules for increased or decreased specificity, increased or decreased affinity, and changes in lipophilicity. Furthermore, C-F bonds are very stable (around 116 kcal/mol) and alter a drug's metabolism profile due to the strength of this bond.[13]

It is estimated that 20% of drugs with FDA approval contain fluorine, and that number is expected to rise considerably in the future.[10] Furthermore, of the top five

selling drugs in 2011, three contained fluorine.[12] A number of highly successful drug

molecules currently on the market, including lanzoprazole (regulating gastric acid),

fluoxetine (a common anti-depressant), and ciprofloxacin (an antibiotic), contain

fluorine atoms (Figure 1.1).[10] Many drugs which act on the central nervous system,

such as anti-depressants, utilize fluorine to increase hydrophobicity in order to cross

the blood-brain barrier. Antibiotics containing fluorine often show increased oral

bioavailability over non-fluorinated molecules.[13] Other areas of drug use and

development include anti-retroviral drugs for the treatment of HIV/AIDs,

chemotherapies for the treatment of cancer, and protein kinase inhibitors.[3]

Lanzoprazole   Fluoxetine   Ciprofloxacin
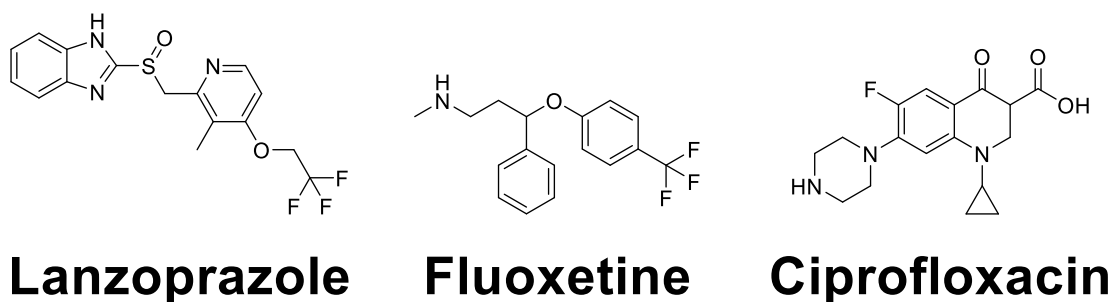
Figure 1.1   A selection of FDA approved, commonly described drugs containing
             fluorine. Fluorinated drugs span many different drug classes and serve
             many roles in drugs including increased lipophilicity, increased
             metabolic stability, and modulation of electronics.

Another common medical application of fluorine is positron emission

tomography (PET) imaging. PET imaging utilizes radioactive tracers, such as ^{18}F, to

monitor cellular activity through uptake of the tracers. One example is the use of ^{18}F

labeled glucose, fludeoxyglucose (^{18}F-FDG) to detect tumors via PET imaging (Figure

1.2).[14] Tumors have an increased glucose metabolism relative to other cells, and therefore take up the glucose-based tracer at a much faster rate, increasing the concentration of ^{18}F in the tumor. While very effective, the use of ^{18}F and PET imaging is very expensive, in part due to the very short half-life of ^{18}F which is around 110 minutes.[14] The short half-life of ^{18}F requires all PET imaging agents to be prepared on site prior to use. In order to facilitate the shortened half-life and the preparation near or on site also greatly limits the chemistry that can be used to prepare PET imaging agents to fast, robust reactions to install the ^{18}F atoms. Furthermore, radioactive materials are hazardous and pose potential health risks to both patients and medical personal.[14] Despite all of these drawbacks, PET imaging is a commonly used technique. We would like to design new probes for use in clinical imaging which do not require the use of radioactive materials to complement current imaging technologies.
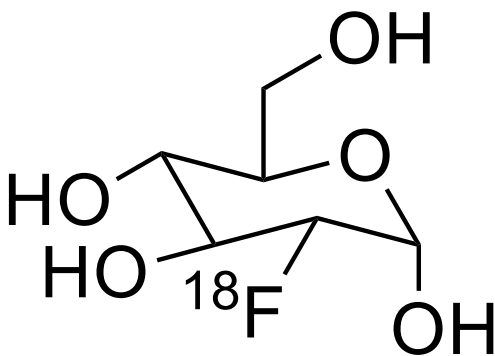
Figure 1.2 The structure of fludeoxyglucose which is used to detect tumor growth due to upregulated glucose metabolism in cancer cells.

While ^{1}H MRI is currently one of the most used clinical imaging techniques, there is significant interest in developing ^{19}F-based magnetic resonance imaging for use in the clinic.[2] ^{1}H MRI suffers from a low signal-to-noise ratio due to the large volume of ^{1}H found within biological systems. Traditional magnetic resonance imaging measures the changes in relaxation rates of water molecules in different types of tissue. While solid structures can be imaged, contrast agents, often caged gadolinium (III) -based molecules, are used to increase the relaxation rate of water to improve images.[15] It is important to note, however, that current MRI technology does not typically image *cellular processes*, but rather anatomical structures. There is significant interest in using ^{19}F MRI for imaging of solid structures as well as specific cellular processes. There are a few known examples of using ^{19}F MRI in pre-clinical and clinical settings.

A few examples of early ^{19}F-based MRI probles include lung imaging using biologically inert fluorine-containing gases which was used in the 1980's.[2] These early studies were completed using sulfur hexafluoride, hexafluoromethane, tetrafluoromethane, and fluoropropane mixed with oxygen to obtain images of the lungs. More recently, perfluorocarbons (PFC) are being explored for detection of cellular activity and cell tracking by ^{19}F MRI.[6] Some recent studies in animal models have demonstrated that ^{19}F MRI from PFCs or macrophages modified with perfluoro-15-crown-5-ether can be used to track inflammatory responses and cellular damage during myocardial infarction.[6] Unlike ^{1}H, the only signal obtained by ^{19}F MRI is from the PFC probe which was introduced into the system. In most animal studies, PFCs were injected via the tail vein of the rats and the PFCs collected in the damaged tissue of interest which then showed an increased ^{19}F signal upon imaging. This shows the

potential power of ^{19}F MRI for use within the clinic. We believe the current state of the art can be improved upon to give even more specific cellular information exploiting pathways and proteins that are known to be overexpressed in disease states such as cancers, Alzheimer's disease, and diabetes. While these methods are effective for imaging damage, they lack the ability to provide cellular and pathway specific information. We want to specifically target interactions, post-translational modifications, and pathways which are known to be involved in disease states.

1.1.2 Fluorinated Amino Acids Have Been Used To Increase Stability and Modulate Biological Activity in Proteins

Our goal is the direct incorporation of fluorinated amino acids into peptides and proteins for applications in molecular imaging. There are a large number of fluorinated amino acids found in the literature (Figure 1.3).[16] These amino acids span multiple classes of amino acids (aliphatic, aromatic, cyclic, etc.) with many different fluorine modifications, from a single fluorine atom to six fluorine atoms in various configurations. Fluorinated amino acids have been used for a multitude of different applications including as magnetic resonance probes within biological systems.[5] Fluorinated amino acids have been exploited in protein engineering to increase stability due to the fluorophobic effect, which is the self-association of fluorinated molecules.[17] One example of increased fluorophobic stabilizing structure is the use of trifluoroleucine and hexafluorleucine into coiled coil peptides.[18-22] The fluorinated leucine derivatives all align on the same face of the coiled coils, allowing them to pack together creating a fluorophobic core, stabilizing the structure. Fluorine has also been used to study protein folding in living cells. Trifluorovaline and pentafluorophenylalanine have been used to study the kinetics of folding in proteins

and increasing stability via burying the fluorinated amino acids in the hydrophobic core of the protein.[23] Trifluorovaline has been used to study folding in the NTL9 protein which is a globular α-ß protein.[24] Pentafluorophenylalanine has been used to study folding in the chicken Villian Headpiece protein and in 4-helix bundles.[25,26] Fluorinated amino acids can also be used to modulate protein activity within cells by increasing lipophilicity of a protein, similar to the effects seen in small molecules. One example of increased lipophilicity can be seen by the introduction of hexafluoroleucine into a bee venom protein, melittin, which increased partitioning of this protein into liposomes.[23]
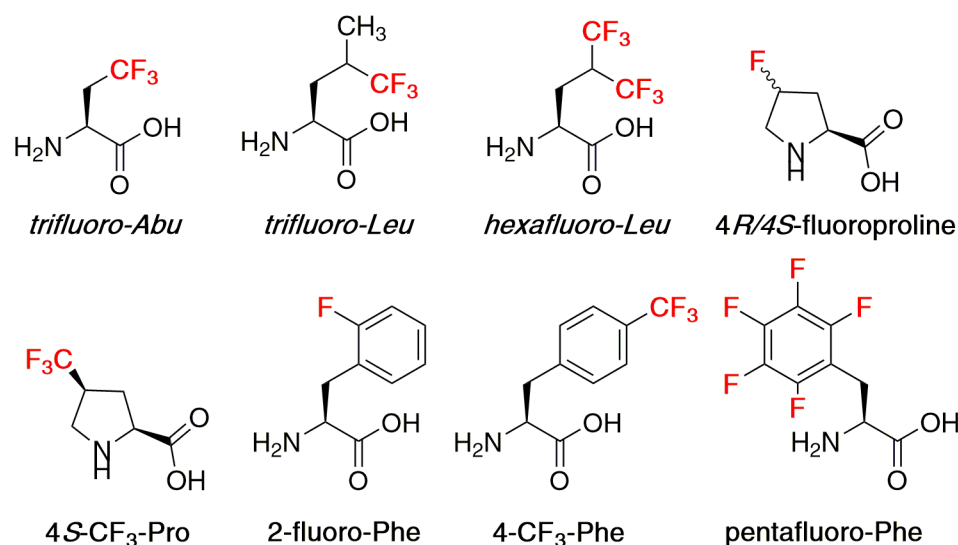
Figure 1.3   A subset of previously described fluorinated amino acids.

1.1.3   Fluorinated Amino Acids Impact Structure

When a large number of fluorine atoms are introduced into folded peptides and proteins with hydrophobic cores, increased stability is observed due to the

fluorophobic effect.[17] The fluorophobic effect is similar to the hydrophobic effect, but employs fluorine rather than simple non-polar atoms and molecules. The fluorophobic effect can be observed in both synthetic polymers and peptides. There are numerous examples of leucine-based fluorinated amino acids, including trifluoroleucine and hexafluoroleucine, stabilizing coiled coil proteins via a fluorinated core.[11,16,21] One example of fluorinated amino acids stabilizing coiled coil proteins via a fluorinated core is the incorporation of hexafluoroleucine into the synthetic leucine zipper A1 protein. The protein with all leucines replaced with hexafluoroleucine showed significantly increased stability under denaturing conditions including thermal stability (m.p. = 76 °C compared to wildtype m.p. = 54 °C) and increased stability in 8 M urea (less than 50% denaturation vs 100% denaturation for wildtype).[22] These proteins showed increased stability over their hydrocarbon-based counterparts due to the increased hydrophobicity observed with fluorine atoms. Furthermore, these amino acids have been incorporated into proteins using bacterial expression systems.[8]

Despite increased stability due to the fluorophobic effect, proteins containing multiple substitution of a single fluorinated amino acid favor extended conformations, strongly disfavoring α-helices.[27] Stability in coiled coil systems is due to a large hydrophobic core consisting of multiple fluorinated amino acids. When single fluorinated amino acids are substituted into a peptide sequences, helical structures are strongly disfavored, leading to a reduction in α-helical structure.[27] Reduction in α-helical structure has been demonstrated using the same fluorinated amino acids that have been used to stabilize protein structures. A number of fluorinated amino acids and their non-fluorinated equivalents were examined in the peptide Ac-YGGKAAAAKAXAAKAAAAK-NH$_2$, where X is the site of modification.[27] Both

alanine and 2-amino-butryic acid demonstrate a strong preference for α-helical structure (f_{helix} = 0.594 and 0.522, respectively) compared to the fluorinated analog, 2-amino-4-trifluorobutryic acid (f_{helix} = 0.208). Similarly, hexafluoroleucine (f_{helix} = 0.251) shows a significant reduction in α-helicity relative to leucine (f_{helix} = 0.502). While aromatic amino acids are less α-helical than some aliphatic amino acids, phenylalanine is more α-helical (f_{helix} = 0.426) than pentafluorophenylalanine (f_{helix} = 0.264). The strong preference of fluorinated amino acids to disfavor α-helical structure is problematic for design of fluorinated peptide and protein probes, because many 40% of protein structures within the PDB are α-helical. Furthermore, helical structures including α-helices and polyproline helices are often found at protein-protein interfaces.[28,29] Given that fluorinated amino acids can have such a strong impact on structural features within proteins and peptides, it is key to determine the secondary structure preferences, which greatly impacts the systems in which fluorinated can be incorporated.

1.1.4 Fluorine Is Superior to Other Heteroatoms For Magnetic Resonance Applications

Currently, a number of nuclei are being explored for imaging in biological systems and complex mixtures. ^{13}C and ^{15}N have both been used to image proteins and in cells.[30-33] Both of these atoms have a very low natural abundance (Table 1.1) and must be enriched to produce high enough concentrations for NMR, at great expense. Furthermore, the gyromagnetic ratios of ^{13}C and ^{15}N are significantly lower than those of ^{1}H and ^{19}F. The gyromagnetic ratio of an atom is directly proportional to the NMR signal sensitivity. The gyromagnetic ratio of each atom is proportional to the magnetic moment of the atom as well as the angular spin momentum. The only nuclei that has a

gyromagetic ratio similar to ^{1}H is ^{19}F, with the gyromagnetic ratio of both ^{13}C and ^{15}N being significantly lower (Table 1.1). The low gyromagnetic ratio significantly decreases sensitivity of these atoms, even when enriched to contain more than the natural abundance. In turn, much longer acquisition times (hours to days) are required for proteins, even for samples containing enriched ^{13}C and ^{15}N either *in vitro* or *in vivo*, for high resolution data.[30] In many cases, complex two-dimensional NMR experiments are required for structural information about proteins.[34,35] Finally, both ^{13}C and ^{15}N are found throughout biological systems, leading to background signal associated with endogenous carbon and nitrogen atoms. [19]F suffers from none of these problems as it is 100% naturally abundant and has a gyromagnetic ratio similar to [1]H.[2] This similarity to ^{1}H means that no special equipment is required to detect ^{19}F, unlike ^{13}C and ^{15}N, for MRI applications. [19]F is also absent from biological systems, leading to no competing background signal. Simple experiments can be used to detect changes by ^{19}F NMR due to chemical shift changes and chemical shift anisotropy due to the large chemical shift range for [19]F.[1,2,5-7,9,36-38] For these reasons, there is significant interest in developing ^{19}F-based magnetic resonance sensors for use within biological systems.

Nuclei	Nuclear Spin	Gyromagentic Ratio (rad s^{-1} T^{-1})	Natural Abundance (%)
^{1}H	½	267.522	~100
^{19}F	½	251.815	~100
^{13}C	½	67.238	1.1
^{15}N	½	19.338	0.37

Table 1.1   Spin, gyromagnetic ratio, and natural abundance of discussed nuclei. ^{19}F is the most similar to ^{1}H which is commonly used for medical imaging.

1.1.5   Design of Highly Sensitive Fluorinated Magnetic Resonance Probes.

While a number fluorinated amino acids have been synthesized, there are still substantial improvements that can be made.[16,39] Many of these amino acids require high concentrations, in the mid-micromolar range, for fast data acquisition times at high signal-to-noise ratios due to the use of a single fluorine or three equivalent fluorines.[2,5,9,11,19,36,38,40] Amino acids containing more than three fluorines, rarely have more than three equivalent fluorines. Fluorine atoms couple to other fluorines as well as to protons, which can lead to complex patterns, many of which produce spectral overlap which complicates analysis.[9,21,22,38,40] Figure 1.3 shows a selection of previously described fluorinated amino acids, many of which suffer from some or all of the problems described. 4*R*- and 4*S*-fluoroproline all demonstrate complex pattersn due to both coupling with protons as well as *cis-trans* isomerization which will be addressed later. 2-Fluoro-Phe also demonstrates coupling with neighboring protons. In all three of these cases, detection of a single fluorine requires at least mid-micromolar

concentrations and proton-coupling leads to a loss in sensitivity, requiring increased concentrations or the use of proton decoupling. Pentafluoro-Phe has more fluorine atoms; however, they are non-equivalent fluorines, which also leads to complex patterns by ^{19}F NMR, reducing the signal-to-noise ratio. Some of the more successful probes incorporate three chemically equivalent fluorines, including 4-CF$_3$-Phe. Dalvit *et al*. demonstrated that 4-CF$_4$Phe could be used to detect both phosphorylation events and proteolysis of peptides in solution at mid-micromolar concentrations.[41] Furthermore, 4-CF$_3$-Phe has successfully been incorporated into protein via the Amber suppression method and was successfully detected by ^{19}F NMR in both nitroreductase and histidinol dehydrogenase to monitor protein folding and *in vivo* activity.[42] While no concentration is reported for these protein systems, significant broadening of the ^{19}F peaks are observed and the ^{19}F-labeled proteins were not always visible by ^{19}F NMR. These data demonstrate a need for different and more sensitive ^{19}F probes.

Our design addresses issues described above by incorporating a perfluoro-*tert*-butyl moiety onto amino acids via an ether linkage (Figure 1.4). The perfluoro-*tert*-butyl moiety contains nine equivalent fluorines which produces a single peak by ^{19}F NMR, which is detectable at nanomolar concentrations.[43,44] Compounds containing a perfluoro-*tert*-butyl group are commercially available as the alcohol and the halides (I, Br, Cl). Previously, perfluoro-*tert*-butanol has been installed as an ether on primary alcohols using the Mitsunobu reaction and on aromatic rings using diazonium coupling.[45,46] The first example of the Mitsunobu reaction being used to install a perfluoro-*tert*-butyl ether was in 1996 to synthesize both a benzodioxane and a geraniol with a perfluoro-*tert*-butyl ether moiety. More recently, perfluoro-*tert*-butyl ethers have been introduced into aromatic systems, specifically bis-styrylbenezenes,

for the detection of ß-amyloids in Alzheimer's disease mouse models. Prior to the work by Pandey *et al.* this chemistry had never been used to fluorinate alpha-amino acids. The Mitsunobu reaction has been used to install a perfluoro-*tert*-butyl ether into ß-amino acids which were successfully introduced into tripeptides.[47] A large singlet peak was observed by ^{19}F NMR when incorporated into 4*R*-perfluoro-*tert*-butyl hydroxyproline. In the case of 4*S*-perfluoro-*tert*-butyl hydroxyproline, two peaks were visible due to *cis-trans* isomerization at the amide bond. ^{19}F NMR spectra showed 2 peaks, representing the *cis* and *trans* conformations in slow exchange.[43] These data indicate that ^{19}F NMR with a perfluoro-*tert*-butyl ether moiety can allow detection of backbone structure.
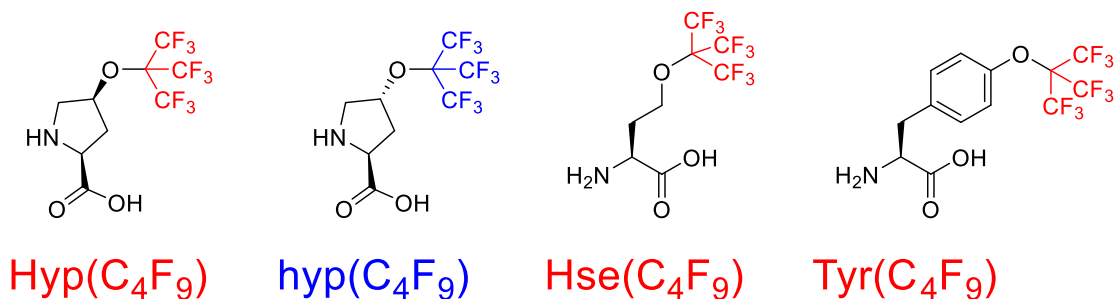
Figure 1.4   Proposed amino acids to be synthesized containing a perfluoro-*tert*-butyl ether.

The primary goal of this work is to develop novel amino acids containing perfluoro-*tert*-butyl ethers which will access different applications. This can be done by installing the perfluoro-*tert*-butyl ether on a variety of different classes of amino acids, including cyclic, aliphatic, and aromatic amino acids. Each of these amino acids

can then be introduced into appropriate peptide-based model systems to exploit their unique characteristics. For example, there are many instances in biology of recognition sequences specifically selecting for proline due to its unique ability to access polyproline helices and turn structures.[29,48-50] Alternatively, aliphatic amino acids more readily adopt α-helical structures.[51,52] Aromatic amino acids are often used in recognition sequences due to their large size and hydrophobic nature. Another benefit to installing the perfluoro-*tert*-butyl on multiple amino acids is it accesses a variety of sizes of amino acid. The proposed hydroxyproline derivatives are smaller in size than the tyrosine or the homoserine derivative. The hydroxyproline derivatives are also potentially more sensitive to perturbations to nearby amino acids due to the cyclic nature of proline as demonstrated by the detection of *cis* and *trans* amide bonds described by Pandey *et al*.

1.1.6 Peptide-Based Model Systems to Determine Structural Effects of Novel Amino Acids

As discussed in the last section, perfluoro-*tert*-butyl alpha-amino acids have not previously been synthesized and fully characterized to determine potential effects on structure. In order to properly design applications for the designed amino acids, characterization of how the addition of the perfluoro-*tert*-butyl ether impacts structure was necessary. While fluorine atoms generally disfavor compact secondary structure, the incorporation of the perfluoro-*tert*-butyl group via an ether linkage has the potential to minimize the effect of the fluorine atoms on secondary structure. The first example of alpha-amino acids containing a perfluoro-*tert*-butyl ether was published by Pandey *et al.* in the Ac-TYXN-NH$_2$ peptide model system (Figure 1.5). This system was derived from the peptide Ac-TYPN-NH$_2$ to examine the *cis* amide bond

propensity in type VI ß-turns. Aromatic-proline sequences play a key role in protein folding events, with these specific sequences showing a higher propensity for *cis* amide bond formation. The Ac-TYPN-NH$_2$ context was developed to maximize the amount of *cis* amide bond formation. Tyr-Pro sequences form more stable *cis* amide bonds due to a favorable C-H/π interaction between the proline H$_\alpha$ and the tyrosine aromatic ring. All twenty conical amino acids were screened at the first and fourth position in the peptide amide bond propensity was determined by measuring the $K_{trans/cis}$. The $K_{trans/cis}$ is a measurement of the equilibrium between the *trans* and *cis* amide bonds, which are in slow exchange. The slow exchange can be observed by ^{1}H NMR where both *cis* and *trans* peaks are visible in the NMR spectrum of the peptide. In the case of unsubstituted proline the $K_{trans/cis} = 2.7$ ($\Delta\Delta G = 0.00$ kcal mol^{-1}). The *trans* amide bonds are stabilized by the *exo* ring pucker of the proline ring while the *cis* amide bonds are stabilized by the *endo* ring pucker. This system is ideal for examining both the effects of proline and aromatic derivatives on peptide structure.

4*R* and 4*S* perfluoro-*tert*-butyl hydroxyproline were originally synthesized on solid phase in the Ac-TYXN-NH$_2$ peptide via a method termed "proline editing". Proline editing allows quick and easy access to a large number of 4-substituted prolines for various applications including bio-orthogonal chemistry, novel sensors, and reactive handles. The full length peptide Ac-TYHypN-NH$_2$, where Hyp = 4*R*-hydroxyproline, is synthesized with a trityl-protected hydroxyproline. The trityl protecting group can be cleanly and easily removed using 2% TES and 2% TFA in CH$_2$Cl$_2$ without cleavage from resin or deprotection of other residues in the peptide. The hydroxyproline can then be converted into a large number of derivatives using the alcohol as a reactive handle. All 123 4-subtitued derivatives synthesized were

characterized by ^{1}H NMR to determine the *cis* amide bond propensity. In general, 4*R*-derivatives favor *trans* amide bonds due to a preference for *exo* ring pucker, while 4*S*-derivatives of proline favor *cis* amide bonds due to a preference for *endo* ring pucker. This preference is observed in the 4*R*-hydroxyproline ($K_{trans/cis}$ = 5.6, $\Delta\Delta G$ = -0.43 kcal/mol^{-1}) and 4*S*-hydroxyproline ($K_{trans/cis}$ = 2.7, $\Delta\Delta G$ = 0.00 kcal/mol^{-1}). The same trend is observed in with the simple modification of the alcohol to a single fluorine modification 4*R*-fluoroproline is significantly more *trans* favoring ($K_{trans/cis}$ = 7.0, $\Delta\Delta G$ = -0.56 kcal/mol^{-1}) than 4*S*-fluoroproline ($K_{trans/cis}$ = 1.5, $\Delta\Delta G$ = +0.35 kcal/mol^{-1}). In the case of perfluoro-*tert*-butyl ethers, the same trend is seen. 4*R*-Perfluoro-*tert*-butyl hydroxyproline favors the exo ring leading to compact dihedral angles and *trans* amide bonds ($K_{trans/cis}$ = 4.0, $\Delta\Delta G$ = −0.23 kcal/mol^{-1}) while the 4*S*-perfluoro-*tert*-butyl hydroxyproline favors an endo ring pucker leading to extended dihedral angles and *cis* amide bonds ($K_{trans/cis}$ = 1.2, $\Delta\Delta G$ = +0.48 kcal/mol^{-1}) (Figure 1.5). The 4*R*-perfluoro-*tert*-butyl hydroxyproline is unremarkable in terms of its *cis* amide bond propensity, following the trend of other electron-withdrawing 4*R*-subsitituted proline derivatives. The 4*S*-perfluoro-*tert*-butly hydroxyproline, however, was the most *cis* favoring derivative synthesized, demonstrating these two amino acids to not only be conformationally different amino acids, but may also have the ability to modulate protein structure. Furthermore, both *cis* and *trans* peaks were visible by ^{19}F NMR. The $K_{trans/cis}$ measured from ^{1}H NMR was identical to that calculated from the ^{19}F NMR. The ^{19}F NMR demonstrates the potential of these probes to be used to detect backbone conformation by ^{19}F NMR.
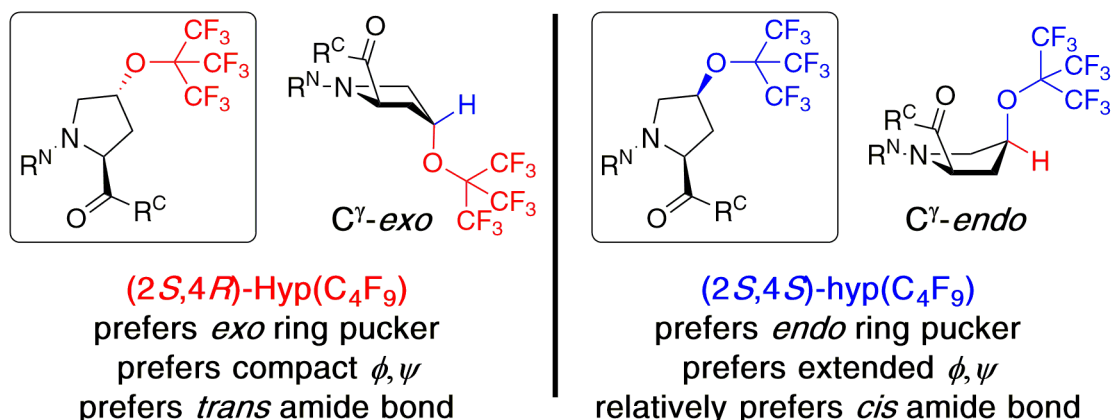
Figure 1.5   (2S,4R)- and (2S,4S)-Perfluoro-*tert*-butyl 4-hydroxyprolines and their expected conformational preferences, in which the 4-substituents are in a pseudoaxial position on the pyrrolidine ring due to stereoelectronic effects of the 4-substitution, based on data in Ac-TYXN-NH$_2$ peptides (X = 4-substituted proline).[43] Hyp = 4R-substituted (*trans* relative stereochemistry) hydroxyproline, indicated by use of capitalized three-letter code and red color; hyp = 4S-substituted (*cis* relative stereochemistry) hydroxyproline, indicated by lower-case three-letter code and blue color. Proline exhibits a mixture of *exo* and *endo* ring puckers. R^N and R^C indicate the N-terminal and C-terminal peptide sequences, respectively.

The Ac-TYPN-NH$_2$ model system can also be used to analyze the electronic effects of aromatic residues by modifying the Tyr to other amino acids.[53] The C-H/π interaction between the aromatic proline H$_\alpha$ and the aromatic ring can be modulated by introducing electron-donating or electron-withdrawing modifications to the aromatic ring. Electron-withdrawing groups favor *trans* amide bonds by weakening the C-H/π interaction between the proline H$_\alpha$ and the aromatic ring by removing electron density from the aromatic ring (Figure 1.6). The electron-poor aromatic system has a significant reduction in driving force for the C-H/π interaction. Electron-donating groups favor *cis* amide bonds by strengthening the C-H/π interaction

17

between the proline H_α and the aromatic ring (Figure 1.6). The electron-donating substitution increases the electron density in the aromatic ring, leading to increased driving force for the C-H/π interaction. A subset of unnatural 4-substituted aromatic derivatives have been examined in the Ac-TXPN-NH$_2$ peptide model system. For example, a classic example of electron-withdrawing substituents is the addition of a nitro-group. When 4-NO$_2$-Phe was introduced into the Ac-TXPN-NH$_2$ peptide model system $K_{trans/cis}$ = 5.5 (ΔG = -1.00 kcal mol^{-1}) which is significantly more *trans* favoring than Tyr $K_{trans/cis}$ = 2.7 (ΔG = -0.58 kcal mol^{-1}). In comparison, a classic example of electron-donating substituents is the addition of an amino-group. In the case of 4-NH$_2$-Phe, the $K_{trans/cis}$ = 2.1 (ΔG = -0.44 kcal mol^{-1}) which is significantly more *cis* favoring than Tyr. Furthermore, a linear free energy relationship was determined between the $K_{trans/cis}$ and the Hammett sigma constant (σ). The Hammett constant for substituents which have not previously been measured, can be calculated using the linear free energy relationship which was determined by Thomas *et al*. Prior to the work presented herein, there was only one published example of a perfluoro-*tert*-butyl ether, which did not include any characterization of the electronic effects of the perfluoro-*tert*-butyl ether modification on the aromatic system. Due to the nine fluorines, it is critical we understand the electronic effects of the designed 4-perfluoro-*tert*-butyl Tyr prior to incorporation into proteins for potential applications, which will be discussed further in the results and discussion sections of this chapter.
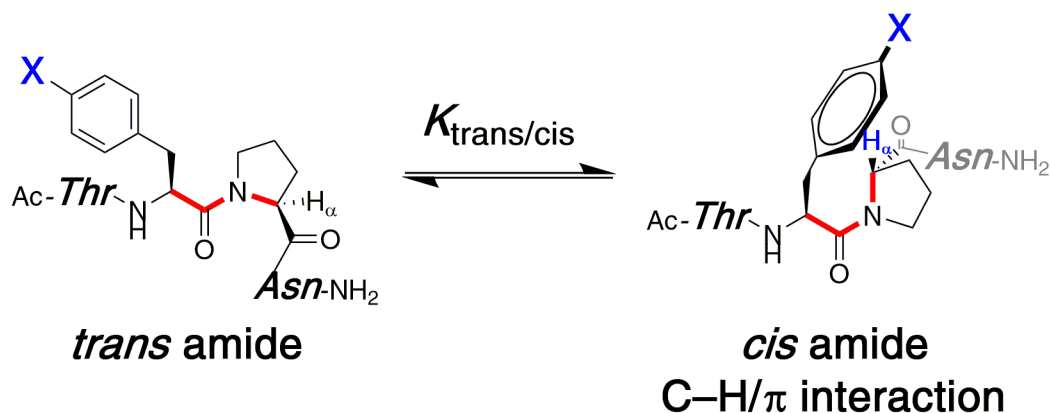
Figure 1.6   *Cis-trans* isomerism of the aromatic-proline amide bond (red) in Ac-TXPN-NH$_2$ peptides (X = aromatic amino acid) can be used to quantify aromatic electronic effects, via $K_{trans/cis}$ and via the δ of proline H$_\alpha$ when in the *cis* conformation (with smaller $K_{trans/cis}$ and more upfield Pro$_{cis}$ H$_\alpha$ δ indicating a more electron-rich aromatic and a stronger C–H/π interaction).

Despite the importance of aromatic-proline interactions in protein folding, this system is not applicable to all amino acids. In exploring potential applications for the designed perfluoro-*tert*-butyl amino acids, it is important to consider applications which incorporate important secondary structures within proteins. Two common compact secondary structures including α-helices and polyproline II (PPII) helices. Both of these structures are common within proteins and are often found at the interfaces of protein-protein interactions. As noted in Section 1.1.3, fluorine negatively impacts propensity for α-helical structure; thus it is important to understand how the perfluoro-*tert*-butyl ether modification will impact both α-helical and polyproline helix structure.

Some aliphatic amino acids as well as proline are found in α-helical structures within proteins. α-Helical propensity of an amino acid can be measured using alanine-

rich Baldwin-type helical peptides.[51] This is a well-established system which has been used to measure helical propensities at various points throughout the helix (beginning, middle, and end) of all canonical amino acids as well as a large set of unnatural amino acids.[51,52,54] In general, aliphatic amino acids such as Ala, Leu, and Lys readily form α-helices while aromatics, such as His and Trp, as well as Gly and Pro, less readily form α-helices. α-Helix propensity is measured using circular dichroism by minima at 222 and 208 nm. Proline is known to disrupt or stop α-helices when introduced in the middle or end of an α-helical sequence, but it is also able to signal the start of the helix.[52] Given the extensive amount of characterization of amino acids within Baldwin model α-helical peptides to determine α-helical propensity, this approach will be used to examine the perfluoro-*tert*-butyl hydroxyproline and the perfluoro-*tert*-butyl homoserine derivatives proposed in Figure 1.4.

PPII helix propensity can be measured using the Ac-GPPXPPGY-NH$_2$ peptide model system.[55,56] This system was developed by Brown and Zondlo and used to measure the polyproline helical propensity of all the canonical amino acids as well as a large subset of unnatural amino acids. The PPII helix is characterized using circular dichroism with a maximum at 228 nm. Unsurprisingly, Pro is the most PPII helix promoting of the canonical amino acids. Long chain aliphatic amino acids such as Leu, Arg, and Lys are also highly PPII helix promoting. Alternatively, ß-branched amino acids, such as Ile and Val, as well as aromatic amino acids significantly less PPII helix promoting. The Ac-GPPXPPGY-NH$_2$ peptide model system will be used to characterize the PPII helix propensity for the perfluoro-*tert*-butyl hydroxyproline and the perfluoro-*tert*-butyl homoserine derivatives proposed in Figure 1.4.

Both α-helix and PPII helix peptide model systems will be used to characterize aliphatic perfluoro-*tert*-butyl amino acids to determine their propensity for compact secondary structures. Utilizing this series of unnatural amino acids will allow us to understand how the perfluoro-*tert*-butyl ether modification will affect secondary structure of peptides incorporating the designed amino acids. This approach will also allow us to appropriately design applications for these unique amino acids based on the structural impact.

1.2 Results

1.2.1 Synthesis of (2*S*,4*R*)-Perfluoro-*tert*-butyl-4-hydroxyproline

4*R*-Perfluoro-*tert*-butyl-4-hydroxyproline (Hyp(C$_4$F$_9$)) has been synthesized on solid-phase within a peptide. Briefly, 4*R*-Hydroxyproline was incorporated by solid phase peptide synthesis into Ac-TYXN-NH$_2$, utilizing a trityl protecting group on the hydroxyproline.[43] The trityl group was selectively removed using 2% TFA before subjecting the peptide on resin to a Mitsunobu reaction to install the perfluoro-*tert*-butyl group. While this reaction did result in product, it was the lowest yielding reaction of the over 100 hydroxyproline derivatives synthesized.

When this reaction chemistry was applied on the larger peptides described in Chapter 3 (twelve residue Abltide peptide), it was unsuccessful, resulting in very low yields or no product observed, most likely due to the large steric demand of the perfluoro-*tert*-butyl group. While installation of the perfluoro-*tert*-butyl ether by the Mitsunobu reaction has previously been established, there are a significant number of challenges with the reaction chemistry. The reaction mechanism requires a phosphine, in this case triphenylphosphine, to react with the azo reagent, DIAD. The DIAD

21

intermediate is a zwitterion which can then deprotonate the acidic nucleophile, in this case perfluoro-*tert*-butanol. The protonated DIAD intermediate can in turn react with the primary or secondary alcohol of the substrate, in this case hydroxyproline, can add to, forming a leaving group. A nucleophile, such as perfluoro-*tert*-butanol, can then perform an S_N2 reaction, displacing triphenylphosine oxide to form the product. Perfluoro-*tert*-butanol, despite having a relatively low pK_a, ~ 5, is large and sterically demanding, leading to lower yields than better nucleophiles.[45,57] Furthermore, hydroxyproline is a secondary alcohol which is sterically more demanding than aliphatic primary alcohols. A common side reaction as a result of this mechanism is elimination of the triphenylphosphine oxide, without successful nucleophilic attack leading to dehydroproline.[58] The mechanism of the Mitsunobu reaction requires attack, of the nucleophile in an S_N2 manner on the substrate. In the case of 4*R*-perfluoro-*tert*-butyl hydroxyproline, the starting material is in the 4*S* configuration leading to an *endo* ring pucker and relatively higher amounts of *cis* amide bonds than the 4*R* configuration. The 4*S* conformation of the starting material makes it less accessible to the nucleophile, especially in the case of a very large, sterically demanding nucleophile such as perfluoro-*tert*-butanol. There is also a possibility that the sequence of the peptide (Ac-KKGEAIYAA(Hyp(C_4F_9))FA-NH$_2$ prevented the solid-phase Mitsunobu reaction from preceding. The nearby Phe residue potentially prevented the reaction from proceeding due to the steric demand on resin. In order to quickly and cleanly access a number of peptides to examine both the structural effects of the perfluoro-*tert*-butyl group as well as the ability to screen peptides for potential applications, a solution-phase synthesis to the Fmoc-protected amino acid was developed.

The solution-phase synthesis of Fmoc-4*R*-perfluoro-*tert*-butyl hydroxyproline utilized the same chemistry as Pandey *et al.* on the hydroxyproline monomer (Figure 1.7). 4*R*-Perfluoro-*tert*-butyl hydroxyproline was synthesized from commercially available 4*R*-hydroxyproline, which is inexpensive due to its isolation from collagen found in chicken feathers. The free amine was Boc protected and the carboxylic acid was methyl protected by well-established protocols.[59] Due to the inversion of the stereocenter upon instillation of the perfluoro-*tert*-butyl ether via the Mitsunobu reaction, the stereocenter must first be inverted to the 4*S* configuration via a Mitsunobu reaction with *para*-nitrobenzoic acid to generate compound **2**. The *para*-nitrobenzoic acid protecting group was then removed via sodium azide cleavage to generate Boc-4*S*-hydroxyproline-OMe (compound **3**). Compound **3** was then subjected to a second Mitsunobu reaction to install the perfluoro-*tert*-butyl ether, generating compound **4** in 25% yield. This compound is particularly problematic for purification as it has a similar R_f value to the desired product. In order to maximize the amount of desired product, a base was added to the reaction to catalyze the formation of the DIAD-adduct and to generate perfluoro-*tert*-butoxide more readily as the nucleophile. Compound **4** was then deprotected on the C-terminus via LiOH cleavage of the O-methyl ester to generate Boc-4*R*-perfluoro-*tert*-butyl-hydroxyproline (**5**) which is compatible with Boc solid-phase peptide synthesis. Alternatively, a global deprotection was used to generate the free amino acid (**6**) using 2 M HCl. compound **6** was then Fmoc-protected using previously established conditions.[43]
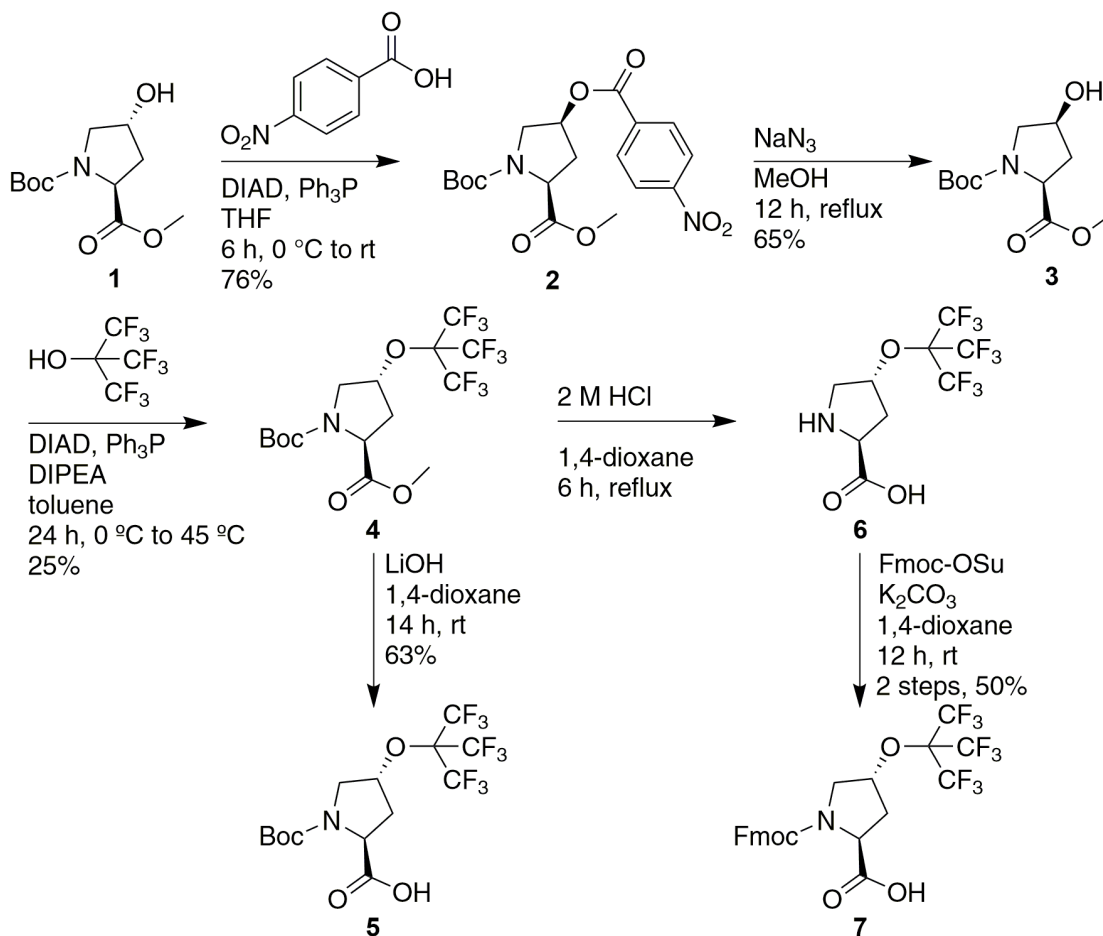
Figure 1.7   Synthesis of (2*S*,4*R*)-perfluoro-*tert*-butyl 4-hydroxyproline as free, Boc, and Fmoc-amino acid.

1.2.2   Synthesis of (2*S*,4*S*)-Perfluoro-*tert*-butyl-4-hydroxyproline

Due to the failed attempt at synthesizing the 4*R*-perfluoro-*tert*-butyl hydroxyproline on solid phase, no solid phase chemistry was attempted in the preparation of peptides with this derivative. 4*S*-Perfluoro-*tert*-butyl hydroxyproline was originally synthesized on solid phase as reported by Pandey *et al.* 4*S*-Perfluoro-*tert*-butyl-hydroxyproline was synthesized from commercially available 4*R*-hydroxyproline (Figure 1.8). The free amine was Boc protected and the carboxylic

acid was methyl protected by well-established protocols to generate compound **1**.[59] The perfluoro-*tert*-butyl group was installed via Mitsunobu reaction with inversion of the stereocenter to produce protected Boc-4*S*-perfluoro-*tert*-butyl hydroxproline (**8**). Compound **8** was either be selectively deprotected with LiOH to produce the Boc free acid (**9**) or globally deprotected using 2 M HCl to generate compound **10**. The free amino acid was then Fmoc-protected (**11**) for use in solid phase peptide synthesis using previously established protocols.[43]
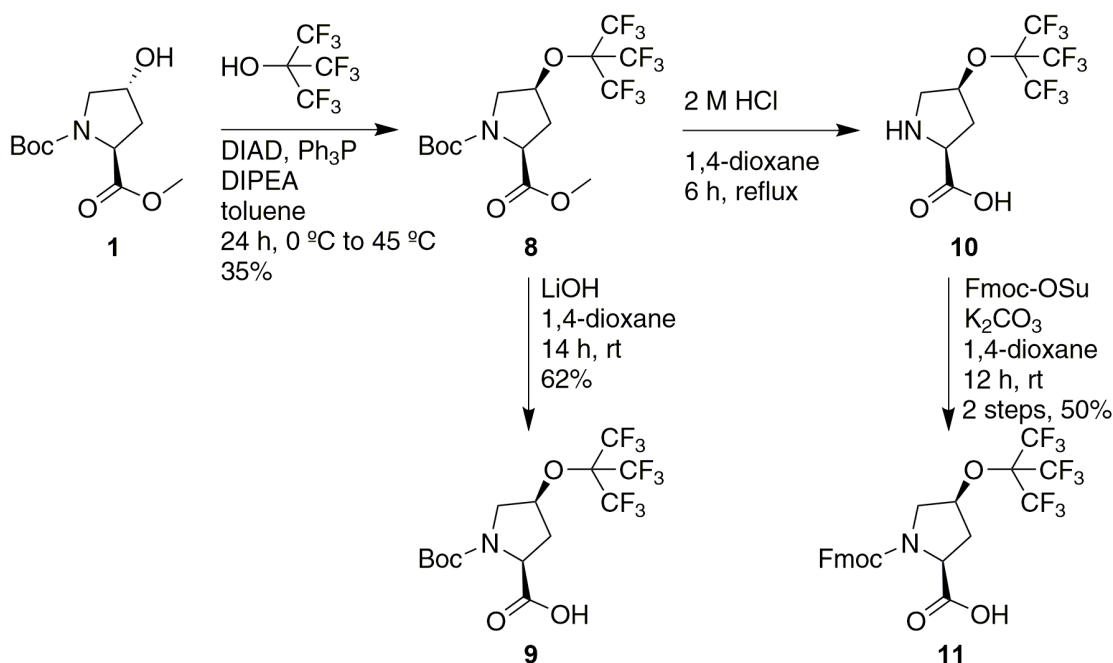
Figure 1.8   Synthesis of (2*S*,4*S*)-perfluoro-*tert*-butyl 4-hydroxproline as free, Boc, and Fmoc amino acids.

Fmoc-4*S*-perfluoro-*tert*-butyl hydroxyproline was synthesized in five steps, with the critical Mitsunobu reaction installing the perfluoro-*tert*-butyl ether resulting in the same challenges as discussed in the previous section. The exception here was

the overall higher yield (35%, compared to 25% for the 4R configuration) due to the confiiguration of the hydroxyproline. The increased yield was also observed when these amino acids were synthesized in a peptide on solid phase.[43]

1.2.3   Synthesis of Perfluoro-*tert*-butyl Homoserine

While the hydroxyproline derivatives are useful due to the unique structure of proline, which will be addressed in Chapter 3, there are many applications where proline and proline derivatives are not accommodated. To address biological applications of perfluoro-*tert*-butyl amino acids which do not accommodate proline, perfluoro-*tert*-butyl homoserine was synthesized. Perfluoro-*tert*-butyl-homoserine was synthesized via the Mitsunobu reaction as described by Bauer and Marsh (Figure 1.9).[60] The reported synthesis, however, used diazomethane to methyl protect the carboxylic acid. The protecting group chemistry used was changed due to the explosive nature of diazomethane. This reaction would have introduced unnecessary danger in the synthesis of the amino acid. The Boc and methyl groups used to protect the amine and carboxylic acid, respectively, were installed to generate compound **14**.[61] The perfluoro-*tert*-butyl group was installed using the Mitsunobu reaction under reflux conditions to generate compound **15**. The Mitsunobu conditions for the installation of the perfluoro-*tert*-butyl ether differ slightly from the hydroxyproline derivative syntheses. A base was not required for the reaction and THF was used as the solvent for the reaction. Instillation of the perfluoro-*tert*-butyl group onto the homoserine side chain is a less sterically demanding reaction due to the use of a primary alcohol over a secondary alcohol. The homoserine alcohol is less sterically congested than the hydroxyproline, leading to higher yields than either hydroxyproline derivative. Toluene and THF have similar reactivity in Mitsunobu reaction. Toluene was not used

for this synthesis due to the literature precedent utilizing THF.[45,58,60,62,63] The methyl

ester and Boc protecting groups were then removed using LiOH and HCl in succession

to generate compound **16**. Global deprotection using 2 M HCl was unsuccessful on

perfluoro-*tert*-butyl homoserine, so a stepwise deprotection of ester hydrolysis

followed by Boc deprotection was used. Finally, the amine was protected using Fmoc-

OSu to generate compound **17** to be compatible with Fmoc solid-phase peptide
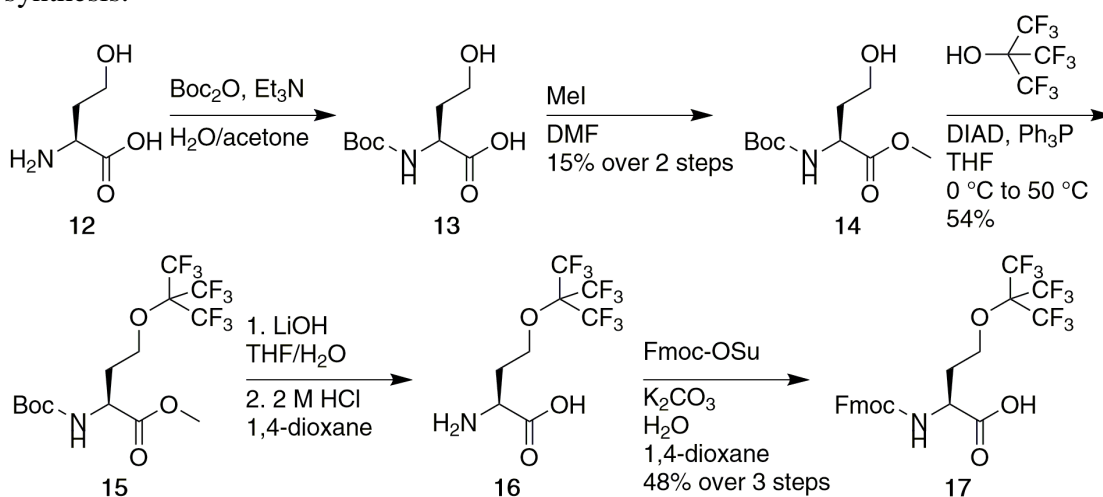
synthesis.

Figure 1.9   Synthesis of perfluoro-*tert*-butyl homoserine as Boc, free, and Fmoc
amino acids.

Synthesis of perfluoro-*tert*-butyl serine was also attempted unsuccessfully.

Serine is a particularly difficult substrate for the Mitsunobu reaction despite being a

primary alcohol.[58] A number of different N-terminal protecting groups was screened,

including Boc, Benzyl, and Cbz, but no product was observed in any case. Solvents

including toluene, tetrahydrofuran, and dichloromethane were also screened without

any successful product formation. Finally, a number of organic bases including

DIPEA, triethylamine, and 2,6-lutidine, as well as different reaction temperatures were

screened. None of these reactions ultimately yielded product. The most common side product from these reactions was dehydroalanine. While Mitsunobu reactions are known to occur on serine, they are often very low yielding and require more robust reagents such as ADDP or TMAD.[62] Mitsunobu conditions are primarily used to produce dehydroalanine and as for intramolecular reactions to create lactones.[64] After a series of test reactions, serine was abandoned in favor of homoserine. The combination of a poor substrate with a poor nucleophile indicate a different synthetic route will most likely be required to obtain a perfluoro-*tert*-butyl serine derivative.

1.2.4 Synthesis of Perfluoro-*tert*-butyl Tyrosine

To synthesize 4-perfluoro-*tert*-butyl tyrosine, diazonium coupling chemistry was used, due to tyrosine being an unsuitable substrate for the Mitsunobu reaction. The Mitsunobu reaction requires the use of a primary or secondary sp^3 hybridized alcohol as a substrate. While there are some examples of generating aryl ethers using the Mitsunobu reaction, the reaction would require the use of perfluoro-*tert*-butanol as the substrate, and there is poor precedent for Mitsunobu reactions on tertiary carbons. Previously, diazonium coupling has been reported on an aromatic system with perfluoro-*tert*-butanol.[46] Diazonium couplings have been reported in the literature with 4-aminophenylalanine to synthesize 4-chloro-, 4-azido-, 4-tetrazole-, 4-thiol-, and 4-selenol-phenylalanine derivatives, as well as phenylalanine derivatives containing nitrogen and sulfur mustards..[65-73] While this work focuses on the incorporation of the perfluoro-*tert*-butyl ether, 4-diazonium-Phe is potentially a very useful reactive handle which can be exploited for quick access to a variety of unique unnatural amino acids using . This reaction was also demonstrated on the peptide in solution phase. This amino acid was synthesized from the commercially available Fmoc-4-NH_2-Phe using

28

sodium nitrite and tetrafluoroboric acid using literature protocols (Figure 1.10). The tetrafluoroborate salt was used as it is more stable than other diazonium salts. Diazonium salts, however, should be handled with care, and this intermediate was not characterized to minimize potential hazards. The Fmoc-4-(N_2-BF_4)-Phe was extracted and the solvent was removed but it was not dried completely. The product was also used immediately in the next reaction. Compound **19** was subjected to reflux overnight in neat perfluoro-*tert*-butanol with an open reflux condenser to prevent explosion. Fmoc-perfluoro-*tert*-butyl tyrosine (Compound **20**) was also deprotected to produce the free amino acid. The synthetic route to this amino acid for peptide use is one of the highest yielding syntheses to an amino acid containing a perfluoro-*tert*-butyl ether. The higher yield is, in part, due to the installation of the perfluoro-*tert*-butyl ether via diazonium coupling over the Mitsunobu reaction, leading to higher yields, and the use of Fmoc-protected starting materials, preventing the need for extra synthetic steps. The Fmoc-perfluoro-*tert*-butyl tyrosine is synthetically accessible in less than 24 hours due to the short synthetic route. All other derivatives require at least five days to synthesize due to the number of synthetic steps required.
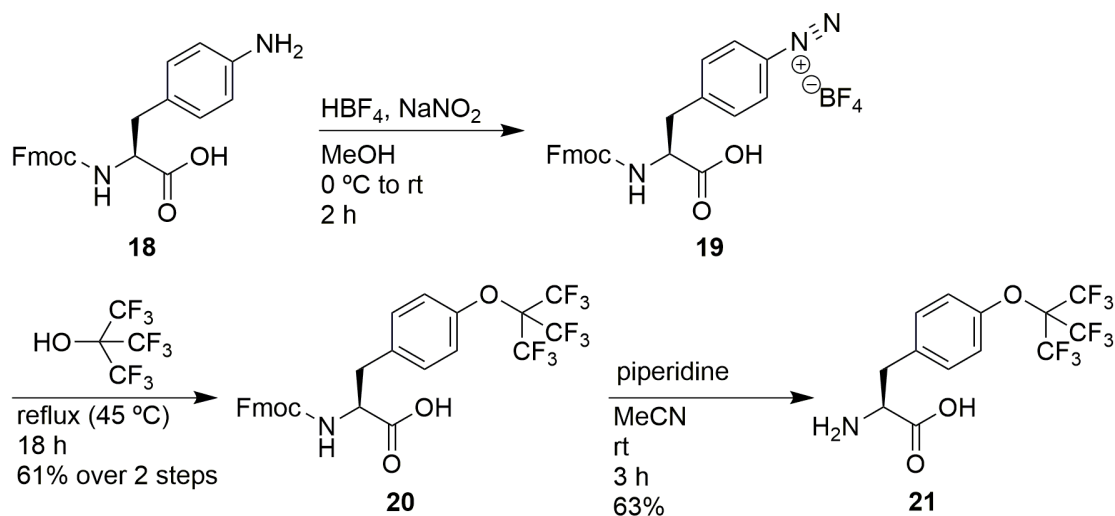
Figure 1.10 Synthesis of perfluoro-*tert*-butyl tyrosine as Fmoc and free amino acid.

Due to the potential for epimerization of compound **21**, the other enantiomer was also synthesized and both amino acids were subjected to chiral HPLC. Chiral columns were used which can separate enantiomers unlike standard HPLC columns. Both the L-Fmoc-perfluoro-*tert*-butyl tyrosine and the D-Fmoc-perfluoro-*tert*-butyl tyrosine were dissolved in methanol and subjected to chiral HPLC individually (Figure 1.11, top and middle). The two solutions of L-Fmoc-perfluoro-*tert*-butyl tyrosine and D-Fmoc-perfluoro-*tert*-butyl tyrosine were also co-injected to verify the chiral HPLC results (Figure 1.11, bottom).
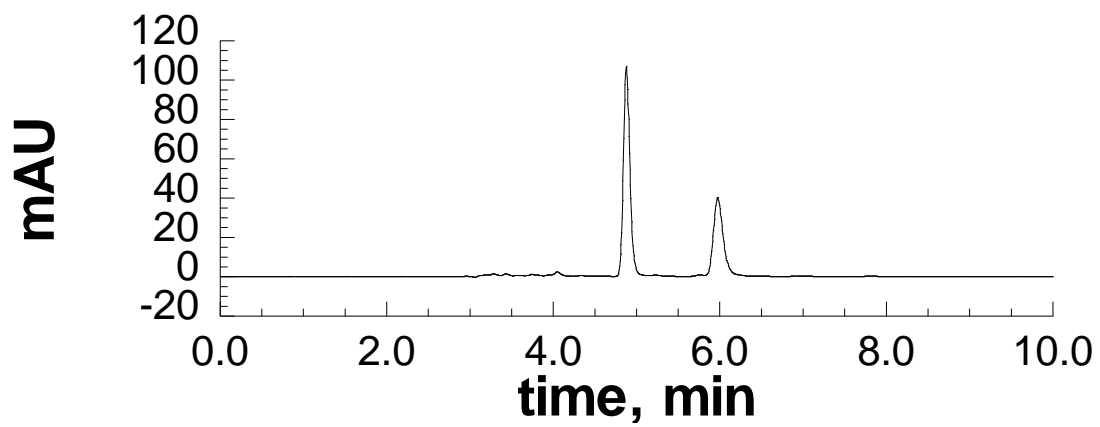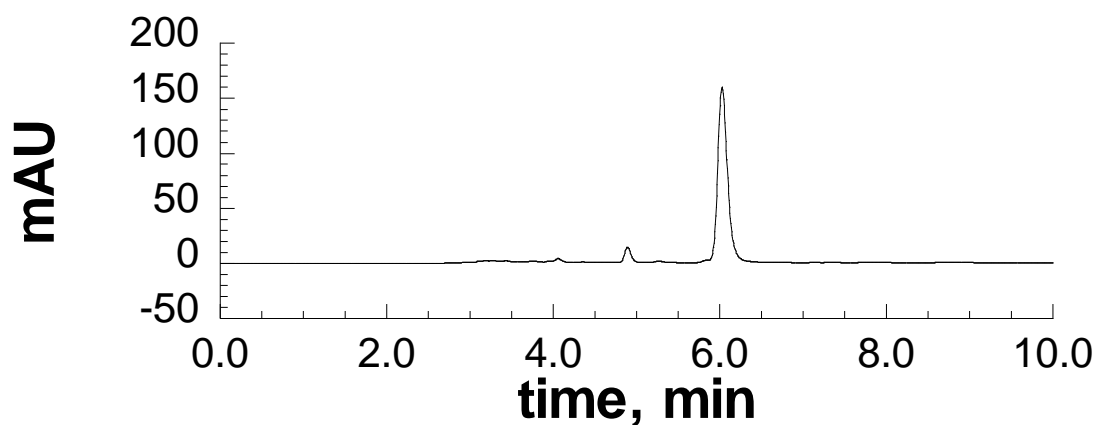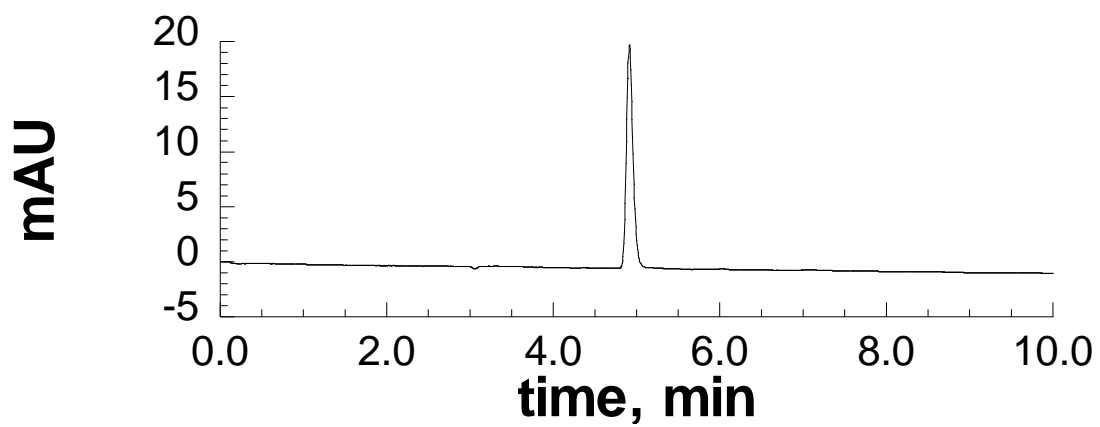
Figure 1.11 Analytical chiral HPLC chromatograms of purified **20** Fmoc-L-Tyr(C$_4$F$_9$)-OH (top), Fmoc-D-Tyr(C$_4$F$_9$)-OH (middle), and co-injection (bottom) (UV detection at 254 nm) using isocratic 20% isopropanol in hexanes over 30 minutes on a Daicel ChiralPak 1A column (250 × 4.6 mm, 5 µm particle, 1.0 mL/min). Compound **20** exhibited greater than 98% ee by integration.

The perfluoro-*tert*-butyl Tyr was also synthesized in solution on a fully synthesized peptide (Figure 1.12). Purified Ac-T(4-NH$_2$-Phe)PN-NH$_2$ was subjected to diazonium salt formation as described above. The solvent was then removed and the peptide was redissolved in perfluoro-*tert*-butanol. The solution was heated to 45 °C for 18 hours. The peptide was then purified by HPLC to yield Ac-T(Tyr(C$_4$F$_9$))PN-NH$_2$ (Figure 1.13). While this reaction was easily accomplished on a four residue peptide, it was not a viable method for installing the perfluoro-*tert*-butyl tyrosine on larger peptides. This reaction chemistry was also attempted on the MDM2 binding peptide discussed in Chapter 2. No product was isolated by HPLC yielding peptide containing perfluoro-*tert*-butyl tyrosine. This may be, in part, due to the steric demand of the larger peptides preventing the perfluoro-*tert*-butanol from reacting with the diazonium salt. Another potential explanation is these larger peptides may potentially be self-associating, preventing access of the perfluoro-*tert*-butanol which may be addressed using different peptides.
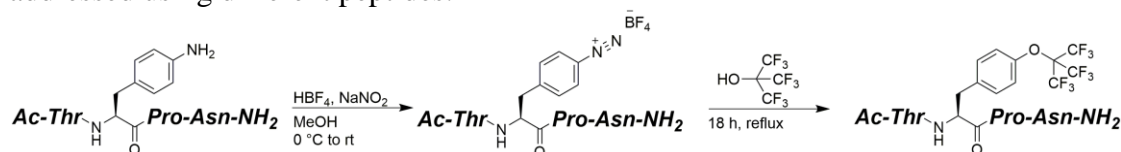
Figure 1.12 Solution-Phase diazotization and synthesis of Tyr(C4F9) within peptides
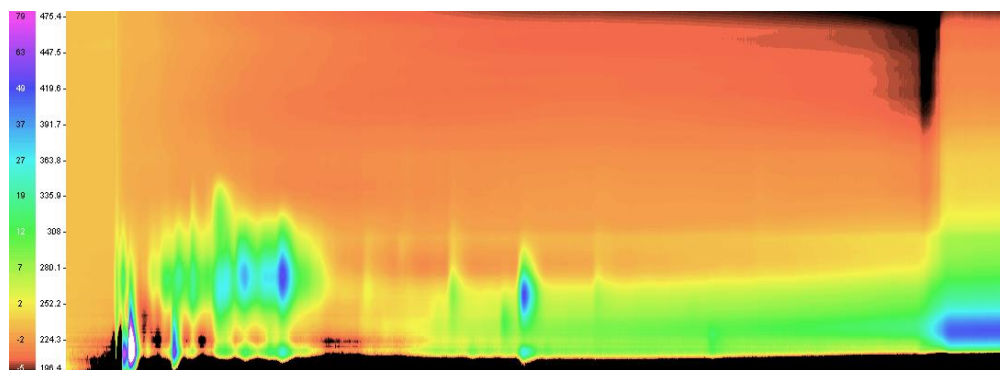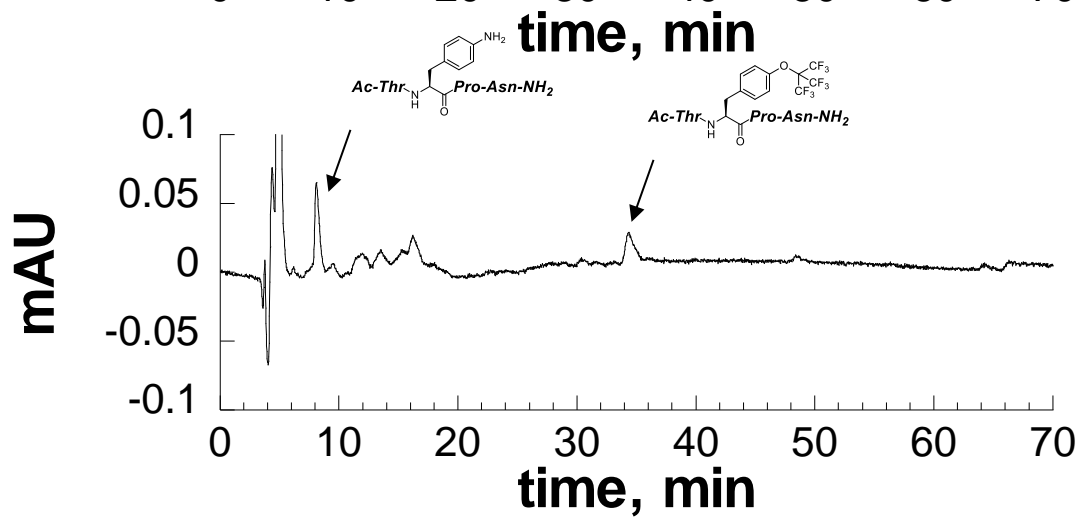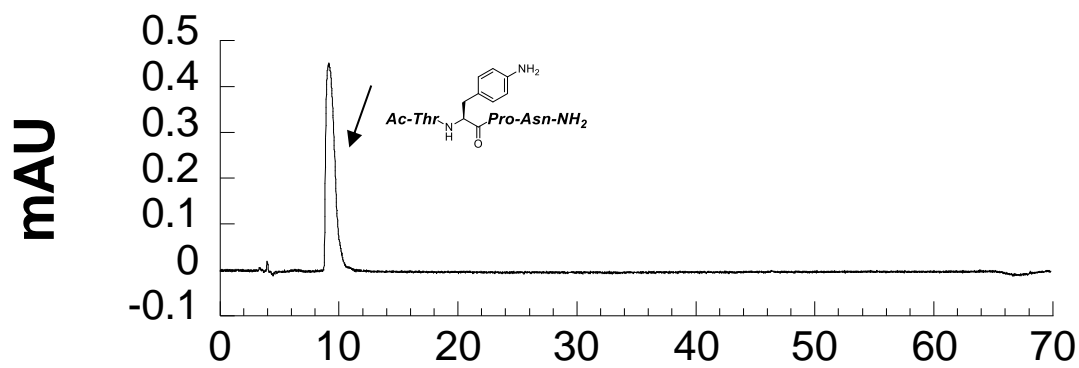
Figure 1.13   Purified Ac-T(4-NH$_2$-Phe)PN-NH$_2$ (top) was converted to Ac-T(Tyr(C$_4$F$_9$))PN-NH$_2$ via diazonium formation followed by heating at 45 ˚C with perfluoro-*tert*-butyl alcohol (middle, HPLC chromatogram, absorbance at 215 nm). Diode array analysis (bottom) of this chromatogram indicates that the other peaks visible in the HPLC chromatogram lack the UV signature of peptides (maximum absorbance below 220 nm) and thus are likely not the result of side reactions leading to other peptide products. As observed, Ac-T(Tyr(C$_4$F$_9$))PN-NH$_2$ has a λ$_{max}$ below 230 nm, associated with absorbance of the amide bonds. Thus, any compounds with a maximum absorbance greater than 230 nm are not associated with the peptide, and are likely due to small molecule impurities.

Other chemistries were explored also for the synthesis of this amino acid, but these attempts were unsuccessful. A number of conditions were examined for use with palladium and copper-catalyzed reactions to install the perfluoro-*tert*-butyl ether (Figure 1.14. Parameters including temperature, solvent, and ligands were explored to optimize reaction conditions all of which resulted in little (less than 5%) or no yield of product. Reactions were screened by mass spectrometry and ^{19}F NMR to determine the presence of product. The basis for exploring metal-catalyzed reaction chemistry was due to the precedent for cross coupling *tert*-butanol to substituted benzene rings, which are sterically similar to aromatic amino acids.[74-77] While palladium- and copper-catalyzed cross coupling can be incredibly robust systems, it is important to note that the pKa of *tert*-butanol is 17 while the pKa of perfluoro-*tert*-butanol is around 5.[78] This very large change in electronics is most likely the source of the lack of reactivity in the case of perfluoro-*tert*-butanol. The reaction chemistry is dependent on the basicity of *tert*-butoxide. Conversion of perfluoro-*tert*-butanol to the perfluoro-*tert*-butoxide sodium salt, as is often used in *tert*-butanol coupling, did not result in

substantial yield. This challenge was easily avoided by the use of diazonium coupling, which resulted in 61% yield under atmospheric conditions.
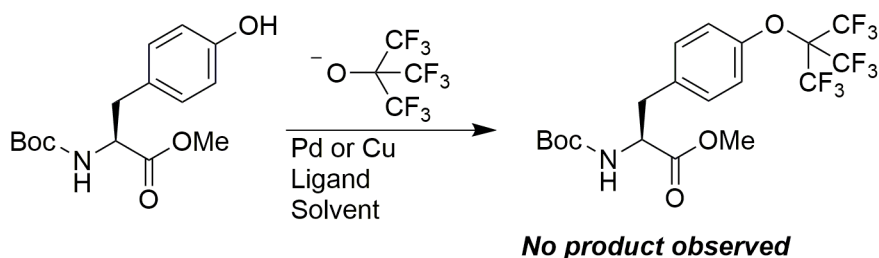
Figure 1.14 Attempts to generate perfluoro-*tert*-butyl tyrosine utilizing metal-catalyzed reaction chemistry were unsuccessful. Both palladium and copper were screened as potential catalysts, along with a number of ligands, solvents, and temperatures.

1.2.5 Incorporation of Amino Acids Containing Perfluoro-*tert*-butyl Groups into Peptides by Solid-Phase Peptide Synthesis

The addition of the perfluoro-*tert*-butyl ether adds significant steric bulk and hydrophobicity to the amino acids synthesized. Under typical amide coupling conditions (i.e. 8% DIPEA in DMF for 1 hr with HBTU) on Rink amide resin, no or minimal coupling was observed with the perfluoro-*tert*-butyl hydroxyproline derivatives while synthesizing peptides discussed in Chapter 3 (Abltide peptides). In short, the purified Fmoc-perfluoro-*tert*-butyl hydroxyproline derivatives were coupled using 2 equivalents of amino acid and HBTU. These coupling reactions led to very little coupling as observed by HPLC and mass spectrometry. A number of conditions were screened to optimize amide coupling reactions with these amino acids, including variations of coupling reagents, coupling times, and resin. Both HBTU and HATU were insufficient for coupling of the perflouro-*tert*-butyl hydroxyproline derivatives.

35

This low reactivity was believed to be, at least in part, due to the hydrophobic nature of these derivatives, as well as the steric demand of the derivatives. COMU has been demonstrated to be a superior coupling reagent, especially for difficult to couple amino acids such as Aib and Abu.[79] For this reason, it was used to couple both hydroxyproline derivatives containing perfluoro-*tert*-butyl ethers. Amide coupling reactions were carried out for 24 hr. While some coupling was observed after 12 hr, to maximize yield, reactions were allowed to proceed at room temperature for 24 hr. Despite better coupling reagents and increased reaction times, amide coupling reactions were still problematic. Therefore, to increase peptide yield, NovaGel PEG-polystyrene graft Rink amide resin (EMD Millipore) was used in place of Rink amide resin.[80] Solvation is key for difficult coupling reactions and allowing larger, more hydrophobic amino acids. The introduction of polyethylene glycol increases solvation leading to more access for sterically hindered amino acids. It should also be noted that smaller peptides (i.e. tetramers) could potentially be synthesized as efficiently on Rink amide resin; however, most of the peptides of interest were at least ten or more residues, including helical model peptide systems and peptides for applications discussed in Chapters 3. Furthermore, the amino acids containing the perfluoro-*tert*-butyl ether can also affect coupling of later amino acids producing incomplete sequences which can be difficult to purify by HPLC. To address this issue, which was commonly observed in longer peptide sequences, HATU was used as a coupling reagent after the perfluoro-*tert*-butyl amino acids were coupled and particularly at β branched amino acids and aromatics.

Amide coupling reactions with perfluoro-*tert*-butyl homoserine and perfluoro-*tert*-butyl tyrosine were not as rigorously screened as the perfluoro-*tert*-butyl

hydroxyproline derivatives. Both of these amino acids were incorporated into peptides described in Chapter 2. Utilizing the experience of the perfluoro-*tert*-butyl hydroxyproline derivatives, the PEG-polystyrene graft Rink amide resin was used without exploring the use of Rink amide resin. Longer coupling times (12 to 24 hours) were also used without examining shorter coupling times. Due to the less strained nature of the homoserine and tyrosine derivatives, however, HATU was examined, rather than COMU. Both perfluoro-*tert*-butyl homoserine and perfluoro-*tert*-butyl tyrosine were coupled into peptides described in Chapter 2 as well as model peptide systems discussed in later sections of this chapter utilizing HATU, as sufficient coupling was observed via HPLC.

1.2.6   Purification of Peptides Containing Perfluoro-*tert*-butyl ethers

Peptides synthesized on solid phase are typically cleaved from resin and deprotected, precipitated in cold ether, redissolved in aqueous solvent (e.g. buffer, methanol, etc.), filtered, and then purified by HPLC. In peptides containing perfluoro-*tert*-butyl ethers, however, ether precipitation was problematic. The purpose of ether precipitation is to remove the organic impurities generated by deprotection of the amino acids. Ether precipitation was first attempted with hydroxyproline derivatives in the Abltide peptides discussed in Chapter 3 and model peptide systems. Upon HPLC of these peptides, very little product was observed, even with the optimization of amide bond couplings previously discussed. For these peptides, the ether layer used to precipitate the peptide was dried and injected on the HPLC. A significant amount of peptide product was observed by HPLC and confirmed by mass spectrometry. In many cases, namely peptides discussed in Chapters 2 and 3, the peptides did not fully precipitate, leaving a significant amount of material in the ether layer (as observed by

HPLC) which is usually discarded. Precipitation was attempted in other solvents utilizing the MDM2-binding peptides discussed in Chapter 2, including hexanes and pentane; however, the problem of incomplete precipitation persisted. One possible explanation for this low extent of precipitation is the presence of residual TFA, which can interact with the perfluoro-*tert*-butyl group preventing ether precipitation. Another potential explanation is that the perfluoro-*tert*-butyl group is non-specifically associating with organic impurities in the mixture due to hydrophobic interactions. Unless otherwise noted, peptides were not precipitated with ether prior to injection on the HPLC.

1.2.7   Stability of Perfluoro-*tert*-butyl Ethers

There are very few compounds containing perfluoro-*tert*-butyl ethers in the literature and little is known about their stability. While ethers are fairly robust in their stability, the nine fluorines do introduce a significant electron withdrawing effect on the C-O bond, potentially acting as a leaving group to reform perfluoro-*tert*-butanol under acidic conditions. Decomposition of the perfluoro-*tert*-butyl ether was not observed under any conditions with either hydroxyproline derivative or with perfluoro-*tert*-butyl homoserine in any of the peptides synthesized throughout this body of work.

In contrast, perfluoro-*tert*-butyl tyrosine was stable under most conditions, with the exception of certain cleavage and deprotection conditions. Substituted aromatic amino acids have the potential to decompose under TFA cleavage and deprotection conditions. Upon subjection to cleavage and deprotection in 95% TFA with 2.5% thioanisole and 2.5% water for 90 min, the product Ac-T(**Tyr(C$_4$F$_9$**))PN-NH$_2$ was afforded in good yield (Figure 1.15). Cleavage and deprotection in 90% TFA

with 5% triisopropyl silane (TIS) and 5% water, however, led to elimination of the perfluoro-*tert*-butyl ether (Figure 1.15). This product decomposition is most likely due to hydride delivery via the TIS, allowing reaction with the perfluoro-*tert*-butyl alcohol as a leaving group. The peptide side-product Ac-TFPN-NH$_2$ was isolated from the cleavage and deprotection reactions utilizing TIS. The elimination product was purified and characterized by HPLC, mass spectrometry, and ^{1}H and TOCSY NMR spectroscopy, with the observed product identical to the previously synthesized peptide Ac-TFPN-NH$_2$ (Figure 1.16). These results indicate that TIS should be avoided during TFA cleavage and deprotection reactions with perfluoro-*tert*-butyl tyrosine.
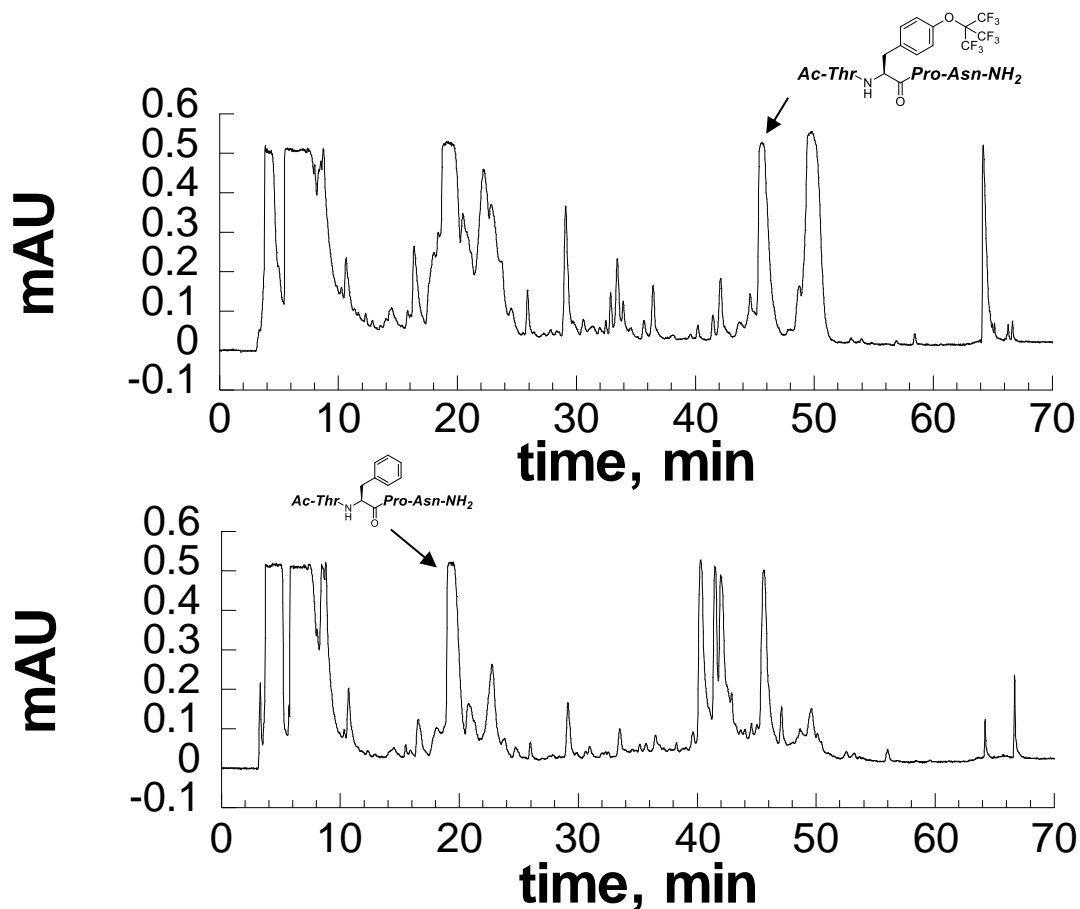
Figure 1.15 HPLC chromatograms of cleavage and deprotection reactions of Ac-
T(Tyr(C$_4$F$_9$))PN-NH$_2$ in 95% TFA with 2.5% thioanisole and 2.5% water
for 90 min (top) and in 90% TFA with 5% TIS and 5% water for 3 hours
(bottom). No ether precipitation was used after TFA cleavage and
deprotection, which results in additional products derived from the
cleavage/deprotection reagent cocktail being observed by HPLC.
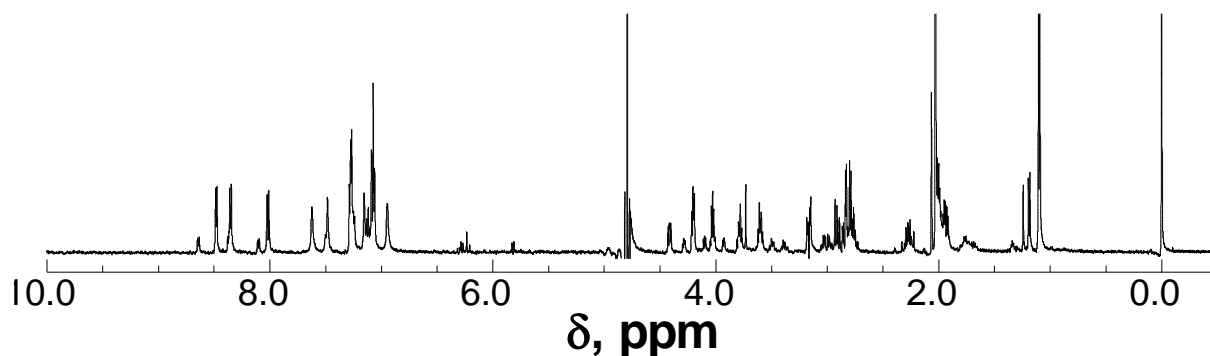
Figure 1.16 ^{1}H NMR spectrum of the peptide Ac-TFPN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl that was isolated by HPLC at t_R = 19.2 min from the cleavage and deprotection reaction of Ac-T(**Tyr(C$_4$F$_9$)**)PN-NH$_2$ with 90% TFA, 5% TIS, and 5% H$_2$O for three hours.

Due to the instability of the perfluoro-*tert*-butyl tyrosine under some cleavage and deprotection conditions, we verified that the peptide was stable under optimized cleavage and deprotection conditions as well as standard peptide synthesis conditions. Purified Ac-T(**Tyr(C$_4$F$_9$)**)PN-NH$_2$ was subjected to the optimized cleavage and deprotection conditions as well as the cleavage conditions that promote elimination of the perfluoro-*tert*-butyl group. After three hours in 90% TFA with 5% TIS and 5% water, Ac-TFPN-NH$_2$ was observed by HPLC (Figure 1.15). When subjected to optimized TFA cleavage/deprotection conditions (95% TFA with 2.5% thioanisole and 2.5% water), no elimination was observed after six hours (Figure 1.17). The peptide was also stable for 24 hours under standard amide coupling (8% DIPEA in DMF) (Figure 1.18) and standard Fmoc deprotection (20% piperidine in DMF) (Figure 1.19), with no identifiable side products observed by HPLC. Collectively, these results indicate perfluoro-*tert*-butyl tyrosine is stable under standard peptide

41

synthesis conditions, with the exception of certain cleavage and deprotection

conditions, specifically utilizing TIS as a scavenger.
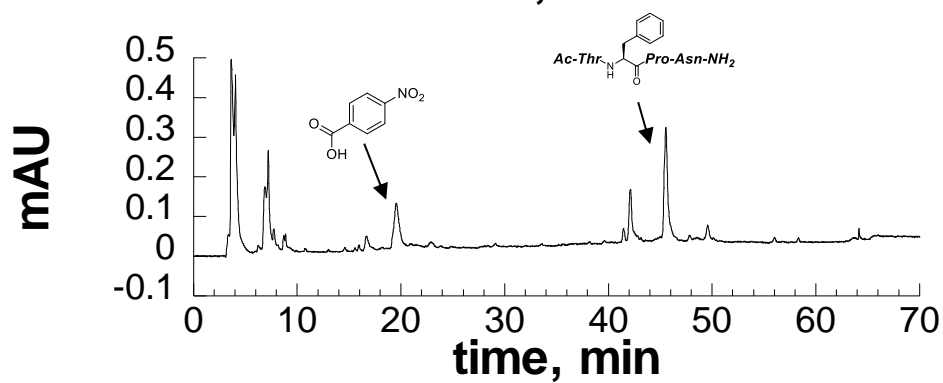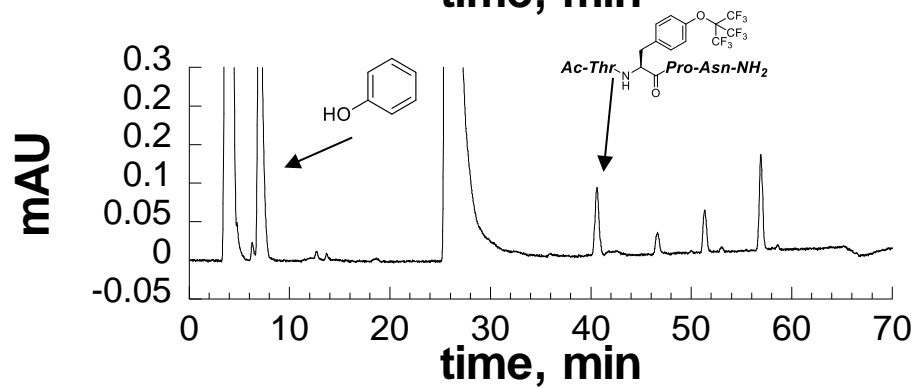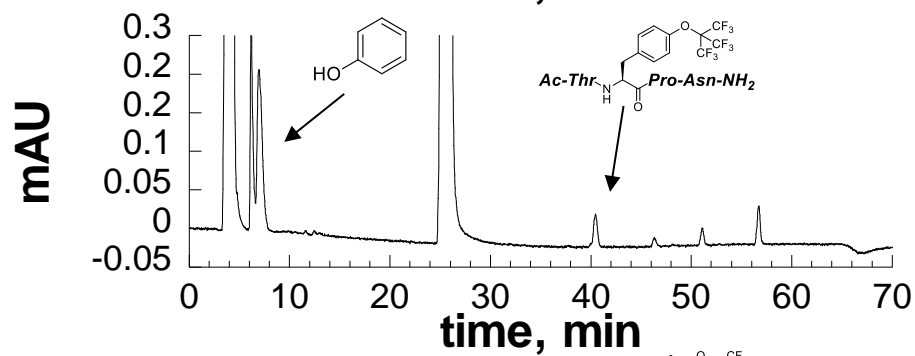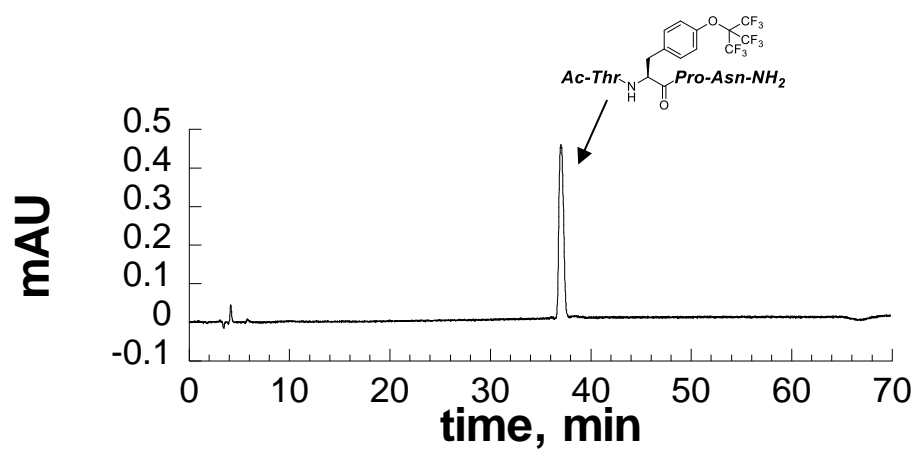
Figure 1.17 HPLC chromatograms of cleavage and deprotection reactions on the purified peptide Ac-T(Tyr(C_4F_9))PN-NH$_2$ (top) and peptide in 95% TFA with 2.5% thioanisole and 2.5% water at 0 hours (t_0) (second) and 6 hours (third) (HPLC conditions: 0-60% buffer B in buffer A over 60 min at 1 mL/min). Peptide and *p*-nitrobenzoic acid (internal standard) peaks were integrated and indicated that minimal side reaction had occurred over 6 hours. (bottom) Peptide in 90% TFA with 5% TIS and 5% water for 3 hours (HPLC conditions: 0-60% buffer B in buffer A over 60 min at 4 mL/min).
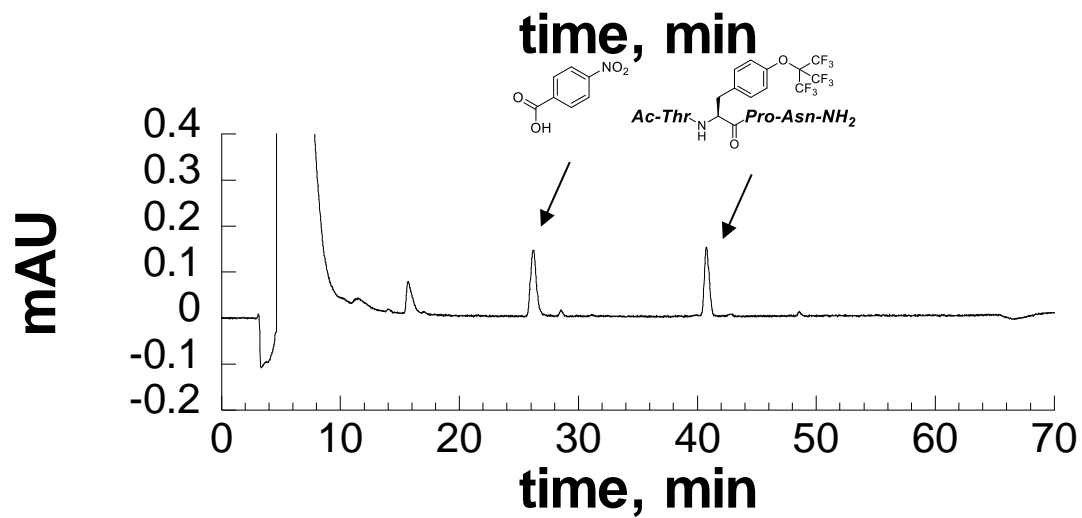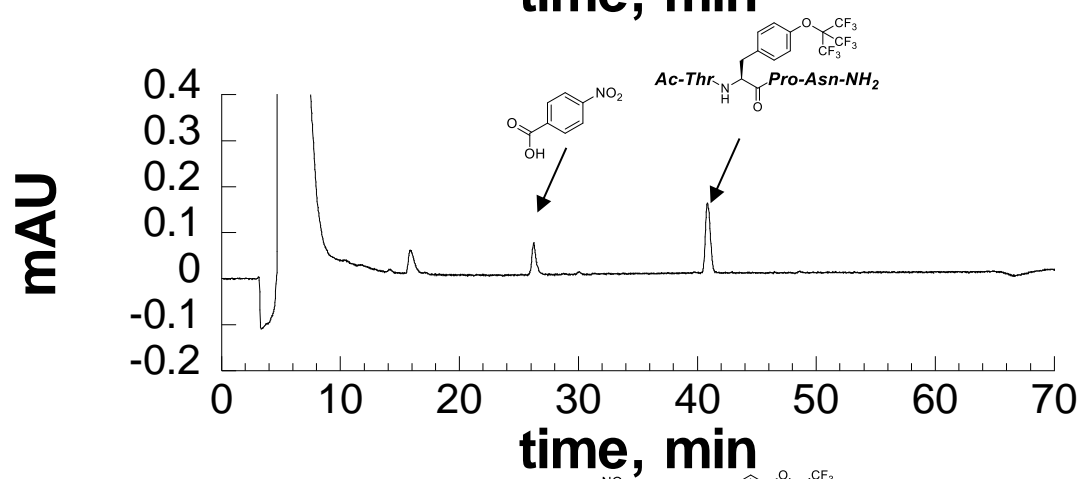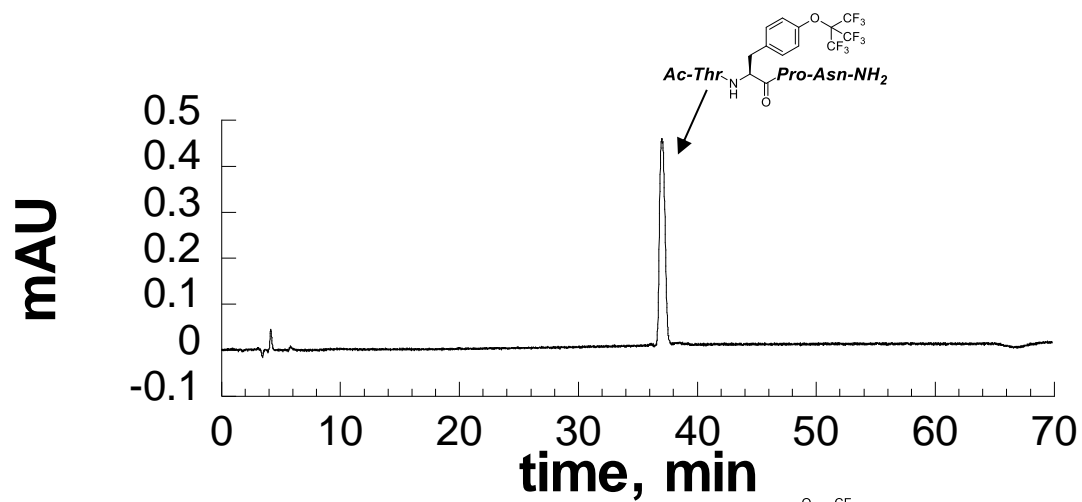
Figure 1.18 Ac-T($\textbf{Tyr(C}_4\textbf{F}_9\textbf{)}$)PN-NH$_2$ is stable in DMF with 8% DIPEA for a minimum of 24 hours at room tempearture. Purified Ac-T($\textbf{Tyr(C}_4\textbf{F}_9\textbf{)}$)PN-NH$_2$ (top) was subjected to a solution of 8% DIPEA in DMF with *p*-nitrobenzoic acid as a standard. HPLC chromatograms were recorded at time = 0 (center) and 24 hours (bottom) using a 60 minute linear gradient of 0 to 70% buffer B (20% H$_2$O, 80% acetonitrile, 0.05% TFA) in buffer A (98% H$_2$O, 2% acetonitrile, 0.06% TFA). Integration of the standard and the peptide peaks, as well as the absence of new peak formation, demonstrate that the peptide Ac-T($\textbf{Tyr(C}_4\textbf{F}_9\textbf{)}$)PN-NH$_2$ is stable under these conditions.
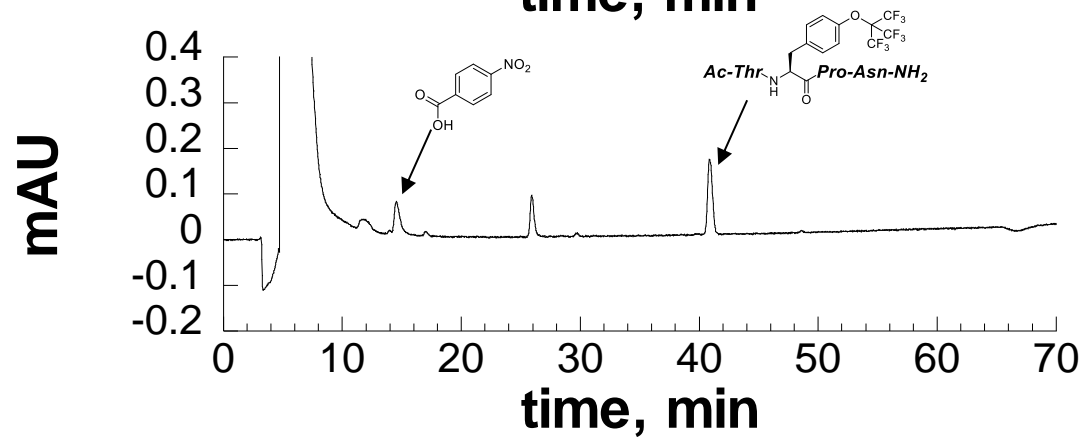
Figure 1.19 Ac-T(**Tyr(C₄F₉)**)PN-NH₂ is stable in DMF with 20% piperidine for a
minimum of 24 hours at room temperature. Purified Ac-
T(**Tyr(C₄F₉)**)PN-NH₂ (top) was subjected to a solution of 20% piperidine
in DMF with *p*-nitrobenzoic acid as a standard. HPLC chromatograms
were recorded at time = 0 (center) and 24 hours (bottom) using a 60
minute linear gradient of 0 to 70% buffer B (20% H_2O, 80% acetonitrile,
0.05% TFA) in buffer A (98% H_2O, 2% acetonitrile, 0.06% TFA).
Integration of the standard and the peptide peaks, as well as the absence
of new peak formation, demonstrate that the peptide Ac-
T(**Tyr(C₄F₉)**)PN-NH₂ is stable under these conditions.

1.2.8 Detection Limit of Peptides Containing a Perfluoro-*tert*-butyl Group via ^{19}F NMR

In order to demonstrate that peptides containing perfluoro-*tert*-butyl ethers
might be effective ^{19}F-based probes, we examined two synthesized peptides to
determine the detection limit of these systems. The designed amino acids exhibit a
significant advantage, allowing for detection at much lower concentration than other
fluorinated probes. Many other fluorinated probes require at least mid-micromolar
levels or long acquisition times for detection by ^{19}F NMR. All ^{19}F NMR data were
collected without the use of a proton decoupler. The 4*R*-perfluoro-*tert*-butyl
hydroxyproline in the peptide Ac-((**Hyp(C₄F₉)**))KAAAAKAAAAKAAGY-NH₂ was
examined by ^{19}F NMR. The peptide was diluted to 200 nM in 5 mM phosphate buffer
at pH 4 with 25 mM NaCl and 10% D_2O. The ^{19}F NMR experiment was performed
with a 7 ppm sweep width and 128 scans (5 minutes and 23 seconds) (Figure 1.20).
The ^{19}F NMR spectrum shows a single, narrow peak at -70.6 ppm, clearly
demonstrating that this peptide can be detected in approximately 5 minutes with good
signal-to-noise (S/NR = 7).

-70.3   -70.4   -70.5   -70.6   -70.7   -70.8   -70.9

δ, ppm
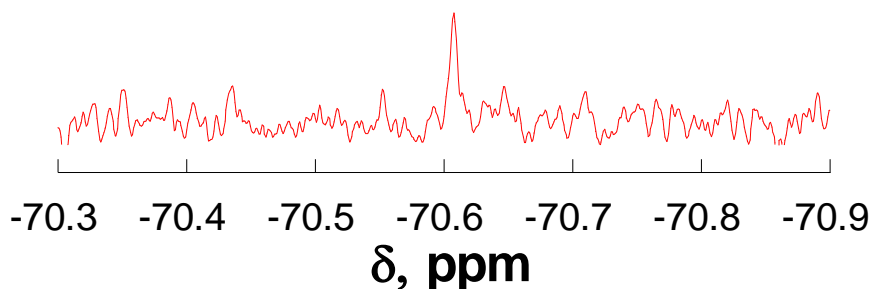
Figure 1.20   200 nM of Ac-(**(2S,4R)-Hyp(C₄F₉)**))KAAAAKAAAAKAAGY-NH₂ in water with 5 mM phosphate buffer (pH 4) with 25 mM NaCl and D₂O. Data were collected in 5 minutes and 23 seconds with 128 scans with 7 ppm sweep width. The data was collected with 7228 points (AQ = 0.91) and zero filled to 8192 points. Data were processed with an exponential multiplier and a line broadening of 2 Hz. Signal to noise ratio was calculated by the SNR peak calculator in MestReNova version 8.

The peptide containing perfluoro-*tert*-butyl tyrosine was also examined under the same conditions as described above. The peptide Ac-T(**Tyr(C₄F₉)**)PN-NH₂ was diluted to either 200 or 500 nM under the same conditions and subjected to the same ^{19}F NMR experiment described above. Again, a sharp singlet was visible by ^{19}F NMR in 5 minutes with good signal-to-noise (S/NR = 7) (Figure 1.21). To further expand on the detection limit, this peptide was also examined by ^{19}F NMR spectrum at 500 nM in 5 mM phosphate buffer at pH 4 with 25 mM NaCl and 10% D₂O, in order to determine the minimum amount of time required for detection in the nanomolar regime (Figure 1.22). Three different NMR experiments were performed using 4 dummy scans, 0.8 second acquisition time, and 3.0 second relaxation delay with decreasing number of scans (time). The first experiment was completed with 32 scans (2 minutes 18 seconds) which produced a large singlet with signal-to-noise = 19.6.

The second experiment was completed with 16 scans (1 minute 16 seconds) which produced a large singlet with signal-to-noise = 14.3. The final experiment was completed with 8 scans (46 seconds) and produced a large singlet with a signal-to-noise = 8.9. One final experiment was used to determine the minimum amount of time that could be used to detect the Ac-T(Tyr(C_4F_9))PN-NH$_2$ peptide at 500 nM. The ^{19}F NMR spectrum was completed with 8 scans, with, only 2 dummy scans used, and a 2.0 second relaxation delay, in order to reduce experiment time down to 28 seconds (Figure 1.23). A singlet was visible by ^{19}F NMR with signal-to-noise = 10.1 These data demonstrate that the perfluoro-*tert*-butyl ether is easily detectable at 500 nM in 30 seconds by ^{19}F NMR.
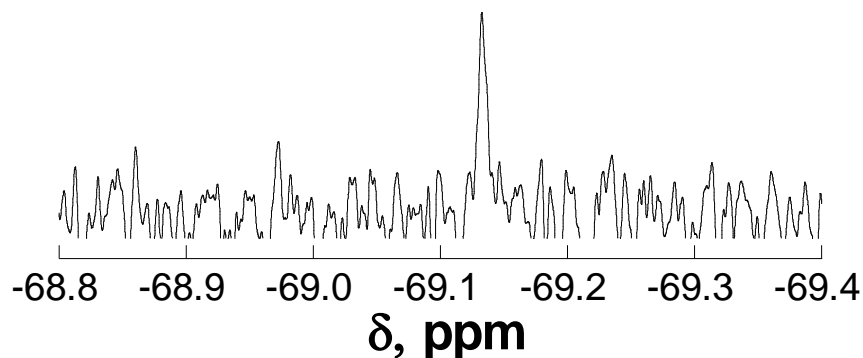
Figure 1.21 ^{19}F NMR spectrum of the peptide Ac-T(**Tyr(C₄F₉)**)PN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl at 200 nM peptide concentration. Data were collected in 5 minutes and 23 seconds with 128 scans, S/N = 7.4. The data were collected with 0.91 seconds acquisition time and a 3.0 second relaxation delay. Data were processed with exponential multiplication and a line broadening of 2 Hz. Signal-to-noise ratio was calculated by the SNR peak calculator in MestReNova version 10.
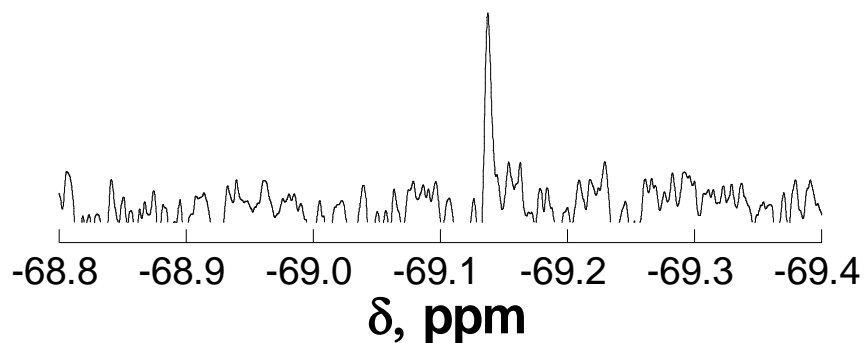
Figure 1.22 ^{19}F NMR spectrum of the peptide Ac-T(**Tyr(C₄F₉)**)PN-NH₂ in 90% H₂O/10% D₂O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl at 500 nM peptide concentration. Data were collected as above using 8 scans (28 sec), S/N = 10.1. Experiments were conducted with 2 dummy scans, 0.8 sec acquisition time, and a 2.0 sec relaxation delay. Signal-to-noise ratio was calculated by the SNR peak calculator in MestReNova version 10.
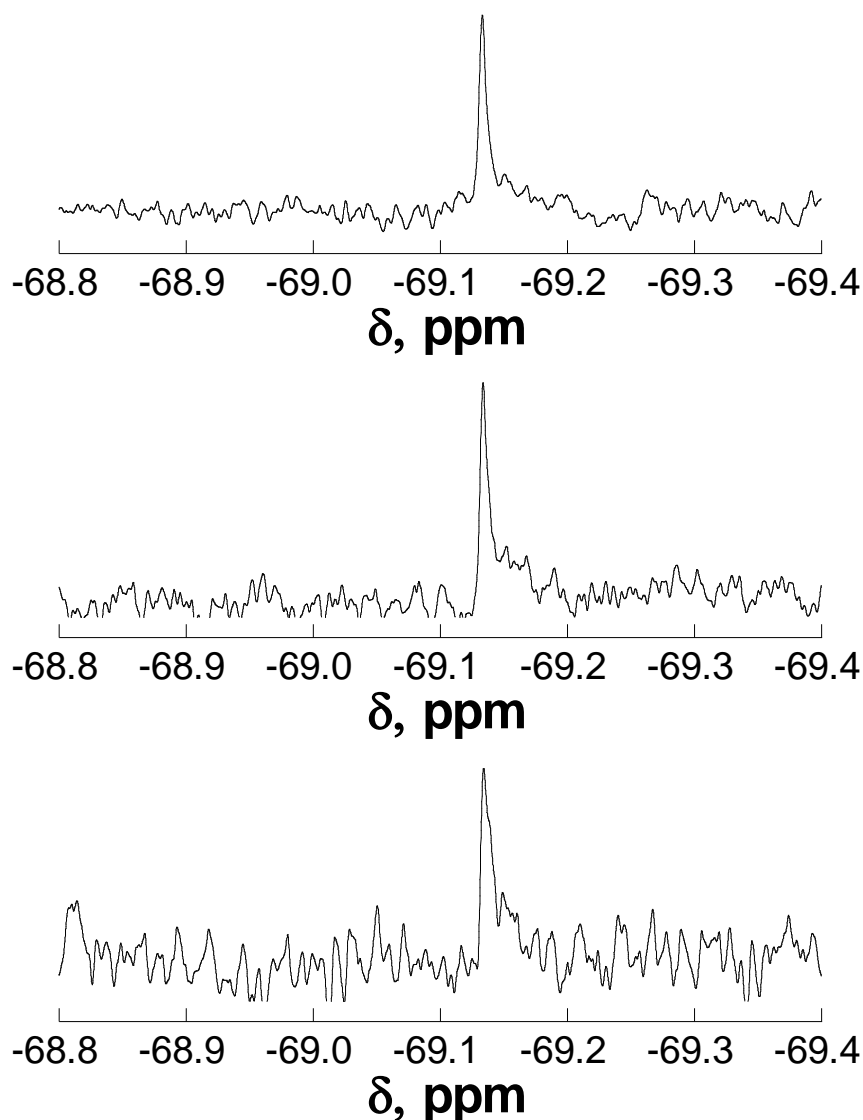
Figure 1.23 ^{19}F NMR spectra of the peptide Ac-T(**Tyr(C₄F₉)**)PN-NH₂ in 90% H₂O/10% D₂O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl at 500 nM peptide concentration. Data were collected as above using 32 scans (2 min, 18 sec), signal-to-noise ratio (S/N) = 19.6 (top); 16 scans (1 min, 16 sec), S/N = 14.3 (middle); or 8 scans (46 sec), S/N = 8.9 (bottom). Experiments were conducted with 4 dummy scans, 0.8 sec acquisition time, and a 3.0 sec relaxation delay. Signal-to-noise ratio was calculated by the SNR peak calculator in MestReNova version 10.

One notable characteristic of the ^{19}F NMR spectra is the presence of only one peak (Figures 1.21 to 1.23). In the case of both perfluoro-*tert*-butyl hydroxyproline derivatives in the Ac-TYXN-NH$_2$ peptide system, two peaks were visible corresponding to the *cis* and *trans* conformations. In the Ac-TXPN-NH$_2$ system where X = Tyr(C$_4$F$_9$), only one peak is visible. This single peak is most likely a result of the *cis* and *trans* peaks being magnetically equivalent. The chemical shift dispersion between the *cis* and *trans* peaks is less than the peak width leading to a single peak by ^{19}F NMR. While being able to differentiate between *cis* and *trans* populations can be useful, the two peaks being magnetically equivalent will simply the spectrum when the perfluoro-*tert*-butyl tyrosine is applied to various biological systems.

These ^{19}F NMR experiments demonstrate the power of the perfluoro-*tert*-butyl ether to be used as a ^{19}F NMR or MRI probe. While all the experiments presented here are NMR-based, these probes can be used directly with ^{19}F MRI experiments. One common application for ^{19}F MRI is as a tracer in which higher concentrations of fluorinated probes collect in targeted areas. One example of this is the use of fluorinated curcumin derivatives which associate with ß-amyloid structures in Alzheimer's Disease mouse models (Figure 1.24). Utilizing trifluoromethyl groups, these structures could be detected at 100 mg/kg when injected into the mouse tail vein.[81] Using a perfluoro-*tert*-butyl group would allow detection at less than a third of that concentration as the perfluoro-*tert*-butyl ether contains nine equivalent fluorines rather than three. This improvement to the probe would potentially significantly reduce the cost required for ^{19}F MRI due to a significant decrease in the amount of material required and the amount of instrument time required to collect the data with the same signal-to-noise. This approach also would greatly increase the sensitivity of

[19]F MRI. In the previous example using ^{19}F labeled curcumin derivatives, experiments were performed at higher concentrations (200 mg/kg) resulting in much sharper images showing a greater area of ß-amyloid formation than lower concentrations. The perfluoro-*tert*-butyl ether would be detectable faster at lower concentration, resulting in potentially superior images.

Figure 1.24   Previously synthesized curcumin derivatives utilizing trifluoromethyl groups could potentially benefit from improved detection utilizing perfluoro-*tert*-butyl ether groups.

1.2.9   Determination of the Conformational Preferences of Amino Acids Containing Perfluoro-*tert*-Butyl Ethers within an α-Helical Model Peptide Context

The α-helix constitutes 40% of folded structures within the protein data bank. These α-helices are known to be involved in recognition at protein-protein interfaces, tertiary structure of proteins, and in trans-membrane segments of membrane proteins.[28,48] One of the major drawbacks to incorporating fluorine into amino acids is that the α-helical propensity is greatly decreased compared to non-fluorinated analogs.[16,27] The reduction in α-helical propensity is problematic, especially for protein engineering, as it can prevent proper folding of proteins or preventing proper binding to a protein target. In order to consider applications involving compact secondary structures like α-helices, we need to quantify α-helical propensity for the perfluoro-*tert*-butyl hydroxyproline and homoserine derivatives. Aromatic amino

acids such as Phe and Tyr, generally have similar propensities for α-helical structures.
[51] The perfluoro-*tert*-butyl tyrosine was not examined because the perfluoro-*tert*-butyl
ether is sufficiently far from the peptide backbone that it most likely would not impact
α-helical structure.

Aliphatic amino acids are known to promote α-helical structure. Leu, Ile, Lys,
and Ala are commonly found in α-helical segments in proteins. While proline usually
disrupts α-helices when found in the middle of the helix, proline can also signal the
start of a helix.[51,52] There is a substantial body of work looking at model systems
utilizing alanine-rich peptides. All of the canonical amino acids have been examined at
various points in the helix, as have a large number of unnatural amino acids. Amino
acids which favor α-helices in decreasing order as follows: Ala, Leu, Arg, Met, Lys,
Glu, Ile, Trp, Ser, Tyr, Phe, Val, His Asn, Thr, Cys, Asp, Gly.[82] Using this scale, we
can quantify the propensity of our designed amino acids to determine if α-helical
based applications are possible using amino acids containing a perfluoro-*tert*-butyl
ether. 4*S*-Perfluoro-*tert*-butyl hydroxyproline, 4*R*-perfluoro-*tert*-butyl hydroxyproline,
and perfluoro-*tert*-butyl homoserine were all incorporated into the Baldwin model
peptide Ac-XKAAAAKAAAAKAAGY-NH$_2$ and examined by CD (Figure 1.25).
This peptide model system has previously been used by the Zondlo lab to explore the
effect of phosphorylation and OGlcNAcylation on α-helical structure. We found that a
full range of α-helical propensities was possible with the three perfluoro-*tert*-butyl
containing amino acids examined. The perfluoro-*tert*-butyl homoserine (36% helical),
while not as helix promoting as some of the more helix-promoting residues such as
Ala, was very similar to proline (38% helical) at the N-terminus (Table 1.2).[52]
Furthermore, in very similar peptides, perfluoro-*tert*-butyl homoserine has a similar α-

helical propensity to Met, demonstrating the potential of this particular amino acid being used as a Met mimic. The perfluoro-*tert*-butyl homoserine, based on these data, may potentially be used as a fluorinated amino acid that can stabilize, rather than destabilize, α-helical structure. The perfluoro-*tert*-butyl hydroxyproline derivatives were less α-helical than proline, 24% for the 4*R*- and 19% for the 4*S*-perfluoro-*tert*-butyl hydroxyproline, but interestingly these amino acids were not the same (Table 1.2). These data, while interesting, are unsurprising. The preferences of 4*R*- and 4*S*-substituted proline residues have been established. As stated previously in Figure 1.5, 4*R*-substitued proline derivatives favor the *exo* ring pucker and more compact structures while 4*S*-subsitiued proline derivatives favor the *endo* ring pucker and more extended structures. Based on this analysis, we expect, and observe, that 4*S*-perfluoro-*tert*-butyl hydroxyproline is less α-helical promoting. In contrast, the 4*R*-perfluoro-*tert*-butyl hydroxyproline is more α-helix promiting relative to the 4*S*. Both hydroxyproline derivatives, however, are less α-helix promoting than Pro. In fact, these amino acids are less α-helical than any of the canonical amino acids examined within this context and similar peptides. The reduction in α-helical propensity is similar to previous data indicating fluorine destabilizes α-helices. There is also potential for the side chain, which in the case of the perfluoro-*tert*-butyl hydroxyproline derivatives but not the perfluoro-*tert*-butyl homoserine, to project forward or back, interacting with neighboring residues. While these data demonstrate that perfluoro-*tert*-butyl homoserine can be used in an α-helix, the hydroxyproline data shows these two amino acids to be unique from each other despite their structural similarities as per previous observations in Pandey *et al.* previous observations.[43]
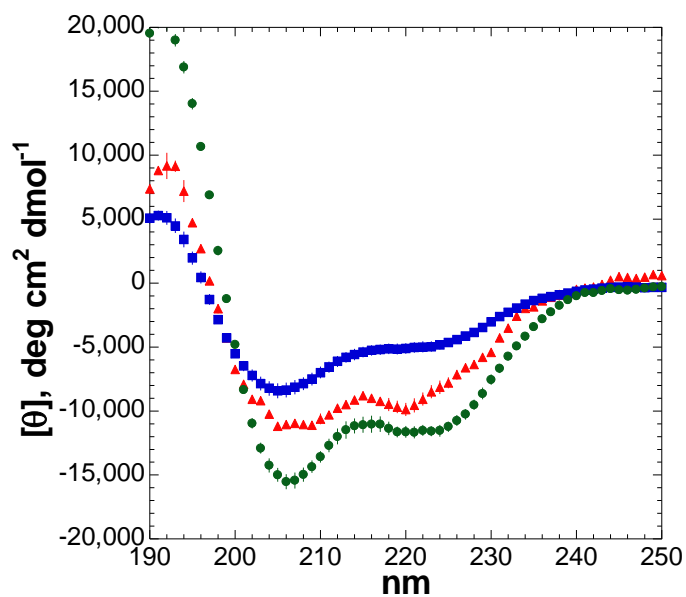
Figure 1.25 CD spectra of peptides with (2*S*,4*R*)-Hyp(C$_4$F$_9$) (red circles), (2*S*,2*S*)-hyp(C$_4$F$_9$) (blue squares), Hse(C$_4$F$_9$) (green diamonds) as the guest (X) residue Ac-XKAAAAKAAAAKAAGY-NH$_2$ model α-helical peptides. Population of α-helix is indicated by the magnitude of the negative band at 222 nm or the ratio [θ]$_{222}$/[θ]$_{208}$.

X =	$[\theta]_{222}$ deg cm^2 dmol^{-1}	$[\theta]_{208}$ deg cm^2 dmol^{-1}	$[\theta]_{190}$ deg cm^2 dmol^{-1}	$[\theta]_{222}/[\theta]_{208}$	$-[\theta]_{190}/[\theta]_{222}$	% α-helix
Ser[a]	-16621	-16434	32243	1.01	1.96	50%
Thr[a]	-14863	-15423	27161	0.96	1.76	45%
Pro[a]	-12030	-13891	16191	0.87	1.17	38%
Hse(C$_4$F$_9$)	-11513	-14962	19519	0.77	1.70	36%
Hyp(C$_4$F$_9$)[b]	-6840	-11042	7380	0.67	1.08	24%
hyp(C$_4$F$_9$)[b]	-4990	-7846	5088	0.64	1.01	19%

Table 1.2   Circular dichroism data for peptides with the indicated guest residues **X** in the α-helical model peptide host system Ac-**X**KAAAAKAAAAKAAGY-NH$_2$. Extent of α-helix is identified by mean residue ellipticity at 222 nm ($[\theta]_{222}$) or via the ratios $[\theta]_{222}/[\theta]_{208}$ or $-[\theta]_{190}/[\theta]_{222}$. % α-Helix was determined via the method of Baldwin, as described previously. Ser and Thr are highly stabilizing to α-helices at their N-terminus due to the hydrogen bonding with the unsatisfied amide N-H hydrogen bond donors.

1.2.10 Determination of Conformational Preferences within a Polyproline II Helical Context

Another common compact secondary structure observed in proteins, especially at protein-protein interfaces, is the polyproline II (PPII) helix. PPII helices are defined by all *trans* amide bonds, with 3 residues per turn in an extended left-handed helix.[83] These structures are more common in proline-rich sequences such as in the tau protein and in SH3 binding domains.[48,84,85] We chose to examine all three of the aliphatic amino acids within this system. Previous data from Brown and Zondlo demonstrate that some aliphatic amino acids aside from proline, most notably Leu, readily form a stable PPII helix as observed in the Ac-GPPXPPGY-NH$_2$ peptide model system.[56] In contrast, β-branched aliphatic amino acids, such Ile, Val, and *tert*-Leu, strongly disfavor PPII helices, most likely due steric clash between the side chain and the side chain of the following residue. To date, there are no data on the propensities of highly fluorinated amino acids in PPII helices. Based on data in sterically demanding amino acids, however, highly fluorinated amino acids would be expected to destabilize PPII structures. We examined both hydroxyproline derivatives as well as perfluoro-*tert*-butyl homoserine in the Ac-GPPXPPGY-NH$_2$ peptide model system to determine the PPII helix propensities of these amino acids. 4*R*-Perfluoro-*tert*-butyl hydroxyproline has a higher propensity for PPII helices than proline, while 4*S*-perfluoro-*tert*-butyl hydroxyproline has a significantly lower propensity than proline (Figure 1.26). These

59

data are again consistent with the data observed by Pandey *et al*. which demonstrated that 4*R*-perfluoro-*tert*-butyl hydroxyproline favors compact structures and *trans* amide bonds while 4*S*-perfluoro-*tert*-butyl hydroxyproline favors more extended structures and *cis* amide bonds. These data are consistent with the PPII helix-promoting trends of 4*R*- and 4*S*-fluoroproline, where 4*R*-fluoroproline is more PPII helix-promoting than Pro, and 4*S*-fluoroproline is less PPII helix-promoting than Pro. The perfluoro-*tert*-butyl homoserine is significantly less PPII favoring than Leu; however, it is in the same range as other similar aliphatic amino acids, including Met and unmodified homoserine (Table 1.3). These data indicate the perfluoro-*tert*-butyl ether modification, in the case of homoserine, is having little impact on the ability of the amino acid to adopt a PPII helix structure. Despite the sterically demanding nature of the perfluoro-*tert*-butyl ether modification, this amino acid is far more PPII helix promoting than Val, Thr, and Ile. These data again support that perfluoro-*tert*-butyl homoserine could potentially be used as a highly fluorinated amino acid that stabilizes compact secondary structures and could potentially act as a structural Met mimic within applied systems. All of these data taken together demonstrate that a large range of helical propensities are accessible via the designed perfluoro-*tert*-butyl amino acids.
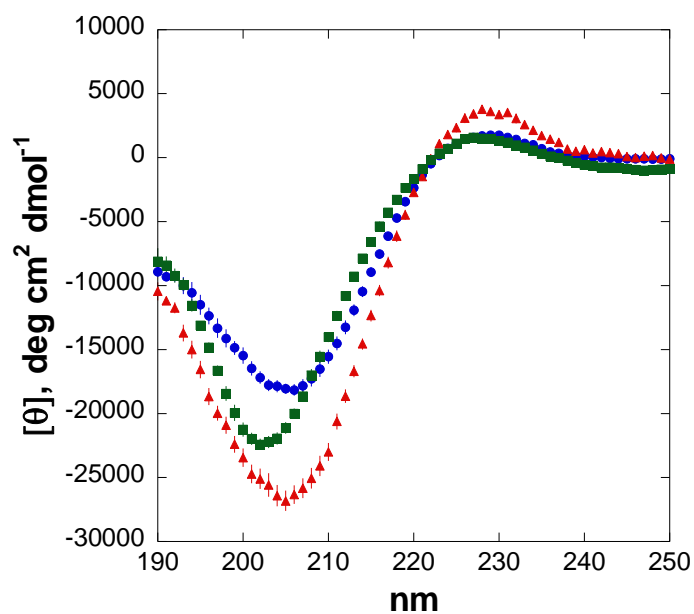
Figure 1.26 CD spectra of peptides with (2*S*,4*R*)-Hyp(C$_4$F$_9$) (red circles), (2*S*,4*S*)-hyp(C$_4$F$_9$) (blue diamonds), Hse(C$_4$F$_9$) (green squares) as the guest (X) residue in Ac-GPPXPPGY-NH$_2$ peptides. Polyproline helix (PPII) is indicated by the magnitude of the positive band at ~228 nm. The peptide with X=Pro exhibits an intermediate structure between the perfluoro-*tert*-butyl hydroxyprolines.

X =	$[\theta]_{228}$ deg cm^2 dmol^{-1}
Hyp(C$_4$F$_9$)	3370[a]
Pro	2950[b]
Leu	2230[b]
hyp(C$_4$F$_9$)	1700[a]
Hse(C$_4$F$_9$)	1496
Hse	1390[b]
Met	1380[b]
Ser	460[b]
Val	-120[b]
tert-Leu	-940[b]

Table 1.3   Circular dichroism data for peptides with the indicated guest residues **X** in the polyproline helix model peptide host system[56] Ac-GPP**X**PPGY-NH$_2$. Extent of polyproline helix is identified by mean residue ellipticity at 228 nm ($[\theta]_{228}$), with more positive numbers indicating greater extent of polyproline helix.

1.2.11 Aromatic Electronic Effects of a Perfluoro-*tert*-butyl Ether

The perfluoro-*tert*-butyl ether has the potential to impact *cis-trans* isomerization, which is critical in protein folding, due to an electronic change of an aromatic system, as discussed in Section 1.1.6. In the case of the perfluoro-*tert*-butyl

tyrosine, we set out to quantify the effect of the perfluoro-*tert*-butyl group on the aromatic electronics. Fluorinated amino acids are known to favor beta structure within a protein context, often disrupting other types of secondary structures. Fluorine is a highly electronegative atom which can affect the conformational preferences within an aromatic-proline sequence, which can be observed in an increase in the $K_{trans/cis}$ ratio. Perfluoro-*tert*-butyl tyrosine was introduced into the Ac-TXPN-NH$_2$ peptide model system. All canonical aromatic amino acids as well as a large set of unnatural aromatic amino acids have previously been examined within this context.[86] ^{1}H NMR spectroscopy of the amide region was used to quantify the $K_{trans/cis}$ of 4.1 (ΔG = -0.83 kcal mol^{-1}) for perfluoro-*tert*-butyl tyrosine, compared to Ac-TYPN-NH$_2$ which has a $K_{trans/cis}$ of 2.7 (ΔG = -0.59 kcal mol^{-1}) (Figure 1.27). Compared to the other fluorinated examples, such as 4-CF$_3$-Phe ($K_{trans/cis}$ of 4.7; ΔG = -0.92 kcal mol^{-1}), the perfluoro-*tert*-butyl ether is intermediate electronically between tyrosine and the fluorinated analogs.

Comparison of the ^{1}H and TOCSY spectra of the peptides peptides Ac-TYPN-NH$_2$ and Ac-T(**Tyr(C$_4$F$_9$)**)PN-NH$_2$ also indicates minimal changes as a result of the addition of the perfluoro-*tert*-butyl group (Figure 1.28-1.35). The protons on the aromatic ring are modestly shifted downfield, as a result of the perfluoro-*tert*-butyl modification. There is also minimal change in chemical shift in the amide region, which is consistent with a modestly electron-withdrawing substitution on the aromatic ring (Table 1.5 and 1.6). Full analysis of the TOCSY spectra also indicates similar chemical shifts for all non-aromatic amino acids.
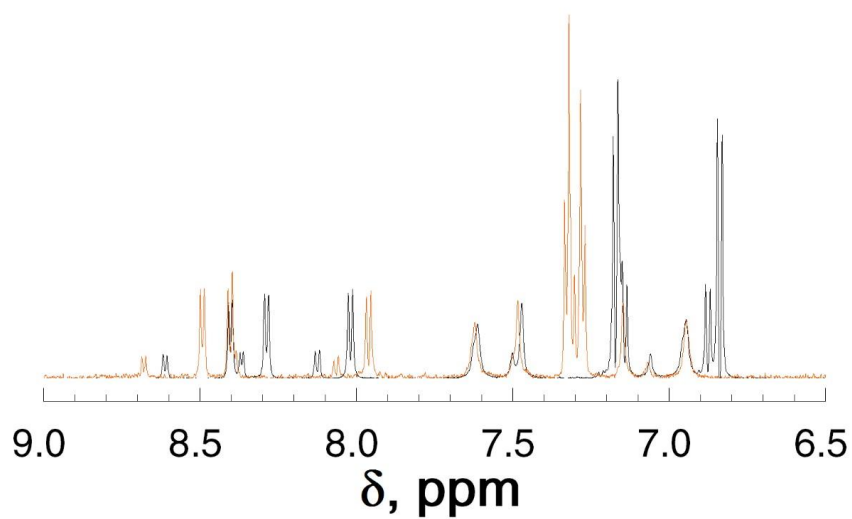
Figure 1.27 ^{1}H NMR spectra (amide-aromatic region, 90% H_2O/10% D_2O, 25 mM NaCl, 5 mM phosphate pH 4) of peptides with X = Tyr (blue) and X = Tyr(C_4F_9) (red). Chemical shift comparisons can be found in Tables 1.5 and 1.6

	δ, Tyr(C₄F₉), *trans*	δ,Tyr(C₄F₉), *cis*	δ, Tyr, *trans*	δ, Tyr, *cis*
H$_N$	8.40	8.39	8.31	8.39
H$_\alpha$	4.92	4.62	4.90	4.90
H$_{\beta 1}$	2.93	3.03	2.86	2.91
H$_{\beta 2}$	3.20	3.20	3.12	3.03
Ring	7.33, 7.29	7.33, 7.28	7.17, 6.89	7.14, 6.88
Pro H$_\alpha$	4.40	4.00	4.42	1.83

Table 1.4   Chemical shift comparisons for Tyr(C$_4$F$_9$) and Tyr protons within the peptides Ac-T(**Tyr(C₄F₉)**)PN-NH$_2$ and Ac-TYPN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.

	Tyr(C$_4$F$_9$) *trans*, ppm	$^3J_{\alpha N}$, Hz	Tyr *trans*, ppm	$^3J_{\alpha N}$, Hz	Tyr(C$_4$F$_9$) *cis*, ppm	$^3J_{\alpha N}$, Hz	Tyr *cis*, ppm	$^3J_{\alpha N}$, Hz
Thr	7.96	8.0	8.20	8.0	8.07	8.0	8.13	8.1
Tyr	8.40	8.4	8.29	8.3	8.39	n.d.	8.37	8.4
Asn	8.49	7.3	8.40	8.4	8.68	7.2	8.61	8.6

Table 1.5   Chemical shift and coupling constant comparisons for amide protons of the peptides Ac-T(**Tyr(C$_4$F$_9$)**)PN-NH$_2$ and Ac-TYPN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl. The Tyr(C$_4$F$_9$) *cis* $^3J_{\alpha N}$ coupling constant was not determined due to spectral overlap with Tyr(C$_4$F$_9$) *trans* peak.
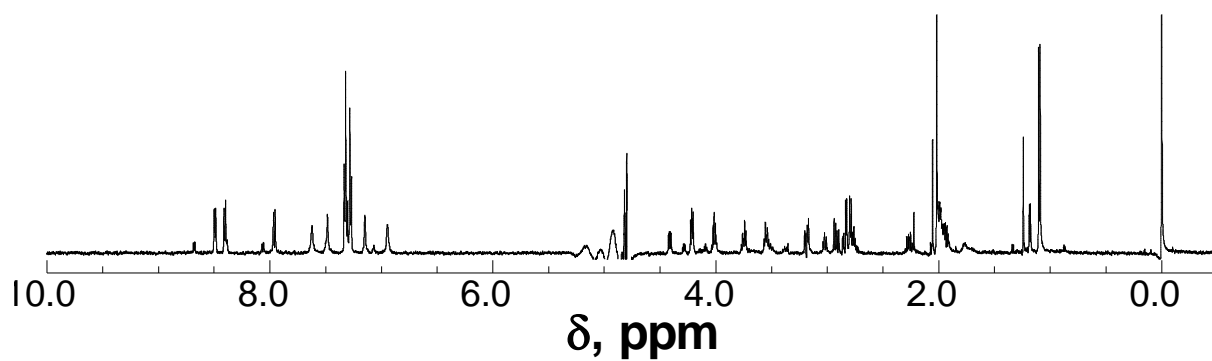
Figure 1.28 ^{1}H NMR spectrum of the peptide Ac-T(**Tyr(C4F9)**)PN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
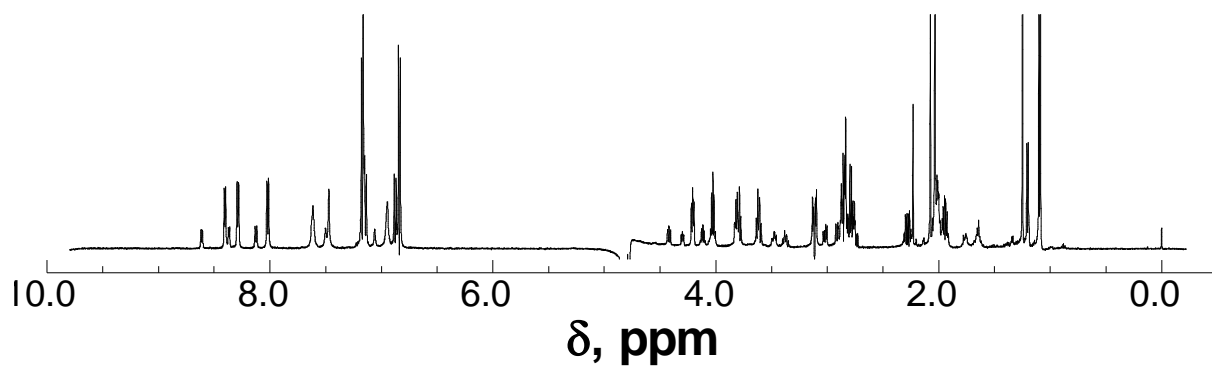
Figure 1.29 ^{1}H NMR spectrum of the peptide Ac-TYPN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.

67
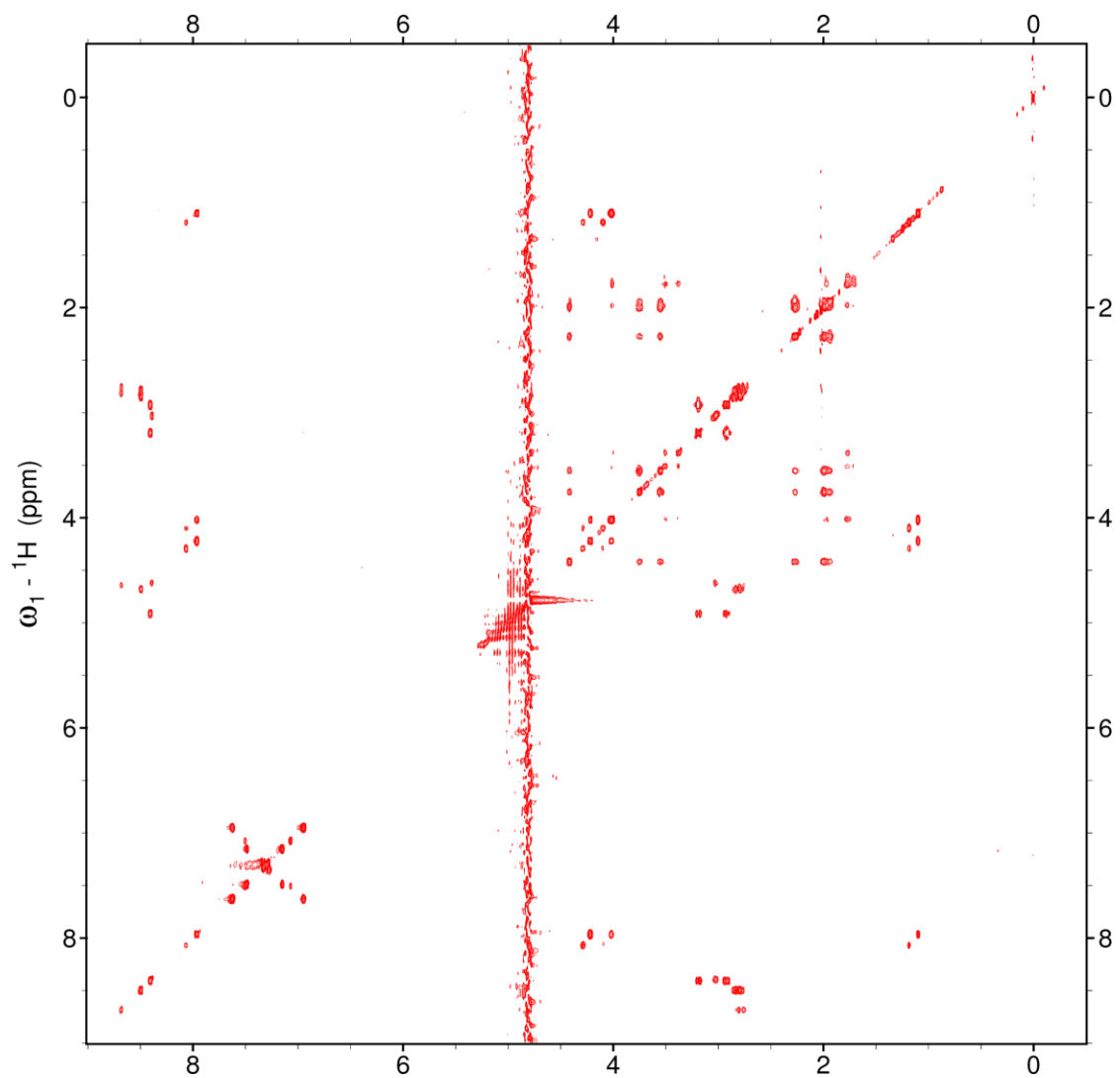
Figure 1.30   TOCSY spectrum of the peptide Ac-T(**Tyr(C₄F₉)**)PN-NH$_2$ in 90%
H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
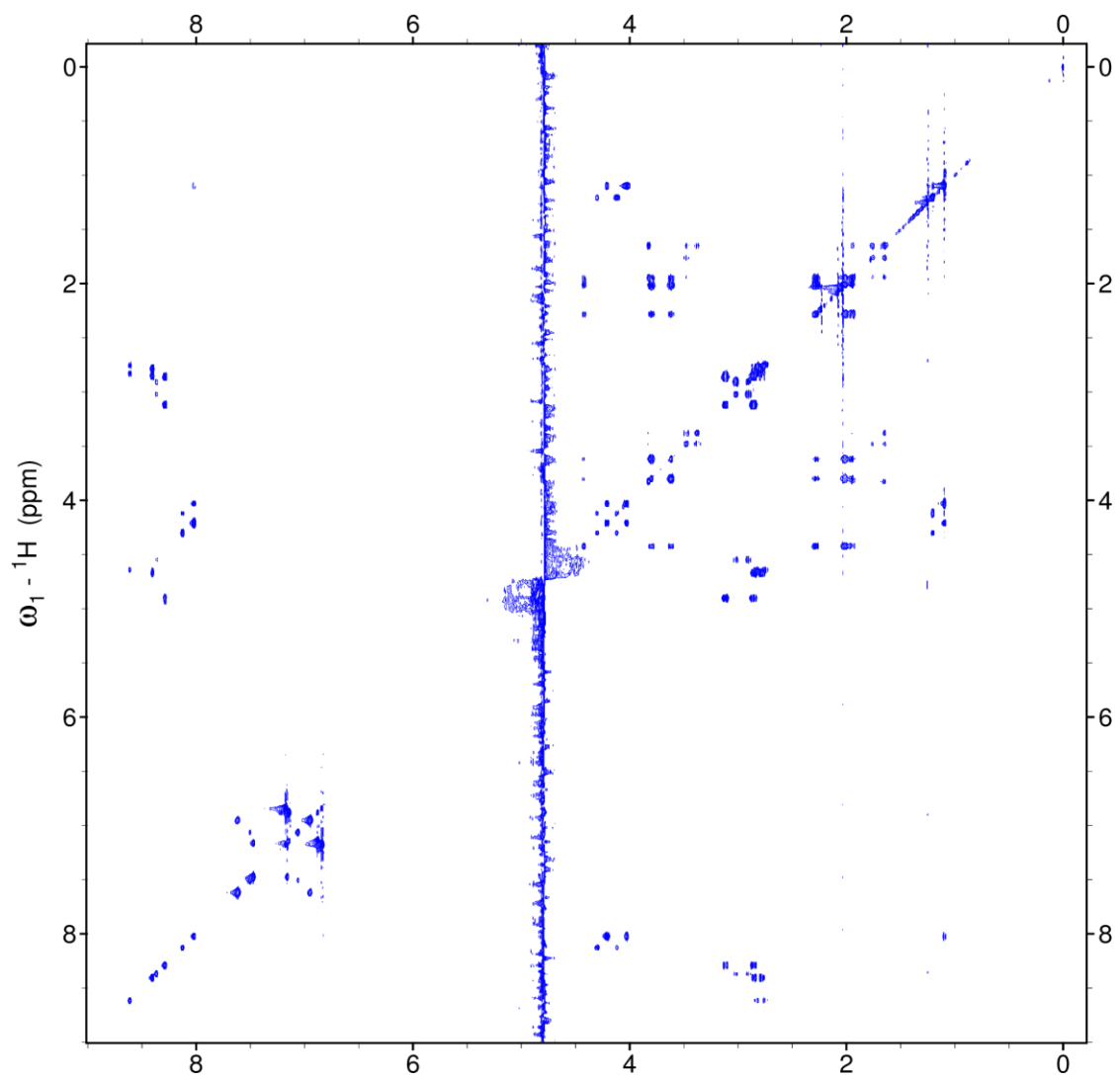
Figure 1.31   TOCSY spectrum of the peptide Ac-TYPN-NH$_2$ in 90% H$_2$O/10% D$_2$O
with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
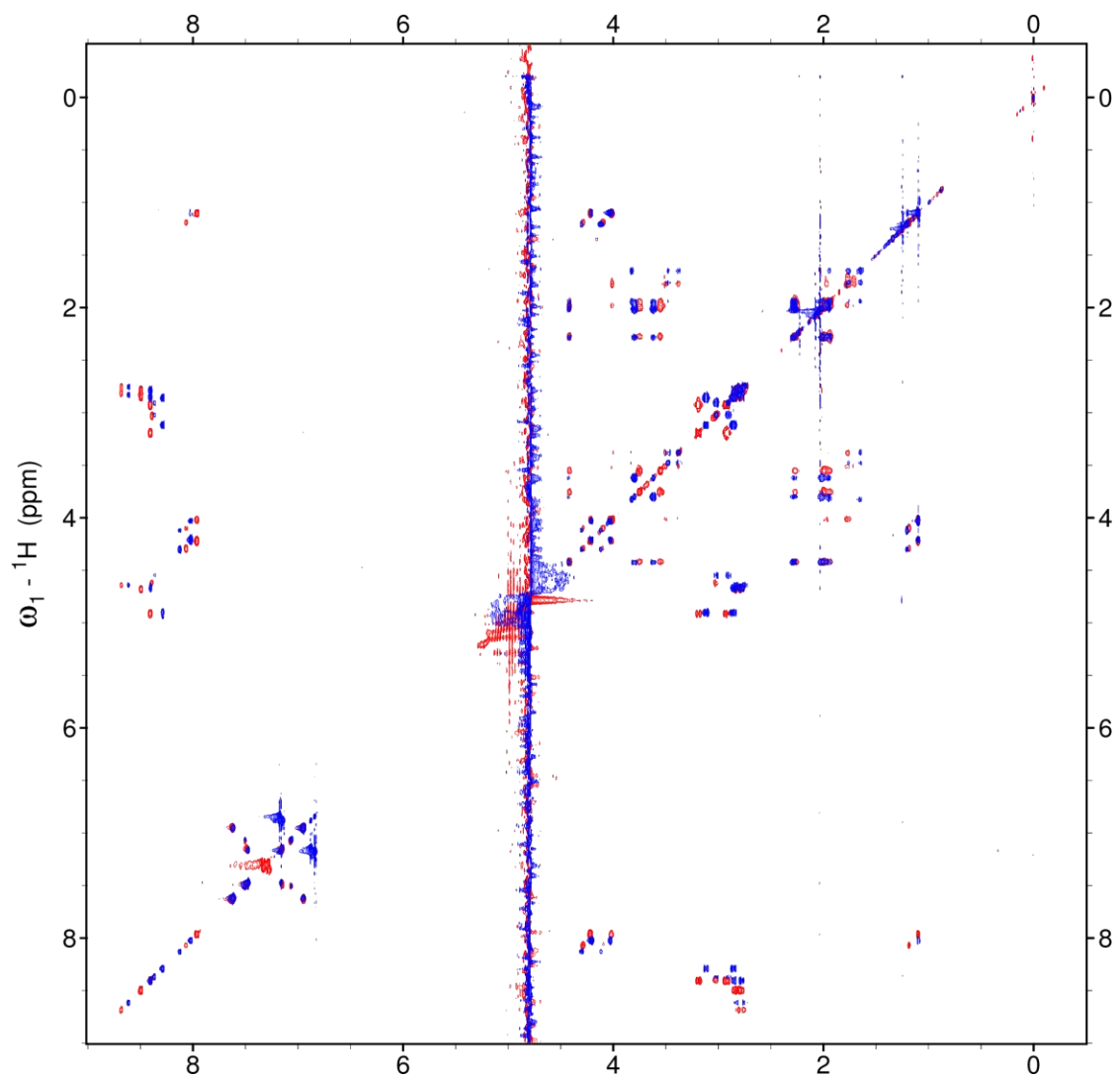
Figure 1.32   Superposition of the TOCSY spectra of the peptides Ac-
T(**Tyr(C₄F₉)**))PN-NH₂ (red) and Ac-TYPN-NH₂ (blue) in 90% H_2O/10%
D_2O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
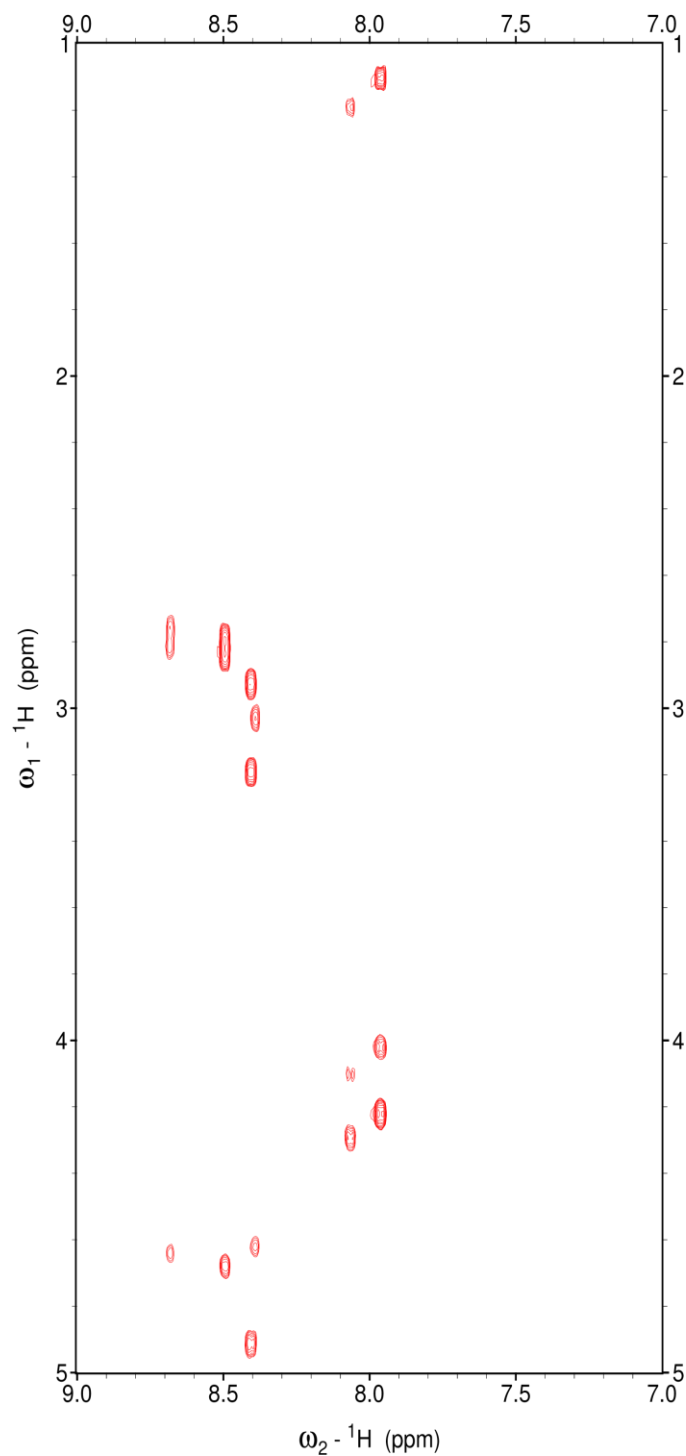
Figure 1.33 Fingerprint region of the TOCSY spectrum of the peptide Ac-T(**Tyr(C₄F₉)**)PN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
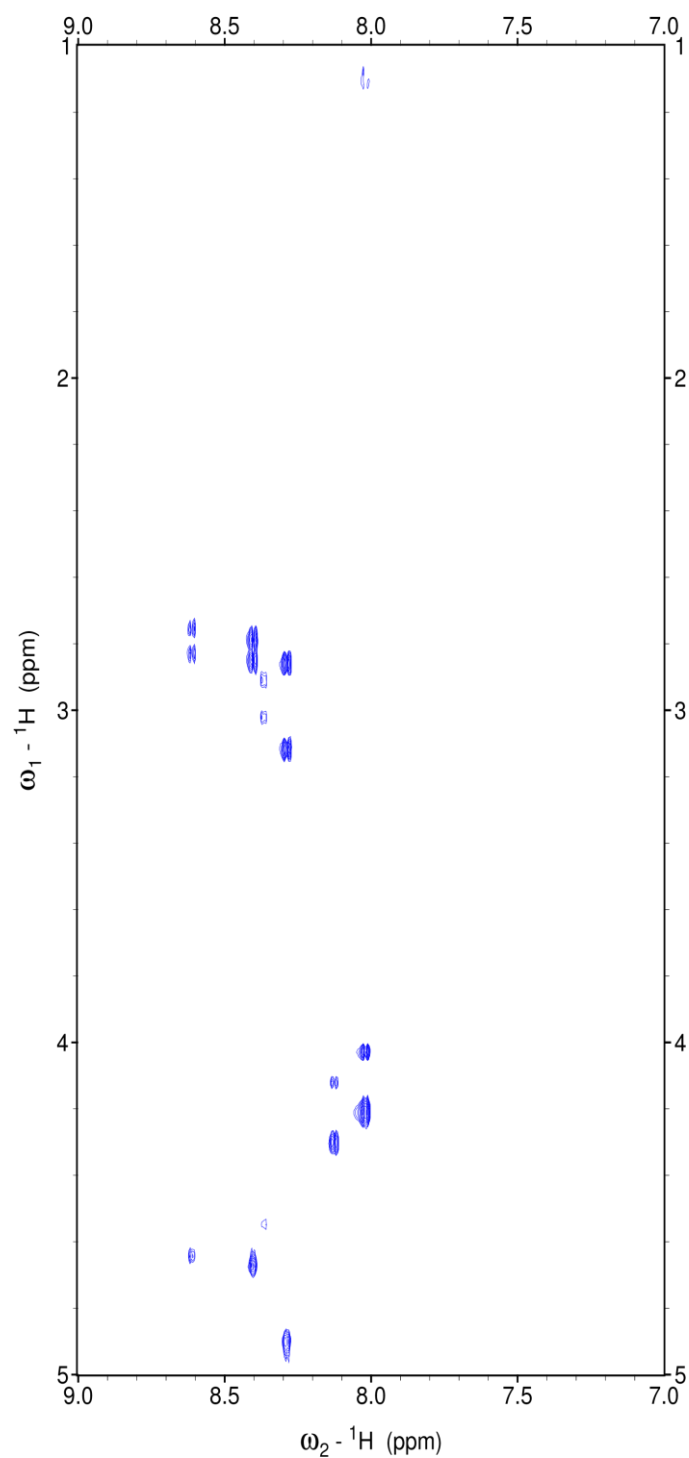
Figure 1.34 Fingerprint region of the TOCSY spectrum of the peptide Ac-TYPN-NH$_2$ in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.
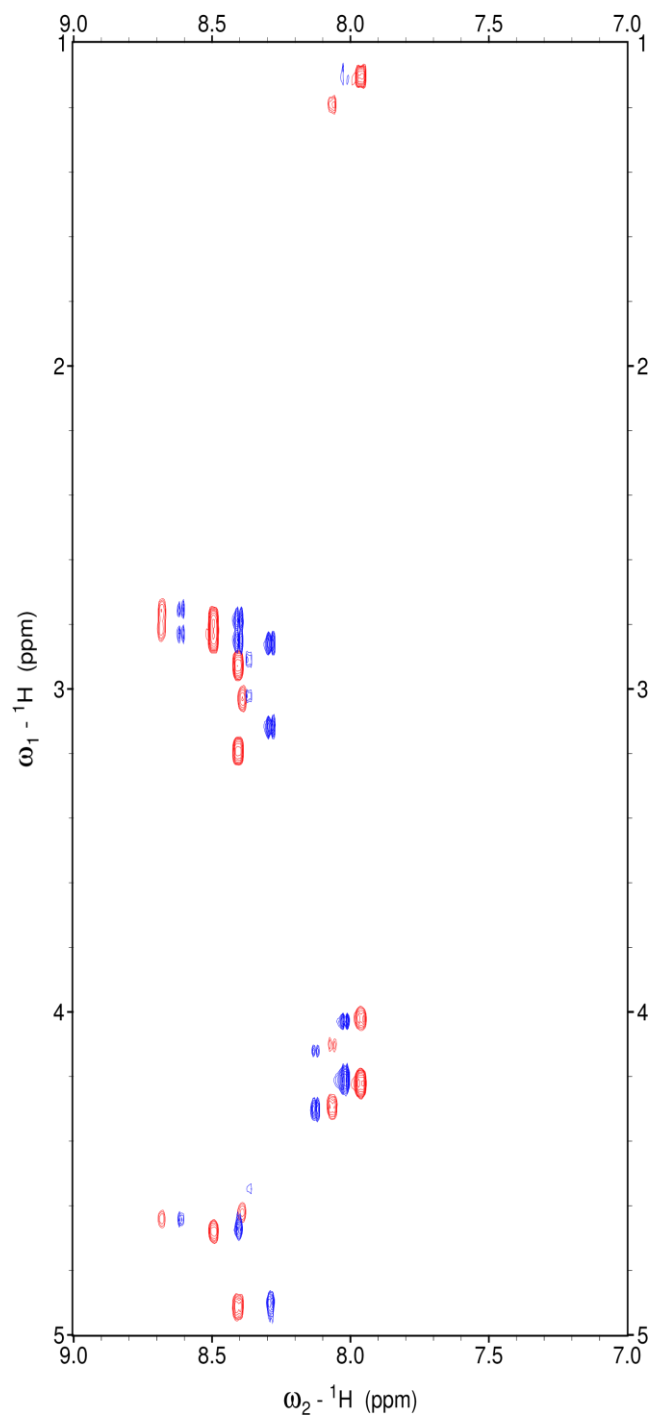
Figure 1.35 Superposition of the fingerprint region of the TOCSY spectra of the peptides Ac-T(**Tyr(C$_4$F$_9$)**)PN-NH$_2$ (red) and Ac-TYPN-NH$_2$ (blue) in 90% H$_2$O/10% D$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.

1.2.12 Calculation of the Hammett Sigma Constant of a Perfluoro-*tert*-butyl Ether

Hammett constants (σ) are a measure of electron-donating or electron-withdrawing effects of functional groups on an aromatic system.[87] The σ value is determined by measuring the pKa of substituted benzoic acids. The location of the substituent on the aromatic ring gives rise to differing σ values, where σ_{para} is a measure of both resonance and inductive effects, while σ_{meta} is primarily a measure of the inductive effects Currently, there is no measured value for the perfluoro-*tert*-butyl ether group. In order to understand the electronic properties of the synthesized amino acid, the σ value for the perfluoro-*tert*-butyl ether was calculated. Previously in the Zondlo lab, a linear free energy relationship between the log($K_{trans/cis}$) in Ac-T(4-Z-Phe)PN-NH$_2$ peptides and σ value of Z of 4-Z-Phe amino acids was established. A series of peptides was synthesized with known σ values of the 4-substituent (Z).[88] The $K_{trans/cis}$ was calculated by measuring the ratio of *trans* and *cis* amide bond populations by ^{1}H NMR. The series of peptides covered both a large range of $K_{trans/cis}$ values, and the data were fit to both σ_{para} and σ_{meta} values. The calculated σ_{para} value for the 4-perfluoro-*tert*-butyl ether is 0.30 using equation (1) with $\rho = 0.295$.[88] The σ_{para} value is seemingly low for a derivative containing nine fluorine atoms; however, it has previously been demonstrated that both distance from the ring and ether linkages minimize the electron-withdrawing effect of multiple fluorines.[87] In this case, all nine fluorines are four bonds separated from the aromatic ring. As seen in Table 1.7, the addition of an ether linkage significantly reduces the electron-withdrawing effect in all previously examined cases. This observation can be attributed to the electron-donating nature of oxygen. Furthermore, as the fluorines move further away from the ring system, they have less of an electron-withdrawing effect on the ring system (Table

1.8). The σ_{meta} value for a perfluoro-*tert*-butyl ether substitution is 0.31 calculated using equation (2) with $\rho = 0.328$.[89]

(1) $\log(K_{\text{trans/cis}}) = 0.2954\sigma_{\text{para}} + 0.523$

(2) $\log(K_{\text{trans/cis}}) = 0.3285\sigma_{\text{meta}} + 0.511$

Substituent	σ_{para}	σ_{meta}	Substituent	σ_{para}	σ_{meta}
CH_3	-0.17	-0.07	CF_3	0.54	0.43
OCH_3	-0.27	0.12	OCF_3	0.35	0.38
O-*t*Bu	-0.29		$O(C_4F_9)$	0.07	0.23

Table 1.6   Trends in both σ_{para} and σ_{meta} values for ether substitutions show similar trends.

Substituent	σ_{para}	σ_{meta}
CHF_2	0.32	0.29
$OCHF_2$	0.18	0.31
CF_2CF_3	0.52	0.47
OCF_2CF_3	0.28	0.48

Table 1.7   The perfluoro-*tert*-butyl butyl ether follows similar trends to other fluorinated substitutions. As the fluorine atoms move further away from the ring, less electron withdrawing nature is observed.

1.3   Summary and Discussion

We have demonstrated the synthesis of four unique highly fluorinated amino acids via the incorporation of a perfluoro-*tert*-butyl ether into multiple classes of amino acids, including cyclic (proline), aliphatic, and aromatic amino acids. This modification was made via the Mitsunobu reaction in the case of the hydroxyproline derivatives and the homoserine derivatives, or via diazonium coupling for the tyrosine derivatives, using the commercially available perfluoro-*tert*-butanol. All four amino acids were synthesized as Fmoc-protected monomers for incorporation into peptides via solid phase peptide synthesis. In synthesizing peptides with these amino acids, we found amide bond coupling reactions with these amino acids are slow. These amino acids can easily be incorporated utilizing longer reaction times, superior coupling reagents, and PEG-polystyrene graft resin for solid-phase peptide synthesis. Furthermore, the perfluoro-*tert*-butyl group can also be installed directly on the peptide as was demonstrated on solid phase by Pandey *et al.* with the hydroxyproline derivatives and in solution with the tyrosine derivative. Due to the sterically demanding nature of the perfluoro-*tert*-butanol and the sterics of incorporation into peptides, it may not be possible to utilize this chemistry on peptides in all cases.

All four amino acids containing perfluoro-*tert*-butyl were placed into previously established peptide model systems to determine their conformational preferences. Both hydroxyproline derivatives and the homoserine derivative were incorporated into the model peptide contexts Ac-GPPXPPGY-NH$_2$ and Ac-XKAAAAKAAAAKAAGY-NH$_2$, which were used to quantify PPII and α-helix propensity, respectively. These data demonstrated that the designed amino acids all contain unique propensities for secondary structures. The 4*R*-perfluoro-*tert*-butyl hydroxyproline strongly favors PPII helices, but relative to other amino acids,

disfavors α-helices. The 4*S*-perfluoro-*tert*-butyl hydroxyproline, relatively disfavors any compact structure compared to 4*R*-perfluoro-*tert*-butyl hydroxyproline, but is more PPII favoring than many amino acids. The perfluoro-*tert*-butyl homoserine relatively favors α-helical structure and favorably adopts PPII helix structures, making it a unique highly fluorinated amino acid which can be introduced into more compact secondary structures without destabilizing them. These data, taken together, demonstrate a wide range of conformation preferences and, in turn, the potential for a wide range of potential applications.

Perfluoro-*tert*-butyl tyrosine was examined in the Ac-TXPN-NH$_2$ model system, which allowed for characterization of the electronic effect of the perfluoro-*tert*-butyl ether, which has not been previously established. Only one other example of a perfluoro-*tert*-butyl aryl ether was previously synthesized. Through NMR analysis, we have demonstrated that the perfluoro-*tert*-butyl ether modification to an aromatic system is electronically conservative relative to other electron-withdrawing groups, despite the steric bulk associated with the perfluoro-*tert*-butyl ether. We were able to calculate Hammett constants for the perfluoro-*tert*-butyl ether, which was previously not measured due to the lack of perfluoro-*tert*-butyl ethers previously synthesized. These data imply that the perfluoro-*tert*-butyl tyrosine could potentially be used as a mimic for aromatic amino acids and large hydrophobic residues such as Tyr and Trp.

Having characterized all of these amino acids, we would like to design applications for each derivative, emphasizing the unique nature of each amino acid. One potential application for all of these amino acids is protein engineering via incorporation by expression of unnatural amino acids. While the technology for incorporating proline-based unnatural amino acids is currently in development, other

aliphatic amino acids and aromatic amino acids are commonly expressed using a variety of techniques such as expression in auxotrophic bacterial or the Amber suppression method, which utilizes a redundant stop codon and mutant tRNA synthestases to incorporate unnatural amino acids. This approach has been demonstrated for various Met and Leu derivatives, as well as a large subset of Phe and Tyr derivatives, including 4-trifluoromethyl-Phe. These amino acids could potentially be expressed within proteins as tracers/trackers, probes for protein folding, structural sensors, localization probes, posttranslational modification sensors, and protein-protein interaction sensors.

The most likely encodable probe is the perfluoro-*tert*-butyl tyrosine. Based on previous work in other labs, nearly 30 different aromatic amino acids have been expressed, including fluorinated amino acids as well as *tert*-butyl tyrosine.[90-94] Sterically, *tert*-butyl tyrosine and perfluoro-*tert*-butyl tyrosine are very similar. Both of these amino acids have free rotation around the *tert*-butyl ether, which make them ideal magnetic resonance probes (Figure 1.36). Chen *et al*. demonstrated that *tert*-butyl tyrosine could be expressed in DnaB helicase and Sortase A utilizing p-cyanophenylalanyl-tRNA synthetase for site specific incorporation in bacterial cells.[94] The protein containing *tert*-butyl tyrosine was subjected to analysis by ^{1}H NMR. The *tert*-butyl tyrosine displays a large singlet by ^{1}H NMR around 1.2 ppm, which is not present in the wildtype protein. The sortase A was also expressed and observed by ^{1}H NMR. The ^{1}H NMR spectrum of the sortase A also displayed a large singlet around 1 ppm. Sortase A, however, natively binds calcium ions. Upon addition of calcium to the apo-protein, a small chemical shift change is observed. Despite being the largest peak by ^{1}H NMR, both proteins contain a large amount of hydrogen atoms, all of

which are visible by ^{1}H NMR leading to a large of number of peaks in the spectrum. Perfluoro-*tert*-butyl tyrosine is similar enough to *tert*-butyl tyrosine that it could potentially be expressed using the same system without the drawback of competing signals.
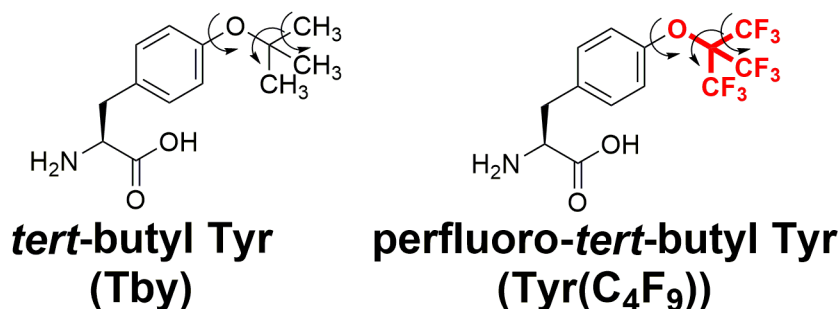
Figure 1.36   *tert*-Butyl tyrosine (Tby) and its fluorinated analogue. Arrows indicate bond rotations that result in isotropic averaging of the NMRactive 1 H nuclei in Tby, even in large proteins

One particularly interesting application of these amino acids in expressed proteins is using antibodies to detect extracellular markers. There are many examples of expression or overexpression of extracellular, membrane associated-receptors in cancer. One of these such examples is the overexpression of nuclear hormone receptors in certain types of cancer including breast cancer and prostate cancer.[95] In some types of breast cancer either or both the estrogen receptor and the receptor tyrosine-protein kinase ErbB-2 (Her2) proteins are overexpressed, leading to an increase in transcription and unchecked proliferation of cells.[96] In the case of prostate cancer, the androgen receptor is overexpressed, also leading to increased transcription.[97] Current therapeutics target these receptors to specifically target cancer

cells. Herceptin is an antibody used for the treatment of Her2-positive breast cancers, and is currently being used very successfully to treat all stages of breast cancer.[96,98] There are also new, very promising immunotherapies both recently approved and currently in development or clinical trials for cancer treatment.[99] The future of many cancer treatments are utilizing antibody-based immunotherapy. We believe that these therapies could be improved with the use of fluorinated amino acids. Current immunotherapies cannot be imaged during treatment. Incorporation of amino acids containing perfluoro-*tert*-butyl groups into these antibodies could be used to detect cancer or track the growth and/or death of cancer cells during immunotherapy treatment. The high signal-to-noise ratio of the perfluoro-*tert*-butyl amino acids would greatly benefit imaging of cancer cells, allowing even small groups of cells to appear on an image where traditional imaging could potentially fail.

Another potential application for expressed amino acids containing perfluoro-*tert*-butyl groups is in the detection of tumor metastasis. Tumor invasion and metastasis is known to activate extracellular matrix metalloproteinases (MMP), which are used in tissue remodeling.[100] MMPs break down proteins to clear space and allow for healthy cell growth, but these pathways are activated by tumor cells as the cells prepare to invade the surrounding tissues. Currently, metastatic cancers are dectected by MRI or PET imaging when solid tumors form throughout the body, away from the original tumor site. The probes containing perfluoro-*tert*-butyl groups could be expressed and delivered to a tumor site. Upon cleavage of the expressed protein, a chemical shift change would be observable indicating active MMPs and potential metastatic cancer before actual tumors would be detectable in other regions of the body.

Another key potential application for these amino acids is in the emerging field of theranostics. Theranostics refers to the development of tools which can be used to both diagnose and treat diseases. These molecules require high specificity to potential targets, as well as some sort of diagnostic probe. These amino acids are perfectly designed to be used as theranotics in the clinic. The perfluoro-*tert*-butyl ether can potentially be designed into recognition sequences and used to increase hydrophobic contacts often required for protein-protein interactions. As already demonstrated, these amino acids can rapidly be detected in the high nanomolar concentration range by ^{19}F NMR, and will also be detectable by ^{19}F MRI. The perfluoro-*tert*-butyl ether could also be incorporated into a small theranostic agent using the same chemistry described above which would gain the same benefits (high signal-to-noise, low concentration detection, potentially increased specificity, and increased hydrophobics), while utilizing a small molecule scaffold which has the potential to be more stable and target different interactions than a peptide. There are a large number of protein kinases which are currently high profile targets for the pharmaceutical industry in the prevention and treatment of disease. The first example of a protein kinase drug target was Gleevec which was approved for treatment of chronic myelogenous leukemia (CML). Gleevec kills cancer cells by binding to Abl kinase, a tyrosine protein kinase, which activates a cascade upon phosphorylation which is required for cell survival. Gleevec "turns off" Abl kinase, causing tumor cell death. The success of this drug demonstrated the power of small molecule-based kinase inhibitor therapeutics, but much like immunotherapies, there is no method by which to image the effectiveness of these treatments. Incorporation of a perfluoro-*tert*-butyl group into a small molecule would increase lipophilicity, potentially increase binding to the protein of interest, and

allow for imaging by ^{19}F magnetic resonance similar to the antibody-based theranostics described above.

Our interest is in placing the designed amino acids into peptide systems that will allow for specific cellular information, which can indicate disease states via ^{19}F NMR. The goal is to design perfluoro-*tert*-butyl amino acids to demonstrate their ability to be designed into specific recognition sequences for ^{19}F imaging within biological systems. Due to the large hydrophobic handle created by the addition of the perfluoro-*tert*-butyl moiety, its ability to be incorporated into recognition sequences needs to be demonstrated. While these amino acids can potentially be expressed in proteins and be used as tracers, the real innovation behind this work is demonstrating that these amino acids can be tolerated within biological systems in key locations. We have explored these amino acids being tolerated by proteins and enzymes which will be demonstrated through biologically relevant systems (to be explored in Chapters 2 and 3), which were designed based on the structural preferences of each amino acid. Perfluoro-*tert*-butyl tyrosine was found to be electronically and conformationally similar to tyrosine and other aromatic amino acids, despite an increase in steric bulk and electron-withdrawing capability. Perfluoro-*tert*-butyl homoserine is conformationally similar to other aliphatic amino acids such as leucine/isoleucine respectively based on structural data. We intend to use these amino acids in recognition sequences for protein-protein interactions which accommodate large hydrophobic amino acids including aromatic and aliphatic side chains. These systems include the estrogen receptor coactivator binding site and the p53•MDM2 binding interaction. Perfluoro-*tert*-butyl homoserine is ideal for the estrogen receptor coactivator binding motif due to its α-helical propensity. The coactivator binds to the

estrogen receptor through the α-helical binding motif LXXLL where X is any amino acid. Perfluoro-*tert*-butyl tyrosine will be incorporated in p53-based peptides which bind to MDM2 via an FXXΦW motif. This motif is α-helical in nature and depends on the increased hydrophobicity of aromatic amino acids for binding. The perfluoro-*tert*-butyl hydroxproline derivatives will be incorporated into protein kinase recognition sequences. While these amino acids could theoretically be applied to protein-protein interactions, the proline ring is unique in that it is cyclized with the backbone of the peptide. The cyclic nature of proline allows the perfluoro-*tert*-butyl group to "feel" the *cis-trans* isomerization at the amide bond in the Ac-TYXN-NH$_2$ model system. Our design exploits this unique property of proline-based amino acids to potentially detect post-translational modifications whose electronic properties could potentially be felt through the peptide backbone structure. We will incorporate 4*R*- and 4*S*-perfluoro-*tert*-butyl hydroxyproline into recognition sequences for two well studied protein kinases: Protein Kinase A (PKA) and Protein Kinase B (Akt). The recognition sequences for protein kinases are based on specific amino acid site chains rather than secondary structure, making the hydroxyproline derivatives ideal for this application.

In this chapter, we have demonstrated the synthesis and key characterization of four novel amino acids containing a perfluoro-*tert*-butyl ether moiety. These amino acids will be further explored in chapters 2 and 3 for potential applications as ^{19}F based imaging probes for detection within biological systems.

1.4　Experimental

1.4.1　Synthesis of (2*S*, 4*S*)-Perfluoro-*tert*-butyl-4-hydroxyproline

Boc-(2*S*,4*S*)-perfluoro-*tert*-butyl-4-hydroxyproline methyl ester (8). Compound **1**

(3.09 g, 12.6 mmol) and Ph_3P (3.96 g, 15.1 mmol) were dissolved in toluene (126 mL)

under a nitrogen atmosphere. The solution was cooled to 0 ºC and stirred on ice for 10

minutes. DIAD (3.05 g, 2.98 mL, 15.1 mmol) was added dropwise to the solution over

15 minutes. Perfluoro-*tert*-butanol (5.95 g, 3.52 mL, 25.2 mmol) and DIPEA (3.18 g,

4.38 mL, 25.2 mmol) were added to the solution, which was stirred on ice for another

5 minutes. The solution was removed from the ice bath, warmed to 45 °C in an oil

bath, and stirred for 24 hours. The solvent was removed under reduced pressure, and

the crude product was dissolved in ethyl acetate (50 mL). The crude product was

washed with brine (2 × 75 mL) and dried over sodium sulfate. The solvent was

removed under reduced pressure and the crude product was redissolved in CH_2Cl_2 (50

mL). The crude product was purified via column chromatography (0-7% ethyl acetate

in hexanes v/v) to obtain compound **8** (2.07 g, 4.47 mmol) as a colorless oil in 36%

yield. ^1H NMR (400 MHz, $CDCl_3$) δ 4.84 (br s, 1H), 4.52-4.50 (dd, *J* = 8.3, 3.7 Hz,

0.4H, minor), 4.41-4.38 (dd, *J* = 9.0, 3.1 Hz, 0.6H, major), 3.84-3.78 (dd, *J* = 12.5, 5.6

Hz, 0.6H, major), 3.76-3.75 (d, *J* = 6.5 Hz, 0.4H, minor), 3.72 (s, 3H), 3.68-3.65 (d, *J*

= 13.0 Hz, 0.6H, major), 3.58-3.55 (d, *J* = 13.0 Hz, 0.4H, minor), 2.53-2.38 (m, 2H),

1.48 (s, 4H, major), 1.43 (s, 5H, minor). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 172.9,

172.8, 154.2, 154.0, 129.4, 129.3, 124.8, 124.6, 124.5, 120.1 (q, *J* = 293 Hz), 105.1,

80.9, 80.8, 66.6, 66.2, 57.5, 57.1, 52.3, 37.6, 36.6, 28.3, 28.2. ^{19}F NMR (376.3 MHz,

CDCl$_3$) δ −70.42 (minor), −70.44 (major). HRMS (LIFDI-TOF) m/z: [M]$^+$ calcd for C$_{15}$H$_{18}$F$_9$NO$_5$ 463.1041, found 463.1042.

Boc-(2*S*,4*S*)-perfluoro-*tert*-butyl-4-hydroxyproline (9). Compound **8** (1.75 g, 3.78 mmol) and LiOH (0.1086 g, 4.53 mmol) were dissolved in a solution of water (20 mL) and 1,4-dioxane (20 mL). The solution was stirred at room temperature for 14 hours. The reaction mixture was acidified to pH 2 with dilute HCl and extracted with ethyl acetate (2 × 75 mL). The solvent was removed under reduced pressure and the crude product redissolved in CH$_2$Cl$_2$. The crude product was purified via column chromatography (0 to 2% methanol in CH$_2$Cl$_2$ v/v) to obtain compound **9** (1.05 g, 2.34 mmol) as an off-white solid in 62% yield. ^{1}H (600 MHz, CDCl$_3$) δ 4.86 (s, 1H), 4.51-4.50 (d, *J* = 6.7 Hz, 0.4H, minor), 4.44-4.42 (d, *J* = 7.8 Hz, 0.6H, major), 3.82-3.76 (m, 1H), 3.72-3.67 (m, 0.6H, major), 3.56-3.55 (m, 0.4H, minor), 2.52-2.44 (m, 2H), 1.49 (s, 3.6H), 1.44 (s, 5.4H). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 177.4, 175.5, 154.7, 153.5, 120.1 (q, *J* = 293 Hz), 80.9, 57.5, 57.1, 52.9, 52.7, 52.1, 37.7, 37.6, 36.1, 28.1. ^{19}F NMR (564.5 MHz, CDCl$_3$) δ −70.38 (major), −70.44 (minor). HRMS (LIFDI-TOF) m/z: [M+H]$^+$ calcd for C$_{14}$H$_{17}$F$_9$NO$_5$ 450.0963, found 450.0958.

(2*S*,4*S*)-Perfluoro-*tert*-butyl-4-hydroxyproline hydrochloride (10). Compound **8** (2.0 g, 4.3 mmol) was dissolved in 1,4-dioxane (15 mL) and 4 M HCl (15 mL) was added. The solution was allowed to stir at reflux for 6 hours. The solvent was removed

under reduced pressure. Compound **10** (1.5 g, 4.3 mmol) obtained was used as a crude

reagent in the next step. Alternatively, compound **8** could be subjected to the same

conditions to yield compound **10**. ^{1}H NMR (600 MHz, MeOD-d$_4$) δ 5.20 (s, 1H), 4.62-

4.60 (dd, J = 9.6, 3.4 Hz, 1H), 3.74-3.71 (dd, J = 13.4, 4.6 Hz, 1H), 3.56-3.53 (d, J =

13.9 Hz, 1H), 2.74-2.68 (ddd, J = 14.6, 10.0, 5.0 Hz, 1H), 2.54-2.51 (d, J = 14.6 Hz,

1H). ^{13}C NMR (150.8 MHz, MeOD-d$_4$) δ 169.3, 120.0 (J = 293 Hz), 78.9, 58.3, 52.1,

36.1. ^{19}F NMR (564.5 MHz, MeOD-d$_4$) δ −71.65. HRMS (LIFDI-TOF) m/z: [M]$^+$

calcd for C$_9$H$_9$F$_9$NO$_3$ 350.0439, found 350.0413.

Fmoc-(2*S*,4*S*)-perfluoro-*tert*-butyl-4-hydroxyproline (11). Crude compound **9** (0.50

g, 1.4 mmol) was dissolved in 1,4-dioxane (7 mL). Fmoc-OSu (0.50 g, 1.4 mmol) and

K$_2$CO$_3$ (0.39 g, 2.9 mmol) were added and the resultant solution was stirred for 14

hours at room temperature. The 1,4-dioxane was removed under reduced pressure and

the crude product was acidified with 2 M HCl (10 mL). The crude product was

extracted with ethyl acetate (2 × 20 mL). The solvent was removed and the crude

product was redissolved in CH$_2$Cl$_2$. The crude mixture (0.40 g, 0.72 mmol) was

purified via column chromatography (0-4% methanol in CH$_2$Cl$_2$ v/v) to obtain

compound **11** (0.38 g, 0.70 mmol) as a white solid in 50% yield. ^{1}H (600 MHz,

CDCl$_3$) δ 7.77-7.69 (d, J = 7.6 Hz, 2H), 7.60-7.51 (m, 2H), 7.42-7.28 (m, 4H), 4.92 (s,

0.5H), 4.88 (s, 0.5H), 4.61-4.56 (m, 1H), 4.53-4.49 (m, 0.5H), 4.43-4.31 (m, 2H),

4.16-4.13 (dd, J = 6.3, 6.1 Hz, 0.5H), 3.90-3.86 (m, 1H), 3.76-3.63 (m, 1H), 3.74 (s,

1H), 2.57-2.41 (m, 2H). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 176.3, 175.7, 154.7, 154.3, 143.9, 143.8, 143.60, 143.58, 141.4, 141.3, 127.8, 127.74, 127.70, 127.13, 127.10, 127.0, 125.1, 125.0, 124.8, 121.1, 119.97 ($J = 293$ Hz), 119.96, 78.2, 67.9, 67.8, 67.0, 57.3, 57.0, 53.4, 52.9, 47.1, 37.7, 36.5, 21.9. ^{19}F NMR (564.5 MHz, CDCl$_3$) δ −70.42, −70.49. HRMS (LIFDI-TOF) m/z: [M]$^+$ calcd for C$_{24}$H$_{18}$F$_9$NO$_5$ 571.1041, found 571.1038.

1.4.2 Synthesis of (2*S*, 4*R*)-Perfluoro-*tert*-butyl-4-hydroxyproline

Boc-(2*S*,4*S*)-*p*-nitrobenzoate-4-hydroxyproline methyl ester (2). Compound **1** (3.90 g, 15.7 mmol), triphenylphosphine (Ph$_3$P) (7.40 g, 28.3 mmol), and *p*-nitrobenzoic acid (3.10 g, 18.8 mmol) were dissolved in anhydrous THF (157 mL). The reaction was conducted under a nitrogen atmosphere. The solution was cooled to 0 ºC. Diisopropylazodicarboxylate (DIAD) (6.35 g, 6.20 mL, 31.4 mmol) was added dropwise over 30 minutes. The solution was removed from the ice bath, allowed to warm to room temperature, and stirred for an additional 6 hours. The solvent was removed under reduced pressure. The crude product was dissolved in ethyl acetate (100 mL) and washed with brine (2 × 200 mL). The solvent was removed and the crude product was redissolved in CH$_2$Cl$_2$ (75 mL). The crude product was purified via column chromatography (CH$_2$Cl$_2$) to yield compound **2** (4.7 g, 12.0 mmol) as a colorless oil in 76% yield. The NMR data corresponded to the literature values.[51,52]

Boc-(2*S*,4*S*)-4-hydroxyproline methyl ester (3). Compound **2** (2.0 g, 5.0 mmol) was dissolved in acetone (50 mL). Sodium azide (0.49 g, 7.5 mmol) was added and the solution was heated at reflux for 14 hours. The solution was allowed to cool to room temperature and solvent was removed under reduced pressure. The crude product was dissolved in ethyl acetate (50 mL) and washed with distilled water (2 × 50 mL). The solvent was removed and the crude product was redissolved in CH_2Cl_2 for purification. Compound **3** (0.80 g, 3.3 mmol) was purified via column chromatography in (2% methanol in CH_2Cl_2 v/v) to obtain a colorless oil in 65% yield. The NMR data corresponded to the literature values.[51,52]

Boc-(2*S*,4*R*)-perfluoro-*tert*-butyl-4-hydroxyproline methyl ester (4). Compound **3** (2.23 g, 9.10 mmol) and Ph_3P (2.86 g, 10.9 mmol) were dissolved in toluene (91 mL) under a nitrogen atmosphere. The solution was cooled to 0 ºC and stirred on ice for 10 minutes. DIAD (2.20 g, 2.15 mL, 10.9 mmol) was added dropwise to the solution over 15 minutes. Perfluoro-*tert*-butanol (4.30 g, 2.54 mL, 18.2 mmol) and DIPEA (2.35 g, 3.16 mL, 18.2 mmol) were added to the solution, which was then stirred on ice for another 5 minutes. The solution was removed from the ice bath, warmed to 45 °C in an oil bath, and stirred for 24 hours. The solvent was removed under reduced pressure and the crude product was dissolved in ethyl acetate (50 mL). The crude product was washed with brine (2 × 75 mL) and dried over sodium sulfate. The solvent was removed under reduced pressure and the crude product was redissolved in CH_2Cl_2 (50 mL). The crude product was purified via column chromatography (0-7% ethyl acetate

in hexanes v/v) to obtain compound **4** (0.790 g, 2.28 mmol) as a colorless oil in 25% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.91 (s, 1H), 4.49-4.47 (dd, J = 8.6, 6.4 Hz, 0.4H, *cis*), 4.40-4.36 (dd, J = 7.8, 7.8 Hz, 0.6H, *trans*), 3.84-3.75 (m, 1H), 3.75 (s, 3H), 3.69-3.66 (d, J = 12.5 Hz, 0.6H, *trans*), 3.60-3.58 (d, J = 12.3 Hz, 0.4H, *cis*), 2.47-2.44 (m, 1H), 2.28-2.23 (m, 1H), 1.46 (s, 4H, *cis*), 1.42 (s, 5H, *trans*). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 171.9, 171.7, 153.1, 152.4, 119.1 (q, J = 293 Hz), 79.8, 77.1, 56.5, 56.1, 51.5, 51.4, 51.2, 36.6, 35.6, 27.2. ^{19}F NMR (376.3 MHz, CDCl$_3$) δ −70.47 (*trans* conformation), −70.53 (*cis* conformation). HRMS (LIFDI-TOF) m/z: [M]$^{+}$ calcd for C$_{15}$H$_{18}$F$_9$NO$_5$ 463.1041, found 463.1051.

Boc-(2*S*,4*R*)-perfluoro-*tert*-butyl-4-hydroxyproline (5). Compound **4** (0.36 g, 0.78 mmol) and LiOH (0.22 g, 0.93 mmol) were dissolved in a solution of water (4 mL) and 1,4-dioxane (4 mL). The solution was stirred at room temperature for 14 hours. The mixture was acidified to pH 2 with dilute HCl and extracted with ethyl acetate (2 × 10 mL). The solvent was removed under reduced pressure and redissolved in CH$_2$Cl$_2$. The crude product was purified via column chromatography (0 to 2% methanol in CH$_2$Cl$_2$ v/v) to obtain compound **5** (0.22 g, 0.49 mmol) as an off-white solid in 63% yield. ^{1}H (600 MHz, CDCl$_3$) δ 4.92 (s, 1H), 4.53−4.51 (dd, J = 7.7, 7.0 Hz, 0.6H), 4.42-4.39 (dd, J = 7.4, 7.4 Hz, 0.4H), 3.83-3.80 (m, 0.4H), 3.74-3.66 (m, 1.6H), 2.51-2.48 (m, 1H), 2.44-2.40 (m, 0.6H), 2.35-2.32 (m, 0.4H), 1.48 (s, 5H), 1.43 (s, 4H). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 177.7, 174.7, 155.9, 153.4, 120.1 (q, J = 293

Hz), 82.3, 81.3, 57.4, 52.6, 37.4, 35.7, 28.2. ^{19}F NMR (564.5 MHz, CDCl$_3$) δ −70.49 (*trans* conformation), −70.55 (*cis* conformation). HRMS (LIFDI-TOF) m/z: [M+H]$^+$ calcd for C$_{14}$H$_{17}$F$_9$NO$_5$ 450.0963, found 450.0946.

(2*S*,4*R*)-Perfluoro-*tert*-butyl-4-hydroxyproline hydrochloride (6). Compound **4** (1.3 g, 2.2 mmol) was dissolved in a solution of 1,4-dioxane (15 mL) and 4 M HCl (15 mL). The solution was allowed to stir at reflux for 6 hours. The solvent was removed under reduced pressure. Compound **6** (0.78 g, 2.2 mmol) was used as a crude reagent in the next step without purification. Alternatively, compound **5** could be subjected to identical conditions to yield compound **6**. ^{1}H NMR (600 MHz, MeOD-d$_4$) δ 5.22 (br s, 1H), 4.55-4.51 (dd, *J* = 9.6, 8.8 Hz, 1H), 3.81-3.79 (dd, *J* = 13.5, 5.0 Hz, 1H), 3.46-3.44 (d, *J* = 13.5 Hz, 1H), 2.53-2.52 (m, 1H), 2.51-2.50 (m, 1H). ^{13}C NMR (150.8 MHz, MeOD-d$_4$) δ 168.7, 120.0 (*J* = 293 Hz), 79.0, 58.1, 51.8, 35.7. ^{19}F NMR (376.3 MHz, MeOD-d$_4$) δ −71.56. HRMS (LIFDI-TOF) m/z: [M]$^+$ calcd for C$_9$H$_9$F$_9$NO$_3$ 350.0439, found 350.0420.

Fmoc-(2*S*,4*R*)-perfluoro-*tert*-butyl-4-hydroxyproline (7). Crude compound **6** (1.01 g, 2.90 mmol) was dissolved in a solution of 1,4-dioxane (15 mL) and H$_2$O (15 mL). Fmoc-OSu (1.17 g, 3.48 mmol) and K$_2$CO$_3$ (0.80 g, 5.80 mmol) were added to the solution and the resultant mixture stirred for 14 hours at room temperature. The 1,4-dioxane was removed under reduced pressure and the crude product was acidified with

90

2 M HCl (10 mL). The crude product was extracted with ethyl acetate (2 × 20 mL). The solvent was removed and the crude product was redissolved in CH_2Cl_2. The product was purified via column chromatography (0-4% methanol in CH_2Cl_2 v/v) to obtain the compound **7** as a white solid (0.83 g, 1.45 mmol) in 50% yield. [1]H (600 MHz, $CDCl_3$) δ 7.77-7.76 (d, J = 7.5 Hz, 1.3H, *trans*), 7.74-7.73 (d, J = 7.6 Hz, 0.7H, *cis*) 7.55-7.52 (dd, J = 12.4, 7.6 Hz, 2H), 7.42-7.38 (m, 2H), 7.32-7.29 (dd, J = 7.4, 7.3 Hz, 2H), 4.92 (s, 0.7H, *trans*), 4.89 (s, 0.3H, *cis*), 4.59-4.57 (dd, J = 7.8, 7.4 Hz, 0.7H, *trans*), 4.52-4.43 (m, 2H), 4.38-4.37 (dd, J = 7.4, 7.4 Hz, 0.3H, *cis*), 4.30-4.27 (dd, J = 7.0, 6.6 Hz, 0.7H, *trans*), 4.17-4.16 (dd, J = 6.6, 6.6 Hz, 0.3H, *cis*), 3.83-3.73 (m, 1H), 3.71 (s, 1H), 2.49-2.45 (m, 1.7H, mixture of *cis* and *trans*), 2.32-2.29 (m, 0.3H, *cis*). ^{13}C NMR (150.8 MHz, $CDCl_3$) δ 173.1, 156.1, 143.5, 143.4, 141.4, 127.9, 127.2, 124.9, 124.8, 120.12 (q, J = 293 Hz), 120.09, 77.9, 68.6, 67.6, 57.6, 56.7, 53.0, 52.7, 47.2, 47.0, 37.6, 35.8. ^{19}F NMR (564.5 MHz, $CDCl_3$) δ −70.38 (*trans* conformation), −70.43 (*cis* conformation). HRMS (CI-TOF) m/z: [M]+ calcd for $C_{24}H_{18}F_9NO_5$ 571.1041, found 571.1027.

1.4.3   Synthesis of Perfluoro-*tert*-butyl Homoserine

Boc-L-homoserine (13). L-Homoserine (5.00 g, 42.0 mmol) was dissolved in 25 mL acetone and 25 mL water. Boc anhydride (10.10 g, 46.2 mmol) and triethylamine (6.4 g, 8.8 mL, 63 mmol) were added, and the resultant solution was allowed to stir at room temperature for 14 hours. The solvent was removed under reduced pressure and

the product was used in the next step without purification. The NMR data corresponded with previous literature data.[1]

Boc-ʟ-homoserine methyl ester (14). Compound **13** (9.10 g, 42.0 mmol) was dissolved in 90 mL DMF. The reaction mixture was cooled to 0 °C in an ice bath prior to the addition of methyl iodide (8.9 g, 3.9 mL, 63.0 mmol). The solution was allowed to warm to room temperature, protected from light, and allowed to stir for 14 hours. The crude reaction mixture was diluted with water (100 mL) and the product was extracted with ethyl acetate (2 × 100 mL). Excess ethyl acetate was removed under reduced pressure. The crude product was washed with 1 M HCl (2 × 50 mL), water (2 × 50 mL), saturated sodium bisulfite (2 × 50 mL), saturated sodium bicarbonate (2 × 50 mL), and saturated sodium chloride (2 × 50 mL). The ethyl acetate was removed under reduced pressure to yield compound **14** (1.67 g, 6.30 mmol) as a colorless oil in 15% yield over 2 steps. The NMR data corresponded with previous literature data.[1]

Boc-perfluoro-*tert*-butyl-homoserine methyl ester (15). Compound **14** (1.67 g, 7.17 mmol) and Ph$_3$P (2.82 g, 10.8 mmol) were dissolved in 50 mL THF in a three-neck round bottom flask with a condenser under a nitrogen atmosphere. The solution was cooled to 0 °C before DIAD (2.17 g, 2.11 mL, 10.8 mmol) was added dropwise over 15 minutes. The solution was allowed to stir on ice for 5 minutes before the dropwise addition of perfluoro-*tert*-butanol (3.39 g, 2.00 mL, 14.3 mmol) over 5 minutes. The reaction mixture was allowed to stir on ice for another 5 minutes before heating to 50 °C for 24 hours. Upon completion, solvent was removed under reduced pressure. The crude reaction mixture was purified via column chromatography (0 to 10% ethyl

acetate in hexanes) to obtain compound **15** as a white solid in 54% yield. ^{1}H (600 MHz, CDCl$_3$) δ 5.41 (d, *J* = 6.1 Hz, 1 H), 4.33 (dd, *J* = 6.8, 6.8 Hz, 1H), 4.06 (dd, *J* = 6.0, 5.9 Hz, 2H), 3.65 (s, 3H), 2.22-2.20 (m, 1H), 2.05-2.03 (m, 1H), 1.35 (s, 9H). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 172.2, 167.7, 154.3, 150.4, 119.1 (q, *J* = 293 Hz), 71.3, 65.2, 51.4, 49.6, 31.2, 27.2, 24.0, 20.6, 20.5. ^{19}F NMR (376.3 MHz, CDCl$_3$) δ −70.91. HRMS (ESI-ORBI) m/z: [M + Na]$^+$ calcd for C$_{14}$H$_{18}$F$_9$NO$_5$Na expected mass, 474.09335, observed mass, 474.09122.

Perfluoro-*tert*-butyl-homoserine hydrochloride (16). Compound **15** (1.74 g, 3.86 mmol) and LiOH (0.231 g, 9.65 mmol) were dissolved in 35 mL THF and 15 mL H$_2$O. The solution was allowed to stir at room temperature for 12 hours. The mixture was acidified to pH 2 using HCl, followed by removal of the THF under reduced pressure. The remaining solution was extracted with ethyl acetate (3 × 20 mL). The organic layers were combined and concentrated under reduced pressure as crude Boc-perfluoro-*tert*-butyl-homoserine. The product was redissolved in 10 mL 2 M HCl and 10 mL 1,4-dioxane. The solution was allowed to stir for 6 hours. The solvent was removed under reduced pressure. No purification was performed on compound **4** (1.30 g crude product). Compound **16** was used as a crude reagent in the next step. ^{1}H (600 MHz, D$_2$O) δ 4.33−4.30 (m, 2H), 4.21-4.19 (dd, *J* = 6.2, 6.1 Hz, 1H), 2.38-2.36 (m, 2H). ^{13}C NMR (150.8 MHz, D$_2$O) δ 171.2, 119.1 (q, *J* = 293 Hz), 65.8, 50.0, 30.0. ^{19}F NMR (376.3 MHz, D$_2$O) δ −70.61. HRMS (ESI-ORBI) m/z: [M + H]$^+$ calcd for C$_8$H$_9$F$_9$NO$_3$ expected mass, 338.04332, observed mass, 338.04178.

Fmoc-perfluoro-*tert*-butyl-homoserine (17). Compound **16** (1.30 g, 3.86 mmol) as a crude product from the prior reaction was dissolved in 20 mL H_2O and 20 mL 1,4-dioxane. Fmoc-OSu (1.69 g, 5.02 mmol) and K_2CO_3 (0.80 g, 5.79 mmol) were added, and the resultant solution was allowed to stir for 14 hours at room temperature. The reaction mixture was acidified to pH 1 and extracted with ethyl acetate (3 × 20 mL). The organic layers were combined and the solvent was removed under reduced pressure. The crude reaction mixture was dissolved in CH_2Cl_2 and purified via column chromatography (0 to 2% methanol in CH_2Cl_2). Compound **17** was obtained as a white solid in 48% yield. ^1H (600 MHz, CDCl$_3$) δ 7.76 (d, J = 7.6 Hz, 2H), 7.57 (dd, J = 7.0, 6.8 Hz, 2H), 7.40 (dd, J = 7.2, 7.2 Hz, 2H), 7.31 (dd, J = 7.4, 7.3 Hz 2H), 5.45 (d, J = 6.7 Hz, 1H), 4.51 (d, J = 6.0 Hz, 1H), 4.42 (d, J = 6.6 Hz, 2H), 4.22 (dd, J = 6.9, 6.8 Hz, 1H), 4.16 (s, 1H), 2.37-2.32 (m, 1H), 2.30-2.25 (m, 1H). ^{13}C NMR (150.8 MHz, CDCl$_3$) δ 143.5, 141.3, 127.7, 127.0, 125.0, 120.0, 67.2, 66.0, 50.8, 47.1, 31.4. ^{19}F NMR (376.3 MHz, CDCl$_3$) δ −70.29. HRMS (LIFDI-TOF) m/z: [M]$^+$ calcd for $C_{23}H_{18}F_9NO_5$ 559.1041, found 559.1053.

1.4.4   Synthesis of Perfluoro-*tert*-butyl Tyrosine

Fmoc-L-4-diazonium-phenylalanine tetrafluoroborate (18) Fmoc-L-4-amino-phenylalanine (0.306 g, 0.761 mmol) was dissolved in a 1:1 solution of tetrafluoroboric acid in water (0.2 mL, 48% to 50% w/w) and methanol (0.2 mL) in an open glass vial. The solution was placed in an ice bath before the dropwise addition of aqueous sodium nitrite (0.053 g, 76 mmol in 0.05 mL H_2O, 1.5 M) over two minutes. The reaction was allowed to warm to room temperature over the course of two hours. The crude product was diluted with water (1 mL) and extracted with ethyl acetate (2 × 1 mL). The crude product was dried under vacuum on a rotary evaporator to produce

an orange-red solid (0.483 g, 0.761 mmol) which was used without further purification in the next step. [1,2] *Warning*: diazonium salts are inherently prone to rapid generation of nitrogen gas. It is critical to use an open reaction vessel and low temperature to prevent accidental explosion. It is critical to immediately use the diazonium tetrafluoroborate to prevent complete drying and possible detonation.

Fmoc-L-perfluoro-*tert*-butyl-tyrosine (19) Crude compound **18** (0.483 g, 0.761 mmol) was dissolved in perfluoro-*tert*-butanol (5 mL) in a round-bottom flask with a reflux condenser open to the atmosphere, and the resultant solution was stirred at reflux (bp of perfluoro-*tert*-butanol: 45 °C) for 18 hours. The excess solvent was removed under vacuum. The crude product was dissolved in ethyl acetate (15 mL) and washed with 1 M HCl (15 mL) and water (15 mL). The ethyl acetate was removed under vacuum. The crude product was purified via column chromatography (0 to 2% methanol in CH_2Cl_2) to yield a yellowish solid (0.289 g, 0.465 mmol) in 61% yield.[2] ^{1}H NMR (400 MHz, CD$_3$OD) δ 7.81-7.79 (d, *J* = 7.6 Hz, 2H), 7.64-7.60 (dd, *J* = 7.1, 7.1 Hz, 2H), 7.41-7.38 (dd, *J* = 7.5 Hz, 7.4 Hz, 2H), 7.32-7.28 (m, 4H), 7.15-7.13 (d, *J* = 8.4, 2H), 4.45-4.41 (m, 1H), 4.30-4.27 (m, 2H), 4.18-4.16 (m, 2H), 3.03-2.97 (m, 1H), 2.99-2.93 (m, 1H) ^{13}C NMR (150.8 MHz, CD$_3$OD) δ 173.4, 156.9, 151.4, 143.9, 143.7, 141.2, 136.8, 130.4, 127.3, 126.7, 124.8, 122.1, 121.4, 119.5, 118.5, 66.5, 55.2, 36.4. ^{19}F NMR (376.3 MHz, CD$_3$OD) δ -70.41. HRMS (LIFDI-TOF) m/z: [M]$^+$ calcd for $C_{28}H_{20}F_9NO_5$ 621.1198, found 621.1211.

L-Perfluoro-*tert*-butyl-tyrosine (20) Fmoc-perfluoro-*tert*-butyl-tyrosine (0.100 g, 0.16 mmol) was dissolved in 20% piperidine in acetonitrile (10 mL) and stirred at room temperature for 30 min. The solvent was removed before redissolving in ethyl

acetate. The crude product was extracted from the solvent with water (1 × 5 mL) and 1 M HCl (1 × 5 mL). The solvent was removed under vacuum and dried to produce the product as off-white solid (0.042 g, 0.11 mmol) in 63% yield. ^{1}H NMR (400 MHz, CD$_3$OD) δ 7.35-7.32 (dd, J = 5.5 Hz, 5.2 Hz, 2H), 7.11-7.06 (dd, J = 8.7 Hz, 8.7 Hz, 2H), 3.78-3.75 (dd, J = 6.4 Hz, 6.1 Hz, 1H), 3.28-3.27 (d, J = 4.4 Hz, 1H), 3.06-3.00 (m, 1H). ^{13}C NMR (150.8 MHz, CD$_3$OD) δ 169.7, 163.7, 161.2, 131.1, 131.0, 130.8, 130.2, 123.5, 115.5, 115.3, 53.7, 35.0. ^{19}F NMR (376.3 MHz, CD$_3$OD) δ -70.27. HRMS (HESI) m/z: [M]$^{+}$ calcd for C$_{13}$H$_{11}$F$_9$NO$_3$ 400.05998, found 400.05897.

1.4.5   Peptide Synthesis and Purification

All peptides were synthesized using standard Fmoc solid phase peptide synthesis (SPPS) on Novagel Rink amide resin (EMD Millipore), using HBTU, HATU, or COMU as a coupling reagent. Perfluoro-*tert*-butyl hydroxyproline derivatives were coupled in four equivalent excess of amino acid and COMU for 36 hours. Perfluoro-*tert*-butyl homoserine and perfluoro-*tert*-butyl tyrosine were coupled in four equivalent excess of amino and HATU for 24 hours. All amino acids coupled after the perfluoro-*tert*-butyl hydroxyproline were double coupled (2 hours for first coupling, 1 hour for second coupling) using HATU. All amino acids prior to the perfluoro-*tert*-butyl amino acid derivatives were coupled in four equivalent excess for 2 hours. Peptides were synthesized by hand or on a Rainin PS3 (Rainin Instruments, Woburn, MA). All peptides were acetylated on the N-terminus (5% acetic anhydride in pyridine, 3 mL, 3 × 5 min) and contained a C-terminal amide.

Peptides were cleaved from the resin and deprotected for 3 hours using 1 mL reaction volume of 5% H_2O, 5% triisopropylsilane, and 90% TFA, except peptides containing perfluoro-*tert*-butyl tyrosine. Peptides containing perfluoro-*tert*-butyl tyrosine were deprotected for 90 min using 2 mL reaction volume of 2.5% H_2O, 2.5% thioanisole and 95% TFA. TFA was removed by evaporation and the peptides were dissolved in 1 mL 500 mM phosphate buffer (pH 7.2). Crude peptides were filtered using 0.45 µM syringe filters before injection on HPLC. Peptides were purified by reverse phase HPLC (Vydac semipreparative C18, 10 × 250 mm, 5 µm particle size, 300 Å pore) using a 60 minute linear gradient of 0 to 60% buffer B (20% H_2O, 80% acetonitrile, 0.05% TFA) in buffer A (98% H_2O, 2% acetonitrile, 0.06% TFA) except for Ac-T(4-NH_2-Phe)PN-NH_2 which was purified on a 60 minute linear gradient of 0 to 20% buffer B in buffer A. Peptides purity was verified by reinjection on a 60 minute linear gradient of 0 to 70% buffer A and buffer B. Analytical data for peptides: Ac-GPP(**(2S,4R)-Hyp(C₄F₉)**)PPGY-NH_2 [retention time by HPLC (t_R, 37.8 min; expected mass, 1056.4, observed mass, 1078.4 $(M + Na)^+$], Ac-GPP(**(2S,4S)-hyp(C₄F₉)**)PPGY-NH_2 [retention time by HPLC (t_R, 36.8 min; expected mass, 1056.4, observed mass, 1078.4 $(M + Na)^+$], **Ac-((2S,4R)-Hyp(C₄F₉))**KAAAAKAAAAKAAGY-NH_2 [retention time by HPLC (t_R, 41.9 min; expected mass, 1706.7, observed mass, 853.5 $(M)^{2+}$], **Ac-((2S,4R)-Hyp(C₄F₉))**KAAAAKAAAAKAAGY-NH_2 [retention time by HPLC (t_R, 40.8 min; expected mass, 1706.7, observed mass, 853.6 $(M)^{2+}$]; Ac-T(**Tyr(C₄F₉)**)PN-NH_2 [t_R, 46.7 min; expected mass, 752.5, observed mass, 753.5 $(M + H)^+$], Ac-T(4-NH_2-Phe)PN-NH_2 [t_R, 8.2 min; expected mass, 533.2, observed mass, 534.3 $(M + H^+)$].

1.4.6 NMR Spectroscopy of Peptides

Peptide NMR experiments were performed on a Brüker 600 MHz (^{19}F 564.5 MHz) NMR spectrometer equipped with a 5-mm Brüker SMART probe in 90% H_2O with 5 mM phosphate buffer (pH 4), 25 mM NaCl, and 10% D_2O. TSP was used for referencing in ^{1}H NMR. Residual TFA was used for referencing in ^{19}F NMR. All 1-D ^{1}H NMR spectra were collected with a Brüker w5 watergate pulse sequence with a 2 to 3 second relaxation delay. TOCSY spectra were collected using a watergate TOCSY pulse sequence. ^{19}F NMR data were collected on a Brüker 600 MHz (^{19}F 564.5 MHz) NMR spectrometer equipped with a 5-mm Brüker SMART probe without the use of proton decoupling. A 0.8 second acquisition time was used with a 200 ppm or an 8 ppm sweep width unless otherwise noted.

1.4.7 Circular Dichroism of Peptides

Circular dichroism (CD) experiments were performed in water with 5 mM phosphate buffer (pH 7.4) with 25 mM KF. All experiments were completed on a Jasco model J-810 spectropolarimeter. All data represents at least three background-corrected independent trials. All data were collected using a 0.1 cm cell (Starna Cells, Atascadero, CA). Ac-((**2S,4R**)-**Hyp(C4F9**))KAAAAKAAAAKAAGY-NH$_2$ peptide data was collected using 10 μM peptide. Ac-((**2S,4R**)-**Hyp(C4F9**))KAAAAKAAAAKAAGY-NH$_2$ peptide data were collected using 50 μM peptide. (Ac-(**Hse(C4F9**))KAAAAKAAAAKAAGY-NH$_2$ peptide data were collected using 25 μM peptide. Ac-GPP((**2S,4R**)-**Hyp(C4F9**))PPGY-NH$_2$ peptide data were collected using 20 μM peptide. Ac-GPP((**2S,4S**)-**hyp(C4F9**))PPGY-NH$_2$ peptide data were collected using 50 μM peptide. Ac-GPP(**Hse(C4F9**))PPGY-NH$_2$ peptide data were collected using 50 μM peptide.

Chapter 2

DETECTION OF PROTEIN-PROTEIN INTERACTIONS VIA 19F NMR UTLIZING PERFLUORO-TERT-BUTYL HOMOSERINE AND PERFLUORO-TERT-BUTYL TYROSINE

2.1 Introduction

Cells communicate internally as well as with other cells via complex signaling pathways. These pathways are composed of interactions between signaling molecules and proteins to propagate signals from one cell to another to control cellular activity. Protein-protein interactions are key components of signaling pathways used for cellular communication both inside the cell and with other cells, but are often the hardest to study due to the transient nature of these interactions. Protein-protein interactions are often short-lived, hydrophobic-driven interactions at protein-protein interfaces, with additional contributions and specificity from electrostatic interactions. Aliphatic amino acids, such as Leu and Ile, and aromatics, such as Tyr and Trp, are often utilized at protein-protein interfaces. A large number of protein-protein interactions have been identified as important interactions in cell signaling pathways involved in disease, such as the estrogen receptor/coactivator binding interaction to increase transcription and the MDM2•p53 interaction, which are both important in some cancer pathways. There are also many unidentified protein-protein interactions,

as they are difficult to investigate due to the fact that they are easily reversible and easily disrupted. New tools are needed to identify protein-protein interactions as well as potential theranostic agents for cancer detection, monitoring, and treatment.

2.1.1 Protein-Protein Interactions Are High Profile Targets for Therapeutics

Protein-protein interactions have long held interest as drug targets for the pharmaceutical industry.[101] Nearly every cell signaling pathway has critical protein-protein interactions, which means nearly every misregulation event in cellular pathways involved in disease states are potential targets. Protein-protein interactions are, however, much more difficult to target than protein-small molecule interactions, such as hormones binding to nuclear hormone receptors.[95] Protein-protein interactions are more difficult to target because a much larger surface area is involved in the interaction and increased flexibility associated with increased degrees of freedom with peptides/proteins.[101-103] Many protein-protein interactions utilize intrinsically disordered proteins (IDPs) at protein-protein interfaces. IDPs are inherently conformationally dynamic and often undergo structural changes as a result of binding at a protein-protein interface.[104] These interactions are generally modest in affinity, in part due to the entropic cost of transitioning from an unfolded to a folded state of the IDP. Many of these interactions exhibit binding affinities in the high nanomolar to mid micromolar regime, leading to very short half-lives, on the order of milliseconds to seconds. Furthermore, there are a large number of undiscovered protein-protein interactions.[28] There are currently efforts to characterize the so-called "interactome" which will potentially uncover a large number of previously undiscovered protein-protein interactions which are key in disease pathways.

Despite the challenges, there have been several breakthroughs in targeting protein-protein interactions.[101] Many of these breakthroughs are involved in cancer signaling pathways, especially in apoptosis pathways whose disruption is a hallmark of cancer cells. Normal cells, upon irreversible damage will undergo apoptosis, also known as programmed cell death, using a series of complex pathways. Cancer cells pathways show increased expression of both anti- and pro-apoptosis proteins, which makes these specific proteins targets in specifically disrupting cancer cell pathways. One example of this is the Bcl-2 family of proteins, which are natively involved in apoptosis pathways.[105] The Bcl-2 family comprises a series of proteins all containing the Bcl-2 homology domains, also known as BH domains, which interact with other proteins leading to cell survival or apoptosis. There are four known BH domains, known as BH1, BH2, BH3, and BH4.[106] Bcl-2, Mcl-2, and A-1 are all similar proteins in this family that lead to cell survival.[107] While the exact mechanisms for these pathways is incompletely understood, mutagenesis studies have shown disruption of the protein-protein interface leads to defective cell signaling pathways.[105] Pro-apoptotic members of the Bcl-2 family contain one more BH domains. Anti-apoptotic members always contain a BH3 domain, but some anti-apoptotic also contain a BH3 domain.[106] Pro-apoptotic members of the Bcl-2 family include Bim, Bid, and Bad, which are all known to lead to apoptosis via a protein-protein interaction with the anti-apoptotic members of the Bcl-2 family. The BH3 domain of the anti-apoptotic Bcl-2 family members interacts other BH domains via protein-protein interactions with other BH domains to form heterodimers, preventing apoptotsis. These proteins are overexpressed, binding to proteins like Bcl-2, ultimately activating programmed cell death.[105] These proteins use α-helical recognition sequences which are approximately

20 residues long.[107] Furthermore, these proteins are often upregulated in cancers which are chemotherapy resistant, making them more attractive targets.[101] There are a number of small molecules which are being studied target the Bcl-2 family of proteins via protein-protein interactions as potential therapeutics.[108] These proteins act by binding to the anti-apoptotic proteins, such as Bcl-2, deactivating the protein, promoting apoptosis pathways. One example of these small molecules is HA14-1 which induces apoptosis in human acute myeloid leukemia cells, which are known to overexpress Bcl-2.[109]

2.1.2   Fluorine Has the Potential to Enhance Protein-Protein Interactions

As increasingly more protein-protein interactions are being targeted for potential therapeutics, one can imagine the use of fluorine in small molecules, peptides, and proteins/antibodies for both the detection and treatment of disease such as cancer. As discussed in Chapter 1, the addition of fluorine has been demonstrated to increase the hydrophobicity of molecules.[29] Many protein-protein interfaces utilize hydrophobic interactions between aromatic and aliphatic amino acids. The addition of fluorine atoms in place of hydrogens thus has the potential to increase favorable interactions via the enhanced hydrophobic effects of fluorine. As previously discussed, fluorine can also increase the stability of proteins via enhanced hydrophobic effects, leading to more stable protein-based therapeutics.[13] Furthermore, the addition of fluorine atoms would allow for imaging of therapeutic agents via ^{19}F magnetic resonance. Thus, if fluorine is introduced into an antibody, the antibody could be traced to its target cells by MRI.

2.1.3 Methods to Detect Protein-Protein Interactions

There are a number of techniques currently used to identify protein-protein interactions. Co-immunoprecipitation (CoIP) and pull-down assays can be used to identify stronger protein-protein interactions against a protein in complex mixtures.[110] In CoIP, a protein of interest is identified and an antibody is developed or applied. The antibody can either be fixed to a solid support or free in solution. The antibody is incubated with the cell lysate, in which the targeted protein binds to other proteins of interest. All of the other cellular components and proteins are removed leaving only the antibody, the protein of interest, and any proteins bound to it (Figure 2.1). Pull-down assays are similar, but a known binding protein is used in place of an antibody. In many cases, however, these assays are unsuccessful in identifying weak or transient interactions due to rapid dissociation of the complex.
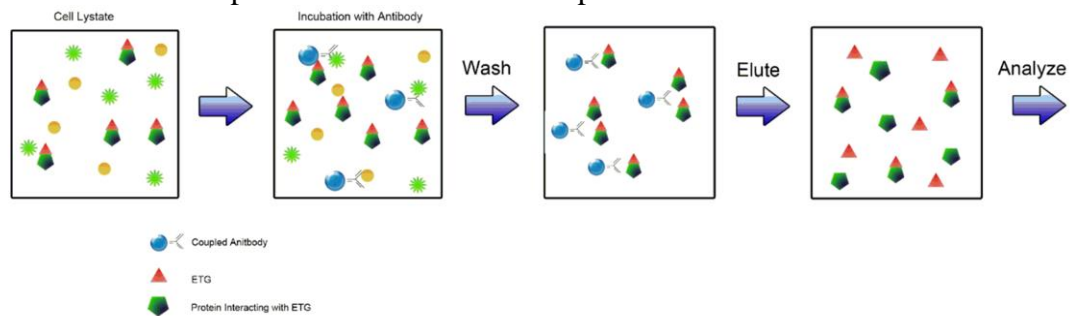
Figure 2.1   A schematic representation of Co-IP assays utilizing antibody technology to isolate interacting proteins. In short, cell lysates are incubated with antibodies which bind to a protein of interest. Unbound proteins and cell debris is washed away, leaving only the protein bound to the antibody and any binding partners. These proteins can be analyzed using Western blot and mass spectrometry.

Weaker interactions can be identified using cross-linking assays and affinity techniques. Cross-linking requires the use of a linker, such as a disulfide bond, which

will link when two proteins come together at an interface.[110] When the protein of interest and its binding partner interact, the linker will attach to the binding partner, covalently linking the two proteins together. Other linkers include photo-inducible linkers and linkers which can pass through a membrane, which allows for cross-linking to occur in living cells. There are multiple groups which can be used in cross linking, including but not limited to primary amines, sulfhydryls, and carbonyls. Many of these functional groups are found on amino acid side chains which can be utilized for crosslinking.[111] In the case of *in vivo* cross linking, however, the cells must be lysed for analysis to isolate cross-linked protein complexes. While both *in vivo* and *in vitro* cross linking are effective methods, proteins of interest must be altered with a cross linker.[110,111] Furthermore, purification methods are still required once the protein of interest is cross-linked with a binding partner.

Affinity techniques utilize a protein of interest fixed to a solid support, with which cell lysates can be incubated to identify potential binding partners. The main issue with this technique is that proteins that bind the protein of interest may or may not be biologically relevant binding partners. While both of these methods have proven effective, there is also a limit to their usefulness.

Another useful technique for the identification of protein-protein interactions is phage display, which was first described in 1985.[110] Phase display incorporates a gene of interest into a bacteriophage fusion protein, which is expressed and displayed on the surface of the membrane of the bacteriophage. The protein of interest is expressed on the surface of a bacteriophage and fixed to a plate. Proteins can then be exposed the expressed protein of interest. Large libraries of proteins can be screened quickly and easily using phage display and is a common technique used to identify potential drug

molecules.[112] Much like the other techniques described, phage display has proven to be an effective technique but requires destruction of the cells of interest. Phage display cannot be used to detect protein-protein interactions in real time. Furthermore, none of these techniques can be used for pre-clinical or clinical imaging.

Finally, a widely-used paradigm in the discovery and characterization of protein-protein interactions is the use of yeast two-hybrid (Y2H) systems.[112] This system utilizes genetically modified yeast which are transformed with a protein of interest (termed the "bait" protein) fused with a DNA-binding protein. A library of "prey" proteins, or potential binding partners, are transformed into a separate plasmid which are fused to a protien activation domain. When the "bait" protein and the "prey" protein come in close proximity, they form a protein-protein interaction, and transcription is activated (Figure 2.2). Common reporter genes for this method include the *lac* repressor system coupled with fluorescent proteins like GFP to produce a fluorescent readout.[113] This method has led to over 60,000 publications as of 2016.[111] Although improvements have been made in the last ten years, false positives continue to be an issue with Y2H systems.[112] Furthermore, Y2H systems do not account for the stoichiometry of protein-protein interactions, which is not always one to one. All current technologies, while commonly used, each come with unique advantages and unique challenges which can potentially be addressed with new techniques.
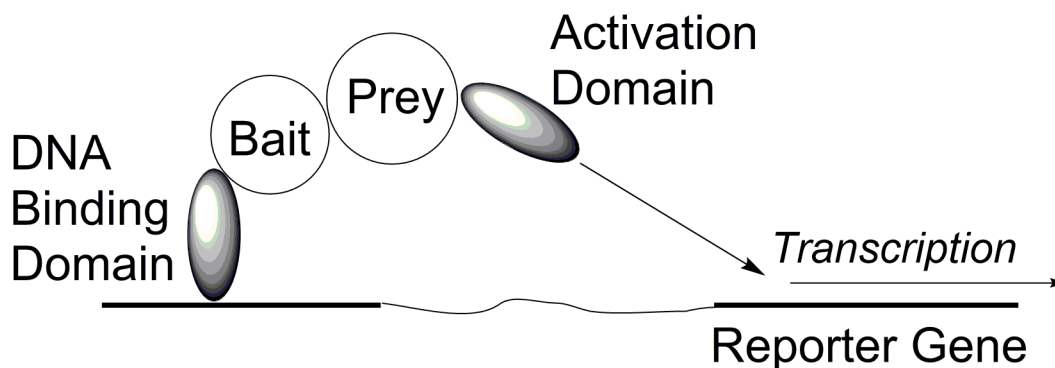
Figure 2.2 A schematic representation of yeast two-hybrid assays to detect protein-protein interactions. A bait protein fused to a DNA binding domain interacts with a prey protein fused to an activation domain via a protein-protein interaction. The activation domain turns on transcription of a reporter gene leading to a fluorescent or colorimetric readout.[113]

2.1.4   Methods of Detection of Protein-Protein Interactions by ^{19}F NMR

There are multiple types of signals which can be obtained by ^{19}F NMR to detect changes in chemical environment. As previously stated in Chapter 1, ^{19}F NMR is very sensitive to changes in chemical environment, due to the nature of the fluorine atom and the large (>200 ppm) chemical shift range. The large chemical shift range allows for detection of changes in the NMR spectrum due to changes in temperature, pH, solvent, etc.[5] When placed into a biological context, there are a number of different potential readouts to gain information about the system.

The first potential readout by ^{19}F magnetic resonance is a change in relaxation rate (T_1 and T_2), as is observed with ^{1}H magnetic resonance as used for MRI.[114] In magnetic resonance experiments, these parameters describe how the signal changes over time. T_1 is the longitudinal relaxation time, which is the amount of time required for the signal to decay over time. When the magnetic field is applied to a system for magnetic resonance, proton nuclei are excited by the radio frequency. The T_1 is a

measurement of how long it takes for the protons in the system to go from the excited state back to equilibrium. The T_1 is the parameter for proton MRI that is changed by the addition of contrast agents, with drastically different T_1 in liquids (around 1500 to 2000 ms) compared to tissue (water-based tissues around 400 to 1200 ms, fat-based tissues 100 to 150 ms).[114] T_1-weighed MRI is used to image anatomical structures.

The second method used to detect changes in magnetic resonance images is T_2. T_2 is the transverse relaxation, which measures how long it takes the excited nuclei to reach equilibrium, or fall out of phase, after a pulse in the transverse direction, which is generally less dependent on the magnetic field than T_1. T_2-weighted images are used to identify pathologies such as lesions. ^{19}F magnetic resonance can also exhibit changes in relaxation as a result in changes in the chemical environment.[115] There are a number of examples demonstrating gadolinium contrast agents can also be used to drastically decrease the relaxation times of nearby fluorine nuclei due to the unpaired electrons of gadolunium.[15] Large, fluorinated polymers have also shown relaxation delay changes for imaging tissue *in vivo*.[116] These probes have the potential to be used in current MRI technology in the clinic with increased sensitivity.

^{19}F chemical shift is also highly dependent on chemical environment. Due to the high prevalence protons within biological systems, it is nearly impossible to look at a single proton signal. Furthermore, due to the relatively small ^{1}H chemical shift range (approximately 15 ppm), significant spectral overlap is seen in complex solutions, preventing observations of small chemical shift changes as a result of changes in the chemical environment. Fluorine chemical shift, however, is documented to respond to small changes in environment with changes in the chemical shift.[38,40] Chemical shift changes are a direct readout of small changes in the chemical

environment of the fluorine atoms. These small chemical shift changes can be used to monitor the chemical environment in complex mixtures such as cell lysates and living cells. One potential application of this approach is changes in the redox environment of a cell. Many cancer cells exist in a state of oxidative stress due to reactive oxygen species (ROS) as a result of deregulation of cellular pathways. The increase of ROS leads to an overall change in the oxidation state of cells.[117] This change could potentially be monitored by ^{19}F containing molecules by chemical shift changes as a result of a change in redox potential of cancer cells compared to healthy cells. These types of applications are ideal for imaging in a pre-clinical and clinical setting. Fluorinated molecules can be injected to the potential site of a tumor and based on small chemical shift changes due to the redox potential of the cells, a tumor could be identified. This application is just one example of the potential use of ^{19}F chemical shift changes to monitor cellular activity. This topic will be discussed further in Chapter 3.

Another method of detection by ^{19}F magnetic resonance is chemical shift anisotropy (CSA).[118,119] Fluorine atoms which can freely rotate in solution and relaxation rates are not affected due to unrestricted movement. Fluorine atoms found within bones and teeth are bound in a matrix which prevents free rotation, shortening the relaxation time dramatically, preventing a signal by ^{19}F NMR/MRI.[2] Without free rotation, signals will broaden due to shortened T_2 relaxation. The same phenomenon can be see when a fluorine atom is bound within a protein structure. Larger molecules, such as proteins, rotate more slowly through space, leading to shorter T_2 relaxation times and, in turn, increased chemical shift anisotropy. Similarly, fluorines which are interfaced with proteins in a hydrophobic pocket, will prevent free rotation, causing

the signal to broaden.[118] The amount of broadening can be quantified by measuring the full width at half height (FWHH) of the peak by ^{19}F NMR. In the case of protein-protein interactions when a fluorine-containing peptide is free in solution, the fluorine atoms can rotate freely, producing a single sharp peak. As the peptide binds to a protein interface, the fluorine atoms are buried at the interface, causing the peaks to broaden. This phenomenon has been used to study the kinetics of protein folding by ^{19}F NMR. Fluorinated amino acids, such as 4-F-Phe, have been introduced into the ligand binding domain of the chemotaxis aspartate receptor in order to study dimerization of this protein by ^{19}F NMR.[120] In short, the 4-F-Phe was introduced into the binding site where it would interface upon dimerization. Dimerization in this system led to the broadening of fluorine signals associated with the 4-F-Phe residue.

Finally, when fluorine is used in imaging applications, ^{19}F signals can be used as tracers.[2] This is one of the most common applications for ^{19}F MRI to date. Expressed proteins containing fluorine atoms or fluorinated small molecules which bind to specific targets can be injected into animal models and imaged as discussed in Chapter 1.[6,81,121,122] As fluorinated molecules collect in the targeted area, a strong signal is produced. This spectrum can be superimposed with traditional ^{1}H MRI images to gain both cell-specific and anatomical information.[81]

2.1.5   Secondary Structure is Often Critical For Protein-Protein Interactions

Protein-protein interactions often interface through helical interfaces such as polyproline II helices and α-helices.[102,104,123,124] Some notable protein-protein interactions include leucine zippers, SH3 domains, nuclear hormone receptors, and E3 ubiquitin ligases.[48,50,84,95,97,125] Many of these interactions can be modified by the presence and absence of post-translational modifications.[102] Protein-protein

interactions are utilized by biological systems because they are rapidly reversible, due to the fact that they are noncovalent interactions, which allows for transient interactions. Protein-protein interactions often take place on the timescale of seconds, which makes them very short-lived.[102,104,123] These short-lived interactions lead to limited duration of signaling as a result of protein-protein interactions. In short, two proteins interface for a matter of seconds, and then dissociated, having sent a short-lived signal activating pathways for a very short period.

In the case of SH3 domains, which are found in a number of notable protein kinases including Abl and Src tyrosine kinases, utilize protein-protein interactions via a polyproline II helix. Both of these kinases contain an SH3 (Src-homology 3) domain, an SH2 domain (Src-homology 2) domain, and a kinase domain. Both the SH3 and the SH2 domains are regulatory domains that control kinase activity.[126] In the Src protein kinase, upon phosphorylation, the SH3 domain is interfaced with the kinase domain due to interaction of the SH2 domain with the kinase, blocking the SH3 domain-mediated protein-protein interaction interface, inactivating the kinase.[127] Upon dephosphorylation, the SH3 domain is exposed allowing for protein-protein interactions with polyproline II helices. The protein-protein interface binds via a PXXP motif, where X is any amino acid.[127] This interaction is key in both Abl and Src kinase activities, which control cell proliferation.[124,128] Src is just one example of an enzyme which uses SH3 domains and binds through polyproline II helix structures.[84,128,129] Over 300 known proteins in the human genome utilize SH3 domains, many of which are involved in regulation of cellular pathways including those involved in cancer, diabetes, and neurological disorders making SH3 domains high profile targets via the protein-protein interface.[128]

Similarly, a number of protein-protein interactions interface through α-helices. A few notable examples include the nuclear hormone receptors such the estrogen receptor and the androgen receptor.[95,97,130] Both receptors are known to control cell proliferation and growth in cancer cells via co-activator protein interactions as an α-helix. These interactions require the recognition sequence to align on one face of the helix to form favorable hydrophobic contacts with the interface, utilizing hydrophobic residues such as Leu, Ile, Tyr, and Phe.[102,131] The amino acids which can be accommodated are dependent on the size of the binding pocket. In the case of the estrogen receptor, aromatic amino acids are too large to be accommodated, favoring Leu and Ile in the binding pocket.[132,133] The androgen receptor, however, has a larger binding pocket, preferring the use of aromatic amino acids in the FXXLF binding motif.[134]

2.1.6 The Estrogen Receptor Promotes Transcription In Cancer Pathways via a Protein-Protein Interaction with an α-Helical Coactivator

The estrogen receptor (ER) is a nuclear hormone receptor which controls cell proliferation via transcriptional activation (Figure 2.3).[135] When estrogen binds to the estrogen receptor via the ligand-binding domain, the receptor dimerizes and enters the nucleus of the cell. Inside the nucleus, the receptor can bind to the estrogen responsive elements (ERE) via the DNA binding domain on the DNA inside the nucleus. Once bound to the DNA, the co-activator binding pocket is exposed via a change in protein structure. The co-activator binding is a shallow hydrophobic pocket, separate from the ligand binding pocket, which binds through an LXXLL motif.[130] Co-activator proteins are associated with the recruitment of the transcriptional machinery, which upregulates transcription (Figure 2.3).[135,136] Some of the proteins which can bind to the estrogen

111

receptor via a co-activator and upregulate transcription include the steroid receptor co-activator 1 (SRC-1), the glucocorticoid receptor interacting protein 1 (GRIP1), and steroid receptor RNA activator (SRA).[137] These proteins all bind through the same hydrophobic interface to upregulate transcription. In estrogen receptor-positive breast cancer, this pathway is upregulated, leading to uncontrolled cell growth and proliferation.[135]
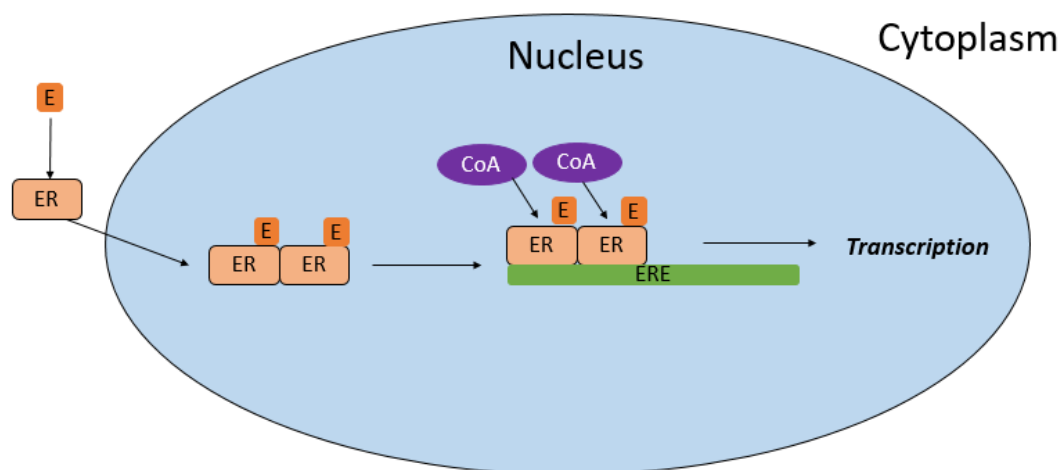
Figure 2.3   A schematic representation of estrogen receptor signaling in cells, demonstrating upregulation of transcription as a result of estrogen (E) binding to the estrogen receptor (ER). Upon dimerization, the estrogen receptor can bind to the estrogen responsive elements (ERE) on the DNA. The co-activator (CoA) can bind in the co-activator binding pocket which allows for the recruitment of the transcription machinery, leading to an increase in transcription inside the nucleus of the cell.

Traditional treatments for breast cancer include the use of estrogen antagonists, which bind to the estrogen receptor in place of estrogens. These drugs are known as

selective estrogen receptor modulators (SERMs). One commonly prescribed SERM is tamoxifen (Figure 2.5).[138] SERMs bind in the ligand-binding pocket, causing structural changes within the ligand-binding domain, preventing estrogen-dependent transcription. Binding of the SERM, as the metabolically active form 4-hydroxytamoxifen, causes a shift in helix 12 of the estrogen receptor, effectively blocking the co-activator binding pocket (Figure 2.4).[133] While SERMS are effective in treating breast cancer, there are a number of issues with the use of drugs like tamoxifen. One of the biggest problems is that some cancers may become tamoxifen resistant; cancers still express the estrogen receptor, but it is often mutated in such a way that prevents tamoxifen from binding to the estrogen receptor.[139] In some cases, mutations can also lead to tamoxifen acting as an estrogen, leading to the upregulation of transcription as a result of tamoxifen binding, resulting in tamoxifen functioning as an ER agonist, rather than an ER antagonist.[138,139] We would like to directly target the estrogen receptor•co-activator interaction as a potential imaging and therapeutic site that can function in a different manner than current traditional methods, such as SERMs. The co-activator binding pocket has great potential as a target for alternative therapies for the treatment of breast cancer and SERM-resistant breast cancer. In addition, demonstration of inhibition at this site would demonstrate the potential of peptides containing perfluoro-*tert*-butyl ethers could be utilized in an α-helical context to target protein-protein interactions to target nuclear hormone receptors.
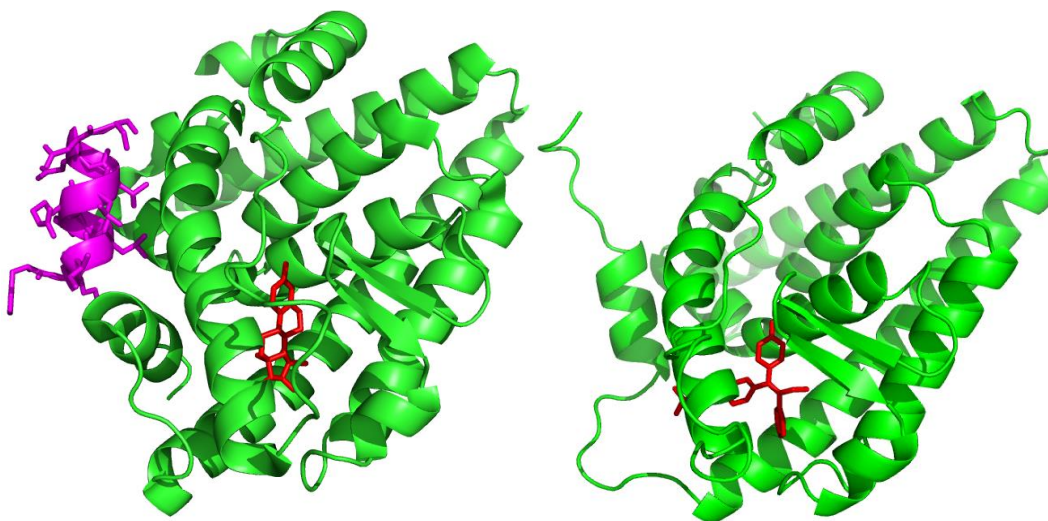
Figure 2.4   The estrogen receptor ligand-binding domain highlighting the two separate binding sites for the ligand (red) and a model co-activator peptide (purple) binding through the LXXLL binding motif (left) and the estrogen receptor ligand binding domain bound to 4-hydroxytamoxifen, preventing co-activator binding (right).
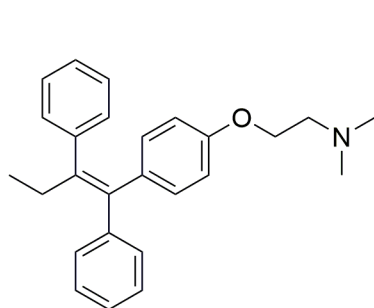
Tamoxifen
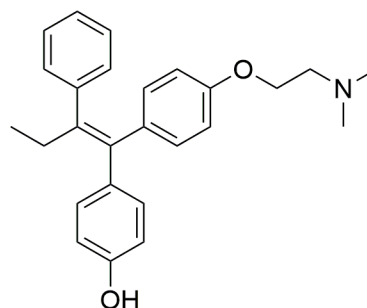
4-Hydroxytamoxifen

Figure 2.5   The structures of tamoxifen, a commonly prescribed anti-estrogen for the treatment of estrogen receptor-positive breast cancer, and 4-hydroxytamoxifen, the metabolic product and biologically active form of tamoxifen.

2.1.7 MDM2 Binds to p53 in Oncogenic Pathways

Another example of an important biologically active α-helical protein-protein interaction involves MDM2, a E3 ubiquitin-protein ligase. MDM2 is a proto-oncoprotein which is involved in the negative feedback loop for p53. p53 is a regulatory transcription factor which controls genes critical to DNA repair pathways.[140] When cells undergo DNA damage, p53 is activated and can bind to p53-responsive elements, which upregulates adjacent genes on the DNA repair pathways or apoptosis pathways. One of the genes activated upon p53-dependent transcription is MDM2. MDM2, in turn, binds to p53, leading to ubiquitination and destruction of p53 (Figure 2.6).[141,142] Furthermore, expression of p53 directly results in the upregulation of MDM2, exhibiting a tightly regulated negative feedback loop. In more than 50% of cancers, mutations of p53 are seen which can lead to a reduction in the rate of transcription rather than DNA repair and apoptosis pathways, via mechanisms including the upregulation of MDM2.[141]
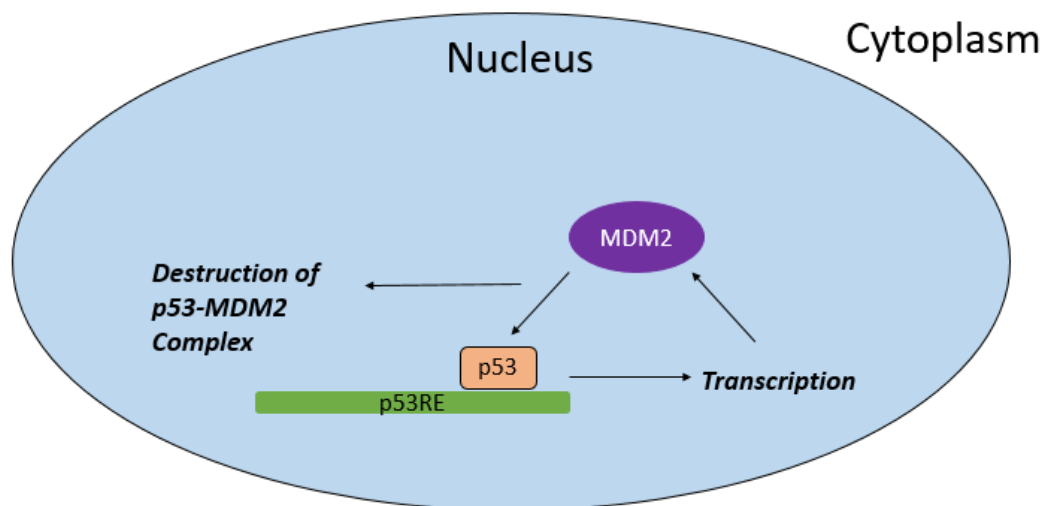
Figure 2.6   A schematic representation of p53•MDM2 negative feedback loop. Upon
cell damage, p53 is activated, binding to p53 responsive elements
(p53RE) leading to an upregulation in transcription of adjacent genes
including MDM2. MDM2 binds to p53, leading to ubiquitination
destruction of p53.

Similar to the nuclear hormone receptors, p53 binds to MDM2 via an α-helical

binding motif FXXΦWXXΦ (Figure 2.7), and this interaction is key in controlling

p53 activity.[143,144] In cancerous cells, p53 is often mutated, leading to loss of function.

Furthermore, in 5 to 10% of cancers, MDM2 is overexpressed, leading to

downregulation of tumor suppressor pathways.[142] Due to the importance of

p53•MDM2 interactions, it has become a well-studied model system of protein-

protein interactions in cancer pathways as well as an important therapeutic target for

the treatment of cancer.
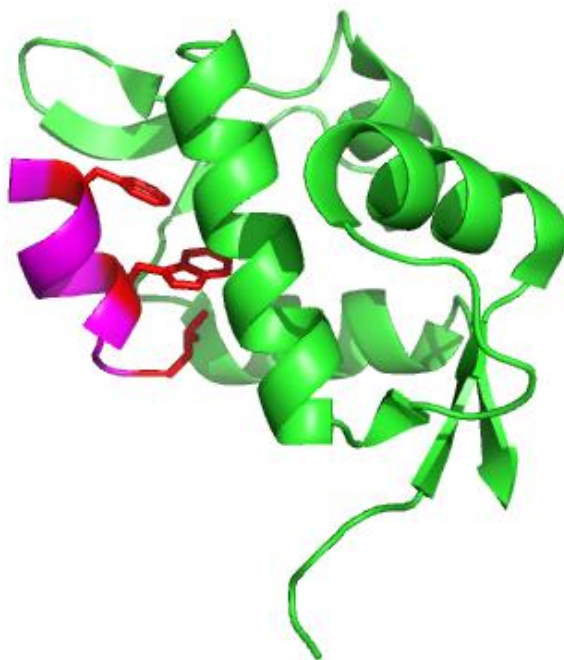
Figure 2.7   The MDM2 crystal structure with a small α-helical peptide bound
              containing the FXXLW.

2.1.8   Small Molecule and Peptide Targeting of α-Helical Protein-Protein Interactions

There are a number of examples in the literature utilizing both small molecules
and synthetic peptides to target protein-protein interactions that form via α-helical
interfaces like p53•MDM2 and the estrogen receptor•co-activator interaction. The key
is to mimic the orientation of the side chains of an α-helix. The orientation of an α-
helix aligns residues at the *i*, *i*+3, and *i*+4 positions along the protein-protein interface
via hydrogen bonding between the *i* CO and the *i*+4 NH leading to the characteristic
3.6 residues per turn and the 5.5 Å pitch (Figure 2.8).[145] In the case of the estrogen
receptor co-activators, the recognition motif commonly incorporates Leu at all of these
positions.[133] In the case of p53, Phe is utilized at the *i* position and Trp is utilized at

the *i*+4 position for recognition.[143] The key to successfully designing either small molecule or peptide-based α-helical mimics is to effectively mimic the geometry and the side chain orientation of the binding protein. There has been a significant effort to target both the p53•MDM2 interaction and the estrogen receptor•coactivator interaction using both small molecules and peptides.

i   *i+4*

LXXLL

i+3

i

i+3

i+4
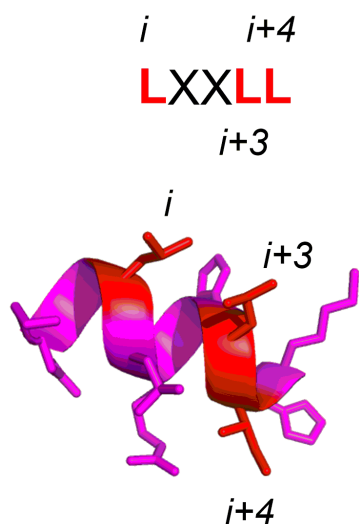
Figure 2.8   Critical binding residues (shown in red) align along one face of the α-helix at the *i*, *i*+3, and *i*+4 positions.

There are a few examples of small molecule inhibitors for the estrogen receptor•co-activator utilizing aromatic scaffolds to attempt to mimic the α-helical geometry (Figure 2.9). One notable example is the development of pyrimidine-based ligands.[137] These molecules are less successful at targeting the estrogen receptor due to the lack of three-dimensional structure, preventing mimicking of α-helical geometry (Figure 2.9). A more successful approach from the Hamilton lab is the biphenyl

scaffold which targeted the estrogen receptor with $K_i = 4.2$ μM.[146] Overall, these compounds suffer from high hydrophobicity, which can lead to non-specific binding. Both of these classes of scaffolds were only assayed against the estrogen receptor, so there is no evidence for the specificity of this interaction.

1   **2**   **3**
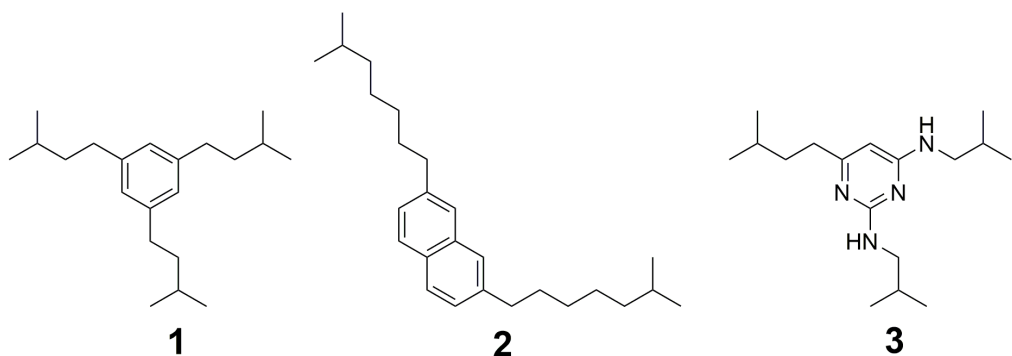
Benzene and Pyrimidine-Based Estrogen Receptor Binding Compounds
K_d = 30 μM (Compound 3)
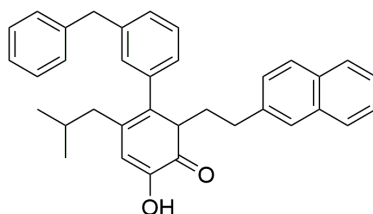
Biphenyl-Based Estrogen Receptor Binding Compounds
K_d = 4.2 μM

Figure 2.9   Small molecule α-helical mimetic compounds targeting the ER•coactivator interaction.[137,146]

There has been a substantial effort to identify small molecules which bind to MDM2, blocking the p53 binding site, with varying degrees of success.[147-151] Similar to the estrogen receptor, a few designed scaffolds have been developed including chalcone and terphenyl, which range from mid-micromolar to high nanomolar affinities (Figure 2.10).[149,151] There are also a large number of drugs currently in development by pharmaceutical industry currently in clinical trials. One of the most promising class of compounds currently being explored for their disruption of the p53•MDM2 interaction are the Nutlins (2.11).[147] These compounds mimic the short α-helical segment of p53 binding to MDM2 via substituted aromatic rings. Furthermore, these compounds bind with low to mid-nanomolar affinity to MDM2. As of 2014, at least seven different small molecules were in clinical trials for use as anti-cancer agents targeting MDM2.[148]
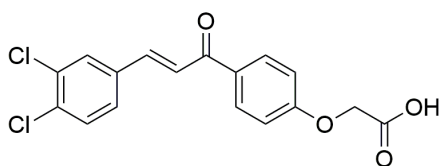
Chalcone Scaffold
K_d = 150 μM
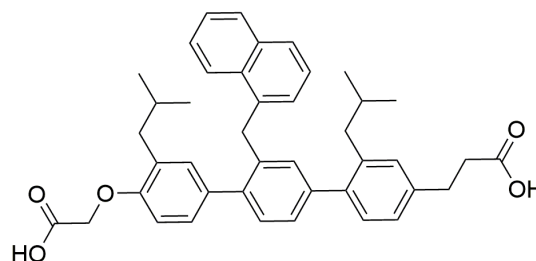
Terphenyl Scaffold
K_i = 0.182 μM

Figure 2.10 Designed α-helical scaffolds to target p53•MDM2 interaction.

Nutlin-1
IC_{50} = 260 nM

Nutlin-2
IC_{50} = 140 nM
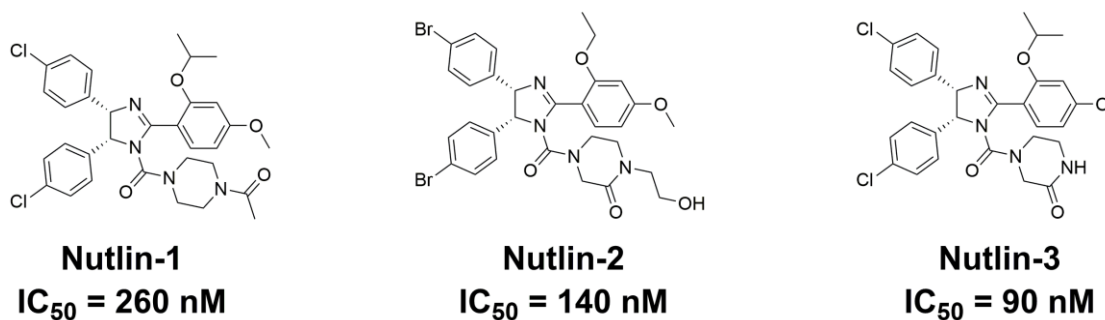
Nutlin-3
IC_{50} = 90 nM

Figure 2.11   The original nutlin compounds discovered in 2004 which have led to a series compounds to target p53•MDM2 which are currently undergoing clinical trials for use as anti-cancer agents.

Peptides have also been used to target the estrogen receptor•co-activator interaction. The estrogen receptor has multiple natural co-activator ligands which could be used to bind to the protein, preventing binding of the natural co-activator proteins *in vivo*. Some of these co-activator sequences have been made as small, synthetic peptides including the SRC Box, NRBox, and TiF-2 Box peptides.[50,125] These all incorporate the relevant LXXLL binding motif in a small peptide. These peptides bind with low micromolar affinity, anywhere from 1 µM in the case of the NRBoxII peptide to 10 µM for the Tif-2 BoxII (Figure 2.12).[50,125] These peptides can be introduced into cells via cell penetrating peptides (CPPs) and are shown to localize in the nucleus of the cell, binding to the estrogen receptor. Another example of synthetic peptides used to target the estrogen receptor are the cyclic peptides developed by Galande *et al*., who developed a series of cyclic peptides incorporating unnatural amino acids to obtain low nanomolar binding. These sequences were based on the PERM-1 sequence and utilized a Cys disulfide bridge to lock the cyclic peptides into an α-helix, obtaining a K_i = 25 nM binding peptide using simply the native LXXLL motif (Figure 2.12).[146] These peptides went a step further,

incorporating a neopentylglycine derivative at the critical Leu positions within the LXXLL binding motif. The neopentylglycine led to increased binding at all three positions over the native PERM-1 sequence; however, the i+3 position within the helix was significantly superior to the others, with $K_i = 7$ pM.[146] These data point to the potential power of small synthetic peptides utilizing unnatural amino acids to target protein-protein interactions.

SRC-1 Box1   `YSQTSHKLVQLLTTTAQQ`
NRBoxII   `LTERHKLLHRLLQE`
Tif-2 Box2   `LKEKHKILHRLLQDSSPV`
PERM1   `K-cyclo(D-CILC)ALLQ`

Figure 2.12   Synthetic peptide sequences based on natural estrogen receptor co-activator sequences. Key Leu resides are underlined.

      MDM2 can also be targeted by small peptides based on the native p53 sequence. The p53-based peptide, utilizing the sequence shown in Figure 2.13, binds with $K_d = 121$ nM.[144] Zondlo e*t al.* examined a series of chimeric peptides to understand the differential binding of p53 and the related p65 transcription factor, which binds with much lower affinity at $K_d = 5.7$ μM. A series of peptides was synthesized, substituting key residues from the p65 binding motif into the p53 peptides to understand how each residue was contributing to the overall binding of the peptide. In the course of this work, one peptide known as p53-P27S was demonstrated to bind with $K_d = 4.7$ nM, which is the highest affinity of any encodable peptide-based ligand.[144] Much like the

estrogen receptor, these data demonstrate that high affinity binding in protein-protein interactions can be obtained using small synthetic peptides.

p53	**CPPLSQETF̲SDL̲W̲KLLPENN**
p65	**CSSIADMDF̲SAL̲L̲QISS**
D21A	**CPPLSQETFS̲AL̲W̲KLLPENN**
K24S	**CPPLSQETF̲SDL̲W̲SLLPENN**
L26I	**CPPLSQETF̲SDL̲W̲KILPENN**
P27S	**CPPLSQETF̲SDL̲W̲KLLSENN**

Figure 2.13 MDM2-binding peptides based on the native p53 sequence, with mutations from the lower affinity p65 sequence. These peptides were examined to determine the site-specific residues which lead to difference in binding affinity of these two native ligands. The FXXΦΦ binding motif is underlined and the mutated residues are shown in red. P27S is the highest affinity encodable peptide-based ligand indentified to date.[144]

While a number of both small molecule and peptide-based examples have been provided here, there is still room for improvement targeting α-helical protein-protein interactions. High binding affinity has been demonstrated in both small molecules and peptides, but we would like to develop a series of peptides which have the potential to not only act as high affinity ligands, and possibly potent therapeutics, but also function as potential imagining agents that would allow imaging of these interactions *in real time* to gain a greater understanding of these protein-protein interactions and their impact in cell signaling pathways.

2.2   Results

2.2.1   Design of MDM2 Binding Peptides Containing Perfluoro-*tert*-butyl Tyrosine

Based on the successful design of the p53- P27S peptide, in order to target the MDM2•p53 interaction utilizing the perfluro-*tert*-butyl group, a series of peptides was designed incorporating perfluoro-*tert*-butyl tyrosine into the FXXΦΦ recognition motif. In the P27S peptide, the critical recognition motif is **FSDLW** (Figure 2.14). Based on crystal structure analysis, we hypothesized there was sufficient space to accommodate the perfluoro-*tert*-butyl tyrosine at both the Phe (*i* position within the helix) and the Trp (*i*+4 position within the helix) residues. As a first approximation, perfluoro-*tert*-butyl tyrosine is only slightly larger than Trp and could potentially mimic any of the aromatic amino acids in a hydrophobic binding motif (Figure 2.15). These peptides could potentially show increased binding affinity due to the increased hydrophobicity of the perfluoro-*tert*-butyl group. It is also possible that the modification to the perfluoro-*tert*-butyl tyrosine will result in a loss of binding affinity. In addition, the P27S peptide has a much higher affinity (K_d = 4.7 nM) over the p53$_{7-36}$ peptide (K_d = 121 nM), such that a loss of binding as a result of the perfluoro-*tert*-butyl modification could still potentially yield a useful ^{19}F NMR/MRI probe for examining the MDM2•p53 interaction.

P27S   Ac-**C**PPLSQETFSDLWKLLSENN-NH$_2$
F19Tyr(C$_4$F$_9$) Ac-**C**PPLSQET**X**SDLWKLLSENN-NH$_2$
W23Tyr(C$_4$F$_9$) Ac-**C**PPLSQETFSDL**X**KLLSENN-NH$_2$

Figure 2.14 Designed MDM2 binding substrates containing perfluoro-*tert*-butyl tyrosine.
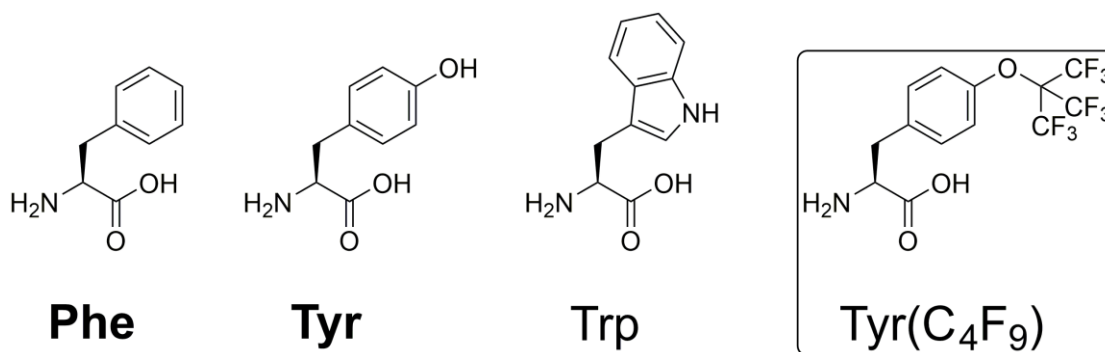
Phe   **Tyr**   **Trp**   **Tyr(C$_4$F$_9$)**

Figure 2.15 Tyr(C$_4$F$_9$) is similar in size to Trp and could potentially act as a mimic of one of the aromatic amino acids including Phe, Tyr, or Trp.

2.2.2   Determination of Dissociation Constants for Designed MDM2 Binding p53-Based Peptides

In order to evaluate the interaction between MDM2 and the designed peptides, the dissociation constant was measured using established assay protocols.[144] Peptides were synthesized on solid phase utilizing Fmoc-Tyr(C$_4$F$_9$)-OH. Synthesis of the perfluoro-*tert*-butyl tyrosine was attempted on the peptide in solution, as was demonstrated in the Ac-T(Tyr(C$_4$F$_9$))PN-NH$_2$ peptide, but this reaction was unsuccessful. Peptides were cleaved from resin and purified without precipitation with diethyl ether. The peptides containing perfluoro-*tert*-butyl tyrosine are extremely

125

hydrophobic and do not precipitate in ether, potentially due to fluorophobic

interactions with residual TFA in ether. Peptides were purified via HPLC and identity

was verified via mass spectrometry. Purified peptides were labeled on the free thiol of

the Cys with 5-iodoacetamidofluorescein. Purified $MDM2_{25-117}$ was obtained from

bacterial expression by established protocols. MDM2 was expressed in *E. coli* in

terrific broth from a pET14b plasmid containing a hexa-histidine tag.[144] The protein

was purified via Ni-NTA affinity chromatography, eluted with imidazole, and

dialyzed to obtain purified MDM2.

Dissociation constants were measured using fluorescence polarization assays,

which are a direct measure of the peptide binding to MDM2. While the perfluoro-*tert*-

butyl tyrosine peptides were based on the P27S design, P27S was not used as a control

for this assay. In this assay, $p53_{12-30}$ was synthesized and used as a control. The P27S

peptide, due to its extremely high affinity, requires the use of significantly less peptide

(5 nM) compared to lower affinity binding peptides such as $p53_{12-30}$ (210 nM).[144] The

measured K_d for this purpose matched previously measured values. The K_d for this

peptide was calculated based on a fit not utilizing the two highest MDM2

concentrations as it appears to potentially be a higher order binding process at higher

concentrations of MDM2. The peptide Phe19Tyr(C_4F_9) was measured to have a

dissociation constant of $K_d = 7.4 \pm 0.3$ µM and the peptide Trp23Tyr(C_4F_9) was

measured to have a dissociation constant $K_d = 9.1 \pm 0.9$ µM (Figure 2.16, Table 2.1).

Both of these peptides were a full order of magnitude lower in affinity than the $p53_{12-}$

$_{30}$ peptide, demonstrating that the perfluoro-*tert*-butyl tyrosine could not be fully

accommodated at either position with the binding motif. Significant loss of binding

has previously been observed with substitutions at the position 23 Trp. The peptide

W23L showed a substantial loss in binding (K_d = 43 ± 2 µM) with the modification of the Trp to a Leu.[144] While sterically, perfluoro-*tert*-butyl tyrosine is more similar to Trp than Leu, in free energy terms, the loss of binding is similar to W23L. These data demonstrate the importance of Trp at position 23 within the MDM2-binding peptide.

There are multiple potential reasons as to why these designs led to a large decrease in binding of the peptide to MDM2. The first reason is the size of the perfluoro-*tert*-butyl tyrosine. While it is approximately the same size as Trp, the shape is significantly different, and the amino acid may orient differently in the binding pocket than the Phe or Trp. Crystal structure analysis also shows the side chains of the Phe and Trp pointing towards each other. It is possible that the peptide cannot adopt the proper α-helical conformation upon binding due to the perfluoro-*tert*-butyl group. It is also known that the Trp makes as many as 10 contacts with residues within MDM2 making it a critical recognition residue.[143,152] The perfluoro-*tert*-butyl group is likely disrupting many of these contacts, preventing necessary specificity from the Trp.
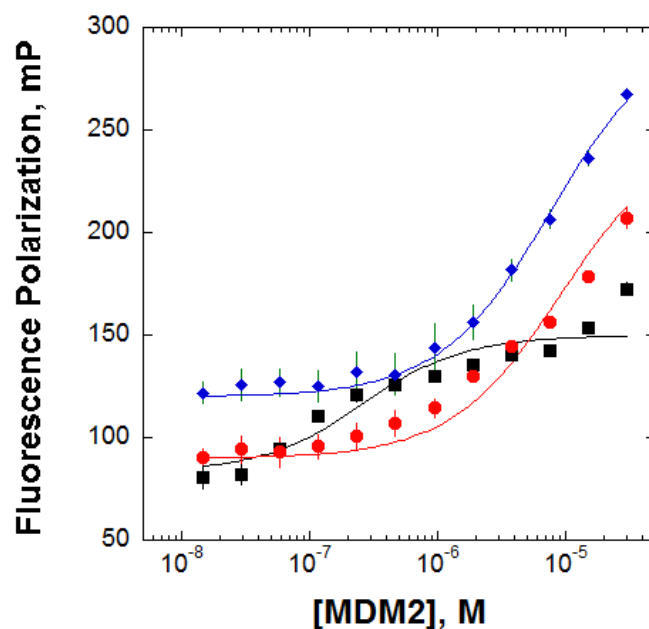
Figure 2.16 Binding isotherms of MDM2 to the fluorescein-labeled peptides p53$_{12-30}$ (black squares), W-Tyr(C$_4$F$_9$) (red circles), F-Tyr(C$_4$F$_9$) (blue diamonds), Assays were conducted using 100 nM fluorescein-labeled peptides in 1× PBS buffer (pH 7.4) with 0.1 mM DTT, and 0.04 mg/mL BSA. MDM2 was diluted using two-fold serial dilutions to final protein concentrations of 30 µM to 0.0045 µM. Fluorescence polarization data represent the average of at least 3 independent trials. Polarization data are in millipolarization units (mP). Each data point is indicative of at least three independent trials. Error bars indicate standard error.

peptide	K_d, μM	error, μM	ΔG, kcal mol^{-1}	$\Delta\Delta G$, kcal mol^{-1}
p53$_{12\text{-}30}$	0.21	0.07	-9.4	0.0
Phe19Tyr(C$_4$F$_9$)	7.4	0.3	-7.0	2.4
Trp23Tyr(C$_4$F$_9$)	9.1	0.9	-6.9	2.5
W23L	43	2	-6.0	3.4

Table 2.1   Dissociation constants (K_d) for the complexes of MDM2 and the indicated peptides. Error bars indicate standard error. K_d values were determined via a non-linear least-squares fit to the fluorescence polarization data.

2.2.3   Design of Estrogen Receptor Coactivator Peptides Containing Perfluoro-*tert*-butyl Homoserine

As demonstrated in Chapter 1, perfluoro-*tert*-butyl homoserine has the potential to be used as a mimic of aliphatic amino acids such as Leu, Ile, and Met (Figure 2.17).[51,52,54] Perfluoro-*tert*-butyl homoserine had the most α-helical propensity of any of the perfluoro-*tert*-butyl amino acids examined. In order to examine the potential applications of perfluoro-*tert*-butyl homoserine as a ^{19}F imaging probe for protein-protein interactions, we designed a series of peptides which would probe the estrogen receptor•co-activator interaction described previously. The estrogen receptor co-activator binding pocket accommodates the **LXXLL** binding motif where X = any amino acid and the Leu align along one face of the co-activator binding region making hydrophobic contacts with the estrogen receptor.[132]
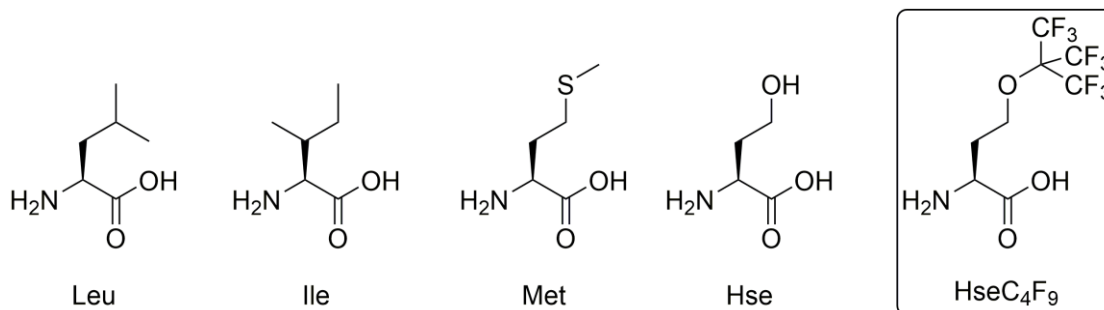
Figure 2.17 Perfluoro-*tert*-butyl homoserine could potentially act as a mimic of Leu, Ile, or Met at a protein-protein interface due to similar size, but with enhanced hydrophobic effect due to the fluorine atoms.

The co-activator binding motif contains three potential Leu residues which can be mutated to the perfluoro-*tert*-butyl homoserine. The native NRBoxII peptide, which was the basis for the design, contains a position 9 Leu (*i* position within the helix) as well as a position 12 Leu (*i*+3 position within the helix) which were mutated to perfluoro-*tert*-butyl homoserine (Figure 2.8). Based on both crystal structure analysis as well as the previous data demonstrating that neopentylglycine could not be accommodated at the position 13 Leu, perfluoro-*tert*-butyl homoserine was not examined at the position 13 Leu (*i*+4 position within the helix).[146] While not explicitly part of the LXXLL binding motif, examination of the crystal structure demonstrates a partial hydrophobic contact of the position 8 Ile (*i*-1 position within the helix), which may also benefit the binding of designed peptides (Figure 2.18). As demonstrated in Chapter 1, Leu is significantly more α-helix-promoting than Ile or perfluoro-*tert*-butyl homoserine. The mutation from Ile to perfluoro-*tert*-butyl homoserine is a structurally more conservative mutation, with the potential to preserve the native structure. Finally, due to the increased hydrophobic effect associated with fluorine, there was a concern there would be a significant increase in non-specific binding of perfluoro-*tert*-

butyl group. To verify the specificity of binding, the position 9 Leu (*i* position within the helix) was mutated to 4*S*-perfluoro-*tert*-butyl hydroxyproline, which was the least α-helix promoting perfluoro-*tert*-butyl amino acid designed (Figure 2.18).[44] This amino acid should reduce the stability of the α-helix formation, inhibiting α-helix mediated binding to the estrogen receptor, while retaining the perfluoro-*tert*-butyl ether group. This peptide functions as a control for the inherent effects of hydrophobicity of the perfluoro-*tert*-butyl ether on binding. Position 9 was chosen over position 12 to reduce α-helical character of the peptide at the beginning of the α-helix.[52,153] All peptides were also designed with an N-terminal Cys for modification with a fluorophore to evaluate binding to the estrogen receptor.

NRBoxII   **CLTERHKILHRLLQE**
Ile8Hse(C₄F₉)   **CLTERHKXLHRLLQE**
Leu9Hse(C₄F₉)   **CLTERHKIXHRLLQE**
Leu12Hse(C₄F₉) **CLTERHKILHRXLQE**
Leu9hyp(C₄F₉)   **CLTERHKIXHRLLQE**

Figure 2.18 Peptide designs for the incorporation of perfluoro-*tert*-butyl amino acids into estrogen receptor coactivator peptides. X indicates the site of modification with a perfluoro-*tert*-butyl amino acid, with either Hse(C₄F₉) (red) or 4*S*-hyp(C₄F₉) (blue) substitution at the indicated residue. All peptides for fluorescence polarization experiments were labeled with 5-iodoacetamidofluorescein on Cys (green). Peptides used in NMR spectroscopy experiments lacked the N-terminal Cys. All peptides are acetylated on the N-terminus and contain C-terminal amides.

2.2.4 Determination of Dissociation Constants for Designed Estrogen Receptor Binding Peptides Containing Perfluoro-*tert*-butyl Ethers

In order to quantitatively examine peptides designed in Figure 2.18 as potential binding partners for the estrogen receptor, peptides were evaluated by determining the dissociation constants. While there are a number of different method to determine binding constants, fluorescence polarization was chosen, which allows for a direct measurement of the peptides binding to the protein of interest.[154] Briefly, this assay measures peptide-protein binding interactions by measuring depolarization of plane-polarized light. When the fluorescently labeled peptide is unbound in solution, the fluorophore can absorb light, and due to the small size and rapid rotation in space, emission from the fluorophore depolarizes the light. When the fluorescently labeled peptide is bound to a large protein, in this case the estrogen receptor, the rotational time is much slower, leading to less depolarization of the light upon emission.[155] These types of assays can be used in a direct or indirect manner. Peptides can be evaluated by adding a fluorescent label, such as fluorescein or rhodamine to the peptide and measure binding directly.[154] Peptides and small molecules can also be measured indirectly via a competition assay with a fluorescently labeled peptide.[155] These types of assays have been used to evaluate NRBoxII•ER binding successfully, meaning there were established conditions on which the assay and the designed peptides could be evaluated.[137,146]

To evaluate binding, peptides were synthesized on solid phase incorporating an N-terminal Cys for fluorescein labeling. The control peptide (Ac-CLTERHKILHRLLQE-NH$_2$), containing only canonical amino acids, was used as it had a known binding constant, to determine the validity of the assay as well as to verify that the fluorophore, fluorescein, would not lead to increased binding due to the

hydrophobic interactions of the protein with the fluorophore.[146] All other peptides were also synthesized on solid phase, with the incorporation of the perfluoro-*tert*-butyl homoserine at the specified positions as denoted in Figure 2.18. Finally, 4*S*-perfluoro-*tert*-butyl hydroxyproline was incorporated into one of the NRBoxII peptides at the Leu$_9$ position to determine that if the perfluoro-*tert*-butyl group was not leading to an increase in affinity via non-specific binding interactions. All peptides were purified and identity verified using mass spectrometry. All peptides were fluorescein labeled using 5-iodoacetamidofluorescein, which conjugates to the free thiol of the Cys and purified via HPLC.

The protein used for all assays was the estrogen receptor alpha ligand binding domain$_{255-595}$. While the estrogen receptor exists in multiple isoforms, not all isoforms are equally expressed in all tissues. Due to our interest in using these probes to detect breast cancer, the estrogen receptor alpha was chosen, as this is the isoform that is overexpressed in breast cancer cells.[132] This protein was expressed from a pET15b plasmid with a hexahistidine tag in BL21 cells using IPTG in terrific broth. Protein was isolated using his-tag resin according to the manufacture's protocols (EMD Millipore) for use in all assays.

All peptides were incubated at 100 nM concentration, with a serial dilution of the purified estrogen receptor protein in buffer. The K_d was measured for each peptide construct (Figure 2.19, Table 2.2) by fluorescence polarization using a fluorescence plate reader. The control NRBoxII peptide, containing no modifications, bound to the estrogen receptor with the previously reported dissociation constant ($K_d = 1.0$ μM).[146] As expected the Leu9hyp(C_4F_9) showed weak binding ($K_d > 150$ μM), which is most likely due to non-specific binding with the protein as this amino acid co-activator

peptide as well as some residual α-helix formation. This peptide should demonstrate a significant reduction in binding to the estrogen receptor due to the 4S-perfluoro-*tert*-butyl hydroxyproline disrupting the necessary α-helical structure required for binding, although, a small amount of α-helix can still form. These data demonstrate that the perfluoro-*tert*-butyl ether does not inherently lead to an increase in binding due to hydrophobicity. The peptide Leu9Hse(C_4F_9) also showed reduced binding to the estrogen receptor ($K_d = 43$ μM) compared to the other peptides containing perfluoro-*tert*-butyl homoserine. These data indicate that, despite the increased hydrophobic effect introduced with fluorine atoms, the perfluoro-*tert*-butyl homoserine may be too large at the initial (i) Leu position. The peptide Ile8Hse(C_4F_9), however, was relatively well tolerated ($K_d = 7.1$). At the Ile position, this side chain is partially solvent exposed, so the primary driving force for binding is most likely the hydrophobic effect, as seen in the crystal structure (Figure 2.4).

Among peptides in this series, the peptide Leu12Hse(C_4F_9) had the highest binding affinity ($K_d = 2.2$ μM), which was similar to the native peptide. These results are unsurprising as both the crystal structure and the data demonstrating high affinity binding with neopentylglycine had previously indicated that this site has the largest space in the hydrophobic binding pocket to accommodate the increased size of the perfluoro-*tert*-butyl group. Due to the high affinity of the Leu12Hse(C_4F_9) peptide, this peptide was further characterized and used in ^{19}F binding experiments to demonstrate the potential of this peptide to be used as a fluorine-based probe of protein-protein interactions. These data indicate the ability to substitute perfluoro-*tert*-butyl homoserine into a recognition motif, in place of Leu, at a protein-binding interface.

134
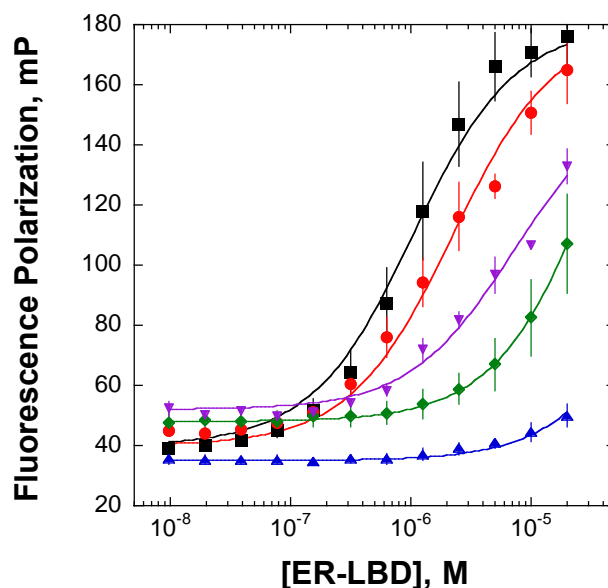
Figure 2.19   Binding isotherms of estradiol-bound ERα LBD to the fluorescein-labeled peptides NRBoxII (black squares), Leu12Hse(C$_4$F$_9$) (red circles), Ile8Hse(C$_4$F$_9$) (magenta inverted triangles), Leu9Hse(C$_4$F$_9$) (green diamonds), and Leu9hyp(C$_4$F$_9$) (blue triangles). Assays were conducted using 100 nM fluorescein-labeled peptides in 1× PBS buffer (pH 7.4) with 20 µM estradiol, 0.1 mM DTT, and 0.04 mg/mL BSA. ERα LBD was diluted using two-fold serial dilutions to final protein concentrations of 10 µM to 0.0098 µM. Fluorescence polarization data represent the average of at least 3 independent trials. Polarization data are in millipolarization units (mP). Each data point is indicative of at least three independent trials. Error bars indicate standard error.

peptide	K_d, μM	error, μM	ΔG, kcal mol^{-1}
NRBoxII	1.0	0.1	−8.2
Ile8Hse(C$_4$F$_9$)	7.1	1.3	−7.0
Leu9Hse(C$_4$F$_9$)	43	4	−6.0
Leu12Hse(C$_4$F$_9$)	2.2	0.2	−7.7
Leu9hyp(C$_4$F$_9$)	>150	n.a.	n.a.

Table 2.2   Dissociation constants (K_d) for the complexes of estradiol-bound ERα LBD and the indicated peptides. Errors indicate standard error. K_d values were determined via a non-linear least-squares fit to the fluorescence polarization data. n.a. not applicable.

2.2.5   Structural Implications of the Perfluoro-*tert*-butyl Ether Group in Estrogen Receptor Binding Proteins

Due to the importance of the α-helical structure upon binding to the estrogen receptor, it is important to characterize changes within the peptide structure due to the perfluoro-*tert*-butyl homoserine modification. While perfluoro-*tert*-butyl homoserine is α-helix promoting at the N-terminus of a peptide, not all amino acids have the same α-helical propensities at all positions within an α-helix.[51,52] For this reason, we decided to examine the α-helical character of both the NRBoxII control peptide and the Leu12HseC$_4$F$_9$ peptide. As seen in Chapter 1, circular dichroism can be used to determine α-helical structure.[51] Peptides were dissolved in buffered water to determine their secondary structure. Unsurprisingly, neither peptide demonstrated any α-helical character by CD. Short peptides are often disordered, and these structures are known to have binding-induced folding to adopt the α-helix upon binding to the estrogen receptor.[50] While short peptides are not necessarily stable α-helices in water, trifluoroethanol (TFE) solutions is known to induce α-helical secondary structure.[156]

Both the NRBoxII control peptide and the Leu12Hse(C_4F_9) were examined in 30% TFE solutions (Figure 2.20). Both peptides formed α-helical structures, as indicated by minima at 222 and 208 nm with the NRBoxII control peptide demonstrating more helical character than the peptide containing a perfluoro-*tert*-butyl group (Table 2.3). The NRBoxII peptide was more α-helical (46 % helical) compared to the peptide Leu12Hse(C_4F_9) (27% helical). As previously noted, Leu is one of the most α-helix promoting amino acids, while perfluoro-*tert*-butyl homoserine is similar to Met and Pro in terms of α-helix promoting character when placed at the N-terminus of a peptide.[52,54] Leu (f_{helix} = 0.546) and, to a lesser degree, Met (f_{helix} = 0.507) are both α-helix promoting in the interior of a peptide, however, Pro (f_{helix} = 0.000) disrupts the α-helix at the interior of a peptide as demonstrated in the Ac-YGGKAAAAKAXAAKAAAAK-COOH peptide.[157] In most cases, Perfluoro-*tert*-butyl homoserine, unsurprisingly, is not as α-helix promoting as Leu in any context but can mimic other aliphatic amino acids in an α-helical context, both at the N-terminus as well as in the interior of α-helical peptides. Furthermore, the perfluoro-*tert*-butyl homoserine is a fluorinated amino acid which does not inherently disfavor α-helical structures, unlike most highly fluorinated amino acids.[16] These data demonstrate the perfluoro-*tert*-butyl homoserine can be introduced at any position within an α-helical context, while maintaining the ability of the peptide to form an α-helix.
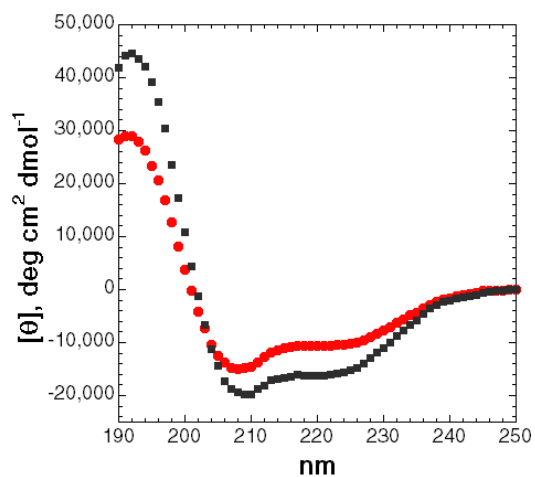
Figure 2.20   CD spectra of NRBoxII (black squares) and Leu12Hse(C_4F_9) (red circles) in 30% TFE with 5 mM phosphate buffer and 25 mM KF at pH 7.4 at 25 °C.

Peptide	$[\theta]_{222}$ deg cm^2 dmol^{-1}	$[\theta]_{208}$ deg cm^2 dmol^{-1}	$[\theta]_{190}$ deg cm^2 dmol^{-1}	$[\theta]_{222}/[\theta]_{208}$	$-[\theta]_{190}/[\theta]_{222}$	% α-helix
NRBoxII	-16133	-19367	41810	0.83	2.59	46%
Leu12Hse(C$_4$F$_9$)	-10719	-14998	28378	0.71	2.65	28%

Table 2.3 Circular dichroism data for NRBoxII and Leu12Hse(C$_4$F$_9$) peptides in 30% TFE. Extent of α-helix is identified by mean residue ellipticity at 222 nm ($[\theta]_{222}$) or via the ratios $[\theta]_{222}/[\theta]_{208}$ or $-[\theta]_{190}/[\theta]_{222}$. % α-Helix was determined via the method of Baldwin, as described previously. Ser and Thr are highly stabilizing to α-helices at their N-terminus due to the hydrogen bonding with the unsatisfied amide N-H hydrogen bond donors.

Both peptides were also examined by NMR to determine any potential perturbations to the peptide structure due to the perfluoro-*tert*-butyl homoserine. Peptides were synthesized without the N-terminal Cys, as the fluorophore label was unnecessary for NMR and CD experiments, and free thiols can potentially form disulfides. All NMR data were recorded at pH 4.0 to slow hydrogen exchange, allowing the amide protons to be observable by ^{1}H NMR. Simple ^{1}H NMR spectrum analysis of both the NRBoxII peptide and the Leu12HseC$_4$F$_9$ peptides showed nearly identical spectra (Figure 2.21). Small changes were observed in the amide region with a peak around 8.5 ppm becoming visible with the substitution of one Leu to perfluoro-*tert*-butyl homoserine. In addition, an additional peaks were observed in the aliphatic

region, around 2 ppm, as a result of the extra methylenes from the perfluoro-*tert*-butyl homoserine.

TOCSY correlation spectra were used to examine through bond coupling of protons in both peptides. Again these spectra were strikingly similar (Figure 2.25, 2.26, 2.27). Superpositions of the fingerprint region of the TOCSY spectra shows a cross peak at 8.5 ppm correlating to a peak at 4.45 ppm as observed in the ^{1}H NMR spectra as a result of the perfluoro-*tert*-butyl homoserine (Figure 2.22, 2.23, 2.24). This peak correlates the perfluoro-*tert*-butyl homoserine amide proton with the H_α proton. Additionally, new cross peaks are visible at 2.0 and 2.4 ppm, corresponding to the H_β and H_γ protons. Examination of the superposition of the full TOCSY spectra show all other regions to be nearly identical between the two peptides. Due to the strong similarities, HSQC correlation spectra were also collected for both peptides which allows for observation of carbon atoms and the corresponding attached protons. Chemical shift index has been correlated to secondary structure, where α-helical secondary structure exhibit an upfield shift relative to unfolded peptides, in the C_α-H_α region.[158] In the event that the perfluoro-*tert*-butyl homoserine is disrupting the α-helix structure of the peptide, we would expect to see a downfield shift in H_α and an upfield shift in C_α associated with the helical residues. Once again, the spectra of the NRBoxII and Leu12Hse(C_4F_9) peptides were nearly identical (Figure 2.28, 2.29, 2.30). Superposition of the two peptides shows new cross peaks associated with the perfluoro-*tert*-butyl homoserine, however, no substantial shifting is observed in the other residues including those which participate in the α-helix (Figure 2.33). As expected, the perfluoro-*tert*-butyl homoserine residue causes shifting of the C_α and the corresponding protons (Figure 2.29). All of these data taken together demonstrate that

these peptides, in an aqueous environment, have very similar structure, demonstrating that perfluoro-*tert*-butyl homoserine has no substantial effect on peptide structure. The NMR data, taken with the CD data in 30% TFE show that the perfluoro-*tert*-butyl ether has a minimal impact on peptide structure.
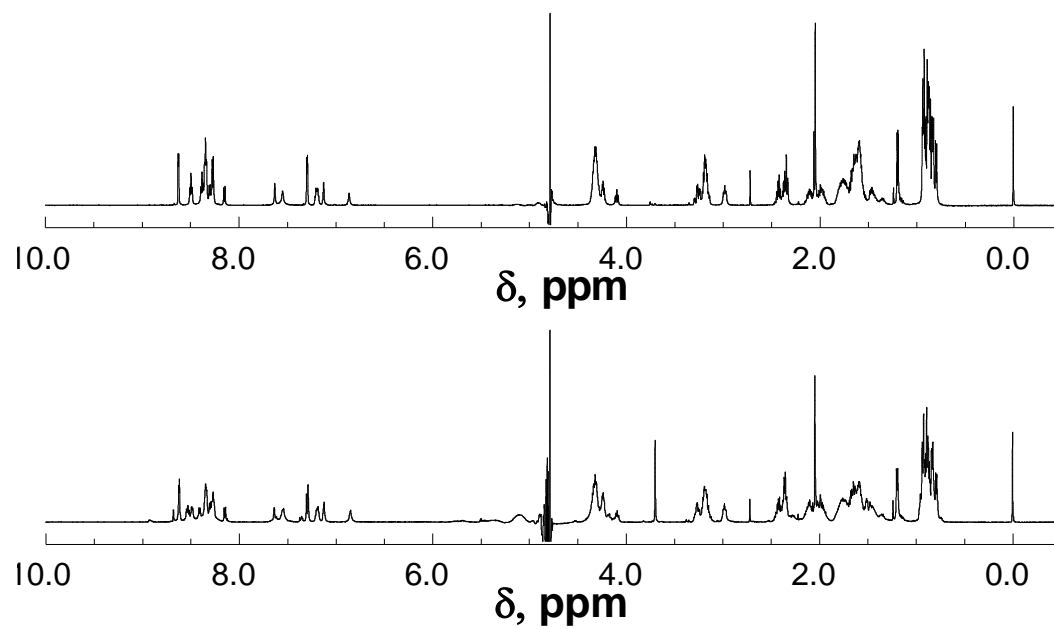
Figure 2.21 ^{1}H NMR spectra of the peptides NRBoxII (top) and Leu12Hse(C$_4$F$_9$) (bottom) in 5 mM phosphate buffer with 25 mM NaCl.

142
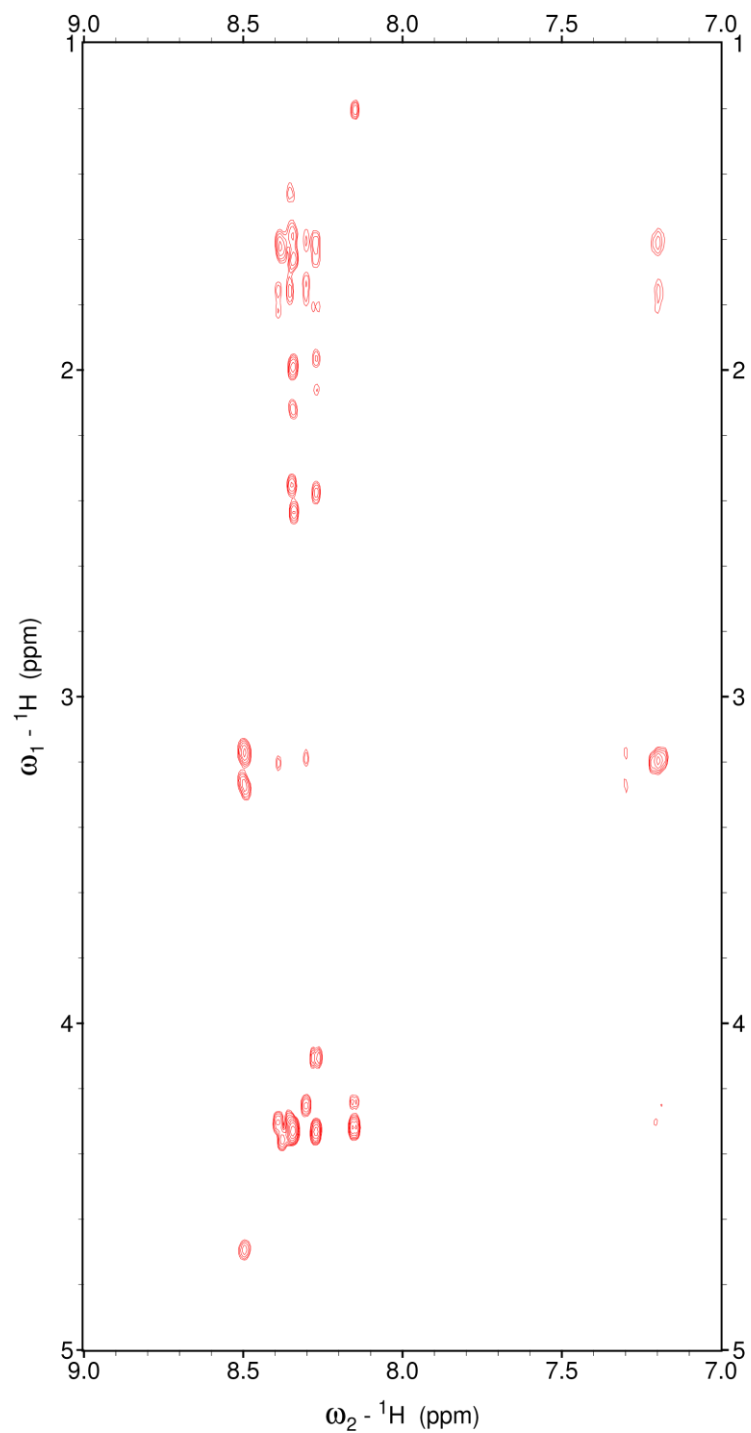
Figure 2.22   Fingerprint region of the TOCSY spectrum of the NRBoxII peptide in 5 mM phosphate buffer with 25 mM NaCl.
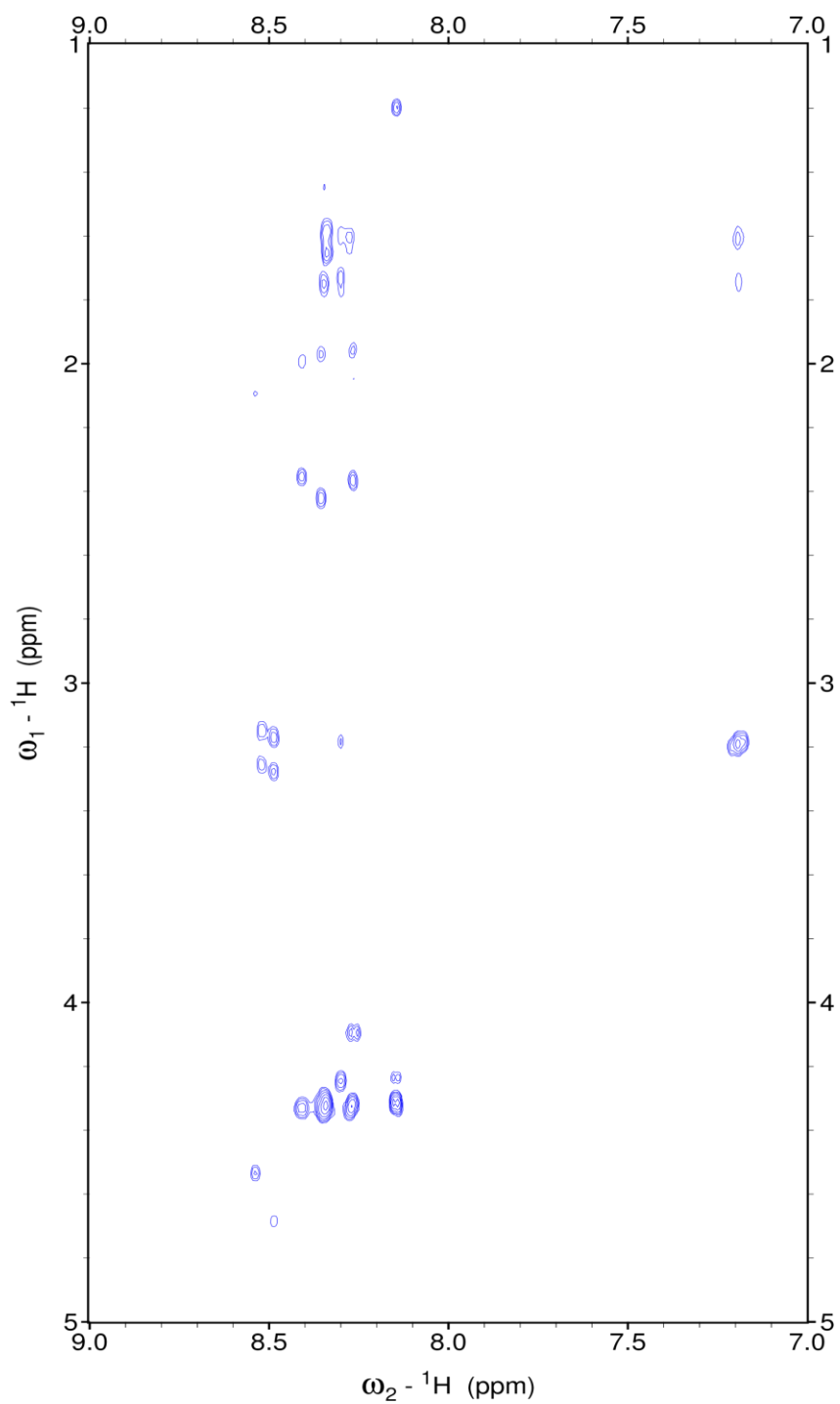
Figure 2.23   Fingerprint region of the TOCSY spectrum of the peptide
Leu12Hse(C_4F_9) in 5 mM phosphate buffer with 25 mM NaCl.
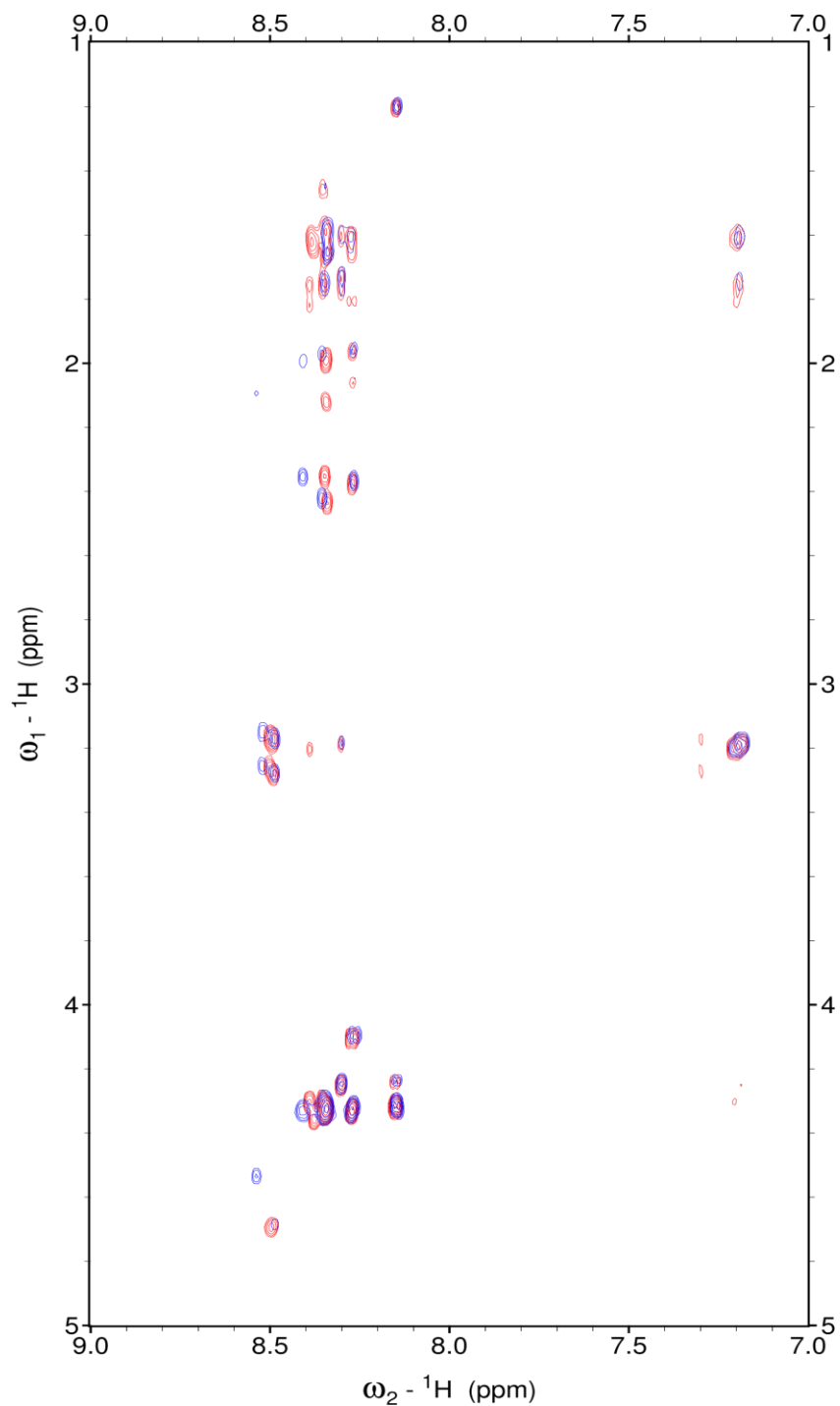
Figure 2.24 Superposition of the fingerprint regions of the TOCSY spectra of the peptides NRBoxII (red) Leu12Hse(C_4F_9) (blue) in 5 mM phosphate buffer with 25 mM NaCl.
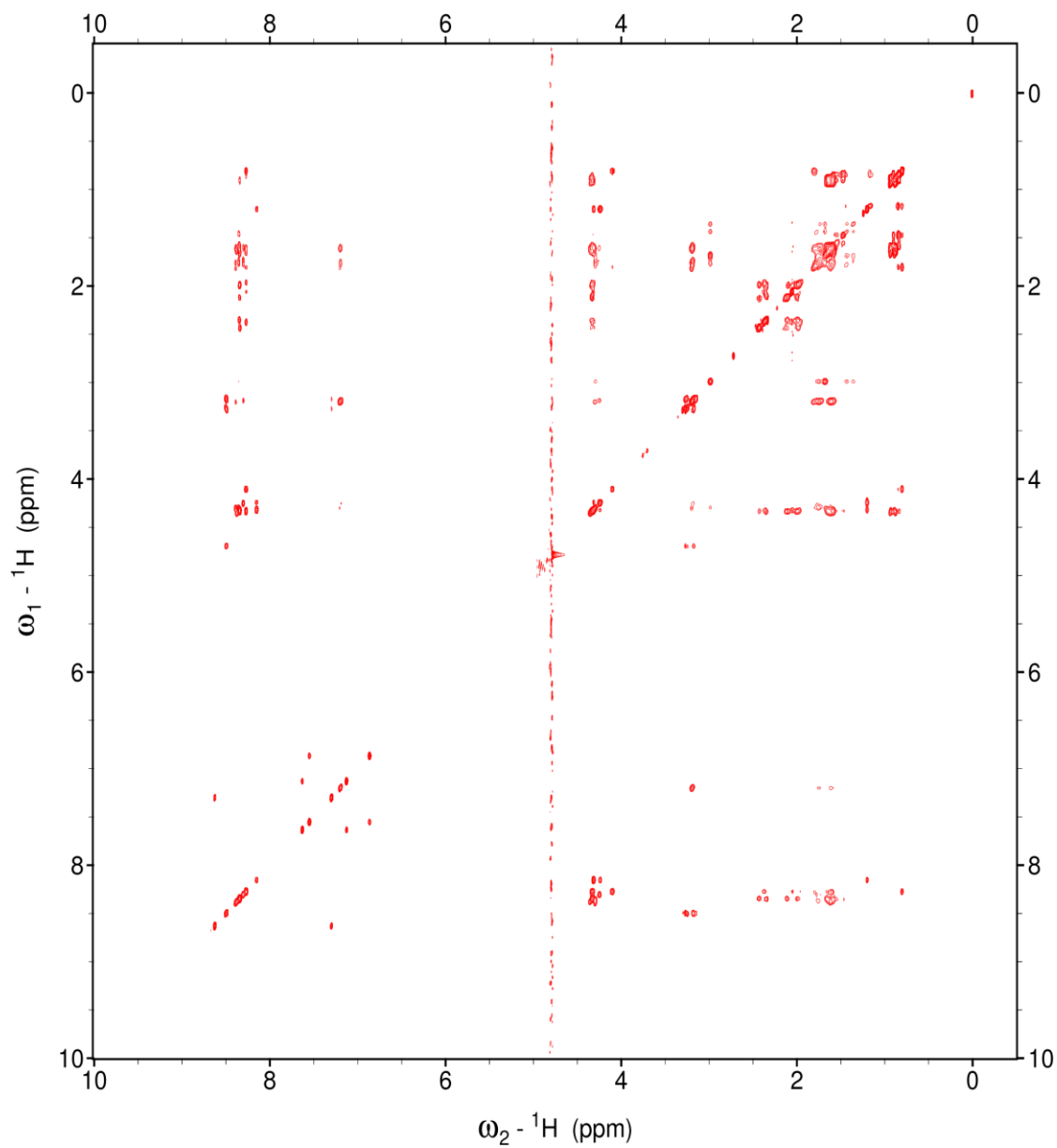
Figure 2.25   Full TOCSY spectrum of the peptide NRBoxII in 5 mM phosphate buffer
with 25 mM NaCl.
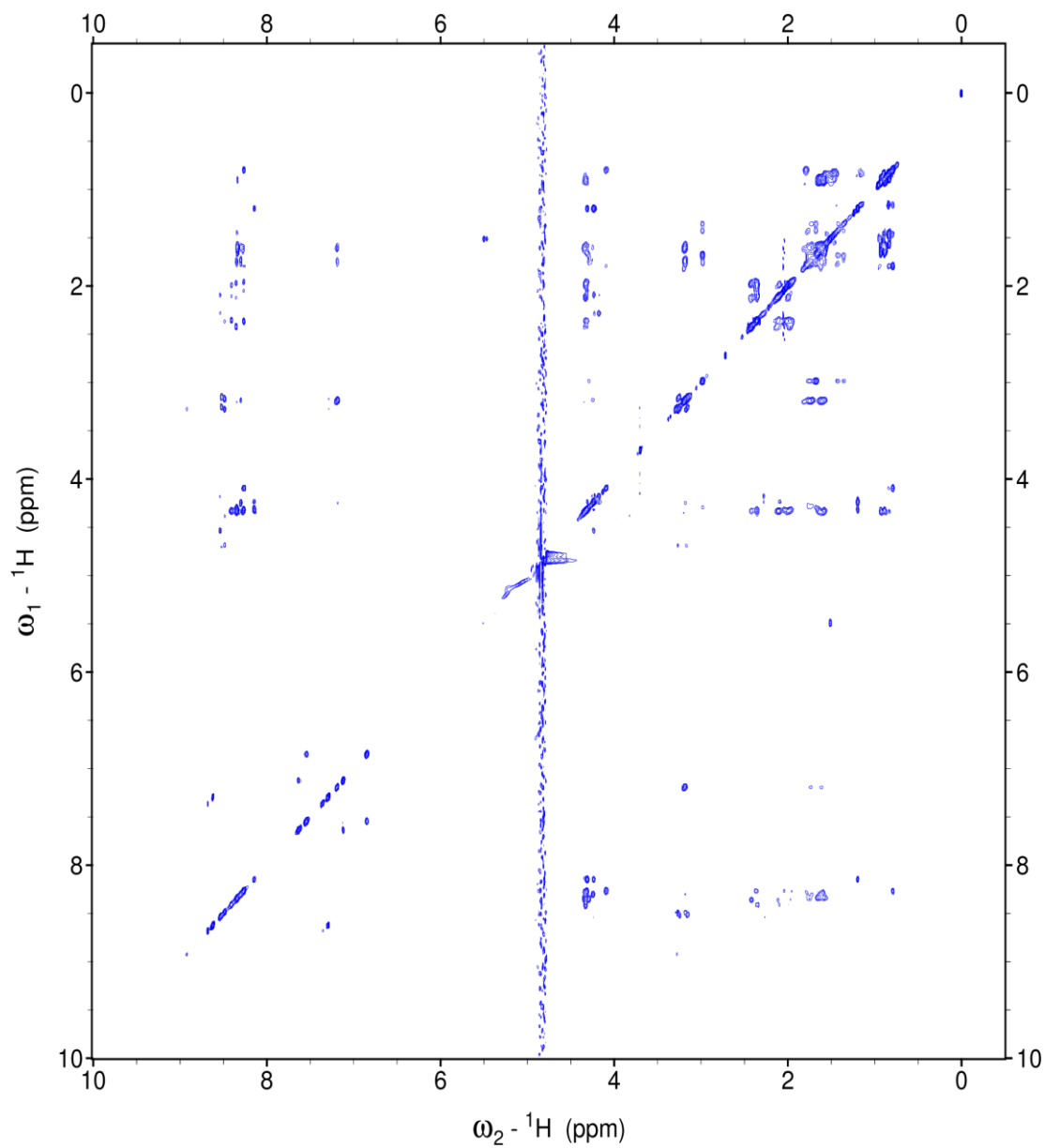
Figure 2.26 Full TOCSY spectrum of the peptide Leu12Hse(C$_4$F$_9$) in 5 mM
phosphate buffer with 25 mM NaCl.
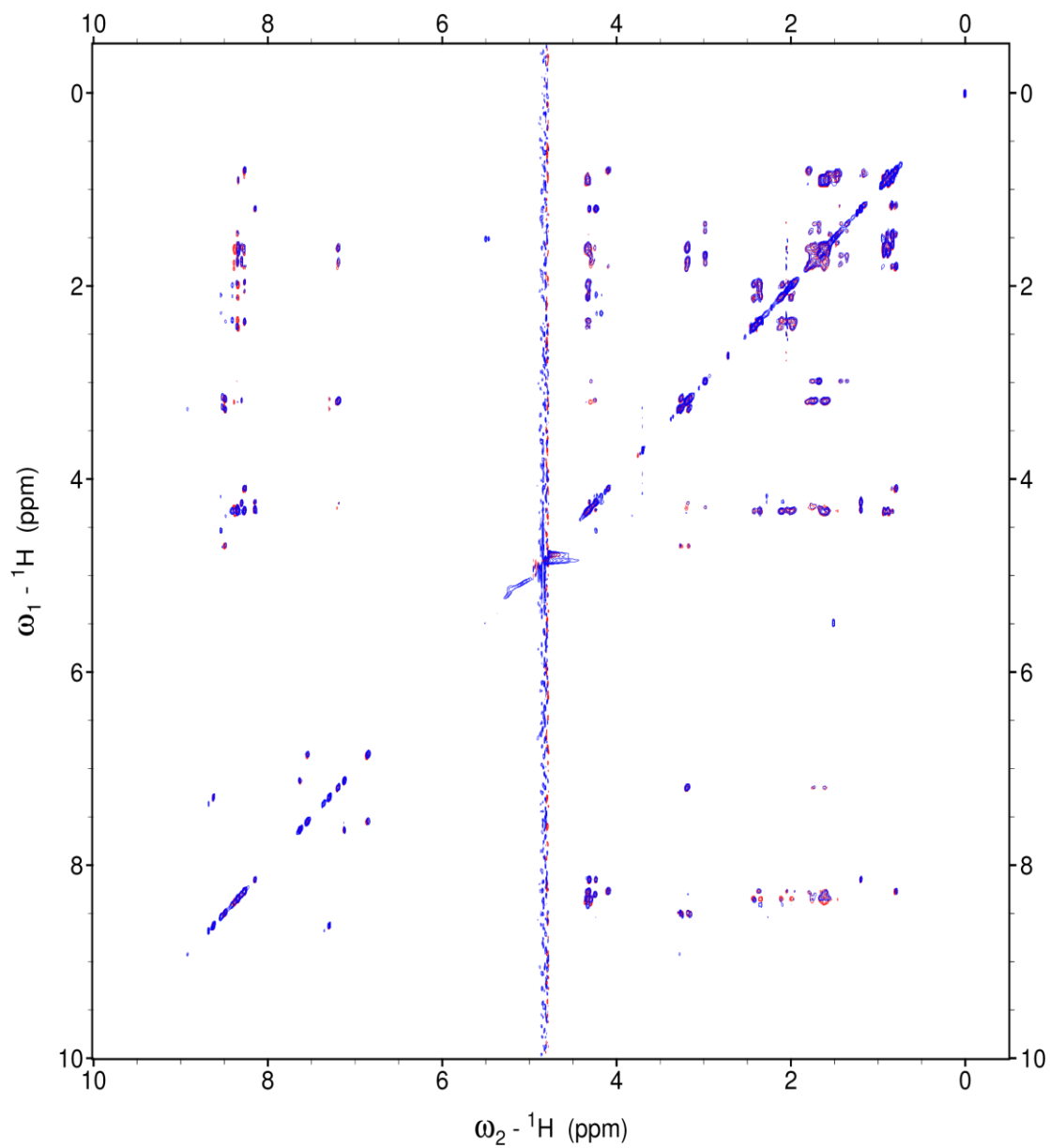
Figure 2.27 Superposition of the full TOCSY spectra of the peptides NRBoxII (red) Leu12Hse(C_4F_9) (blue) in 5 mM phosphate buffer with 25 mM NaCl.

Figure 2.28 Hα-Cα region of the ^{1}H-^{13}C HSQC spectrum of the peptide NRBoxII in 5 mM phosphate buffer with 25 mM NaCl.

149
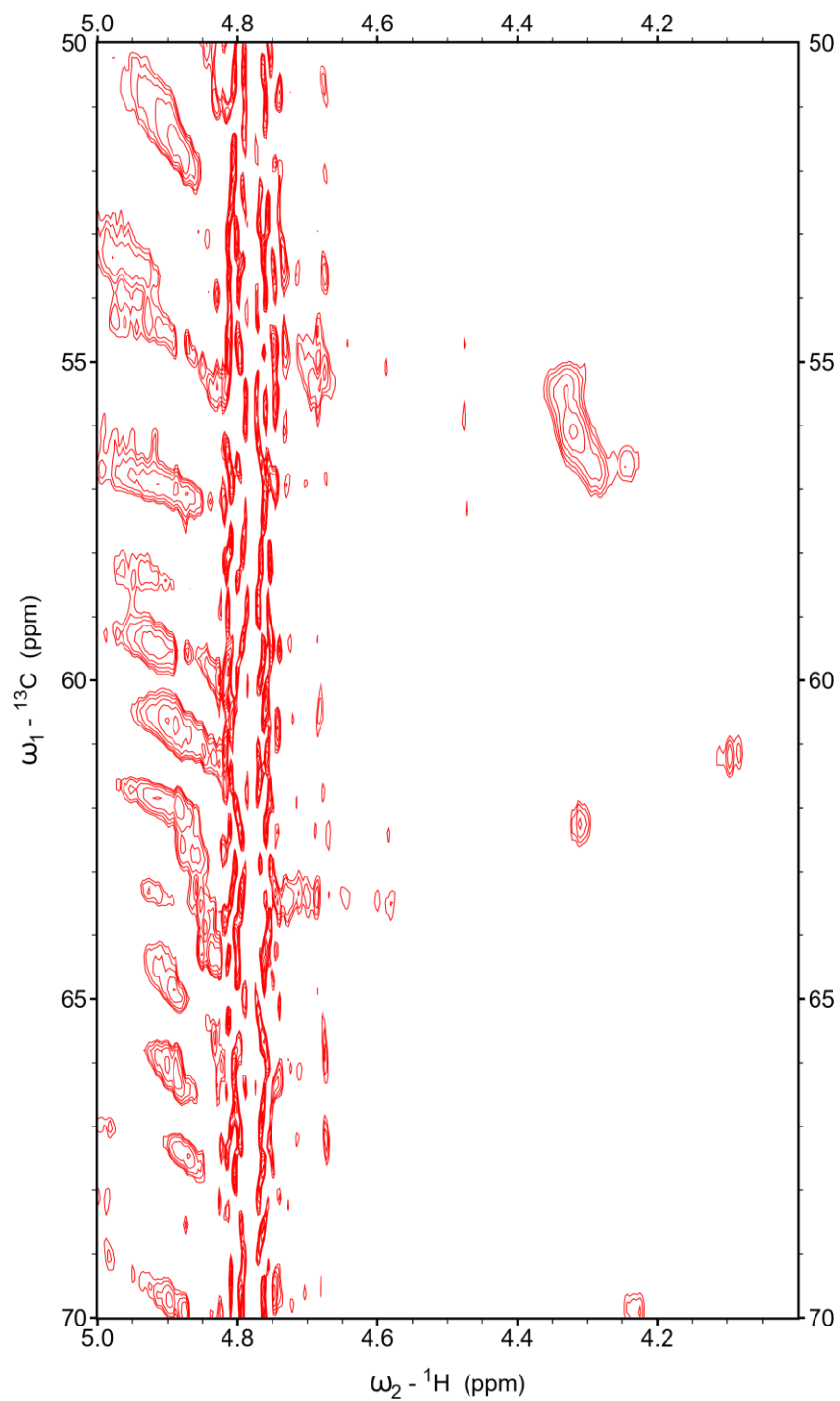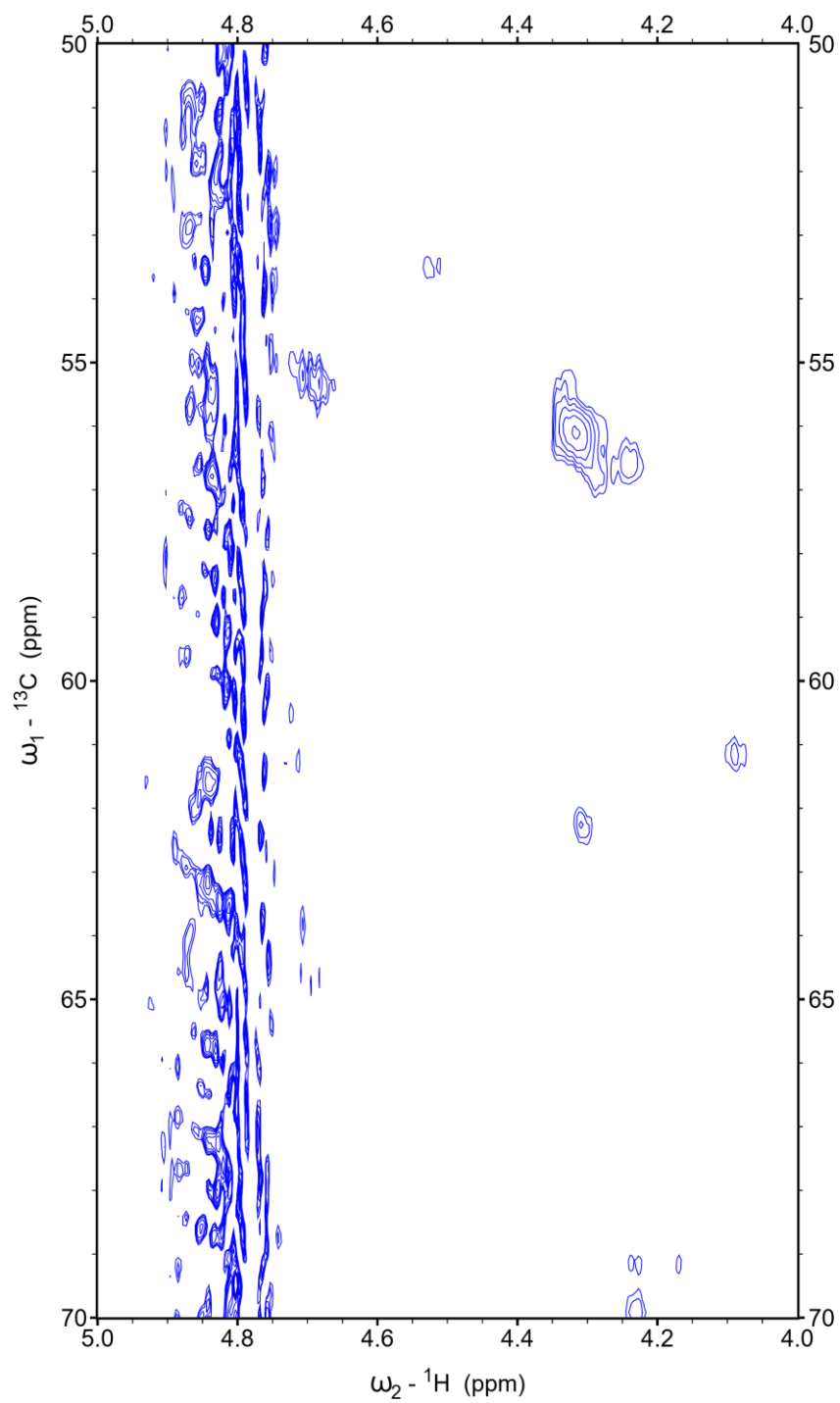
Figure 2.29 Hα-Cα region of the ^{1}H-^{13}C HSQC spectrum of the peptide Leu12Hse(C$_4$F$_9$) in 5 mM phosphate buffer with 25 mM NaCl.
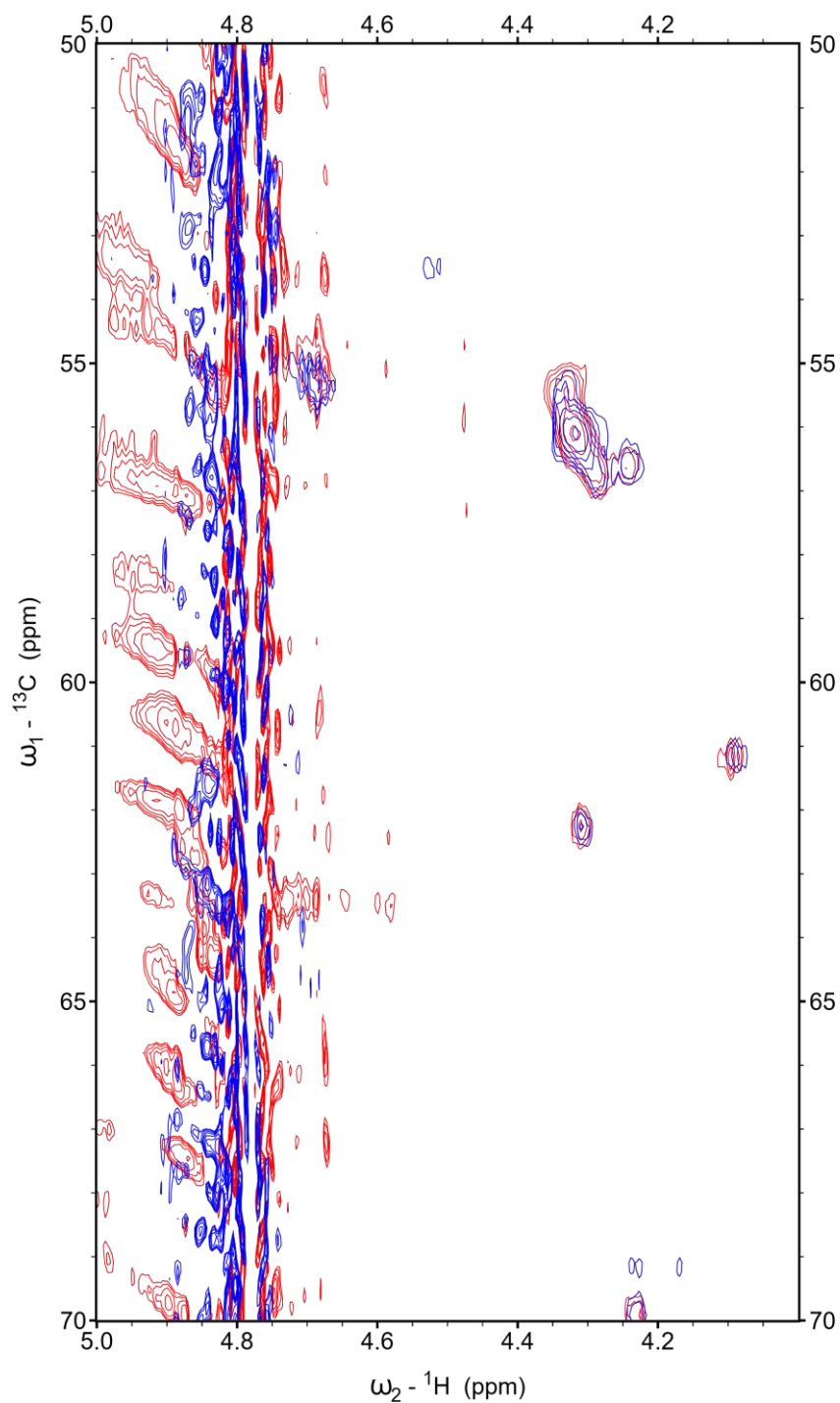
Figure 2.30 Superposition of the Hα-Cα regions of the ^{1}H-^{13}C HSQC spectra of the peptides NRBoxII (red) and Leu12Hse(C_4F_9) (blue) in 5 mM phosphate buffer with 25 mM NaCl.
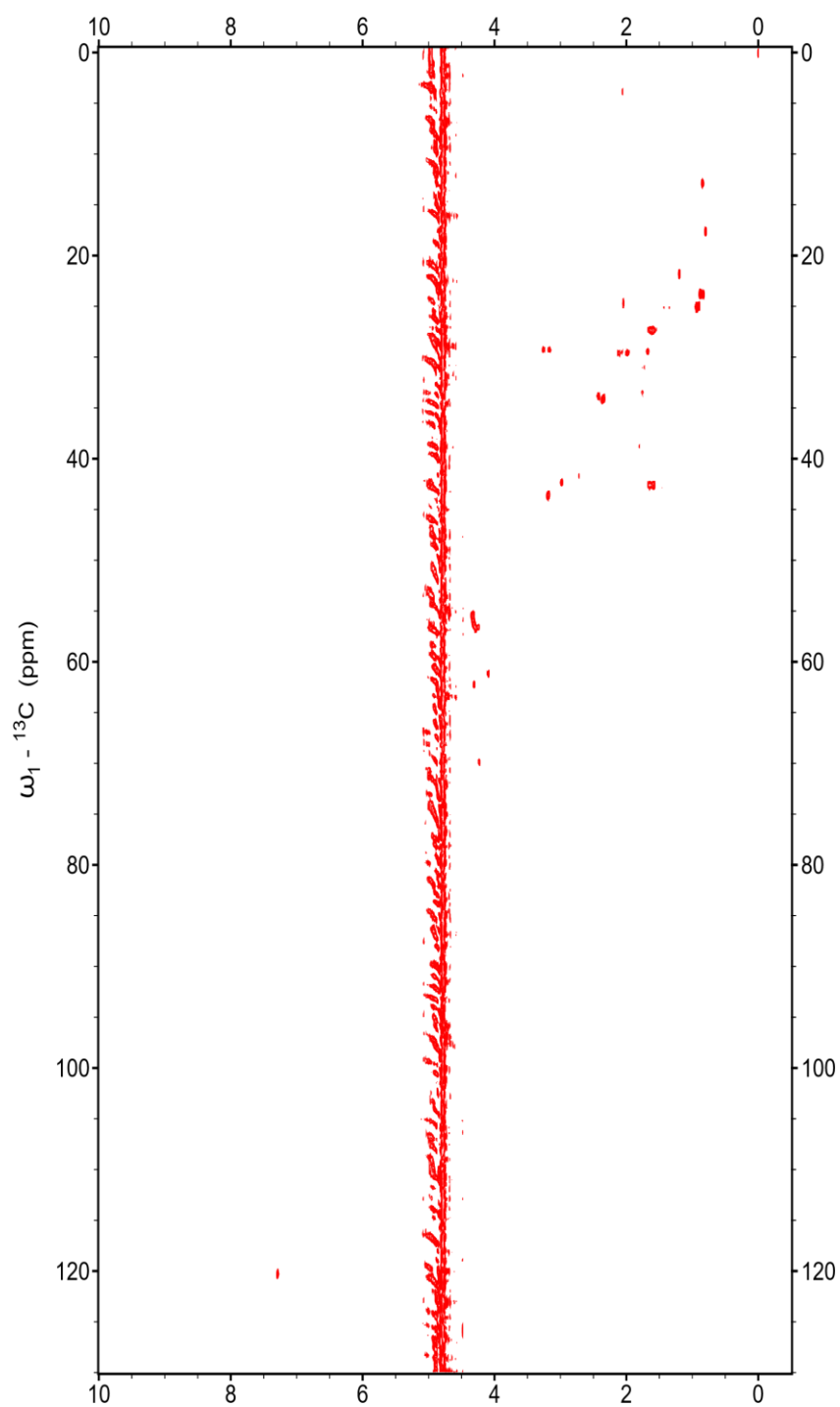
Figure 2.31 Full ^{1}H-^{13}C HSQC spectrum of the peptide NRBoxII in 5 mM phosphate
buffer with 25 mM NaCl.
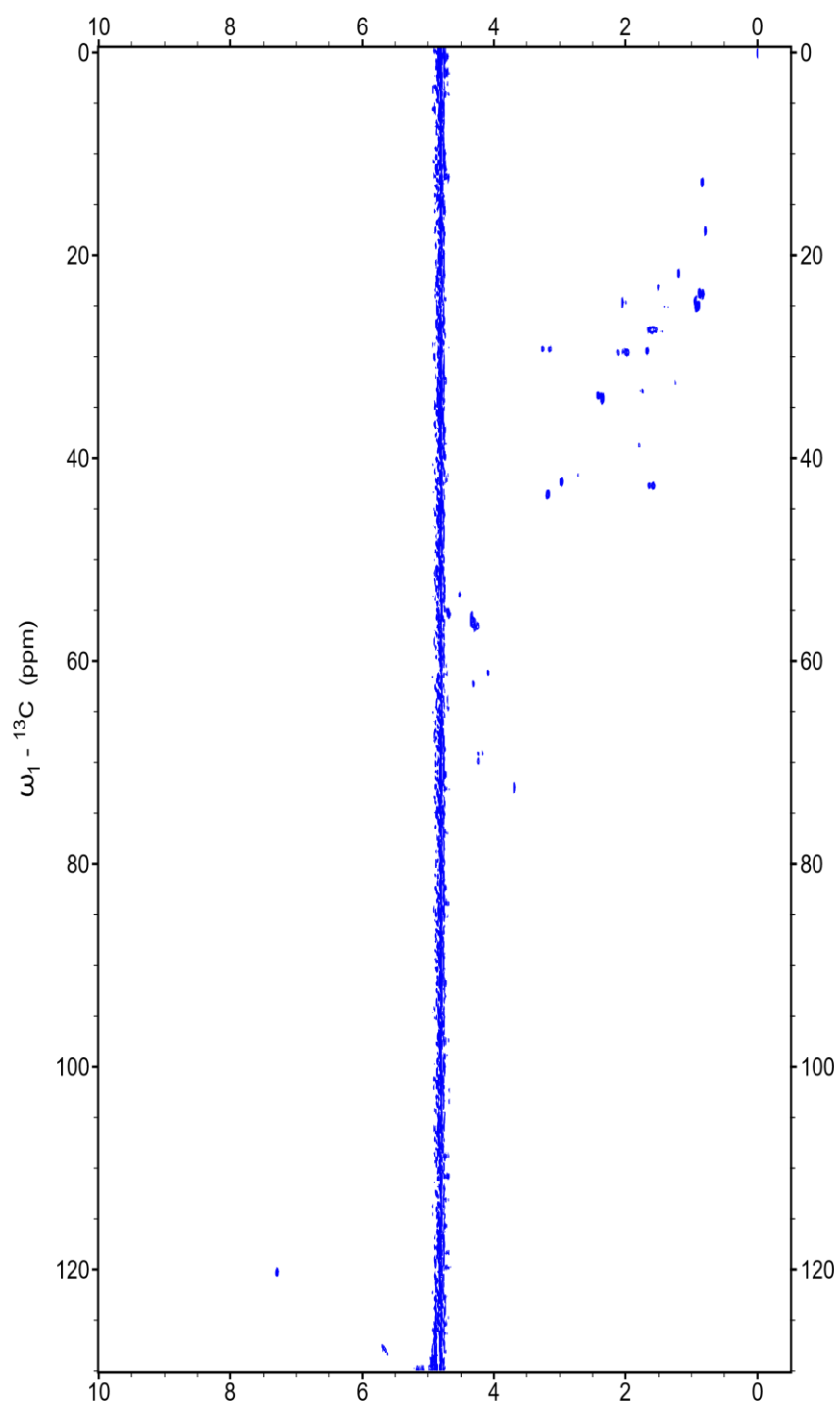
Figure 2.32   Full ^{1}H-^{13}C HSQC spectrum of the peptide Leu12Hse(C$_4$F$_9$) in 5 mM phosphate buffer with 25 mM NaCl.
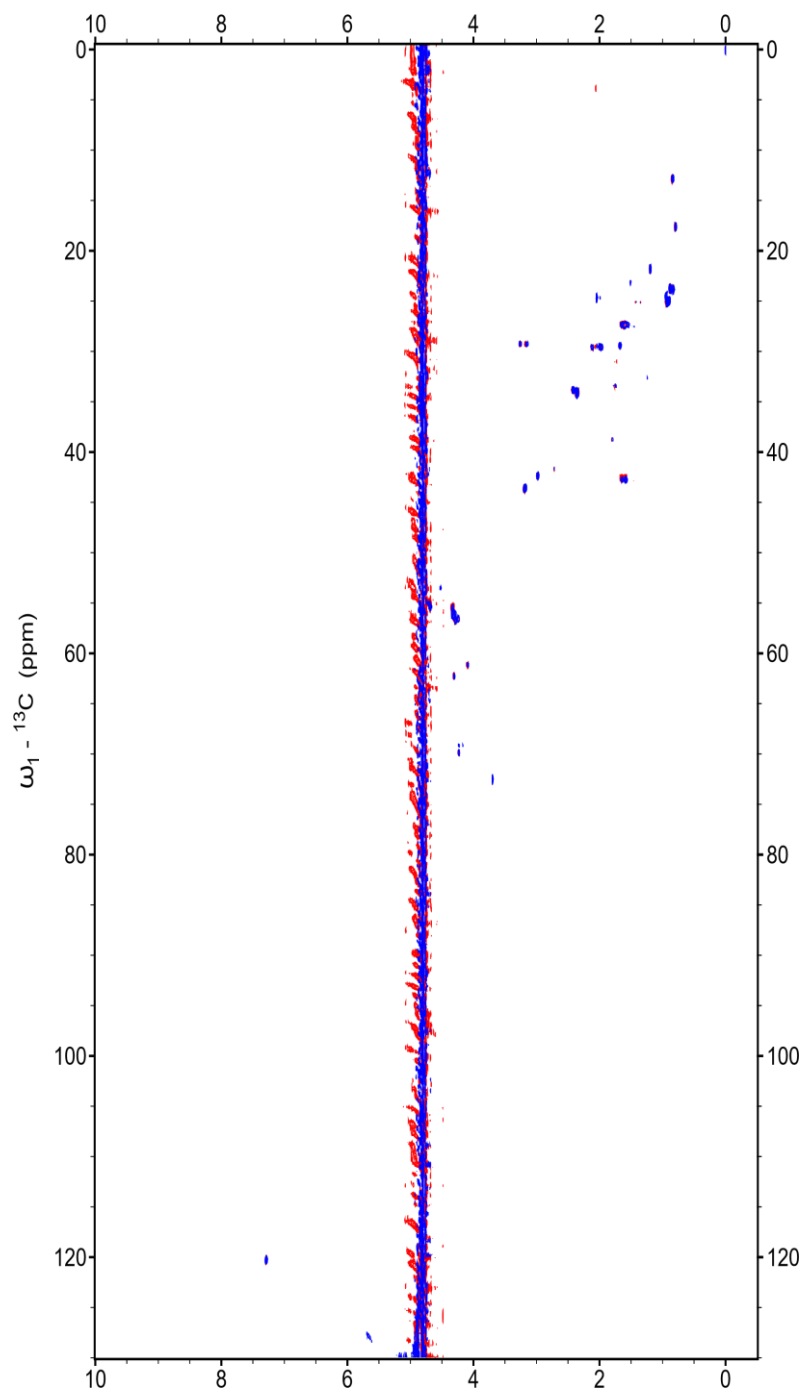
Figure 2.33   Superposition of the full ^{1}H-^{13}C HSQC spectra of the peptides NRBoxII (red) Leu12Hse(C$_4$F$_9$) (blue). The similarities in the NMR spectra of both peptides are consistent with similar structures being adopted by both peptides, and with data indicating that Hse(C$_4$F$_9$) has good propensity for the α-helical conformation.

2.2.6 ^{19}F Detection of Estrogen Receptor-Peptide Binding via Chemical Shift Anistropy or Exchange Broadening

Having demonstrated that the peptide Leu12Hse(C$_4$F$_9$) binds to ERα LBD with near-native affinity and is structurally nearly identical the NRBoxII co-activator peptide, we set out to demonstrate that this peptide could be used to detect estrogen receptor•co-activator binding by ^{19}F NMR. The peptide Leu12Hse(C$_4$F$_9$) was diluted to 5 µM in buffer and the ^{19}F NMR spectrum was collected (128 scans, 5 mins) which, as expected, exhibited a single, sharp peak (Figure 2.34). Based on the fluorescence polarization data, as ERα LBD is titrated into the solution, Leu12Hse(C$_4$F$_9$) binds to the protein. This interaction was also observed by ^{19}F NMR; as ERα LBD was titrated into a buffer solution with 5 µM peptide, chemical shift anisotropy was observed (Column 1, Figure 2.34). While there is a possibility that a chemical shift change is happening as a result of Leu12Hse(C$_4$F$_9$) binding to the protein, this potential change is not observable due to chemical shift anisotropy (CSA). As the peptide binds to the protein, the increase in size (ERα LBD = 39 kDa), two dynamic regimes are expected: in the free peptide, rapid rotation results in a single sharp peak. In constrast, in the peptide•protein complex, slow tumbling combined with the lack of free rotation of the perfluoro-*tert*-butyl group, which is interfacing with the protein, leads to increasing CSA of the singlet observed by ^{19}F NMR. To demonstrate that this interaction is a result of the Leu12Hse(C$_4$F$_9$) peptide binding to the protein at the co-activator binding pocket, 10 µM NRBoxII peptide was added to the NMR samples (< 5 µL) which can displace the Leu12Hse(C$_4$F$_9$), recovering the sharp singlet by ^{19}F NMR (Column 2, Figure 2.34). In all cases, the Leu12Hse(C$_4$F$_9$) ^{19}F signal was restored to a sharp singlet upon incubation with the NRBoxII peptide. These competition data demonstrate that the NRBoxII peptide is displacing the Leu12Hse(C$_4$F$_9$) peptide in the

co-activator binding pocket. If the Leu12Hse(C$_4$F$_9$) peptide was binding elsewhere on the protein, it would not be displaced by the NRBoxII peptide, which is known to bind in the co-activator binding pocket.[50,125,132,133]

Similarly, 4-hydroxytamoxifen (OHT), a commonly used SERM, is known to alter the estrogen receptor, preventing binding of the co-activator to the estrogen receptor.[135,138] We expect no observable CSA of the peptide Leu12Hse(C$_4$F$_9$) as a result of incubation with OHT-bound ERα LBD. The titration of estrogen receptor with 5 µM Leu12Hse(C$_4$F$_9$) was repeated with 50 µM OHT instead of 17ß-estradiol. In fact, no CSA was observed even at the highest concentration of ERα LBD (30 µM) (Column 3, Figure 2.34). There are small satellite peaks visible in the presence 50 µM OHT. These peaks are most likely chemical shift changes associated with hydrophobic interactions with OHT, which is a large, hydrophobic molecule which, while free in solution, could interact with the perfluoro-*tert*-butyl group of the Leu12Hse(C$_4$F$_9$) at high concentration, leading to the observed chemical shift change.
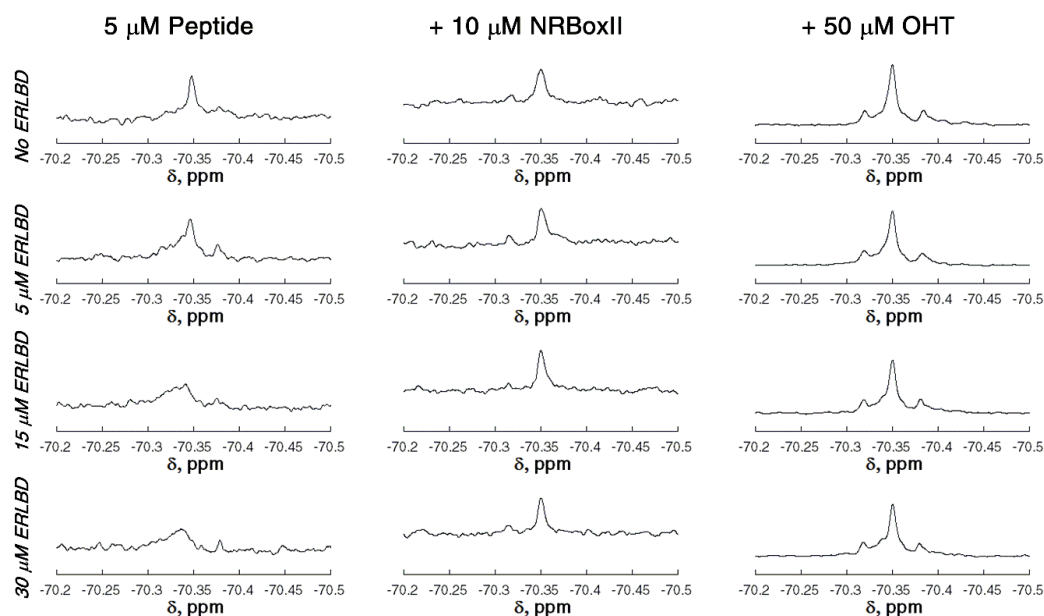
Figure 2.34 ^{19}F NMR titration of the peptide Leu12Hse(C_4F_9) with the ERα LBD. All NMR samples contain 5 µM Leu12Hse(C_4F_9) in PBS with 0.1 mM DTT and 10% D_2O for NMR lock. All samples contain 20 µM estradiol, except for experiments with tamoxifen (OHT), which contain 50 µM tamoxifen instead of estradiol. Competition experiments were conducted by adding 10 µM NRBoxII peptide (< 5 µL volume added). All samples were incubated at room temperature for a minimum of 30 minutes before the NMR experiment. All experiments were conducted on a 564.5 MHz NMR spectrometer with 128 scans, a 10 ppm sweep width, an acquisition time of 0.8 seconds, and relaxation delay of 3.0 seconds (8 minutes total experiment time).

Having established conditions under which we could observe the protein-protein interaction between the ERα LBD and the peptide Leu12Hse(C_4F_9), we optimized the ^{19}F NMR experiments to examine binding. The titration of Leu12Hse(C_4F_9) was repeated at 5 µM peptide with the addition of 0 to 30 µM ERα LBD (Figure 2.35, panels a-d) using 1-hour acquisition times (1024 scans) for enhanced signal-to-noise. These data demonstrate an equilibrium of Leu12Hse(C_4F_9)

binding to ERα LBD via broadening in the presence of ERα LBD. The CSA in these data can be quantified using the full width, half height (FWHH) measurement performed in MNova (version 10). The FWHH is a measurement of the peak width. As anisotropy occurs, broadening is observed, leading to an increase in the measured FWHH. In Figure 2.35a, the FWHH measurement of the free Leu12Hse(C_4F_9) peak is 6 Hz. As the ERα LBD is titrated into the solution, the peak broadens to FWHH = 17 Hz at ERα-LBD = 5 µM vs FWHH = 18 Hz (Figure 2.35, panels c and d) at ERα LBD = 15 and 30 µM. At 5 µM ERα LBD, the Leu12Hse(C_4F_9) is not fully bound to the protein, based on K_d = 2.2 µM, and we therefore, expect, more CSA at higher concentrations of protein, when more Leu12Hse(C_4F_9) is bound. We can also quantify the rescued ^{19}F signal as a result of competition with the NRBoxII peptide (Figure 2.35, panel f) which has a FWHH = 8 Hz. Similary, OHT also rescues the ^{19}F signal by displacing the peptide as measured by (Figure 2.35, panel g) FWHH = 9 Hz. These data indicate the signal is nearly completely recovered as the Leu12Hse(C_4F_9) is displaced from the ERα LBD protein using either the control NRBoxII peptide (20 µM) or OHT (50 µM). We also demonstrate the specificity of this interaction by incubation of the Leu12Hse(C_4F_9) peptide with Bovine Gamma Globulin (BGG). BGG is a 155 kDa protein which does not have a protein-protein interaction interface. If non-specific interactions were the basis of the broadening observed by ^{19}F NMR, we would expect to see CSA as a result of incubation with the Leu12Hse(C_4F_9) peptide. In contrast, a sharp singlet was observed by ^{19}F NMR upon incubation with 30 µM BGG ,with FWHH = 6 Hz, similar to that observed in the free peptide.

Ultimately, the goal of this work is to apply perfluoro-*tert*-butyl amino acids in more complex solutions, including detection of protein function living cells. In order

158

to demonstrate the potential of Leu12Hse(C_4F_9) to detect protein-protein interactions in more complex systems, such as cells, the peptide was incubated in HeLa cell lysates, both in the presence and absence of the estrogen receptor. HeLa cells do not express the estrogen receptor, so purified ERα LBD was added to the HeLa cell lysates.[159] In more complex mixtures, such as cell lysates, the possibility for non-specific interactions increases substantially due to the increased number of proteins in solution. The Leu12Hse(C_4F_9) peptide in HeLa cell lysates did demonstrate some CSA as a result of non-specific interactions (FWHH = 10 Hz) compared to the peptide in buffer (FWHH = 6 Hz) (Figure 2.35, panel a and h). This broadening is more than was observed with the BGG (FWHH = 6 Hz), but still significantly less than observed with ERα LBD (FWHH = 18 Hz). Upon addition of 15 µM ERα LBD in HeLa cell lysates with 5 µM Leu12Hse(C_4F_9), significantly more CSA is observed (FWHH = 21 Hz). These data demonstrate that, despite a small amount of CSA as a result of more complex mixtures, most likely due to an increase in hydrophobic proteins in solution, ERα LBD•co-activator protein-protein interactions are detectable utilizing the Leu12Hse(C_4F_9) peptide in complex mixtures.
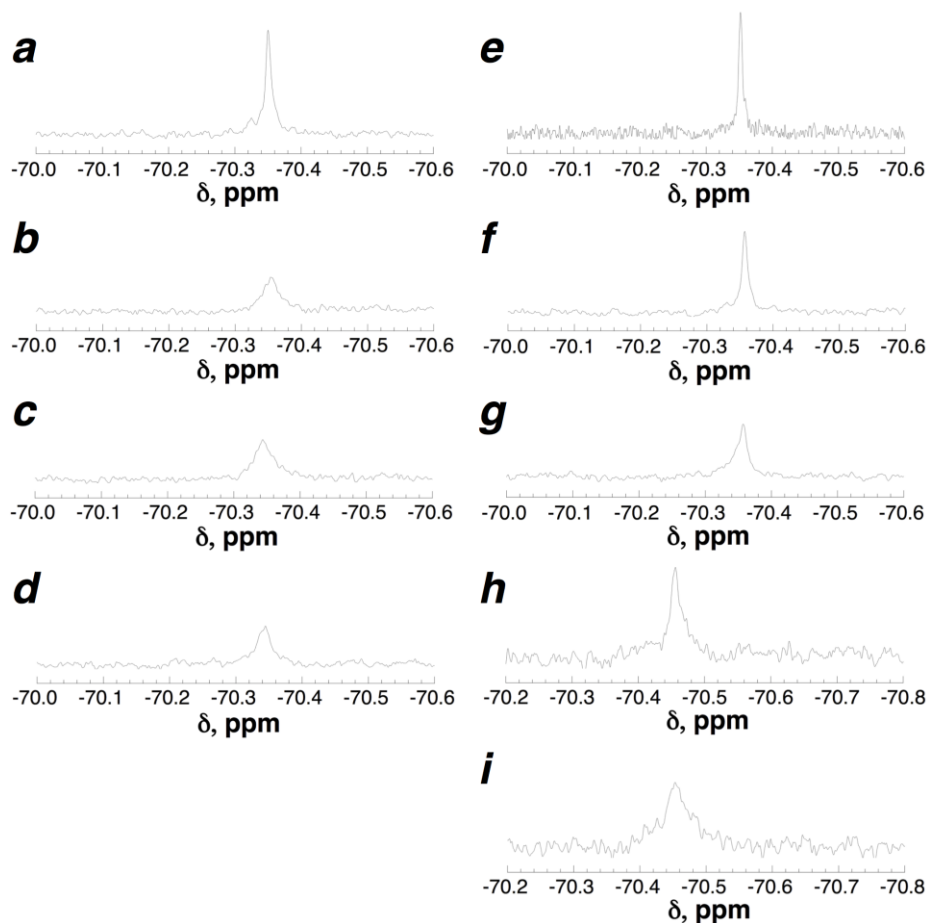
Figure 2.35 ^{19}F NMR spectroscopy of the peptide Leu12Hse(C$_4$F$_9$). Experiments were conducted with 5 µM peptide in 1× PBS (pH 7.4), 0.1 mM DTT, 20 µM estradiol (except g), and 10% D$_2$O. (a-e) The peptide Leu12Hse(C$_4$F$_9$) with (a) no added protein; (b) 5 µM ERα LBD; (c) 15 µM ERα LBD; (d) 30 µM ERα LBD; and (e) 30 µM bovine gamma globulin. (f) 5 µM peptide Leu12Hse(C$_4$F$_9$) with 15 µM ERα LBD and 20 µM NRBox II. (g) 5 µM peptide Leu12Hse(C$_4$F$_9$) with 15 µM ERα LBD and 50 µM 4-hydroxytamoxifen (OHT), without added estradiol. (h-i) The peptide Leu12Hse(C$_4$F$_9$) (h) with 200 µL HeLa cell lysates and (i) with 200 µL HeLa cell lysates and 15 µM ERα LBD (final concentration). All ^{19}F NMR spectra were acquired with a 10 ppm sweep width, acquisition time = 0.8 seconds, 1024 scans, and a relaxation delay of 3.0 seconds. Figures 2.35a-2.35g have identical scales on the y-axes.

2.3   Summary and Discussion

Protein-protein interactions are key steps in cell signaling pathways and are used to regulate protein activity; however, they are very difficult to detect and monitor in real time in complex mixtures. Despite the large number of methods developed to detect protein-protein interactions, new tools are needed to understand the complex role of these interactions in cell signaling pathways.[28,110,112] To begin to address this problem, we have developed novel amino acids to detect protein-protein interactions in complex mixtures by ^{19}F NMR. We have chosen to use both perfluoro-*tert*-butyl tyrosine and perfluoro-*tert*-butyl homoserine in key recognition motifs in the MDM2•p53 and estrogen receptor•co-activator protein-protein interfaces.[50,133,136,143,152]

Perfluoro-*tert*-butyl tyrosine was introduced at the Phe (*i* position within the recognition α-helix) and the Trp (*i*+4 position within the recognition α-helix) within the p53-P27S peptide to determine the ability of the MDM2 hydrophobic binding pocket to accommodate the perfluoro-*tert*-butyl group. The P27S peptide is the highest affinity encodable peptide-based ligand for MDM2 known, with a $K_d = 4.7$ nM.[144] Substitution of the key hydrophobic residues, Phe and Trp, led to a substantial loss of binding ($K_d = 7.4$ μM, $\Delta\Delta G = 2.4$ kcal mol^{-1} and $K_d = 9.1$ μM, $\Delta\Delta G = 2.5$ kcal mol^{-1} respectively). These data demonstrate that perfluoro-*tert*-butyl tyrosine was accommodated with a substantial loss in affinity within the binding pocket of MDM2. There are a number of reasons this amino acid, despite being similar in size to Trp, cannot be incorporated into the P27S sequence while maintaining nanomolar binding affinity. Despite fluorine's superior hydrophobic character, both Phe and Trp are key recognition resides within the p53 sequence. In the case of Phe to Tyr(C$_4$F$_9$), the perfluoro-*tert*-butyl tyrosine is most likely too large to be fully accommodated within

161

the binding pocket without substantial reordering of MDM2. Examination of the crystal structure shows very little space in the binding pocket to accommodate a larger amino acid.[143] Furthermore, the side chains of both the Phe and the Trp point towards each other, with less than a 3 Å distance between the side chains of the two residues. This small distance may hint towards disruption of the α-helix, due to steric clash between the residues, required for binding at the interface. In the case of Trp, contacts are made with 10 different residues in the binding pocket.[143] Perfluoro-*tert*-butyl tyrosine is potentially disrupting many of these interactions, preventing high affinity binding. Furthermore, the shape and geometry of perfluoro-*tert*-butyl tyrosine may be sufficiently different from Trp that it is poorly accommodated at this position. Indeed, energy minimization suggests that the perfluoro-*tert*-butyl ether is substantially out of the plane of the aromatic ring.

While a new high affinity ligand was not successfully designed to target the MDM2•p53 protein-protein interaction, there is still potential, moving forward, for perfluoro-*tert*-butyl tyrosine to be useful in identifying protein-protein interactions. MDM2•p53 is just one model system utilizes large aromatic amino acid residues as recognition residues in protein-protein interactions. Another example is the androgen receptor, which binds to a co-activator through an FXXL/(F/W) motif.[130,134] One potential reason the perfluoro-*tert*-butyl tyrosine was not accommodated within the MDM2•p53 interaction was due to two large aromatic amino acids being utilized in the binding pocket. While we briefly pursued the androgen receptor, insufficient quantities of protein were produced in tests of expression of the androgen receptor, to measure binding dissociation constants, preventing evaluation of peptides containing perfluoro-*tert*-butyl tyrosine in this system.

While we were unable to design high affinity ligands for the MDM2•p53 interaction, we were able to develop an estrogen receptor binding peptide containing perfluoro-*tert*-butyl homoserine. We examined a series of peptides which incorporated perfluoro-*tert*-butyl homoserine at the *i*-1, *i*, and *i*+3 positions within the **ILXXLL** α-helical binding motif. We were able to demonstrate that substitution was accepted at the *i*+3 position (K_d = 2.2 µM) with near-native binding affinity when compared to the NRBoxII peptide control peptide (K_d = 1.0 µM).[132,136,137] These data, combined with both CD and NMR analysis, showed very little perturbation to the native peptide structure with the perfluoro-*tert*-butyl homoserine modification. The native co-activator binding sequence is intrinsically disordered, only adopting the recognition α-helix while binding to the estrogen receptor. Despite the disordered nature of the peptide, α-helical structures can be adopted by the perfluoro-*tert*-butyl homoserine, as demonstrated by CD data in 30% TFE as well as in the high affinity binding demonstrated by the Leu12Hse(C_4F_9) peptide.

Furthermore, we were able to successfully demonstrate detection of the protein-protein interaction between Leu12Hse(C_4F_9) and the estrogen receptor by ^{19}F NMR. Chemical shift anisotropy (CSA) was observed as the estrogen receptor was titrated into a solution of 5 µM Leu12Hse(C_4F_9) peptide by ^{19}F NMR. The CSA was quantified using FWHH measurements to quantify the amount of broadening observed. We also demonstrated upon competitive inhibition of binding, by either the NRBoxII peptide or OHT, that the ^{19}F signal associated with the peptide Leu12Hse(C_4F_9) was recovered, demonstrating that the observed broadening is a direct result of peptide binding to the estrogen receptor. This interaction was also

163

demonstrated to be specific to the estrogen receptor, as no CSA was observed upon incubation with bovine gamma globulin.

Finally, we were able to demonstrate the potential to detect the estrogen receptor•co-activator interaction in complex solutions by observing the same ^{19}F signal broadening phenomenon in simple aqueous buffered solutions and more complex mixtures. The peptide Leu12Hse(C_4F_9) was incubated in HeLa cell lysates, both in the presence and absence of the estrogen receptor. While more CSA was observed in the HeLa cell lysates without the estrogen receptor than in aqueous buffer (FWHH = 10 Hz vs FWHH = 6 Hz, respectively), demonstrating an increased amount of non-specific binding in complex solutions, binding was still clearly visible by ^{19}F NMR CSA in the presence of 15 µM estrogen receptor (FWHH = 21 Hz). These data suggest the possibility of perfluoro-*tert*-butyl homoserine to be used to detect protein-protein interactions in real time in live cells.

The next logical step for the Leu12Hse(C_4F_9) sensor peptide is to introduce it into cells to determine the activity in a living system, for both imaging and therapeutic potential. Short term goals for this work include demonstrating the potential to image the estrogen receptor•co-activator interaction in live cells by ^{19}F NMR. The Leu12Hse(C_4F_9) peptide could be introduced into cells via conjugation with poly-arginine cell penetrating peptides.[160] The Leu12Hse(C_4F_9) could be conjugated to the poly-Arg sequence via a disulfide bond which would transport the Leu12Hse(C_4F_9) peptide into the cell.[160] The disulfide bond would be reduced inside the cell, leaving the Leu12Hse(C_4F_9) peptide inside the cell. The cells could then be imaged by ^{19}F magnetic resonance to determine if CSA could be used to image binding to the estrogen receptor inside cells. Fluorescence microscopy could also be used to

determine localization of the Leu12Hse(C_4F_9) peptide inside the cells. Most of the estrogen receptor protein is found within the nucleus of the cell, and other studies with small synthetic LXXLL motif-containing peptides have demonstrated these small peptides localize into the nucleus as they bind to the estrogen receptor, down-regulating transcription.[50] These experiments would be carried out with the Leu12Hse(C_4F_9) peptide to determine localization of the peptide within the cell.

We would also like to demonstrate down-regulation of transcription due to binding within the cell. While we have demonstrated that peptides containing perfluoro-*tert*-butyl homoserine can bind to the estrogen receptor and ^{19}F magnetic resonance can be used to detect this interaction, we would also like to demonstrate its usefulness as a potential therapeutic agent. The Leu12Hse(C_4F_9) peptide bound to the estrogen receptor would block the binding pocket, preventing the co-activator from binding, downregulating transcription. Estrogen receptor-mediated transcription can be monitored via transcription assays.[50] In short, the peptide is introduced into the cell and transcription is measured using a reporter gene assay to quantify the amount of transcription observed as a result of estrogen receptor-mediated transcription. We expect to see less estrogen receptor-mediated transcription as a result of the Leu12Hse(C_4F_9) peptide, which has been seen with other synthetic peptides.[50]

Another potential future direction for this work is the adaptation of the perfluoro-*tert*-butyl moiety in small molecule drug design to target protein-protein interactions. There is significant interest in targeting protein-protein interactions of the estrogen receptor as a potential therapeutic for the treatment of breast cancer. There are a large number of designed small molecules in the literature which demonstrate binding to protein-protein interfaces via mimicry of α-helical structures.[137,146,147,161]

These small molecule scaffolds utilize hydrophobic cores, like aromatics, with modifications to mimic side chains in the same geometry as an α-helix. The perfluoro-*tert*-butyl ether could potentially be introduced into one of these molecules to mimic the $i+3$ position of an α-helix to target the estrogen receptor co-activator binding pocket. The introduction of the perfluoro-*tert*-butyl ether into small molecules increases not only their hydrophobicity, which would be beneficial for binding, but also could be tracked by ^{19}F magnetic resonance.

There are also multiple directions to improve upon the peptide Leu12Hse(C_4F_9) design for higher affinity binding and biological activity of this peptide when interacting with the estrogen receptor. While Leu12Hse(C_4F_9) was the highest affinity binding peptide, there is potential for improvement of this peptide by utilizing the perfluoro-*tert*-butyl homoserine at the Ile_8 position. Previous work demonstrates higher affinity binding with the addition of a neopentylglycine at the $i+3$ position.[146] The data presented in Figure 2.19 demonstrate that the perfluoro-*tert*-butyl homoserine is well tolerated at the i-1 position, which would allow for modification with both the perfluoro-*tert*-butyl homoserine as a ^{19}F imaging handle, while still gaining increased binding affinity due to the $i+3$ neopentylglycine modification. Alternatively, the estrogen receptor binds to the ERE on the DNA to upregulate transcription as a dimer, allowing for binding of two co-activators.[130,135,136] The dimer could be targeted in a bidentate by linking two of the Leu12Hse(C_4F_9) peptides together via flexible linkers such as polyethylene glycol linkers. These types of systems, utilizing two linked peptides, would allow for binding in a bimolecular fashion, utilizing synthetic peptides. These types of bimolecular systems are also possible via expression. Utilizing unnatural amino acid expression of the perfluoro-

166

tert-butyl homoserine, peptide-based linkers could be used, such as a Gly-based linker, to create expressible fusion proteins which could target the estrogen receptor in a bidentate fashion.[162]

Finally, there is potential for amino acid containing perfluoro-*tert*-butyl group to be used broadly in the detection of protein-protein interactions. While this work specifically targeted a select few α-helical sequences, there are a large number of key protein-protein interactions which could be addressed in this fashion. One potential target are the SH3 domains, which utilize a PPII helical recognition sequence PXXP for protein-protein interactions (Figure 2.38).[84,126,127] These interactions are particularly difficult to target because the binding pocket is very shallow and utilizes small aliphatic amino acids for recognition. Furthermore, SH3 domains are critical in a number of cell signaling pathways, acting as key regulatory domains for important proteins such as proto-oncogene tyrosine-protein kinase Src (Src), the growth factor receptor-bound protein 2 (Grb2), and the proto-oncogene c-Crk (Crk).[84,124,126,128] These sequences could potentially utilize perfluoro-*tert*-butyl homoserine or 4*R*-perfluoro-*tert*-butly hydroxyproline at one of the critical recognition sites. Both of these amino acids readily undergo formation of a PPII helix (Figure 1.26). A potential ligand for these interactions includes the mSOS ligand, which has a known sequence and known low to mid-micromolar binding to SH3 domains, including Src, Grb, and Crk (Figure 2.39).[124,163] These interactions are based on the hydrophobic nature of Pro and Val, with the SH3 domain utilizing a shallow hydrophobic binding pocket.[126] We believe we can gain increased specificity for specific SH3 domains, such as Src over Crk, utilizing perfluoro-*tert*-butyl amino acids in the binding motif due to improved hydrophobic interactions via the perfluoro-*tert*-butyl ether. By utilizing a number of

substrates with mutations to an amino acid containing a perfluoro-*tert*-butyl ether at different locations, we may be able to achieve increased specificity for certain SH3 domains over others due to small variations in the binding pocket.
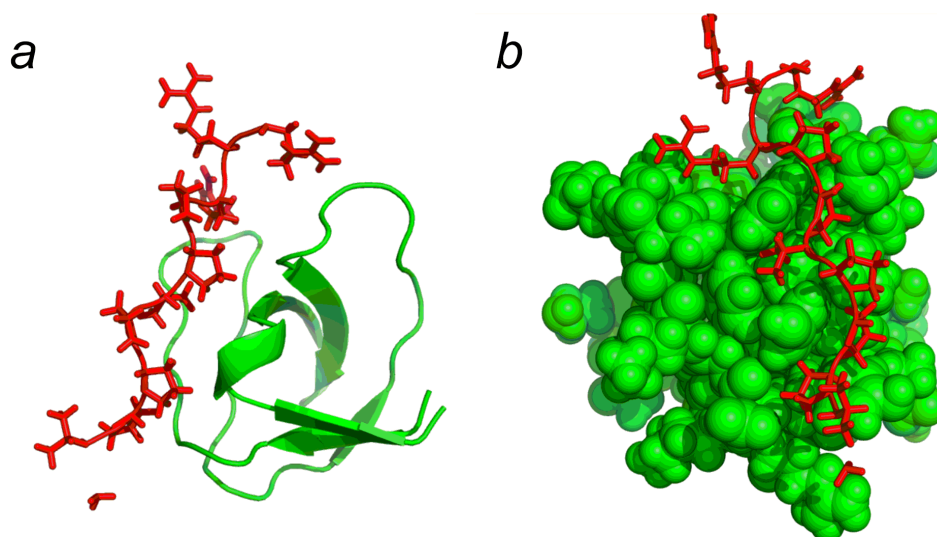
Figure 2.36 (a) Solution NMR structure of Grb2 (green) complexed with the mSOS-derived ligand (red) (1GBQ). (b) Spacefill model of Grb2 complexed with the mSOS ligand demonstrating the hydrophobic binding pocket interacting with proline residues.[164]

mSOS	YEVPPVPPRA
Pro5Hyp(C_4F_9)	YEVP**P**VPPRA
Pro8Hyp(C_4F_9)	YEVPPVP**P**RA
Val3Hse(C_4F_9)	YE**V**PPVPPRA
Val6Hse(C_4F_9)	YEVPP**V**PPRA
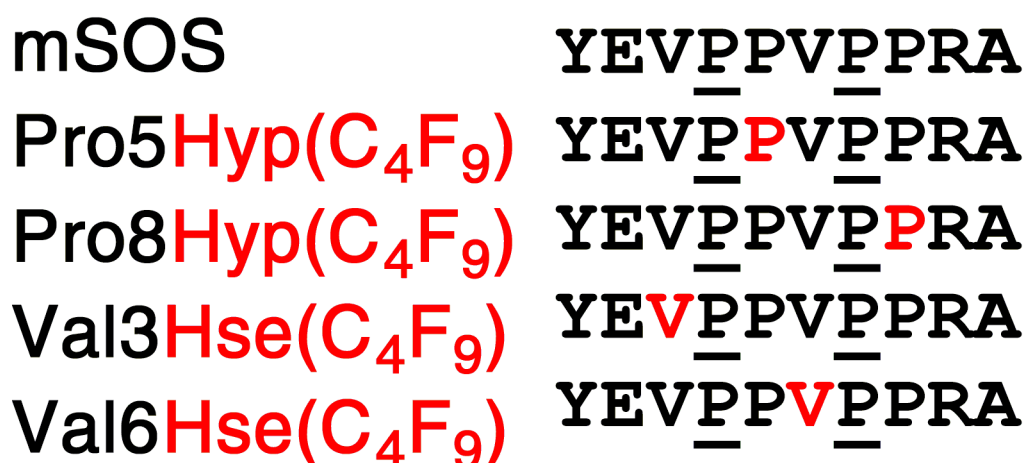
Figure 2.37 Proposed mSOS-based ligands for targeting SH3 domain protein-protein interactions via amino acids containing perfluoro-*tert*-butyl ether.

2.4   Experimental

2.4.1   Peptide Synthesis and Purification

Peptides containing perfluoro-*tert*-butyl homoserine (Hse(C_4F_9)) were synthesized on NovaGel PEG-polystyrene graft Rink amide resin (EMD Millipore, 0.67 mmol/g) by standard solid-phase peptide synthesis using HATU as a coupling reagent on 0.1 mmol scale. Control NRBoxII peptides were synthesized on Rink amide resin (0.70 mmol/g) on 0.1 or 0.25 mmol scale using HBTU as a coupling reagent. Amide coupling reactions with perfluoro-*tert*-butyl homoserine (4 equiv.) were allowed to proceed for 24 hours at room temperature. Subsequent amide coupling reactions were conducted as double coupling reactions (2 hours for the first coupling, 1 hour for the second coupling). All peptides were acetylated on the N-terminus (5% acetic anhydride in pyridine, 3 mL, 3 × 5 min) and contained a C-terminal amide. Peptides were subjected to cleavage from the resin and deprotection for 3 hours using 1 mL reaction volume of 90% TFA with 5% H_2O and 5%

triisopropylsilane. TFA was removed by evaporation, and the peptides were subsequently dissolved in 1 mL 500 mM phosphate buffer (pH 7.2). Peptides containing a cysteine were maintained in a reduced state by addition of approximately 2 mg DTT. Crude peptide solutions were filtered using a 0.45 µm syringe filter before injection on the HPLC. Peptides were purified by reverse phase HPLC (Vydac semipreparative C18, 10 × 250 mm, 5 µm particle size, 300 Å pore) using a 60 minute linear gradient of 0 to 60% buffer B (20% H_2O, 80% acetonitrile, 0.05% TFA) in buffer A (98% H_2O, 2% acetonitrile, 0.06% TFA). Peptide purity was determined by the presence of a single peak by reinjection on analytical HPLC. Analytical data for peptides: Ac-CLTERHKILHRLLQE-NH$_2$ [t_R 35.4 min; expected mass, 1929.1, observed mass, 1930.9 $(M + H)^+$]; Ac-CLTERHK(Hse(C$_4$F$_9$))LHRLLQE-NH$_2$ [t_R 37.9 min; expected mass, 2137.0, observed mass, 1068.8 $(M + H)^{2+}$]; Ac-CLTERHKI(Hse(C$_4$F$_9$))HRLLQE-NH$_2$ [t_R 45.7 min; expected mass, 2137.0, observed mass, 1068.8 $(M + H)^{2+}$]; Ac-CLTERHKILHR(Hse(C$_4$F$_9$))LQE-NH$_2$ [t_R 42.7 min; expected mass, 2137.0, observed mass, 1068.8 $(M + H)^{2+}$]; Ac-CLTERHKI(hyp(C$_4$F$_9$))HRLLQE-NH$_2$ [t_R 39.9 min; expected mass, 2148.0, observed mass, 1075.0 $(M + H)^{2+}$] Ac-LTERHKILHRLLQE-NH$_2$ [t_R 32.7 min; expected mass, 1826.1, observed mass, 914.4 $(M + H)^{2+}$]; Ac-LTERHKILHR(Hse(C$_4$F$_9$))LQE-NH$_2$ [t_R 41.5 min; expected mass, 2032.0, observed mass, 1017.4 $(M + H)^{2+}$].

2.4.2   Fluorescein Labeling of Peptides

Peptides were labeled on the N-terminal cysteine using 5-iodoacetamidofluorescein (5-IAF) (Sigma-Aldrich). Purified peptides were dissolved in 250 mM phosphate buffer pH 4.0 (200 µL) and 10 mg/mL 5-IAF in DMF (200 µL) and incubated in the dark at room temperature for 90 minutes. The reaction mixture

was diluted with water and filtered before purification by HPLC. Peptides were purified by reverse phase HPLC (Vydac semipreparative C18, 10 × 250 mm, 5 μm particle size, 300 Å pore) using a 60 minute linear gradient of 0 to 60% buffer B (20% H_2O, 80% acetonitrile, 0.05% TFA) in buffer A (98% H_2O, 2% acetonitrile, 0.06% TFA) for fAc-CLTERHKILHRLLQE-NH_2 and a 60 minute linear gradient of 15 to 65% buffer B in buffer for all peptides containing perfluoro-*tert*-butyl homoserine. Peptide purity was determined by the presence of a single peak by reinjection on analytical HPLC. Analytical data for peptides: fAc-CLTERHKILHRLLQE-NH_2 [t_R 42.3 min; expected mass, 2316.4, observed mass, 1159.4 $(M + H)^{2+}$]; fAc-CLTERHK(Hse(C_4F_9))LHRLLQE-NH_2 [t_R 37.8 min; expected mass, 2524.3, observed mass, 1262.3 $(M + Na)^{2+}$]; fAc-CLTERHKI(Hse(C_4F_9))HRLLQE-NH_2 [t_R 41.6 min; expected mass, 2524.3, observed mass, 1262.3 $(M + Na)^{2+}$]; fAc-CLTERHKILHR(Hse(C_4F_9))LQE-NH_2 [t_R 40.2 min; expected mass, 2524.3, observed mass, 1262.3 $(M + Na)^{2+}$]; fAc-CLTERHKI(hyp(C_4F_9))HRLLQE-NH_2 [t_R 39.6 min; expected mass, 2536.3, observed mass, 1268.4 $(M + H)^{2+}$].

2.4.3   Circular Dichroism of Peptides

Circular dichroism (CD) experiments were conducted on a Jasco model J-810 spectropolarimeter. NRBoxII and Leu12Hse(C_4F_9) peptides were dissolved in 30% TFE with 5 mM phosphate buffer and 25 mM KF at pH 7.4. Data represent the average of at least 3 independent trials. Error bars indicate standard error. Data were background-corrected but were not smoothed.

2.4.4 NMR Characterization of Peptides

Peptide NMR experiments were performed on a Brüker 600 MHz (^{19}F 564.5 MHz) NMR spectrometer equipped with a 5-mm Brüker SMART probe in water with 5 mM phosphate buffer (pH 4), 25 mM NaCl, and 10% D_2O. TSP was used as a standard for referencing in ^{1}H NMR. All 1-D NMR spectra were recorded with a Brüker w5 watergate pulse sequence with a 2 to 3 second relaxation delay. TOCSY spectra were recorded using a watergate TOCSY pulse sequence. ^{1}H-^{13}C HSQC spectra were recorded in D_2O with 512 t1 increments with 4096 data points and 16 scans in each t1 increment. The data were acquired in phase-sensitive mode using echo/antiecho-TPPI gradient selection with decoupling during acquisition. HSQC spectra were acquired with a 1.75 second relaxation delay. All NMR spectra were recorded at 298 K.

2.4.5 Protein Expression and Purification

Plasmid encoding the human estrogen receptor ligand-binding domain (ERα LBD) (amino acids 255-595) was obtained from the Koh lab at the University of Delaware. The ERα LBD expression construct was derived from pET-15b and contained a Hexa-Histidine tag. The plasmid was transformed into BL21(DE3) pLysS competent cells (Novagen) using heat shock and grown overnight at 37 °C on luria broth (LB) containing 50 mg/L ampicillin (Amp) media. A single colony was selected and grown overnight at 37 °C in a 100 mL LB/Amp culture. Terrific broth expression cultures were inoculated with 25-30 mL of the overnight culture and grown to an optical density of 0.6 to 0.8. Cultures were induced with 1 mM IPTG for 5 hours before centrifugation. Bacterial cell pellets were frozen at –20 ˚C until protein

purification. ERα LBD was purified using His-Bind resin (EMD Millipore) according to the manufacturer's directions, with 200 μM PMSF and 10 μM estradiol added to the binding buffer. Protein identity was verified by SDS-PAGE and MALDI MS. Protein eluents were dialyzed into phosphate-buffered saline (PBS) at pH 7.4 with 5 mM EDTA and 0.5 mM DTT, and protein concentration determined via Bradford assay.

2.4.6 Fluorescence Polarization Assays

Peptides and ERα LBD were diluted in $1\times$ PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM K_2HPO_4, 2 mM KH_2PO_4). Peptides were quantified by absorbance at 492 nm using a fluorescein extinction coefficient of 83,000. Fluorescence polarization assays were conducted in 96 well round-bottom black opaque plates (Costar) using two-fold serial dilutions of ERα LBD, to final protein concentrations from 0.0098 μM to 10 μM, in PBS containing 0.1 mM DTT, 20 μM 17ß-estradiol, 0.04 mg/mL BSA, and 100 nM fluorescein-labeled peptide. Plates were incubated in the dark for 30 minutes at room temperature before reading on a Perkin-Elmer Fusion plate reader using a 485 nm fluorescein excitation filter and a 535 nm emission filter with polarizer. All data are the average of at least three independent trials. All dissociation constants (K_d) were determined using a nonlinear least-squares curve fit (KaleidaGraph 4.1) of the K_d to the fluorescence polarization data, via a binding equation for 1:1 complex formation. All data were fit using equation 1, where $p_{min} =$ the polarization value without protein present, $p_{max} =$ polarization at saturation, and the data fit to calculate K_d.

(1)

$$\text{polarization} = p_{min} + (p_{max} - p_{min}) \times \frac{[\text{ligand}]_{total} + [\text{protein}]_{total} + K_d) - \sqrt{([\text{ligand}]_{total} + [\text{protein}]_{total} + K_d)^2 - 4[\text{ligand}]_{total}[\text{protein}]_{total}})}{2[\text{ligand}]_{total}}$$

2.4.7 ^{19}F NMR Spectroscopy of Peptides Bind with the Estrogen Receptor

Peptide concentrations were quantified by NMR using 1 mM maleic acid as a standard. Peptides and protein were diluted in 1× PBS (140 mM NaCl, 2.7 mM KCl, 10 mM K_2HPO_4, 2 mM KH_2PO_4) pH 7.4 with 0.1 mM DTT and 20 μM 17ß-estradiol. For peptide competition experiments, NMR spectra were recorded in the absence of the NRBoxII peptide. The NRBoxII peptide was added (< 5 μL) and the resultant solution was allowed to equilibrate for a minimum of 30 minutes at room temperature before the NMR experiment was conducted. For experiments using 4-hydroxytamoxifen (OHT), 50 μM OHT was used instead of estradiol. The fluorinated peptide concentration in all NMR experiments was 5 μM. All NMR samples contained 10% D_2O. Solutions were incubated at room temperature for at least 30 minutes to allow for equilibration. Data were collected on a Brüker 600 MHz (^{19}F 564.5 MHz) NMR spectrometer equipped with a 5-mm Brüker SMART probe without the use of a proton decoupler. A 0.8 second acquisition time was used with a 10 ppm sweep width.

Chapter 3

DETECTION OF POST-TRANSLATIONAL MODIFICATIONS VIA 19F NMR

3.1 Introduction

Post-translational modifications are made to proteins to alter their structure and function after the protein has been expressed within a cell. There are many different types of post-translational modifications; however, they generally all involve a covalent modification to the protein and often require an enzyme to install the post-translational modifications.[165] A few examples of post-translational modifications include phosphorylation, glycosylation, and proteolysis.[166-169] Post-translational modifications are generally longer lived than the protein-protein interactions discussed in Chapter 2; however, like protein-protein interactions they are involved in cell signaling pathways, often regulating a protein's ability to participate in cell signaling pathways. While not all post-translational modifications are permanent, as is proteolysis for example; they do often require an enzyme to reverse the modification.[165] Currently, most post-translational modifications are identified through indirect methods.[49,170-176] In order to gain a full picture of cell signaling pathways, direct probes are needed to study enzymes *in real time*.

One common, well-studied example of post-translational modifications is phosphorylation. There are over 500 known protein kinases within the human genome, accounting for approximately 2% of mammalian genomes, which are critical in regulation of cell signaling pathways.[166] Proteins are phosphorylated on a Ser, Thr, or Tyr residues by protein kinases, and can be dephosphorylated by phosphatases (Figure

3.1). Kinases are generally divided into two classes: Tyr kinases and Ser/Thr kinases based on the amino acids which they phosphorylate.[175] These post-translational modifications are key steps in cell signaling pathways which control cell growth, death, metabolism, and cellular activity.[167] Misregulation of phosphorylation events are often seen in cancer pathways, diabetes, and neurological diseases such as Alzheimer's disease.[166] New tools to study phosphorylation events are necessary to understand these pathways and how they are regulated in both healthy and unhealthy cells. Furthermore, imaging probes which detect kinase activity can give cell specific information for disease diagnostics.
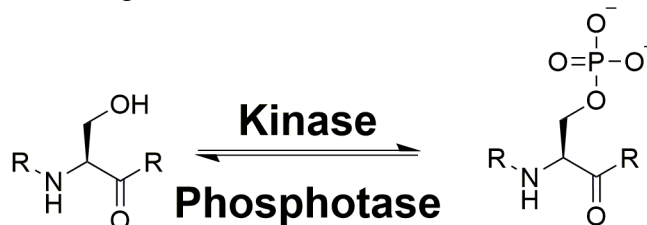
Figure 3.1   Serine (shown here), threonine, and tyrosine can all be phosphorylated by kinases and dephosphorylated by phosphatases.

3.1.1   Methods for the Detection of Post-translational Modifications

There are a number of methods to detect post-translational modifications; however, the vast majority of them are indirect methods. One of the most common techniques for detecting post-translational modifications is Western blotting.[167,173,177-180] Cells are lysed for Western blot analysis and gel electrophoresis is performed on cell lysates. The separated cell lysates are then transferred to a blotting membrane. The membrane is then incubated with a primary antibody which binds to the protein of interest containing the post-translational modification of interest. A secondary antibody is then used to bind to the primary antibody, which is bound to the target

protein. The secondary antibody contains a detection method, such as horseradish peroxidase, which allows visualization of the protein of interest.[181] Despite Western blotting being a standard technique for the detection of kinase activity, it can be problematic. Due to the very small epitope, often a single phosphorylated residue, being detected by the antibody, false positives are common. The high rate of false positives is a result of the detection of a single residue epitope; however, detection at a single residue to necessary, especially in proteins which can contain multiple phosphorylation sites. While the detection limit of a Western blot is in the nanogram regime, it is difficult to quantify the amount of phosphorylation from one Western blot to another. Furthermore, samples must be processed via lysis prior to gel electrophoresis, which can lead to dephsophorylation of the protein of interest, as not all phosphorylation sites are stable under cell lysis and denaturing conditions.[182]

Similar to Western blot analysis, enzyme-linked immunosorbent assays (ELISA) can be used to identify phosphorylation of a target protein. This method utilizes similar antibodies to Western blotting; however, they are adsorbed to a surface which the cell lysate is washed over.[171,173,183-185] There are multiple different detection methods for ELISA, the most common of which are fluorescent or colorimetric. A few common fluorescent dyes utilized are 4-methylumbelliferone, hydroxyphenylacetic acid, and 3-p-hydroxyphenylproprionic acid. A few common colorimetric dyes include 3,3',5,5' tetramethylbenzidine, 3,3',4,4' diaminobenzidine, and 4-chloro-1-naphthol. The dyes react with the bound antibodies to produce either a fluorescent or colorimetric readout, demonstrating the presence of the phosphorylation event on the protein of interest. These types of assays can be calibrated with a standard, leading to more quantitative data than Western blotting; however, samples still require

processing which can result in a loss of phosphorylation.[183] Furthermore, neither of these techniques can be used in real time to monitor kinase activity directly in living systems.

Another common technique for studying kinase activity, as well as post-translational modifications in general, is mass spectrometry.[170,175,182,186] post-translational modifications lead to a change in mass of a protein which makes monitoring changes in mass an excellent method to detect Post-translational modifications; however not all post-translational modifications are easily detected. For example, this technique is commonly used to identify phosphorylated proteins via digestion and analysis of the resultant peptides. Often an enrichment step is required to remove fragments which do not contain any post-translational modifications. Furthermore, this enrichment step can often prove to be unreproducible and can compromise the integrity of the sample.[187,188] This analysis, however, can be complex due to poor ionization of phosphorylated peptides using electrospray ionization as well the phosphate group ionizing with a negative charge rather than a positive charge.[182] This technique also can result in the loss of phosphorylation and can also be problematic for identifying multiple sites of post-translational modifications.[188] Phosphoproteomics is one of the more quantitative methods of detection for phosphorylation. There are a number of techniques including isotope enrichment, chemical labeling, as well as label-free methods which can be used for sample quantification.[188]

Alternatively, a few different methods for detecting kinase activity in living cells have been identified.[171,189,190] One such example uses fluorescent proteins and FRET due to phosphorylation to identify kinase activity (Figure 3.2). In this case, cyan

fluorescent protein (CFP) with a PKA phosphorylation site was expressed as a fusion

protein with a yellow fluorescent protein (YFP) connected with a phosphate binding

protein, 14-3-3τ. Upon phosphorylation, the 14-3-3τ domain would bind to the

phosphate binding protein, which brought the CFP close enough to the YFP for a

Förster resonance energy transfer to occur, increasing the fluorescent signal by

approximately 20 to 50% (Figure 3.2).[191] A similar probe was developed for Akt, with

FHA2 utilized as the phosphate-binding protein.[192] The Akt sensor showed smaller

fluorescence changes than the PKA sensor, with only a 10% to 30% fluorescence

change. Tyrosine kinases can also be detected utilizing these types of FRET-based

sensors. Src, Abl, and EGF receptor sensors were all developed utilizing the same

fluorescent proteins and the phosphotyrosine binding protein.[174] The tyrosine kinase

sensors demonstrated a 20% to 35% change in fluorescence. While the tyrosine kinase

probes have been demonstrated both *in vitro* and *in vivo* to detect tyrosine

phosphorylation, they require the expression and use of large proteins. These sensors

have successfully been expressed in cell culture for fluorescence imaging of live cells.

The benefit of these sensors is they are able to address questions in the lab to

determine the location of signaling and the compartmentalization of kinase activity.[174]

Furthermore, some of these sensors, specifically the tyrosine kinase sensors, have been

utilized in living cells to detect kinase activity in real time. There are, however, a

number of drawbacks to this technology. Fluorescence changes as a result of these

proteins are very small, leading to minimal changes in fluorescence which are difficult

to detect. Both CFP and YFP will absorb and emit light, independent of the FRET

signal. The amount of background fluorescence produced by this system leads to very

small changes, approximately 10 to 50%, depending on the exact system used, in

fluorescence because of the FRET signal, making detection difficult. Furthermore, it cannot be used within animal models and in a clinical setting. Ideally, new smaller probes which are applicable to cells, pre-clinical, and clinical imaging for detection of kinase activity are in development.
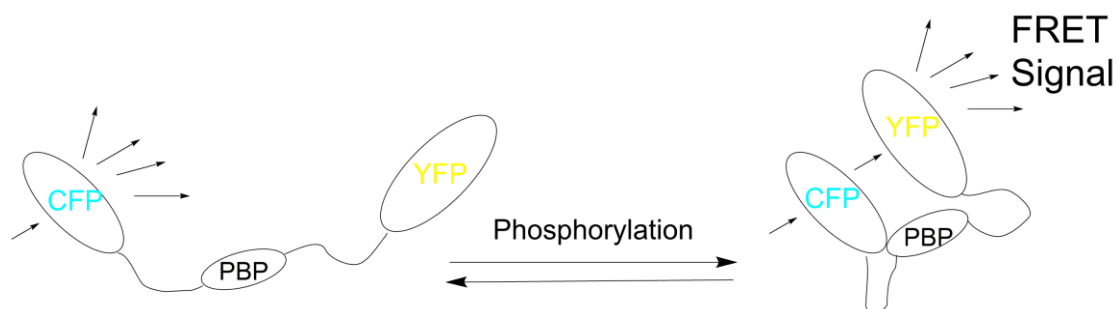
Figure 3.2   A schematic representation of fluorescent protein-based sensors for kinase activity. The phosphate-binding protein (PBP) contains a known recognition sequence for a kinase, which upon phosphorylation can bind to the phosphate binding protein (PBP). This brings the tethered YFP into the Förster radius to produce a FRET interaction between the two fluorescent proteins.

Another example of a fluorescent kinase sensor are the protein-kinase inducible domains (pKIDs) were developed by Balakrishnan and Zondlo to detect protein kinase activity via small peptides.[189,193] These peptides are modeled after EF-hand proteins, which are naturally occurring small calcium-binding proteins.[194] These proteins were exploited by the Imperiali group in the form of lanthanide binding tags (LBTs) which utilize the same recognition sequence to bind lanthanides instead of calcium, which have a similar atomic radius as calcium (Figure 3.3).[195] The key design element utilized in LBTs is the incorporation of a position 7 Trp. The Trp residue

binds the lanthanide metal via the backbone carbonyl, allowing for a resonance energy transfer between the Trp and the metal, which produces a stronger fluorescent signal than the lanthanide metal alone. The pKID design incorporates these metal-binding residues while also incorporating recognition sequences for protein kinases. Critical amino acids for terbium binding include the Asp residues at positions 1, 5, and 9, as well as the Asn residue at position 3, which bind the terbium via the residue side chain. There are two other important design features found in the pKID peptides. The first is a Trp residue in place of the Tyr which allows for a resonance energy transfer with terbium when bound via the backbone carbonyl. The other is the mutation of the critical Glu at residue 12 to a Ser. When the Ser is not phosphorylated, it is a poor Glu mimic, preventing effective binding of the terbium. Upon phosphorylation by a protein kinase, in this example PKA, phosphoserine makes an excellent Glu mimic, allowing for binding of the terbium (Figure 3.4). This places the Trp close enough to the terbium for a resonance energy transfer. When the Trp is excited at 280 nm, it emits at 350 nm, which overlaps with terbium's excitation wavelength. In turn, the terbium emits luminescence at 544 nm.[196] While this system is small and potentially genetically encodable, imaging of live cells can be problematic. Most notably, the cells with the pKID sensor must be soaked in terbium solutions and irradiated with 350 nm light, which can cause substantial damage to the cells. Even utilizing multiphoton imaging to mimic 280 nm light, the laser required causes substantial cellular damage. Furthermore, this system also cannot be applied in pre-clinical or clinical settings due to the requirement for 280 nm light.

Figure 3.3   pKID-PKA sensor peptide based on the EF Hand peptide. The crystal structure of the LBT (PDB: 1TIJ) demonstrates the important Glu residue side chain interactions with the lanthanide binding interaction. The pKID design mutates one of these Glu residues to a Ser/phosphoSer to mimic Glu upon phosphorylation by PKA.
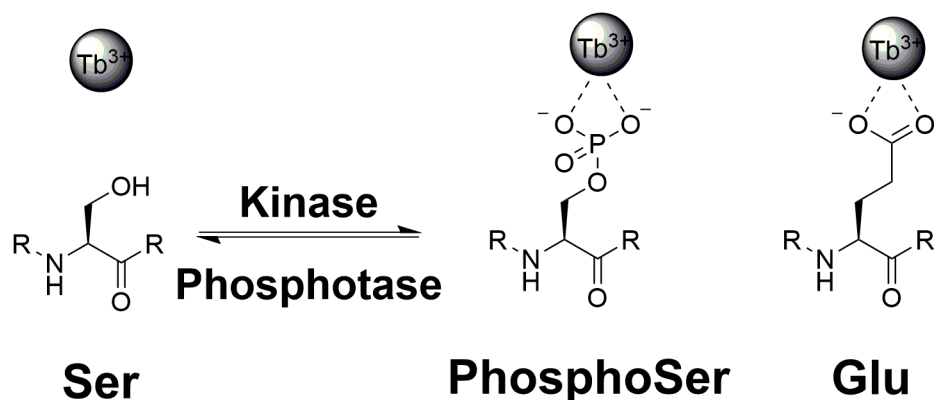
Figure 3.4   pKID sensors utilize the similar size and charge of phosphoserine to mimic Glu8 in peptides upon phosphorylation. Ser, however, is a poor Glu mimic so upon dephosphorylation, no luminescence is observed, as terbium binding is disrupted.

3.1.2 ^{19}F-Based Detection of Enzymatic Activity

While there are a large number of methods to detect enzymatic activity, such as many of the examples provided in the last section, we are interested in developing ^{19}F-based sensors for magnetic resonance detection of enzymatic activity. There are a

few examples of ^{19}F-based sensors which have previously been developed. These sensors span a large range of sizes from small molecules to large proteins to detect various enzymatic activities and post-translational modifications. One example of a small molecule-based ^{19}F sensor is 2-fluoro-ATP.[197] ATP, as previously discussed, is the currency of energy transfer within the cell. ATP is utilized by kinases as both an energy source and as a source of inorganic phosphate.[198] Stockman took advantage of the conversion of ATP to ADP to develop a small molecule-based ^{19}F sensors as shown in Figure 3.5. The chemical shift change from 2-fluoro-ATP to 2-fluoro-ADP is approximately 0.1 ppm.[197] The 2-fluoro-ATP was utilized by 3-phosphoinositide dependent kinase 1 for this specific example but is known to be accepted in a wide range of enzymes.[199] There are a number of drawbacks for this sensor. It cannot be used in complex media which contain more than one potential enzyme of interest, as a large number of enzymes utilize ATP. The sensor also only utilizes a single fluorine modification, requiring a high micromolar amounts of the fluorinated ATP, decoupled pulse sequences, and long acquisition times (8 to 16 mins). Finally, as demonstrated by Stockman, not all enzymes that utilize ATP as a substrate will accept 2-fluoro-ATP.
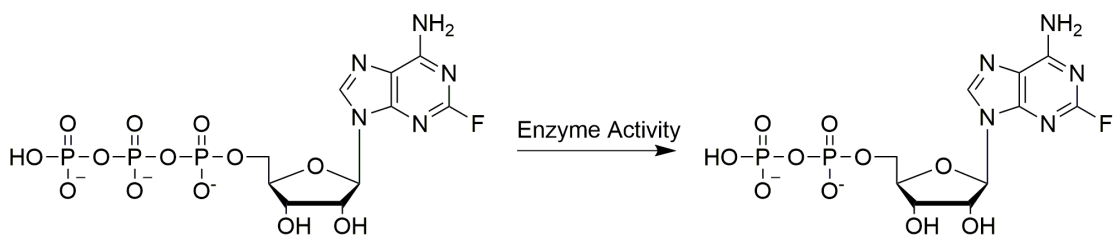
Figure 3.5   2-fluoro-ATP was utilized as a small-molecule ^{19}F-based sensor of kinase activity. A 0.1 ppm chemical shift change was observed by ^{19}F NMR upon enzymatic conversion from 2-fluoro-ATP to 2-fluoro-ADP.[197]

Peptides have also been utilized as ^{19}F-based sensors of enzymatic activity. Dalvit *et al*. introduced both an N-terminal trifluoro-acetyl group and a 3-trifluoro-Phe into a peptide substrate for both Akt and trypsin (Figure 3.6).[41] When both substrates were subjected to phosphorylation by ATP, the N-terminal label showed a 0.02 ppm chemical shift change from non-phosphorylated to phosphorylated. The substrate containing 3-trifluoromethyl-Phe, however, showed a 0.04 ppm chemical shift change upon phosphorylation. These results are unsurprising as the 3-trifluoromethyl-Phe is one position away from the phosphorylation site, while the N-terminal CF_3 modification is eight residues away from the phosphorylation site. Furthermore, the substrate containing 3-trifluoromethyl-Phe was subject to trypsin digestion. Upon cleavage with trypsin, the peptide showed a nearly 0.1 ppm chemical shift change by ^{19}F NMR. These data are encouraging; however, there is significant room for improvement. This system only utilizes three equivalent fluorines, which requires mid-micromolar concentrations and longer acquisition times (at least 6 mins).[41] Furthermore, 3-trifluoromethyl-Phe can only be introduced in positions which accept large aromatic amino acids. We hope to improve upon this design by introducing new probes which will be more sensitive as well as access a wider array of amino acid types (aromatic, aliphatic, and cyclic) for the use in enzyme substrates.

$$CF_3\text{-CO-ARKRERAYSFGHHA}$$

$$\text{ARKRERA(3-}CF_3\text{-F)SFGHHA}$$

Figure 3.6   The ^{19}F-based kinase and protease sensor peptides developed by Dalvit *et al*.[41] The fluorine modifications are shown in red and the phosphorylation site is shown in green.

3.1.3 Misregulation of Cell Signaling Pathways via Phosphorylation Events

There are a number of known phosphorylation events which are important in misregulation of cell signaling pathways leading to disease states. A few of these include cancers, neurodegenerative conditions including Alzheimer's Disease, and diabetes.[167,177,179,180,200,201] Key in understanding and treating these diseases are novel tools to study misregulation as well as imaging agents to detect specific cellular pathways, potentially leading to earlier detection than clinical symptoms would develop such as cognitive disturbances due to Alzheimer's Disease or solid mass tumors in cancers.[201] Although it is known that there are 500 genes which encode protein kinases, many of these enzymes are uncharacterized or minimally characterized, preventing a full understanding of the role of kinases and phosphorylation in many disease pathways. There is currently a large-scale effort to characterize the full human "kinome" in order to determine the full extent of phosphorylation within cellular pathways.[178] New tools are needed in order to fully understand the extent of phosphorylation within cellular pathways.

3.1.4 Protein Kinases are Therapeutic Targets for the Treatment of Diseases

Since misregulation of cell signaling due to phosphorylation events is known to be involved in disease pathways, it is unsurprising that kinases have become targets for the pharmaceutical industry.[202] One success story is that of Gleevec, which targets Abl kinase.[200] Abl kinase is a tyrosine kinase which is involved in cell proliferation. Some individuals carry a mutation which expresses Abl kinase as a fusion protein that makes it constitutively active, causing cells to grow and divide uncontrollably.[203] This pathway leads to chronic myelogenous leukemia (CML). 95% of patients with CML are known to have this mutation. Gleevec was introduced by Novartis in 2001 for

treatment of CML, which has revolutionized the treatment of CML, as well as introduced a new paradigm in cancer treatment by targeting protein kinases for treatment of cancer.[200] Gleevec is known to bind to the active site of Abl kinase, blocking any substrate from entering, preventing phosphorylation (Figure 3.7).[204] This drug is leading the way in developing novel therapeutic agents which disrupt cancer cell signaling pathways via targeting of kinase activity. The first step in this process, however, was identifying the roll of Abl kinase in CML proliferation pathways. In order to develop new therapeutics for the treatment of misregulated pathways in diseases, new tools are needed to study and understand cell signaling pathways including phosphorylation and other post-translational modifications.

Figure 3.7   The x-ray crystal structure (3GVU) of Abl kinase bound to
Gleevec (red), a potent inhibitor of kinase activity.[204,205]
Gleevec binds to Abl kinase via the catalytic binding pocket
blocking the substrate from entering the pocket, preventing
phosphorylation by blocking the substrate from entering the
kinase.

3.1.5   Protein Kinase A is a Well Documented System for Studying Protein Kinases

Protein Kinase A (PKA), also known as cAMP-dependent Protein Kinase, is

one of the most thoroughly studied protein kinases. It is known to have a variety of

targets within biological systems, involved in many different cellular pathways.[178]

PKA has been used as a model to study protein kinases as a class due to its simplicity and ease of handling. PKA contains one of four possible regulatory subunits and a catalytic subunit.[206] The regulatory subunits bind to cAMP, also known as adenosine 3',5'-cyclic monophosphate, which is a metabolic side product of ATP degradation and an important signaling molecule. The catalytic subunit docks to one of the regulatory subunits, which inactivates the kinase in the absence of cAMP. ATP is the primary "currency" of energy transfer within the cell.[198] Many different cellular processes utilize ATP and, in turn, cAMP to propagate signals intracellularly. PKA is found in nearly all organisms and is highly conserved which indicates it is potentially extremely important in cell signaling. PKA is also believed to be important in cell signaling, as it is controlled by cAMP, which is found in every cell.[207] PKA was also the first characterized mammalian second messenger system, but there are still many aspects of PKA signaling which have not been studied, including fully mapping PKA activities and phosphorylation sites within cellular pathways.[207] PKA is a robust system which can be studied *in vitro* in the absence of the regulatory domain (i.e. constitutively active) as well as ease of expression which has led to several crystal structures including apo and substrate-bound.

Unsurprisingly, there are many known targets in a cell signaling pathways which PKA phosphorylate. Some of these targets include pyruvate kinase, glycogen synthase kinase 3 (GSK-3), and various histone proteins, but there have been hundreds of identified targets since the discovery of PKA.[207] The recognition sequence for PKA phosphorylation has been well studied, examining all twenty amino acids as well as phosphoserine and phosphothreonine at all positions five residues prior to the phosphorylation site and four residues after the phosphorylation site. PKA utilizes a

recognition sequence containing basic residues, specifically strongly favoring arginine, at the -2 and -3 positions upstream from the phosphorylation site, either a serine or threonine.[208] PKA also favors an aliphatic residue at the +1 position from the phosphorylation site. Early studies on the PKA catalytic subunit led to the development of a small peptide substrate, deemed the Kemptide peptide, with the sequence Ac-LRRASLGAK-NH_2, which was optimized for PKA recognition and phosphorylation, to study PKA phosphorylation.[208] The Kemptide peptide is an ideal starting place to examine fluorinated amino acids for the detection of kinase activity by ^{19}F NMR.

3.1.6   Protein Kinase B Signaling Regulates Cell Survival and Proliferation

Akt is involved in apoptosis pathways, which regulate cell survival and proliferation. It is commonly invoked in cancer pathways as well as diabetes and neurodegenerative diseases.[209] There are three known isoforms of Akt which are found in different parts of the cell. Each isoform is involved in different signaling pathways; however, they all utilize the same recognition motif. Akt must be phosphorylated for activation, but once it is activated it has a large number of targets all throughout the cell which control metabolism, proliferation, and survival.[210] Much like the MDM2•p53 interaction, Akt upregulation is seen in a variety of cancers including pancreatic, ovarian, and colorectal cancers, leading to an increase in cell proliferation.[209,211,212] One potential mechanism for this is the disruption of the tumor suppressor, PTEN, pathways. Upon loss of activity of PTEN, due to mutation in cancerous cells, PI3K pathways are activated, a downstream target of which is Akt, leading to an increase in Akt activity due to upregulation. This pathway ultimately leads to cell growth and proliferation due to increased Akt activity.[213] Akt is also

involved in diabetes via the insulin signaling pathway.[210] Glucose uptake is, in part, regulated by Akt via the activation and translocation of the GLUT4 glucose receptor. Deregulation of Akt is associated with clinical insulin resistance often observed in type II diabetes.[214] Insulin resistance is also observed via Akt pathways via PI3 kinase pathways leading to kidney and heart disease as a result of diabetes.[211] Akt also, in part, controls neuronal cell death in neurological diseases. In Parkinson's disease, constitutively active, overexpressed Akt leads to neuronal cell death during the progression of the disease.[210] With all of the clinical pathologies as a result of Akt misregulation, it is unsurprising that it is a high-profile drug target for the pharmaceutical industry. To date, no Akt inhibitors have been approved by the FDA; however, as of 2016, 13 different Akt inhibitors were in clinical trials.[215] Furthermore, new probes are needed to study Akt activity to understand the full extent of Akt misregulation in disease.

3.2 Results

3.2.1 Design of Abl Kinase Substrates Containing Perfluoro-*tert*-butyl Ether

We sought to incorporate a perfluoro-*tert*-butyl amino acid near the phosphorylation site of a kinase substrate in order to allow for the detection of phosphorylation via ^{19}F NMR. Due to the cyclic nature of proline and proline derivatives, we decided to examine the perfluoro-*tert*-butyl hydroxyproline derivatives. These derivatives will be most receptive to changes in backbone structure and electronics. Furthermore, the hydroxyproline derivatives favor extended conformations. Kinase substrates do not usually adopt compact secondary structures

such as α-helices, but rather favor loops and extended ß structures, when bound to enzyme.

Due to the historical significance and a strong understanding of Abl kinase activity, Abl kinase is an excellent model system to begin the development of new tools to study kinase activity.[200] Abl kinase has a known recognition sequence, realized in the optimal peptide sequence of Abltide, which makes an excellent starting point for a ^{19}F-based phosphorylation sensor for tyrosine kinases.[185] The Abltide peptide contains the key Abl kinase recognition residues EAI (-3, -2, and -1 relative to the phosphorylation site respectively) directly upstream of the phosphorylation site (Figure 3.8). Furthermore, the Abltide peptide contains a proline at the +3 position relative to the phosphorylation site. While perfluoro-*tert*-butyl hydroxyproline is significantly larger than proline, in this case, previous data show that all of the aromatic amino acids (Phe, Tyr, Trp) are also accepted at the +3 position, demonstrating larger, more hydrophobic amino acids can be accepted at this position.[216] While there is a Phe at the +4 position, this site does not readily accept proline, minimizing the likelihood that perfluoro-*tert*-butyl hydroxyproline would be accepted at this position. Our goal was the observation of a single ^{19}F signal for phosphorylated and non-phosphorylated peptides.

Abltide   **KKGEAIYAAPFA**

4*R*-FAblide **KKGEAIYAAPFA**
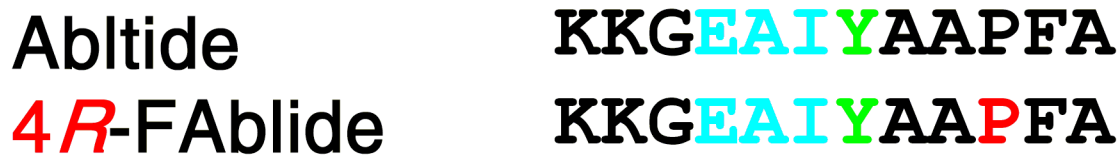
Figure 3.8   Abl kinase substrate containing Hyp(C$_4$F$_9$) based on the Abltide peptide. Residues required for enzyme recognition (cyan) are upstream of the phosphorylation site (green). Hyp(C$_4$F$_9$) (red) will be mutated at the +3 proline site in the sequence which is not part of the recognition motif.

3.2.2 First Generation Abl Kinase Substrate: 4*R*-FAblide Phosphorylation

Peptides in Figure 3.8 were synthesized on solid phase according to established protocols for the incorporation of amino acids containing perfluoro-*tert*-butyl groups as discussed in Chapter 1. In order to demonstrate that the detection of kinase activity is possible by ^{19}F NMR, peptides were chemically phosphorylated on solid phase utilizing established chemistry (Figure 3.9). Both phosphorylated and non-phosphorylated peptides were cleaved under standard conditions and purified by HPLC. Peptides were subjected to ^{19}F NMR upon verification by mass spectrometry. No significant change in chemical shift ($\Delta\delta < 0.01$ ppm) was observed by ^{19}F NMR between the phosphorylated and non-phosphorylated peptides. The lack of change may be a result of an interaction between the perfluoro-*tert*-butyl hydroxyproline and the following Phe (Figure 3.10). Both the perfluoro-*tert*-butyl group and the Phe side chain can freely rotate, allowing them to form a hydrophobic interaction between the fluorine atoms and the aromatic ring. Furthermore, Pro-aromatic sequences are known to stabilize *cis* amide bond formation which would increase interaction between the aromatic ring and the perfluoro-*tert*-butyl ether, as well as potentially prevent the substrate from entering the enzyme active site.[53] This interaction maybe attributed to the lack of chemical shift change associated with this peptide due to the side chain

192

interactions preventing a noticeable chemical shift change due to changes in backbone
conformation and electronics.
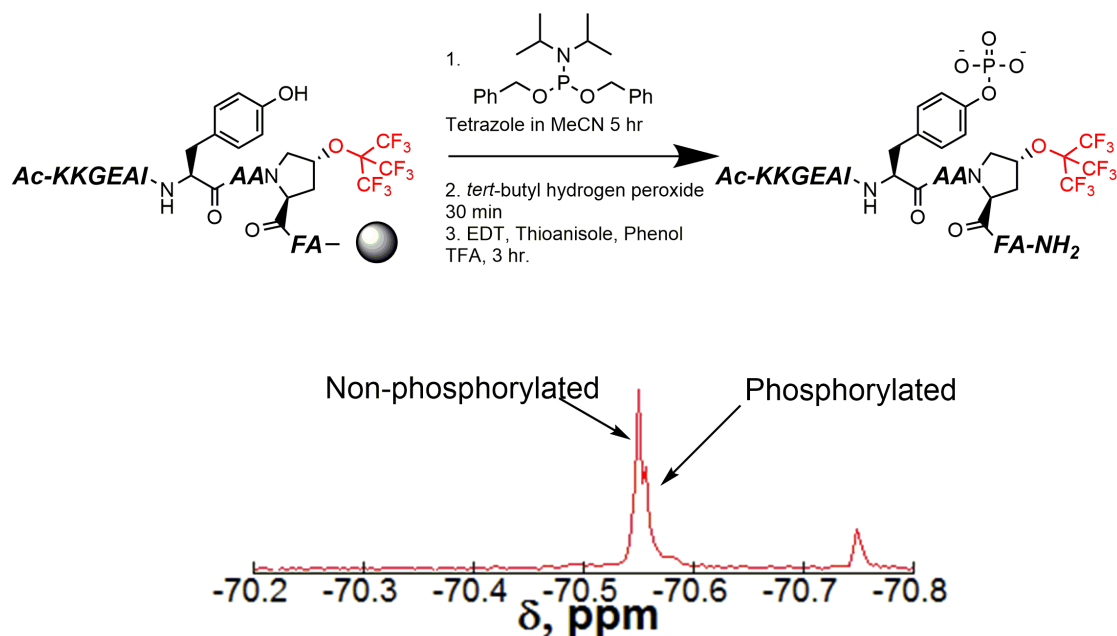
Figure 3.9   First-generation Abl kinase sensor was phosphorylated on solid phase
and purified (top). ^{19}F NMR spectra of purified non-phosphorylated and
phosphorylated FAbltideR in 1× reaction buffer (50 mM HEPES buffer
(pH 7.5)) at room temperature. NMR spectrum was taken with a 10 ppm
sweep width and 128 scans (5 min).
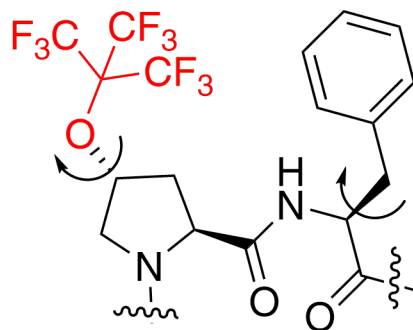
Figure 3.10 Free rotation of the side chains on both the perfluoro-*tert*-butyl hydroxyproline and the Phe potentially allow for a strong hydrophobic interaction between the two side chains, preventing detection of the phosphorylation event by ^{19}F NMR.

3.2.3 Second Generation Abl Kinase Substrate

A second generation Abl kinase substrate was developed replacing the Phe with Arg at the +4 position. Previous data shows that a Phe is not required at the +4 position, as the +4 position is not part of the Abl kinase recognition sequence.[217] Arg was chosen to minimize hydrophobic interactions with the neighboring 4*R*-perfluoro-*tert*-butyl hydroxyproline. Microarray analysis of all 20 canonical amino acids at the +4 position relative to the phosphorylation site indicate that Arg is well tolerated at the +4 position, as well as Phe and Trp, both of which could potentially interact with the perfluoro-*tert*-butyl group via the hydrophobic effect (Figure 3.11).[185] The second generation Abl kinase sensor peptide was synthesized, cleaved from resin and deprotected, and purified by HPLC. Peptides were verified by mass spectrometry (Figure 3.12).
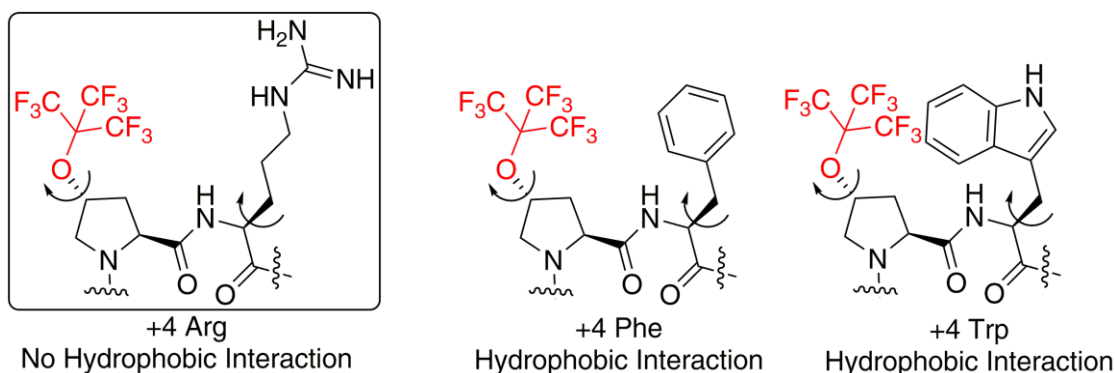
Figure 3.11 Abl kinase accepts Phe, Trp, and Arg at the +4 position from the phosphorylation site. Of these amino acids, Arg was chosen due to minimal potential hydrophobic interaction with Hyp(C_4F_9) in sensor peptides.

Figure 3.12 The second generation Abl kinase substrate design substituting Arg for Phe at the +4 position from the phosphorylation site (green). Recognition residues are cyan and Hyp(C_4F_9) in red.

This substrate was also subjected to chemical phosphorylation on solid phase to verify a change in chemical shift by ^{19}F NMR upon phosphorylation. Upon cleavage from resin, deprotection and purification, both the phosphorylated and non-phosphorylated peptides were analyzed by ^{19}F NMR (Figure 3.13). Unlike the previous design, this peptide did demonstrate a 0.01 ppm chemical shift change upon phosphorylation. While this is an incredibly small chemical shift change, it is similar

195

to previous results using tifluoromethylphenylalanine to detect phosphorylation on the following Ser phosphorylation site.[36] The improvement over previous systems is the presence of nine equivalent fluorines, compared to three equivalent fluorines, leading to an improvement in signal-to-noise. Also, previous systems have only detected phosphorylation in Ser/Thr-containing substrates, which upon phosphorylation, propagate the signal through the backbone moreso than Tyr. Ser and Thr undergo structural changes in backbone conformation upon phosphorylation, which can more easily be detected than Tyr phosphorylation.[218] The aromatic ring of the Tyr side chain minimizes the effects of the phosphorylation on the backbone, despite electronic changes to the aromatic system of the Tyr ring.
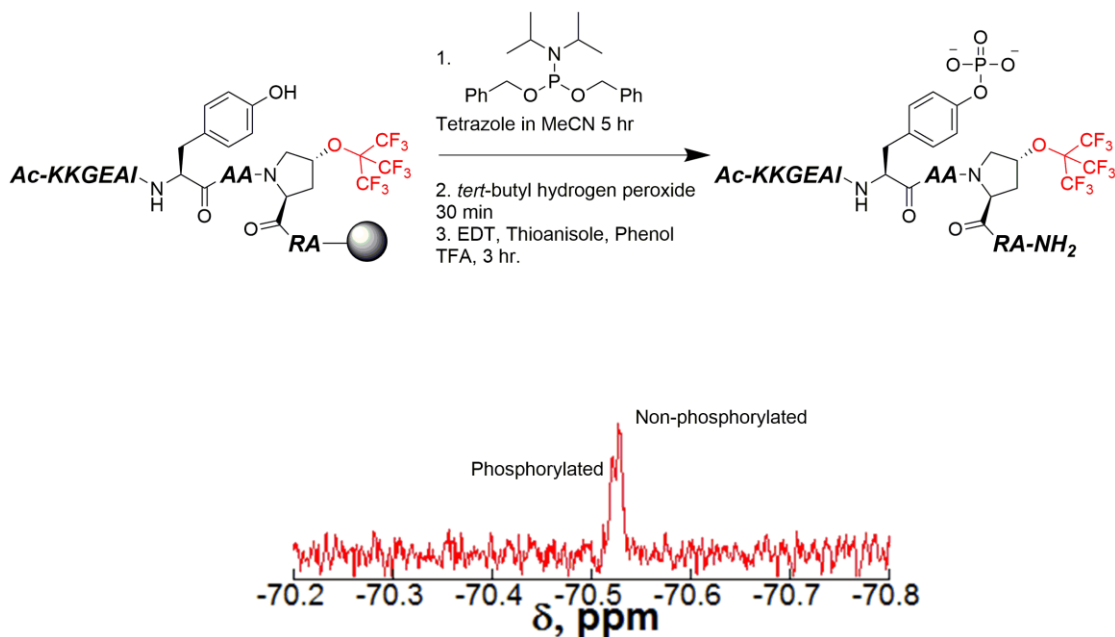
Figure 3.13 Second-generation Abl kinase sensor was phosphorylated on solid phase and purified. ^{19}F NMR spectra of purified non-phosphorylated and phosphorylated FAbltideR in 1× reaction buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT) at room temperature. NMR spectrum was taken with a 10 ppm sweep width and 128 scans (5 min).

Having demonstrated a chemical shift change upon phosphorylation, the peptide was subjected to phosphorylation by Abl kinase *in vitro* to demonstrate that 4*R*-perfluoro-*tert*-butyl hydroxyproline could be accommodated within the recognition site of Abl kinase. Non-phosphorylated peptide was incubated with purified Abl kinase for one hour at 37 °C to determine both if phosphorylation by the enzyme was possible and if it could be detected by ^{19}F NMR (Figure 3.14). The reaction mixture was heated to 100 °C to heat inactivate the kinase and stop the reaction, before

197

diluting the solution with 10% D_2O in H_2O for ^{19}F NMR analysis. While the phosphorylated and non-phosphorylated peaks do not completely resolve by ^{19}F NMR spectra, making integration of the phosphorylated and non-phosphorylated peaks difficult, HPLC demonstrated 60% phosphorylation after an hour incubation at 37 °C (Figure 3.14 and 3.15). This change is similar to the phosphorylation rates seen with the Abl kinase pKID sensor developed by Zondlo *et al.*[189] These data demonstrate 4*R*-perfluoro-*tert*-butyl hydroxyproline can be accommodated within the recognition site of the Abl kinase substrate peptide. Furthermore, these data demonstrate that these sensors can easily be detected at 5 µM concentration in approximately five minutes. Similar systems which utilize ^{19}F sensors to detect kinase activity required 30 µM concentrations to obtain similar results.[41]
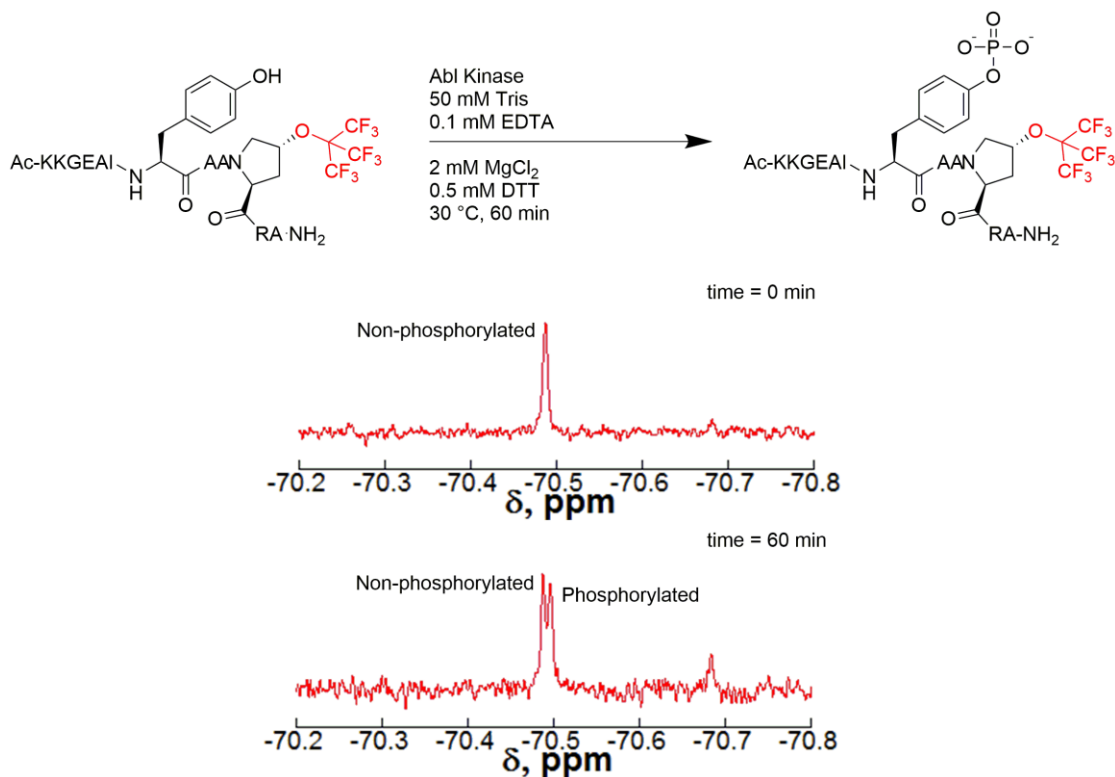
Figure 3.14 Second-generation Abl kinase sensor was phosphorylated via Abl kinase over 60 min at 30 °C. ^{19}F NMR spectra of purified non-phosphorylated and enzymatically phosphorylated FAbltideR in 1× reaction buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT) at room temperature. NMR spectrum was taken with a 10 ppm sweep width and 128 scans (5 min) and 5 µM peptide.
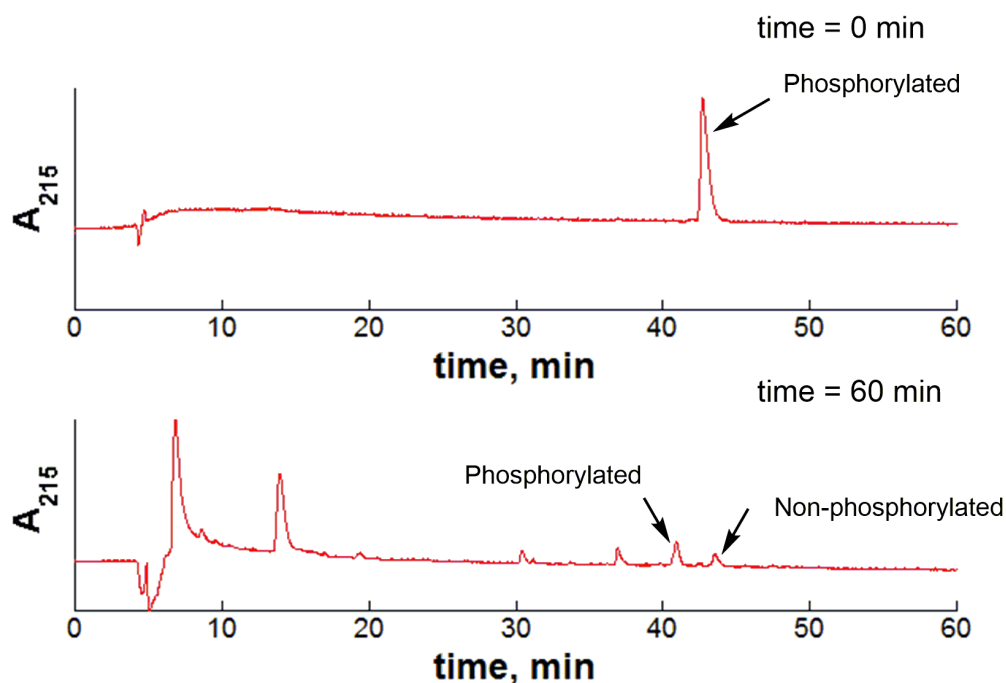
Figure 3.15 HPLC chromatogram of 4*R*-FAbltideR reinject (top) and enzymatically phosphorylated 4*R*-FAbltideR (bottom). The non-phosphorylated peptide was incubated for 60 min with Abl kinase before ^{19}F NMR and injection on the HPLC. HPLC chromatogram was collected over 60 min on a 0-70% ramp of buffer B in buffer A.

One of the benefits of a ^{19}F kinase sensor is the ability to detect phosphorylation events by ^{19}F NMR in complex solutions. In order to demonstrate the ability to detect phosphorylation in biological systems, the peptide substrate was subjected to Abl kinase phosphorylation in HeLa cell lysates. HeLa cells were grown in the presence of or absence of Epidermal Growth Factor (EGF), which is a known activator of Abl kinase activity. The activation of Abl kinase activity by EGF should lead to an increased amount of phosphorylation over the unstimulated lysates. In both EGF-stimulated and unstimulated lysates, however, a new peak was observed by ^{19}F

NMR, independent of either the phosphorylated or non-phosphorylated peptide (Figure 3.16). HPLC and mass spectral analysis of this peptide was demonstrated to be Ac-KKGEAIYAAHyp(C_4F_9), indicating cleavage of the RA-NH_2 on the C-terminus of the peptide (Figure 3.17). These data are consistent with results seen by Dalvit and coworkers who also reported chemical shift changes as a result of proteolysis by trypsin.[189] Furthermore, in EGF-stimulated lysates, both phosphorylated and non-phosphorylated proteolysis products could be observed by ^{19}F NMR. Any of the Aspartyl proteases are potentially acting on the peptide substrate, using the Arg for recognition. Utilizing the ExPASy PeptideCutter software, Arg-C was identified as a potential protease acting on the substrate.[219] While not the desired product of these experiments, they demonstrate the capability of this system to detect other post-translational modifications and enzymatic activities, including protease activity.
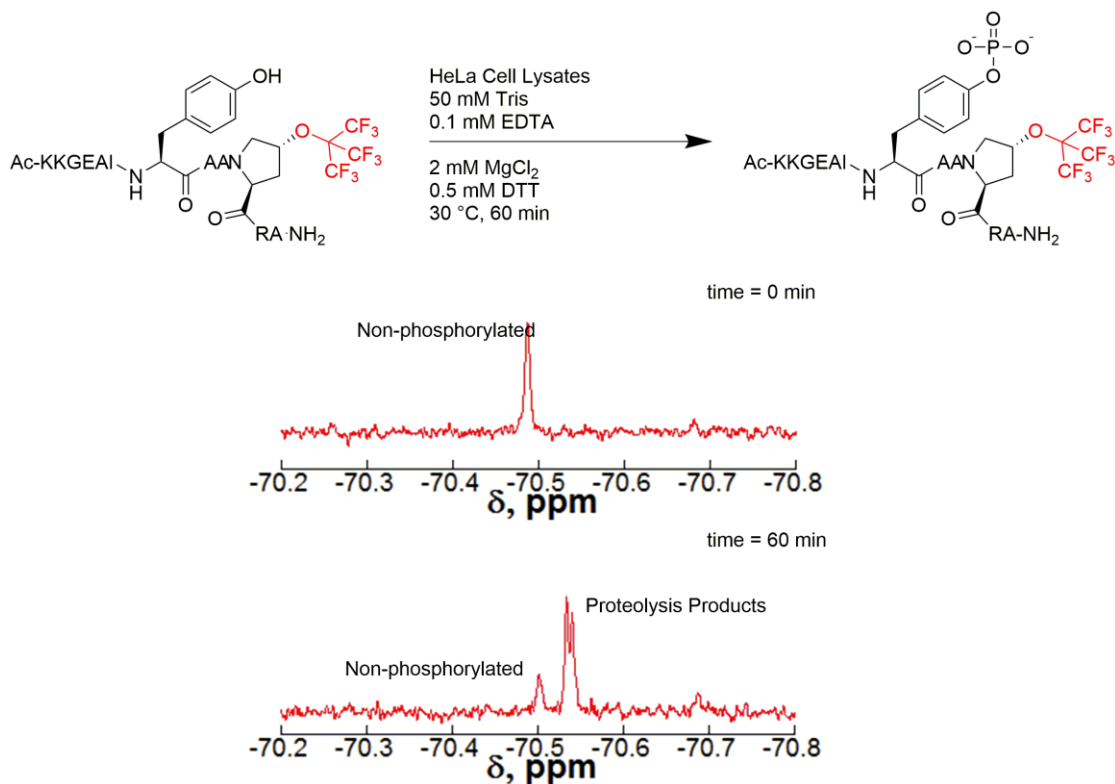
Figure 3.16 Second-generation Abl kinase sensor was phosphorylated via Abl kinase in EGF-stimulated HeLa cell lysates over 60 min at 30 °C. ^{19}F NMR spectra of purified non-phosphorylated and enzymatically phosphorylated FAbltideR in 1× reaction buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT) with 175 µL HeLa cell lysates at room temperature. The NMR spectrum was taken with a 10 ppm sweep width and 128 scans (5 min) and 5 µM peptide.
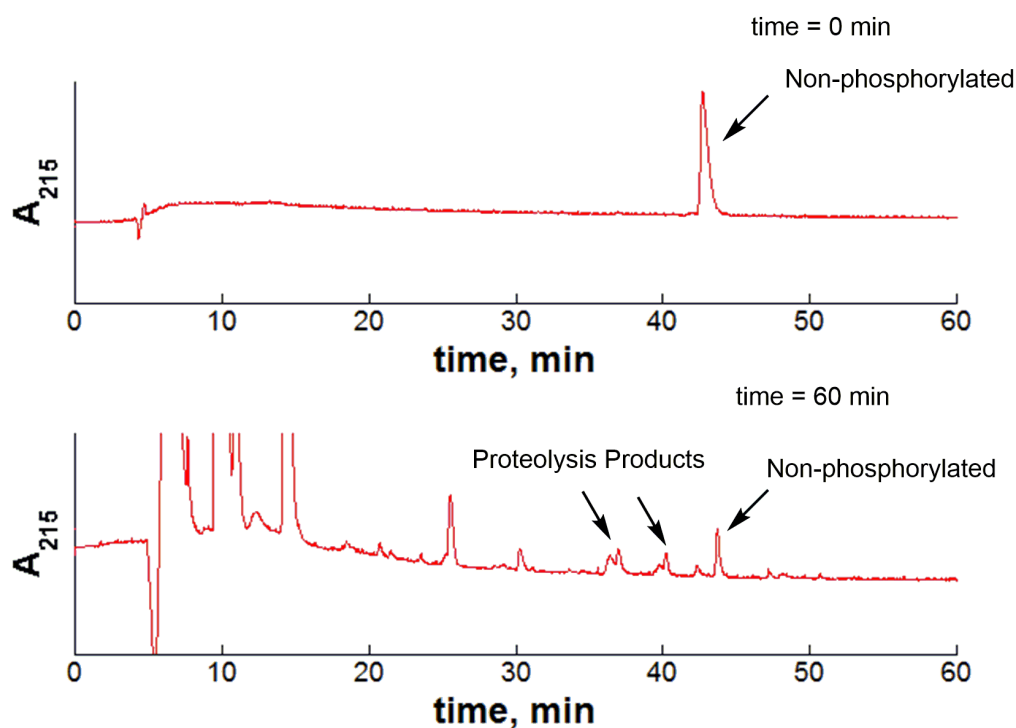
Figure 3.17 HPLC chromatogram of 4*R*-FAbltideR reinject (top) and enzymatically phosphorylated 4*R*-FAbltideR in HeLa cell lysates (bottom). The non-phosphorylated peptide was incubated for 60 min with EGF-stimulated HeLa cell lysates before ^{19}F NMR and injection on the HPLC. HPLC chromatogram was collected over 60 min on a 0-70% ramp of buffer B in buffer A.

3.2.4 Third Generation Abl Kinase Substrate

A third generation Abl kinase sensor was developed to remove the proteolysis site from the designed substrate. This design replaced the Arg residue, which was most likely acting as a recognition residue for proteases, with an Asn residue, which was included to reduce proteolysis, while still minimizing potential hydrophobic

interactions with the perfluoro-*tert*-butyl group (Figure 3.18). Again, based on peptide microarray data, Abl kinase is known to accept Asn at the +4 position.[49]

Abltide	**KKGEAIYAAPFA**
4*R*-FAblide	**KKGEAIYAAPFA**
4*R*-FAblideR	**KKGEAIYAAPRA**
4*R*-FAblideN	**KKGEAIYAAPNA**

Figure 3.18   The third generation Abl kinase substrate replacing Arg with Asn to prevent proteolysis. Recognition residues are shown in cyan and the phosphorylation site is shown in green. $Hyp(C_4F_9)$ is shown in red.

The third-generation Abl kinase substrate also demonstrated a 0.01 ppm chemical shift change upon phosphorylation by chemical means on solid phase. When this peptide was subjected to phosphorylation by Abl kinase *in vitro*, only 40% phosphorylation was observed by ^{19}F NMR, indicating that the +4 Asn is less favorable for phosphorylation, but kinase activity was still detectable by ^{19}F NMR (Figure 3.19). Percent phosphorylation measured by ^{19}F NMR was confirmed by HPLC (Figure 3.20).
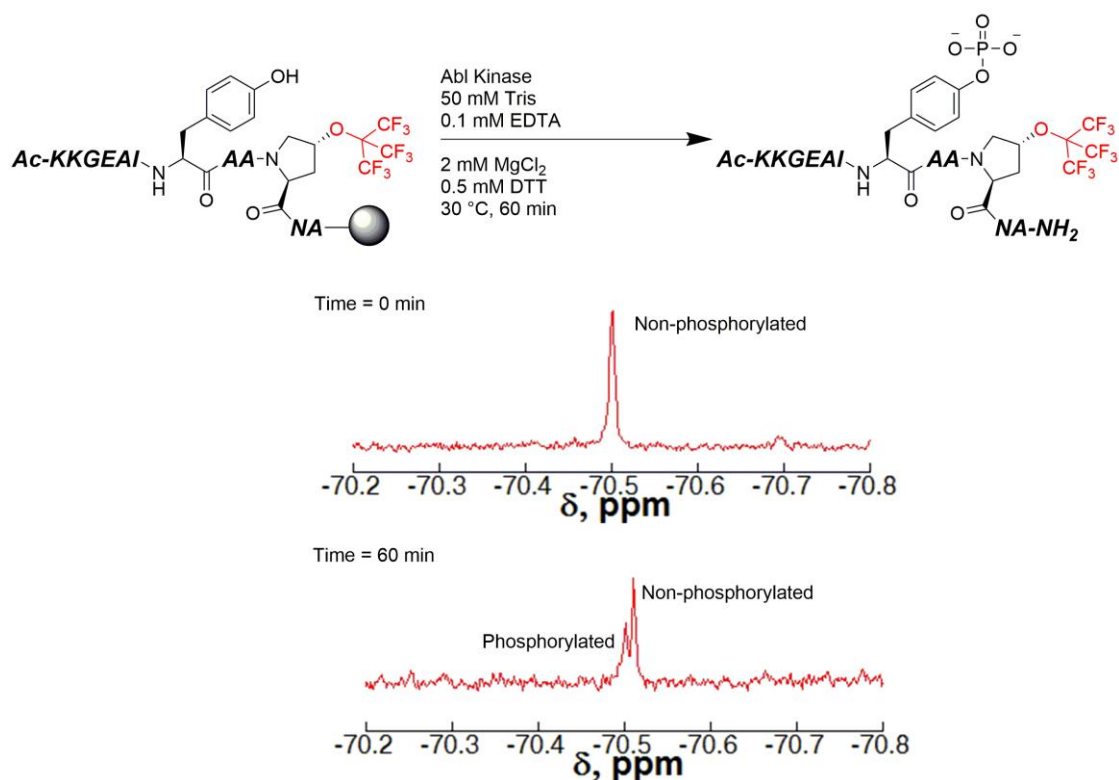
Figure 3.19 Third-generation Abl kinase sensor was phosphorylated via Abl kinase over 60 min at 30 °C. ^{19}F NMR spectra of purified non-phosphorylated and enzymatically phosphorylated FAbltideN in 1× reaction buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT) at room temperature. The NMR spectrum was taken with a 10 ppm sweep width and 128 scans (5 min) and 5 µM peptide.
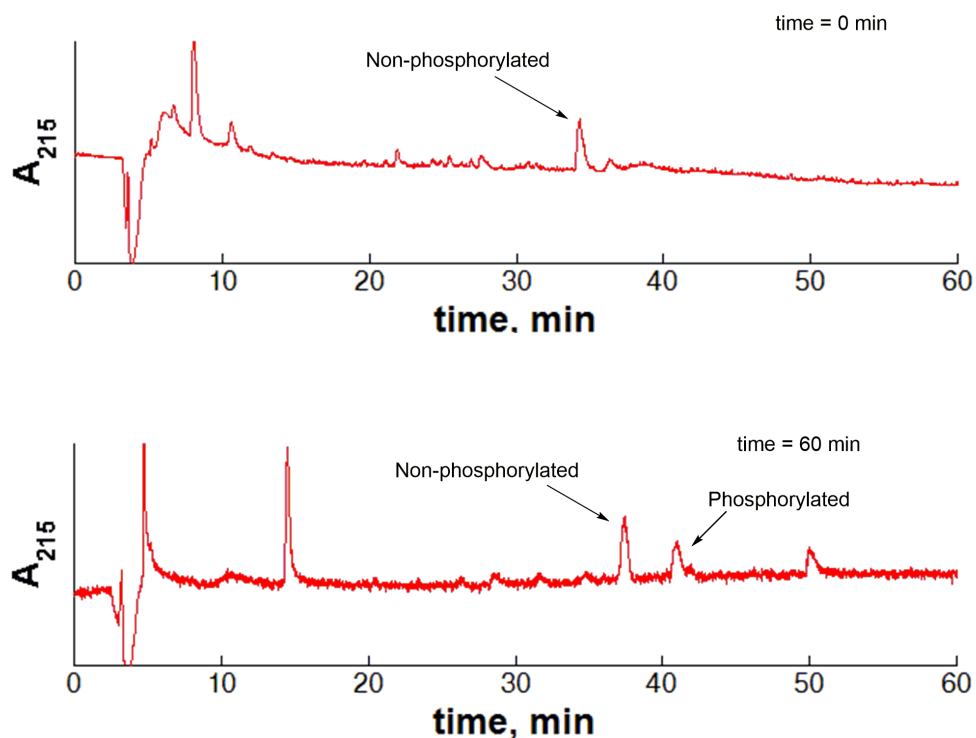
Figure 3.20 HPLC chromatograms of non-phosphorylated 4*R*-FAbltideN (top) and enzymatically phosphorylated peptide after 60 min incubation with Abl kinase and ^{19}F NMR experiment (bottom).

Furthermore, this substrate was subjected to phosphorylation by Abl kinase in HeLa cell lysates. In EGF-stimulated HeLa cell lysates, phosphorylation was observed. However, this peptide was inconsistently phosphorylated upon replication and the percent phosphorylation as integrated by ^{19}F NMR and HPLC were not consistent. This inconsistency is most likely due to insufficient separation by ^{19}F NMR. Based on these substrate challenges, we turned our attention to Ser/Thr kinases to investigate other potential systems for ^{19}F-based detection of kinase activity.

3.2.5 Design of PKA Peptide Substrates Containing a Perfluoro-*tert*-butyl Ether

In order to demonstrate that perfluoro-*tert*-butyl hydroxyproline derivative can be used to universally detect phosphorylation, we also sought to detect Ser/Thr protein kinase activity. In order to circumvent the challenges of fluorescence sensors to study PKA activity, a sensor was designed containing a perfluoro-*tert*-butyl ether group to detect kinase activity by ^{19}F NMR. Since there is an optimized recognition sequence for PKA and tools, such as the fluorescent protein or Trp/Tb-based sensor peptides discussed in section 3.1.1, for fluorescence are not required, the design of a ^{19}F peptide sensor was based on the Kemptide. As previously mentioned, all the positions surrounding the phosphorylation site have been examined with all twenty natural amino acids plus phosphoserine and phosphothreonine. These data indicate that PKA will accept any amino acid, including large, hydrophobic, aromatic amino acids, at the -1 position, relative to the phosphorylation site, as well as proline (Figure 3.21). The perfluoro-*tert*-butyl hydroxyproline derivatives were chosen to probe this system for two primary reasons. The first reason is that no secondary structure is required for the Kemptide substrate to be recognized by PKA. The second reason is that the ^{19}F NMR of both the 4*R* and 4*S* perfluoro-*tert*-butyl hydroxyproline derivatives in the Ac-TYXN-NH$_2$ model system clearly showed peaks corresponding to the *cis* and *trans* conformations of the peptide. These data indicate this amino acid will have a stronger propensity to sense changes in conformation as well as changes in electronics of nearby amino acids due to post-translational modifications. When examining the perfluoro-*tert*-butyl tyrosine in the Ac-TXPN-NH$_2$ system, only one peak was seen in the ^{19}F NMR spectrum, despite the clearly visible *cis* peaks in the ^{1}H NMR spectrum. The single peak observed by ^{19}F NMR is most likely due to the *cis* and *trans* peaks being magnetically equivalent, indicating the perfluoro-*tert*-butyl tyrosine may not be

useful in this application. The perfluoro-*tert*-butyl homoserine was considered. However, early studies showed that the perfluoro-*tert*-butyl was most likely too far away from the backbone to sense any sort of modification to neighboring resides. Both 4*R* and 4*S* perfluoro-*tert*-butyl hydroxyproline were introduced at the -1 position within the Kemptide peptide. No other modifications were made to this peptide from the known Kemptide peptide sequence.

Kemptide   LRRASLGAK
4*R*-FKemptide   LRRPSLGAK
4*S*-FKemptide   LRRPSLGAK
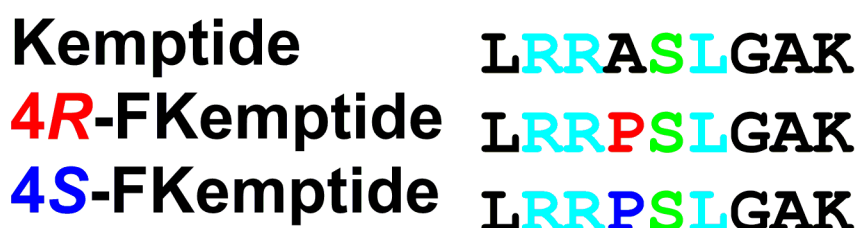
Figure 3.21   PKA substrate design based on the Kemptide peptide with modification at the -1 position to either 4*R* or 4*S* perfluoro-*tert*-butyl hydroxyproline (red). The recognition residues are shown in cyan and the phosphorylation site is shown in green.

3.2.6   Phosphorylation of PKA Substrates Containing Perfluoro-*tert*-buyl Groups

As previously mentioned, kinases generally either phosphorylate Tyr or Ser/Thr. In order to demonstrate the generality of amino acids containing perfluoro-*tert*-butyl ether as ^{19}F probes for phosphorylation, both 4*R* and 4*S* perfluoro-*tert*-butyl hydroxyproline were introduced, at the -1 position relative to the phosphorylation site, into the Kemptide peptide. Both peptides were chemically phosphorylated on solid phase to determine if a chemical shift change was visible upon phosphorylation by ^{19}F NMR. In this case, the chemical shift change upon phosphorylation was much greater than previously observed. The substrate containing the 4*R*-perfluoro-*tert*-butyl

hydroxyproline showed a 0.04 ppm downfield chemical shift change upon phosphorylation while the substrate containing 4*S*-perfluoro-*tert*-butyl hydroxyproline showed a 0.07 ppm upfield chemical shift change upon phosphorylation (Figure 3.22). This chemical shift change is a significant improvement over previous literature sensors and the Abl kinase sensor previously developed. This improvement in chemical shift change is most likely due to structural changes of the Ser upon phosphorylation which can be propagated through the backbone to the perfluoro-*tert*-butyl hydroxyproline. Proline conformation is known to be linked to backbone conformation.[220,221] Both of the introduced amino acids strongly favor either the *exo* (4*R*-perfluoro-*tert*-butyl hydroxyproline) or *endo* (4*S*-perfluoro-*tert*-butyl hydroxyproline) ring pucker. The ring pucker inherent in the proline derivatives dictates the direction of the chemical shift change. Also worth noting, the magnitude of the chemical shift change and the direction of the chemical shift change were not the same for the 4*R* and the 4*S* derivatives. This proof of concept demonstrates that a chemical shift change is observable by ^{19}F NMR upon phosphorylation.

Both peptides were also subjected to enzymatic phosphorylation by PKA *in vitro* to determine their effectiveness as kinase substrates (Figure 3.21). Both peptides were incubated with purified PKA for five minutes before dilution and subjection to ^{19}F NMR and HPLC (Figure 3.22, 3.23, and 3.24). Previously, optimized pKID-based sensors developed by Zondlo *et al*. demonstrated 100% phosphorylation by PKA in five minutes. Similar results were obtained for the substrates containing 4*R*-perfluoro-*tert*-butyl hydroxyproline, where 100% phosphorylation was observed in five minutes by both ^{19}F NMR and HPLC. Interestingly, the substrate containing 4*S*-perfluoro-*tert*-butyl hydroxyproline was only 60% phosphorylated in five minutes by PKA. These

data indicate a stereochemical preference for the substrate 4*R*-perfluoro-*tert*-butyl hydroxyproline.
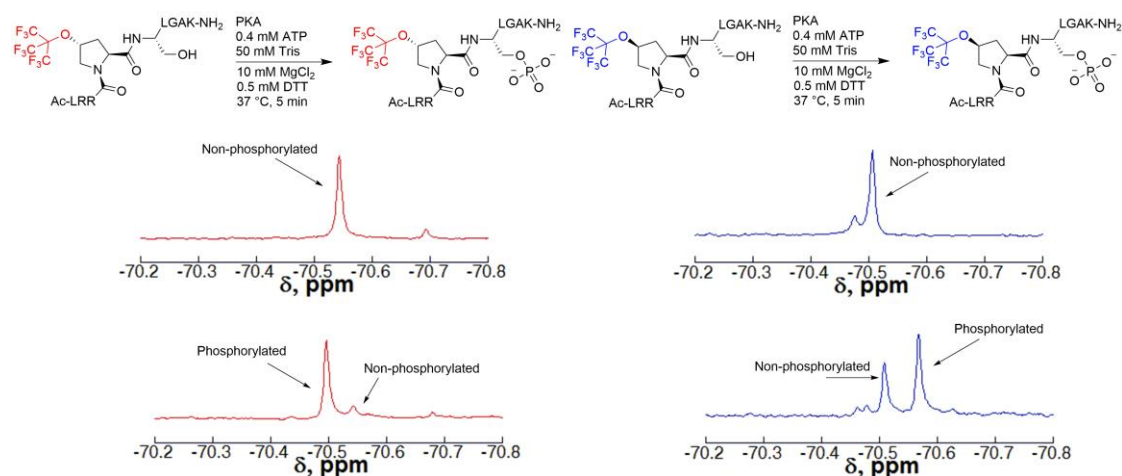
Figure 3.22 ^{19}F NMR spectra for enzymatic assays of both the 4*R*-FKemptide and the 4*S*-Kemptide peptides demonstrating a chemical shift change as a result of enzymatic phosphorylation. In both cases 100 µM non-phosphorylated peptide was incubated with purified PKA kinase in 1× reaction buffer with 400 µM ATP at 37 °C for 5 min prior to ^{19}F NMR experiment. NMR spectra were taken with a 10 ppm sweep width and 128 scans (5 min) and 5 µM peptide.
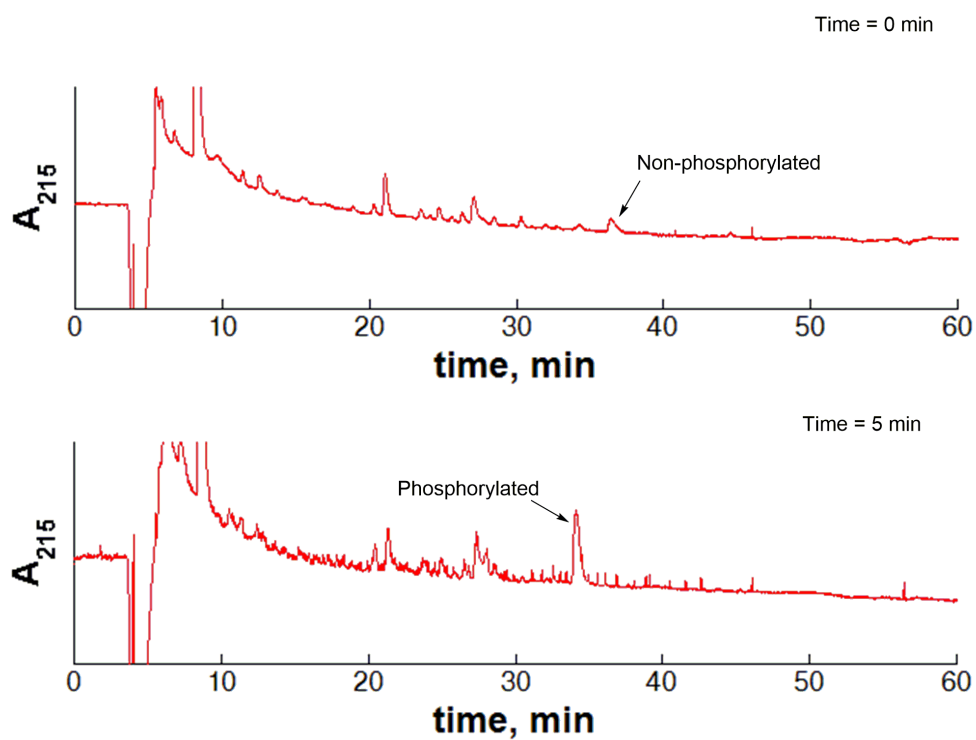
Figure 3.23 HPLC chromatograms of non-phosphorylated 4*R*-Kemptide (top) and enzymatically phosphorylated peptide after 5 min incubation with PKA kinase and ^{19}F NMR experiment (bottom). HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.
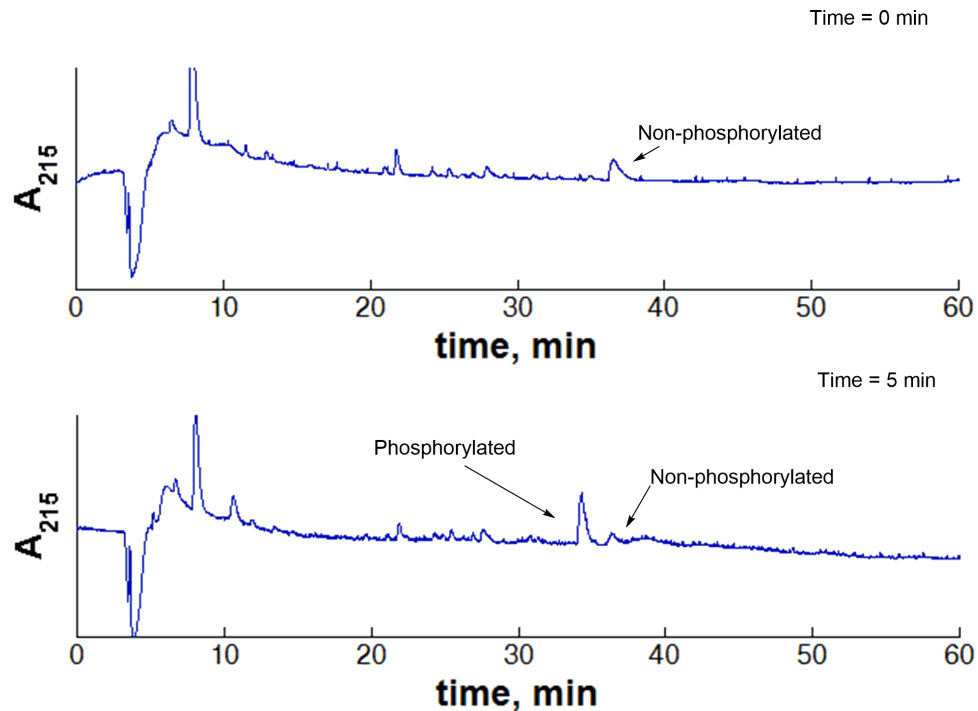
Figure 3.24 HPLC chromatograms of non-phosphorylated 4*S*-Kemptide (top) and
enzymatically phosphorylated peptide after 5 min incubation with PKA
kinase and ^{19}F NMR experiment (bottom). HPLC chromatograms were
measured over 60 min on a 0-70% gradient of buffer B in buffer A.

3.2.7 Design of Akt Peptide Substrates Containing a Perfluoro-*tert*-butyl Ether

Much like PKA, Akt is a well-studied kinase which has a known recognition

sequence and an optimized peptide substrate.[222] This substrate is known at the Aktide

peptide. This peptide incorporates the necessary recognition Arg residues at the -5 and

-3 position from the phosphorylation site as well as a large hydrophobic Phe at the +1

position. Also, similar to PKA, all twenty amino acids plus phosphoserine and

phosphothreonine have been exampled at all positions surrounding the phosphorylation site.[49] These data indicate that hydrophobic amino acids such as His, Phe, and Trp are accepted at the -1 position relative to the phosphorylation site, as well as proline (Figure 3.25). These data are unsurprising as PKA and Akt are closely related kinases, both members of the AGC kinase family, with similar recognition sequences.[223] This similarity leads to a significant amount of cross talk between substrates. Again, both the 4*R* and the 4*S* perfluoro-*tert*-butyl hydroxyproline derivatives were introduced at the -1 position. The hydroxyproline derivatives were chosen because they favor extended structures, relative to the other amino acids containing perfluoro-*tert*-butyl ether, and proline conformation is linked to backbone structure.[220,221]

Aktide   ARKRERTYSFGHHA

4*R*-FAktide   ARKRERTPSFGHHA

4*S*-FAktide   ARKRERTPSFGHHA

Figure 3.25 Designed Akt substrates based on the Aktide peptide incorporating both the 4*R* and the 4*S* perfluoro-*tert*-butyl hydroxyproline derivatives (red or blue, respectively). The recognition residues are shown in cyan and the phosphorylation site is shown in green.

3.2.8 Phosphorylation of Akt Substrates Containing a Perfluoro-*tert*-butyl Ether

The designed Akt substrate peptides were synthesized on solid phase and purified via HPLC. These peptides were also subjected to chemical phosphorylation on solid phase to determine the extent of change in chemical shift as a result of

phosphorylation. Much like the PKA substrates, a large chemical shift change was visible by ^{19}F NMR for both the 4*R* and the 4*S* perfluoro-*tert*-butyl hydroxyproline derivative. The substrate containing 4*R*-perfluoro-*tert*-butyl hydroxyproline showed a 0.05 ppm downfield chemical shift change upon phosphorylation, while the substrate containing 4*S*-perfluoro-*tert*-butyl hydroxyproline showed a 0.06 ppm upfield chemical shift change upon phosphorylation (Figure 3.26). These peptides followed the same pattern as observed with the PKA substrates in which the 4*S* perfluoro-*tert*-butyl hydroxyproline substrate produced a larger chemical shift change than the 4*R* perfluoro-*tert*-butyl hydroxyproline.

Having demonstrated a significant ^{19}F chemical shift change upon phosphorylation, it was demonstrated that these peptides could be phosphorylated by Akt. Both peptides were incubated with purified Akt at 37 °C for an hour (Figure 3.26). The Aktide peptide containing 4*R*-perfluoro-*tert*-butyl was only 40% phosphorylated after one hour. Conversely, the peptide containing 4*S*-perfluoro-*tert*-butyl hydroxyproline was 95% phosphorylated in one hour, as determined by both ^{19}F NMR and HPLC (Figure 3.26, 3.27 and 3.28).
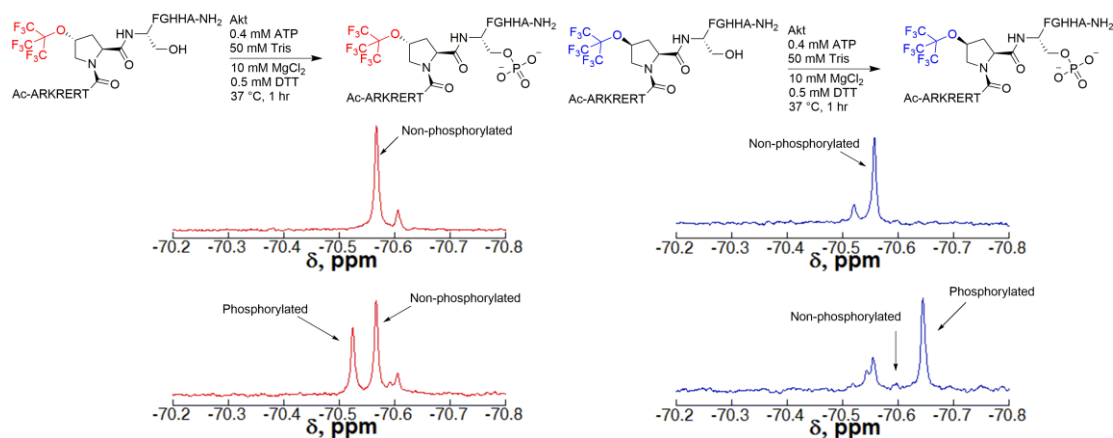
Figure 3.26 ^{19}F NMR spectra for enzymatic assays of both the 4*R*-FAktide and the 4*S*-Aktide peptides demonstrating a chemical shift change as a result of enzymatic phosphorylation. In both cases 100 µM non-phosphorylated peptide was incubated with purified Akt kinase in 1× reaction buffer with 400 µM ATP at 37 °C for 60 min prior to ^{19}F NMR experiment. NMR spectra were taken with a 10 ppm sweep width and 128 scans (5 min) and 5 µM peptide.
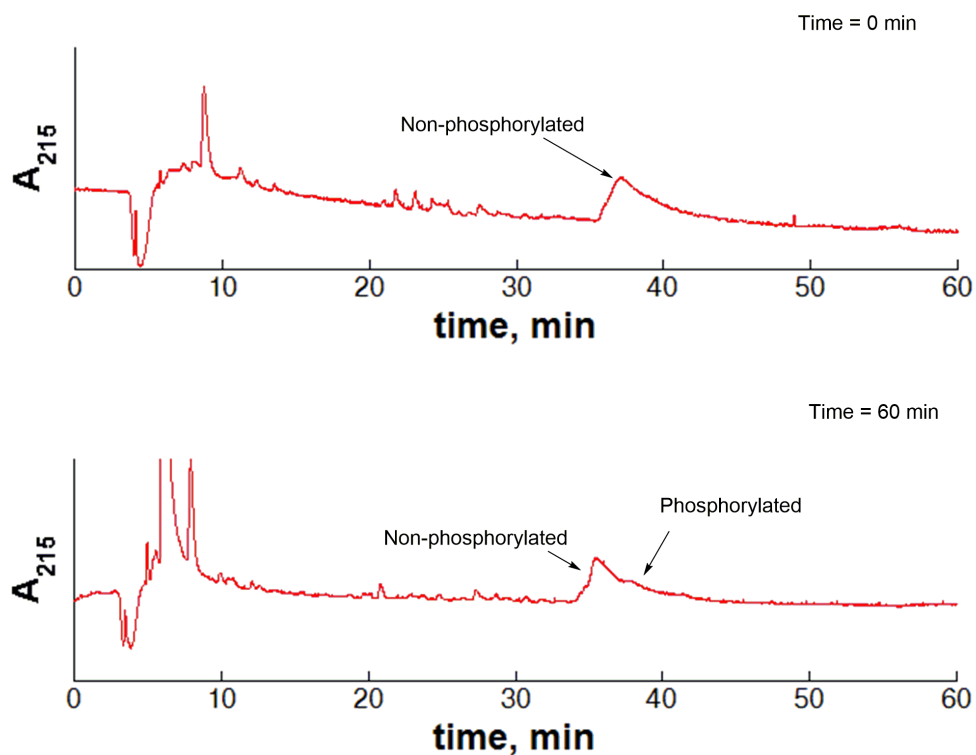
Figure 3.27 HPLC chromatograms of non-phosphorylated 4*R*-FAktide (top) and enzymatically phosphorylated peptide after 60 min incubation with Akt kinase and ^{19}F NMR experiment (bottom). HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.
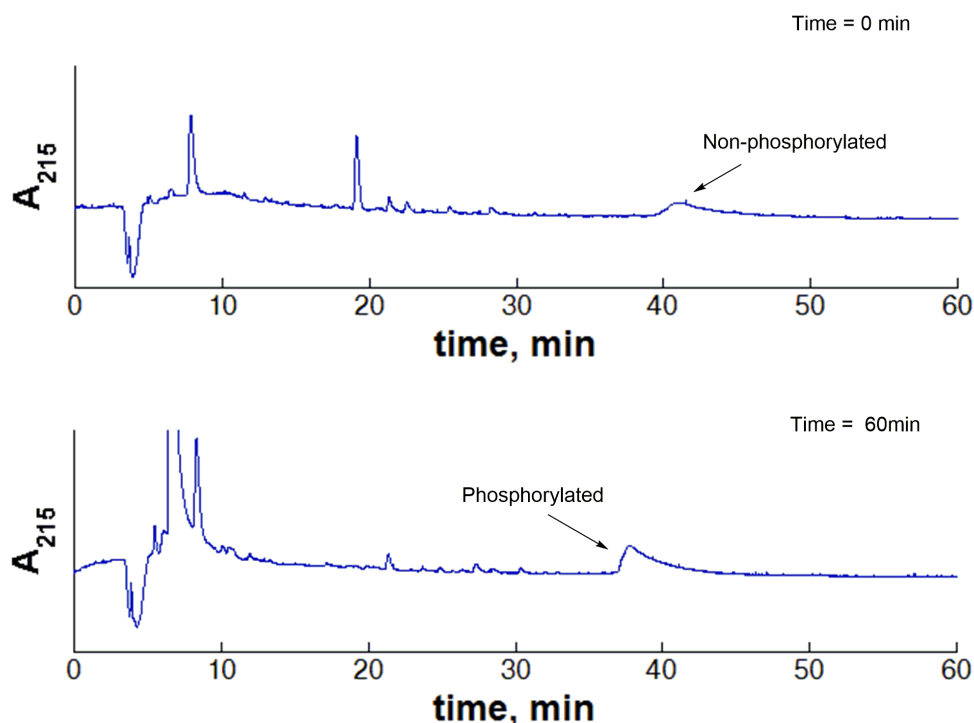
Figure 3.28 HPLC chromatograms of non-phosphorylated 4*S*-FAktide (top) and enzymatically phosphorylated peptide after 60 min incubation with Akt kinase and ^{19}F NMR experiment (bottom). HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.

3.2.9   PKA and Akt Exhibit Differing Stereochemical Preferences in Substrates Containing Perfluoro-*tert*-butyl Hydroxyproline Derivatves

PKA and Akt are similar kinases with very similar recognition sequences, requiring basic residues upstream of the phosphorylation site. These kinases belong to the AGC kinases and can potentially phosphorylate similar targets due to the similarity of recognition motifs for both PKA and Akt.[223] For this reason, the designed substrates are particularly interesting, as PKA and Akt demonstrate differing stereochemical

preferences in the substrates with perfluoro-*tert*-butyl hydroxyproline. While there are a number of potential explanations for this distinction, based on modeling of the substrates within the kinase binding pockets, the stereochemical preference is most likely due to solvation of the perfluoro-*tert*-butyl group in the substrate. In the case of PKA, the 4*R*-perfluoro-*tert*-butyl group is buried within the kinase while in the active site, minimizing solvation of the fluorine atoms (Figure 3.29). A similar interaction is observed for Akt upon modeling of the substrate containing 4*S*-perfluoro-*tert*-butyl (Figure 3.30). The orientation of perfluoro-*tert*-butyl ether, in both cases, maximizes the hydrophobic interactions with the binding pocket of the enzyme. Alternatively, in both cases, the opposite diastereomer is more solvent exposed, leading to increased interaction with bulk water.
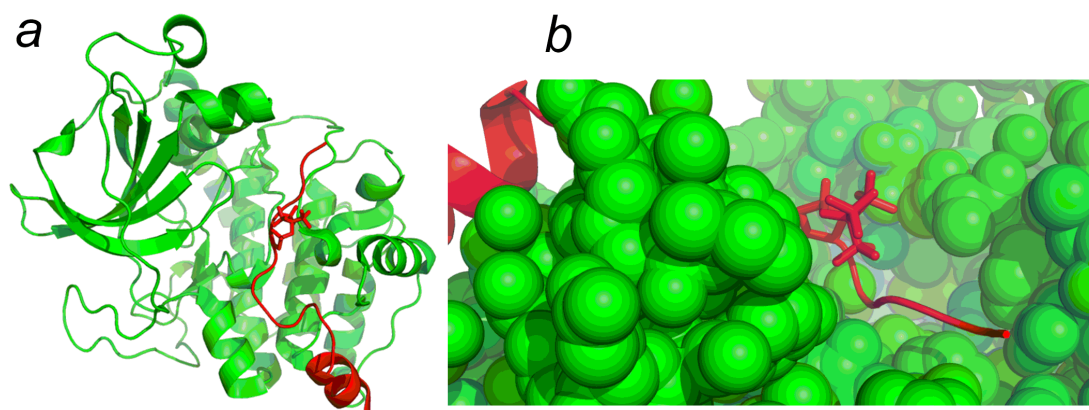
Figure 3.29 (a) crystal structure (1JLU) of PKA bound with substrate which has been mutated to 4*R*-perfluoro-*tert*-butyl hydroxyproline to demonstrate potentially beneficial hydrophobic interactions with the kinase demonstrating a preference for the substrate containing 4*R*-perfluoro-*tert*-butyl hydroxyproline.[224] (b) Close-up of the substrate binding pocket with the mutated 4*R*-perfluoro-*tert*-butyl hydroxyproline demonstrating the perfluoro-*tert*-butyl ether is buried within the hydrophobic cleft of the enzyme.
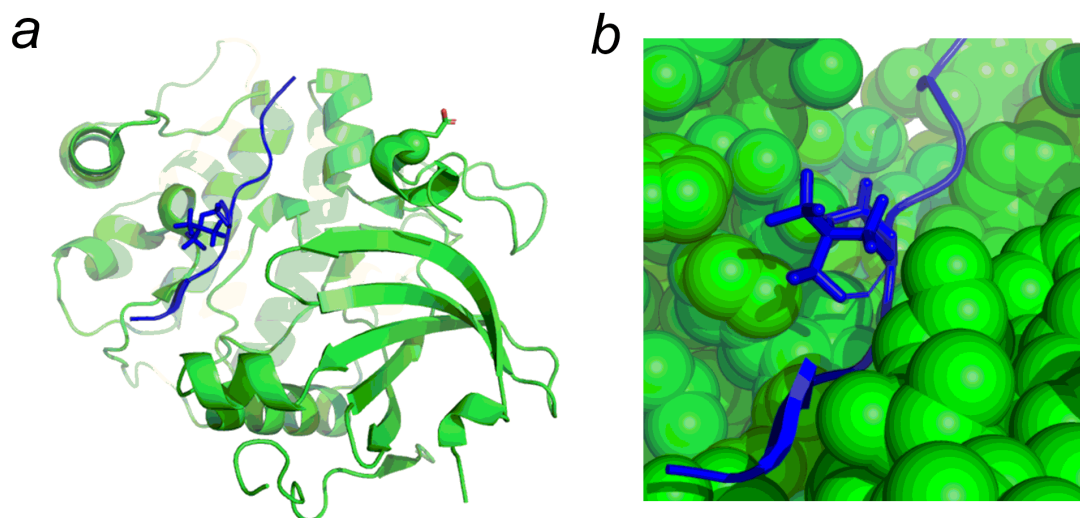
Figure 3.30 (a) crystal structure of Akt (3CQU) bound with substrate which has been mutated to 4*S*-perfluoro-*tert*-butyl hydroxyproline to demonstrate the potentially beneficial hydrophobic interactions with the kinase, demonstrating a preference for the substrate containing 4*S*-perfluoro-*tert*-butyl hydroxyproline.[225] (b) Close-up of the substrate binding pocket with the mutated 4*S*-perfluoro-*tert*-butyl hydroxyproline, demonstrating the orientation of the perfluoro-*tert*-butyl ether in the binding pocket to maximize hydrophobic contacts.

3.2.10 Real Time Detection of Phosphorylation of PKA Substrate Containing 4*R*-Perfluoro-*tert*-butyl Hydroxyproline by ^{19}F NMR

Ultimately, these systems have the potential to be used as a real-time probe to monitor phosphorylation in live cells and animals. To demonstrate the ability to detect phosphorylation in real time, *in vitro* kinase assays were performed by ^{19}F NMR. The Kemptide substrate containing4*R*-perfluoro-*tert*-butyl hydroxyproline was subjected to phosphorylation by PKA (Figure 3.31). ^{19}F NMR experiments were used to monitor the phosphorylation over the course of thirty minutes. Each experiment consisted of 128 scans, totaling five minutes. Over the time course of this experiment, the ^{19}F NMR peak corresponding to the non-phosphorylated peptide decreased, while the

corresponding phosphorylated peak increased over the experiment. At the end of the time-course, the NMR sample was injected on the HPLC to verify the amount of phosphorylation (Figure 3.32)
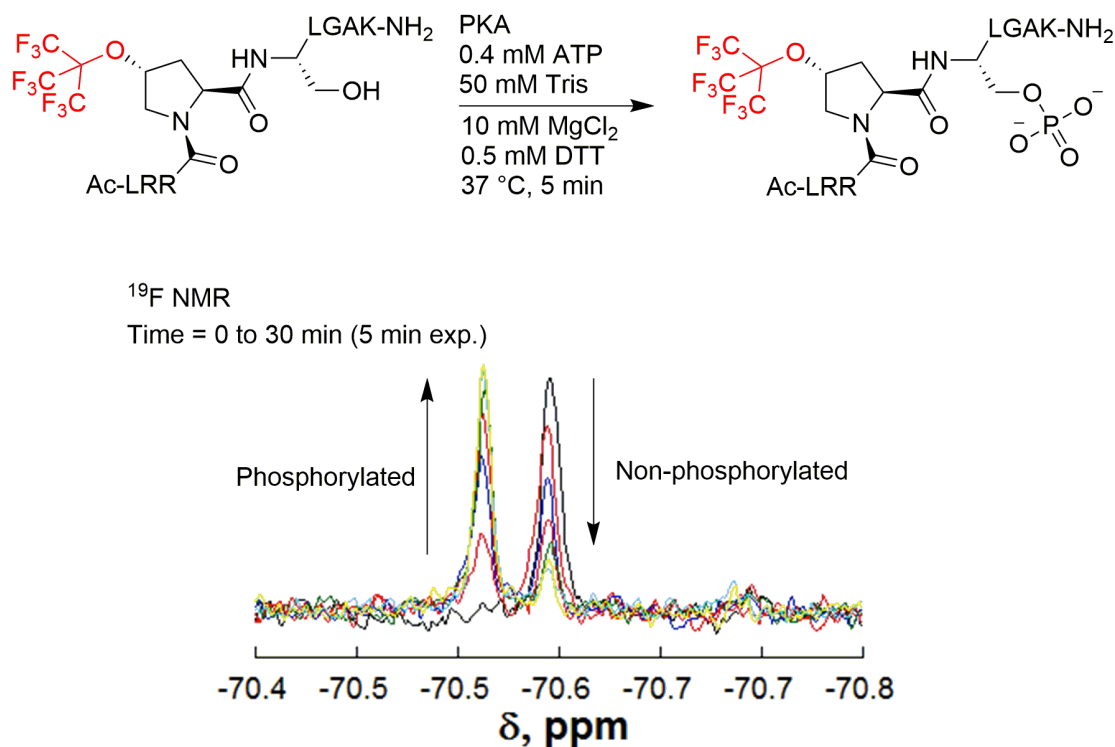
Figure 3.31 Real time ^{19}F NMR monitoring of phosphorylation of 4R-Kemptide over 30 min. Peptide was diluted to 20 µM in 1× buffer with 400 µM ATP to a final volume of 500 µL. Data were collected at 0 min (black), 5 min (red), 15=0 min (blue), 15 min (pink), 20 min (green), 25 min (yellow), and 30 min (cyan).

Time = 0 min

Non-phosphorylated

Time = 5 min
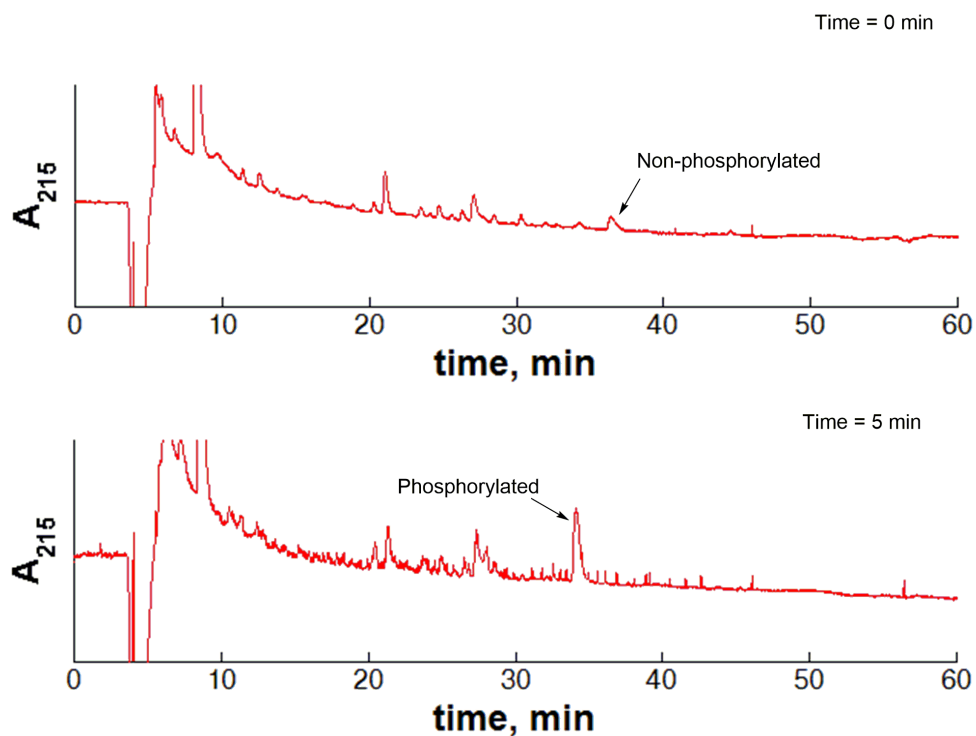
Phosphorylated

Figure 3.32 HPLC chromatograms of non-phosphorylated 4*R*-Kemptide (top) and enzymatically phosphorylated peptide after 30 min incubation with PKA kinase and ^{19}F NMR experiment (bottom). HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.

3.2.11 Real-Time Detection of Phosphorylation of PKA Substrate Containing 4*R*-Perfluoro-*tert*-butyl Hydroxyproline by ^{19}F NMR in HeLa Cell Lysates

Ultimately, this system would ideally be applied to the imaging of kinase activity in real time in live cells. The next step towards that goal was to collect real-time data in complex mixtures, such as HeLa cell lysates. PKA is endogenously expressed in HeLa cell lysates, so no stimulation was required, unlike prior work with

Abl kinase. The Kemptide substrate containing 4*R*-perfluoro-*tert*-butyl was subject to PKA phosphorylation in HeLa cell lysates. This work was done by incubating the non-phosphorylated peptide with HeLa cell lysates in an NMR tube. An initial ^{19}F NMR experiment was taken before the addition of ATP to the reaction mixture, to verify that no initial phosphorylation or other modifications had occured to the peptide. The ATP was added to the NMR tube, and a series of five minute ^{19}F NMR experiments was taken over the course of an hour (Figure 3.33). This experiment was done in the presence of phosphatase inhibitors to prevent dephosphorylation of the sensor peptide. Over the course of an hour, the peak corresponding to the non-phosphorylated peptide disappeared, while the peak corresponding to the phosphorylated peptide proportionately appeared. These data indicate sensor peptides containing perfluoro-*tert*-butyl groups can be used to monitor phosphorylation events in real time by ^{19}F NMR in complex mixtures.
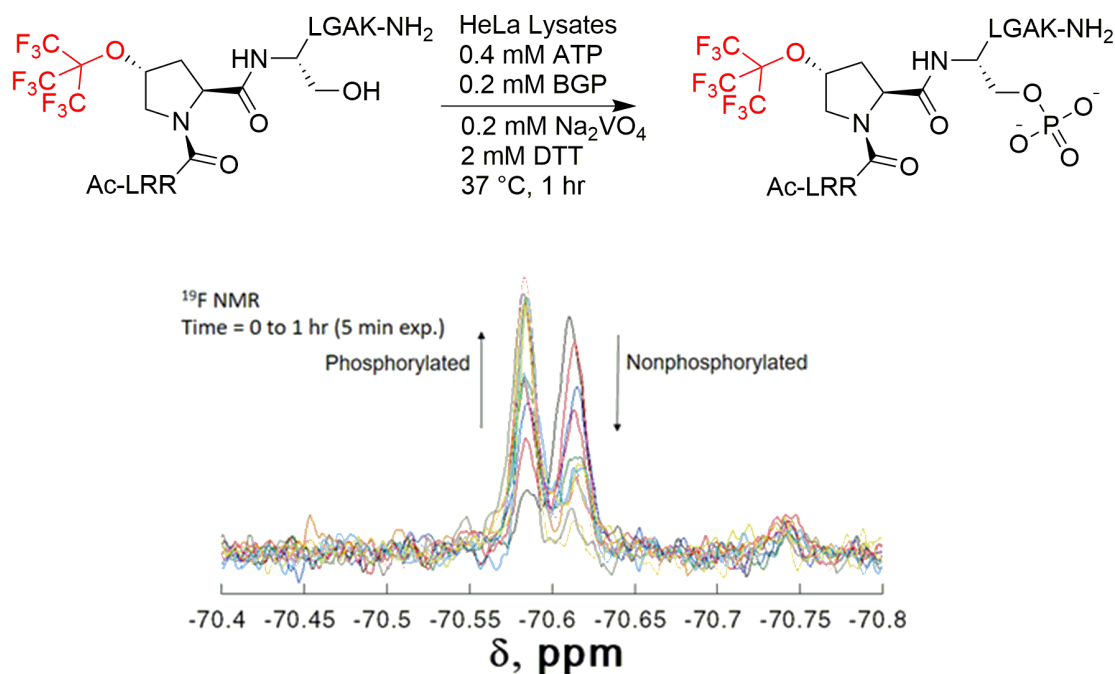
Figure 3.33 One hour real time phosphorylation of 4*R*-Kemptide peptide by HeLa cell lysates. Peptide was diluted to 20 µM in 1× buffer with 175 µL HeLa cell lysates with 160 µM ATP 400 µL β-glycerophosphate, and 200 µM Na_3VO_4 to a final volume of 500 µL. Data were collected at 0 min (black), 5 min (red), 10 min (blue), 15 min (pink), 20 min (green), 25 min (cyan), 30 min (yellow), 35 min (orange), 40 min (light blue), 45 min (grey), 50 min (purple), 55 min (yellow).
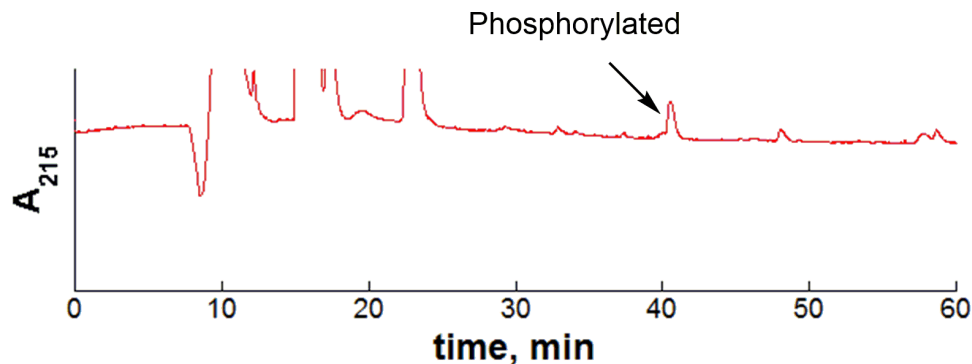
Figure 3.34 Enzymatically phosphorylated peptide after 60 min incubation with HeLa cell lysates in the presence of protease inhibitors and ^{19}F NMR experiment. HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.

The above experiment was repeated in the absence of the phosphatase inhibitors (BGP and Na$_3$VO$_4$) to determine the amount of dephosphorylation that could potentially occur (Figure 3.35). The data from this experiment demonstrate almost no effect from dephoshorylation, despite the lack of phosphatase inhibitors. These data are encouraging as phosphatase inhibitors can be problematic when used with live cells. The phosphatase inhibitors used in the previous experiment are β-glycerophosphate and sodium orthovanadate. β-Glycerophosphate is a broad-spectrum Ser/Thr phosphatase inhibitor, which inhibits a large range of Ser/Thr phosphatases.[226] Sodium orthovanadate is a tyrosine phosphatase inhibitor, which inhibits most tyrosine phosphatase.[227]
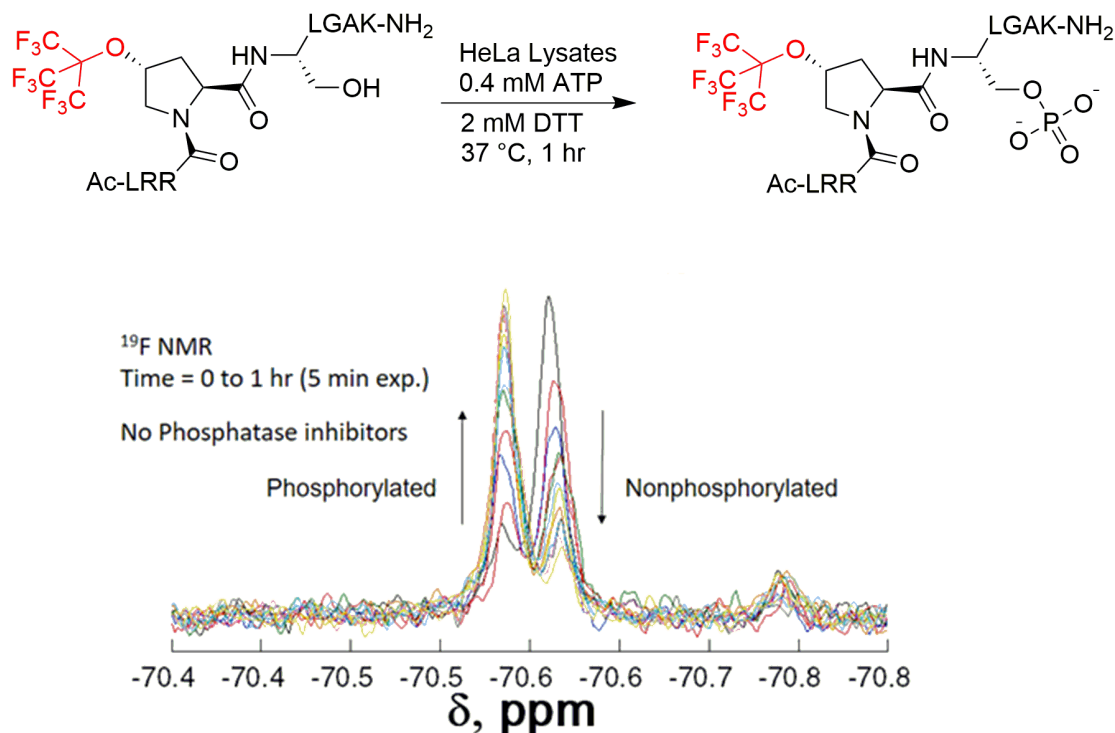
Figure 3.35 One hour real time phosphorylation of 4*R*-Kemptide peptide by HeLa cell lysates without the presence of phosphatase inhibitors. Peptide was diluted to 20 µM in 1× buffer with 175 µL HeLa cell lysates with 160 µM ATP 400 µL to a final volume of 500 µL. Data were collected at 0 min (black), 5 min (red), 10 min (blue), 15 min (pink), 20 min (green), 25 min (cyan), 30 min (yellow), 35 min (orange), 40 min (light blue), 45 min (grey), 50 min (purple), 55 min (yellow).
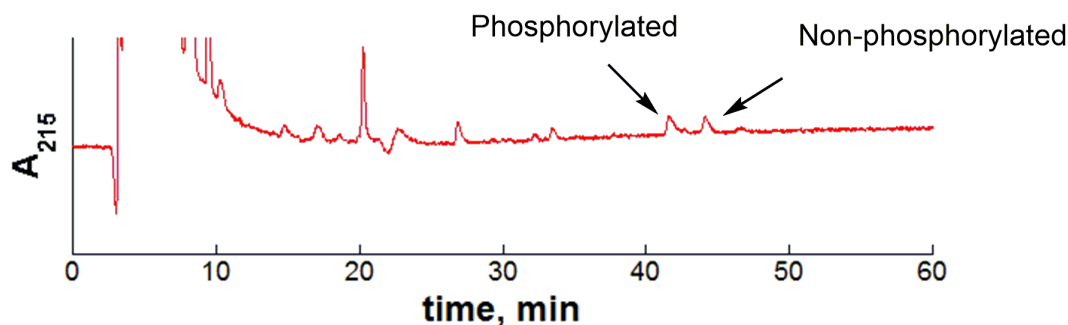
Figure 3.36 Enzymatically phosphorylated peptide after 60 min incubation with HeLa cell lysates in the absence of protease inhibitors and ^{19}F NMR experiment. HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.

Finally, to verify that phosphorylation is a result of PKA activity, this experiment was repeated in the presence of H-89, a known PKA inhibitor (Figure 3.37).[228] Over the course of an hour, less than 20% phosphorylation was observed. This change in the extent of phosphorylation is due to the H-89 inhibiting PKA activity. H89 binding to PKA has a $K_i = 48$ nM.[229] There is still a small amount of phosphorylation because H-89 does have some cross reactivity with related kinases, including Aktα, MSKI, AMPK, and CHK1, meaning there may be a small amount of active PKA or a small amount of phosphorylation from another kinase. [228]
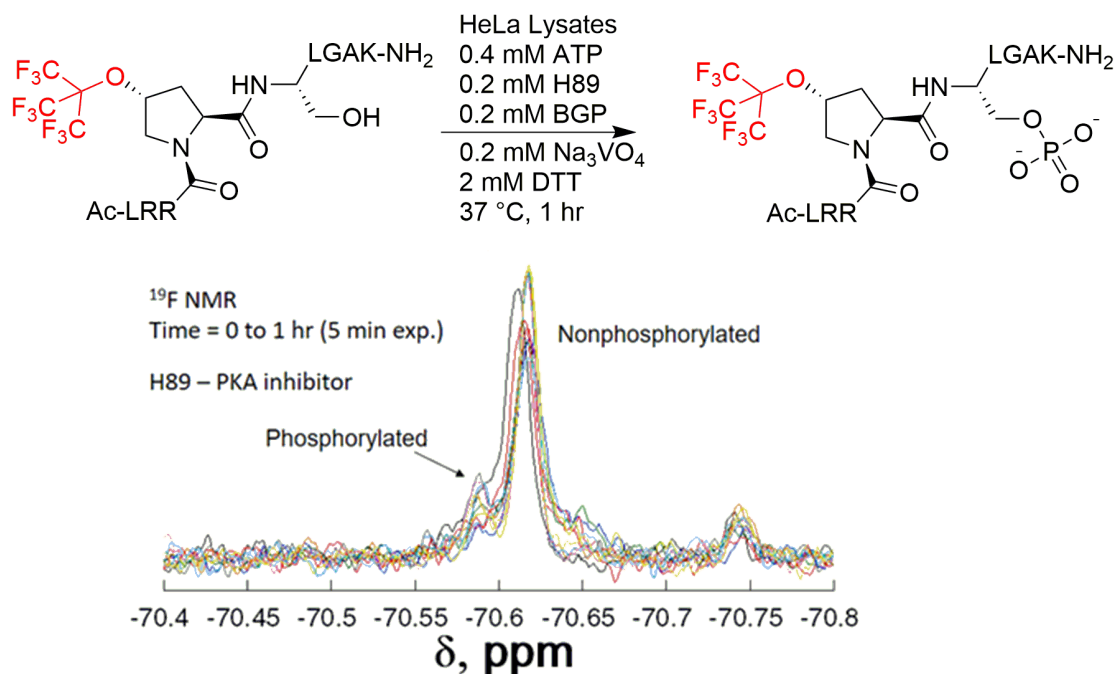
Figure 3.37 Less than 20% phosphorylation is observed in HeLa cell lysates over the course of an hour by ^{19}F NMR with H-89 (200 µM) inhibition. Peptide was diluted to 20 µM in 1× buffer with 175 µL HeLa cell lysates with 160 µM ATP 400 µL β-glycerophosphate, and 200 µM Na$_3$VO$_4$ to a final volume of 500 µL. Data were collected at 0 min (black), 5 min (red), 10 min (blue), 15 min (pink), 20 min (green), 25 min (cyan), 30 min (yellow), 35 min (orange), 40 min (light blue), 45 min (grey), 50 min (purple), 55 min (yellow).
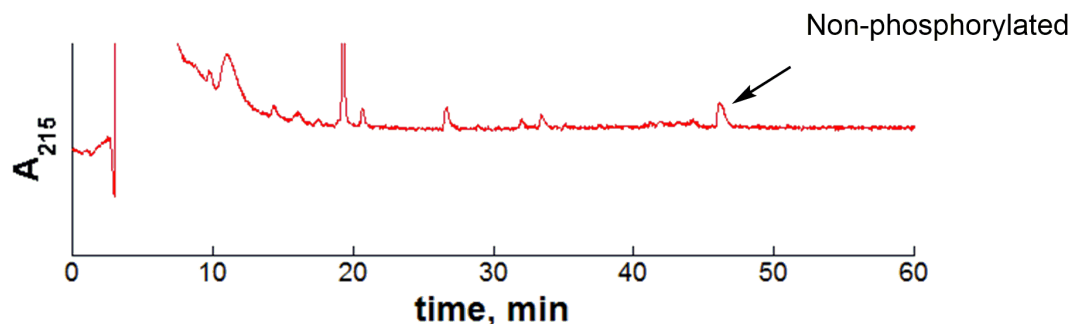
Figure 3.38 Enzymatically phosphorylated peptide after 60 min incubation with HeLa cell lysates in the presence of H89, a PKA inhibitor, and ^{19}F NMR experiment. HPLC chromatograms were measured over 60 min on a 0-70% gradient of buffer B in buffer A.

All of these data can be quantified based on the extent of phosphorylation over time (Figure 3.39). While these data are not a true measure of the kinetics of phosphorylation, they do point to this system's potential to be used in measuring enzyme kinetics by ^{19}F NMR which has previously been demonstrated in other systems.[189] Kinetic analysis of phosphorylation requires significantly more rigor than the data presented here. Quantitative analysis of the phosphatase-inhibited and the non-inhibited experiments demonstrate very little difference in the extent of phosphorylation between the conditions, confirming little or no dephosphorylation as a result of phosphatases under these solution conditions.
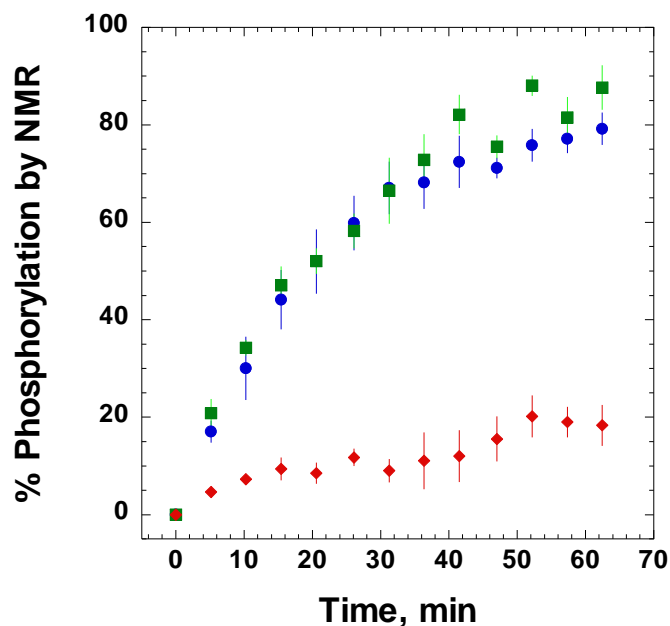
Figure 3.39 Quantification of real time ^{19}F NMR data in HeLa cell lysates for phosphate-inhibited (green squares), non-inhibited (blue diamonds), and H-89 inhibited (red diamonds) of the 4*R*-Kemptide peptide. Data represent three independent trials, and error bars indicate standard error.

3.3   Summary and Discussion

In this work, we demonstrated that perfluoro-*tert*-butyl hydroxyprolines could be used to detect protein kinase activity in complex mixtures. The 4*R*-perfluoro-*tert*-butyl hydroxyproline was introduced into Abl, PKA, and Akt kinase recognition motifs, and successfully allowed detection of phosphorylation on all three substrates. A chemical shift change by ^{19}F NMR was detected upon phosphorylation in all substrates. A 0.01 ppm chemical shift change was observed in the Abl kinase substrates containing 4*R*-perfluoro-*tert*-butyl hydroxyproline, while a 0.05 ppm chemical shift change was observed for both the PKA and the Akt sensor peptides.

The 4*S*-perfluoro-*tert*-butyl hydroxyproline was introduced into PKA and Akt kinase recognition motifs and successfully detected phosphorylation on both substrates. Similarly, a 0.06 ppm chemical shift change by ^{19}F NMR was detected upon phosphorylation in all substrates. Despite the similarity between PKA and Akt, different stereochemical preferences between the 4*R* and the 4*S* perfluoro-*tert*-butyl hydroxyprolines. These data indicate a potential mechanism to differentiate related kinases with similar recognition sequences. Significant cross-talk is observed for PKA and Akt substrates both in sensors and in biological systems due to the similarity is substrate recognition sequences.[230] This similarity is problematic in designing sensors specific to an individual protein kinase. Both Abl and PKA kinase substrates exhibited phosphorylation in HeLa cell lysates which was detectable by ^{19}F NMR. Furthermore, the extent of phosphorylation could be quantified based on integration of ^{19}F NMR peaks, demonstrating these substrates containing the perfluoro-*tert*-butyl moiety to be potential probes for enzymatic kinetics by NMR as was previously done by Dalvit *et al*.[38,41]

While there are examples of using ^{19}F labelled amino acids in the literature for kinase, proteolysis, and kinetic probes by NMR, a number of improvements have been made over prior systems.[38,40,41] The first is increased sensitivity due to the nine equivalent fluorines compared to one or three equivalent fluorines. Probes developed for this work were easily detected at 5 µM in five minutes, while other systems require 30 µM or higher quantities for similar assays with ten minute or longer acquisition times, as is seen in work by Dalvit *et al.* with the Akt sensor ARKRERAF(3-CF_3)SFGHHA under similar assay conditions (buffers, temperatures, etc). Furthermore, less than 0.02 ppm chemical shift change was observed in the Dalvit *et*

al. system, as compared to our 4*S*-perfluoro-*tert*-hydroxyproline substrate, which has a 0.06 ppm chemical shift change upon phosphorylation. In the course of this work, it was also demonstrated that kinases not only tolerate these amino acids in the active site, but phosphorylate with similar rates to other kinase sensors containing all canonical amino acids. While this technology may not be applicable to all kinases, in properly designed substrates, kinases which utilize proline and large hydrophobic residues near the phosphorylation site (within a few residues) should be able to be monitored by ^{19}F NMR. Furthermore, these sensors were able to demonstrate a larger chemical shift change than previously observed with other fluorinated amino acids in the detection of Ser/Thr phosphorylation. Finally, by utilizing both synthesized diastereometers, different phosphorylation rates were observed with similar kinases, demonstrating the potential for these amino acids to differentiate kinase activity by closely related kinases.

We have clearly demonstrated substrates containing perfluoro-*tert*-butyl amino acids can be used to detect enzyme activity in complex media, including HeLa cell lysates. These data indicate the potential for these ^{19}F-based peptide sensors could be utilized to detect enzyme activity in live cells in real-time. Due to the amount of background signal, standard ^{1}H NMR experiments cannot easily detect enzyme activity. A number of groups have explored monitoring phosphorylation by ^{31}P NMR.[231-233] While possible, ^{31}P suffers from similar problems as ^{13}C and ^{15}N as discussed in Chapter 1. Furthermore, ATP cannot be used in mixtures containing more than one enzyme which utilized ATP. Multidimensional NMR has often been used, including ^{13}C/^{15}N and ^{1}H/^{15}N 2D experiments, have been used to examine Ser/Thr phosphorylation.[30,32,34,35,234,235] In many cases, downfield chemical shift changes are

observed upon phosphorylation.[234,235] However, these techniques require complex pulse sequences and long acquisition times, which may or may not be compatible with live cells. Data interpretation can also be complex. ^{19}F NMR has the greatest potential for monitoring enzyme activity by NMR; however, the field is still in its infancy. While work by Dalvit *et al.* and Stockman demonstrate the ability to detect enzyme activity, they do not move beyond assays employing purified enzymes.[38,40,41,197] Furthermore, there is a significant body of work utilizing ^{19}F NMR and MRI as tracers or as structural probes in live cells.[1-3,5,6,20,40,42,81,121,236-239] Our system is a significant improvement over previous NMR-based methods of kinase activity detection. The HeLa cell lysate data demonstrate that, not only can kinase activity be detected quickly, in real-time, but the data are also quantitative, which can be used to monitor kinetics of enzymes in live cells.

Moving forward, there are a number of applications for these particular amino acids in the detection of post-translational modifications. While proteolysis was detected in HeLa cell lysates, it was an undesired side product. As detection of proteolysis has already been demonstrated, one logical application is the detection of proteolysis in cells and animals. Matrix metalloproteins, as previously mentioned, are extracellular proteases which are activated in metastatic cancers.[100] These probes could potentially be introduced into MMP substrates to image pathways relating to metastatic cancers for both study and pre-clinical imaging. MMP-9 is one of the most well studied of the matrix metalloproteins, which has a known recognition sequence of PxxHy, where Hy is a hydrophobic amino acid.[240] 4*R*-perfluoro-*tert*-butyl hydroxyproline could potentially be introduced at either the Pro or the hydrophobic residue in the peptide **SGPLFYS**VTA for detection of MMP-9 activity. The

hydrophobic residue could also be replaced with either perfluoro-*tert*-butyl tyrosine or perfluoro-*tert*-butyl homoserine if the hydroxyproline is not accepted. The chemical shift change previously observed due to proteolysis was 0.04 ppm. These data indicate the introduced perfluoro-*tert*-butyl hydroxyproline should be close, within a few amino acids, to the site of proteolysis. Furthermore, the 4*S*-perfluoro-*tert*-butyl hydroxyproline consistently showed a larger chemical shift change, which could potentially be exploited in this context.

Another potential application is imaging the dynamics between glycosylation and phosphorylation. There are a number of phosphorylation sites in various proteins which are known to also be glycosylation sites.[85,170,179] These probes could potentially be used to study the interplay between phosphorylation and glycosylation. One example is studying phosphorylation and *O*-GlcNAcylation in the tau protein. Tau is known to be hyperphosphorylated in Alzheimer's Disease. Furthermore, it has been demonstrated that phosphorylation and *O*-GlcNAcylation have opposing structural effects in tau peptide model systems.[85] Perfluoro-*tert*-butyl hydroxyproline derivatives could potentially be introduced into this system to study the interplay of phosphorylation and *O*-GlcNAcylation in the tau system to examine the mechanism of tau misfolding in Alzheimer's Disease. Initially, the 4*R*-perfluoro-*tert*-butyl hydroxyproline could be introduced into the model peptide Ac-KTPP-NH$_2$ and Ac-KSPP-NH$_2$ to determine chemical shift changes with Ser/Thr, OGlcNAcylation, and phosphosphorylation. These experiments will determine if and how the chemical shift changes as a result of OGlcNAcylation and phosphorylation. Both Pro sites should be examined to determine if both sites are viable for the perfluoro-*tert*-butyl hydroxyproline, which will allow for more sites which could be examined to

potentially allow for enzyme recognition. Numerous sites in the tau protein are known to be phosphorylated by GSK-3β. To determine if the 4*R*-perfluoro-*tert*-butyl hydroxyproline is tolerated by GSK-3β, it would be introduced into Ac-KRREILSRRPpSYR-NH$_2$ peptide at the Pro position and subjected to enzymatic phosphorylation.

There is also potential to use perfluoro-*tert*-butyl hydroxyproline derivatives as oxidation sensors within cells. Oxidative stress is seen in cells due to pathologies including diabetes, cancer, and neurodegenerative diseases.[241-243] These pathways are often poorly understood, due to non-enzymatic post-translational modifications like nitration and oxidation events. Nitration, for example, often takes place on exposed Tyr residues. When nitration occurs, the electronics of the Tyr ring change due to the extremely electron-withdrawing nature of the nitro group. Furthermore, many sites of tyrosine nitration occur in extended loop structures.[244,245] Proline residues are also often found in loop structures.[53] These oxidative changes should be detectable by ^{19}F NMR and, in the case of non-enzymatic modifications, no recognition sequence would be required to detect modifications (Figure 3.40).
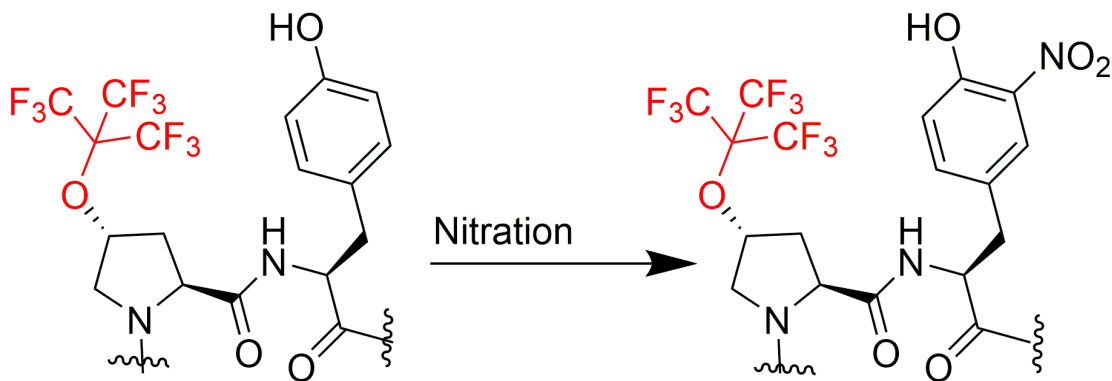
Figure 3.40 4*R*-perfluoro-*tert*-butyl hydroxyproline could potentially be introduced into a small β-turn peptide adjacent to a tyrosine as a tyrosine nitration sensor.

Beyond detection of post-translational modifications, these substrates demonstrate that substrates containing perfluoro-*tert*-butyl hydroxyproline could be used to study kinetics and potential therapeutic agents. As already demonstrated, not only are kinases high profile drug targets, but these sensors can be used to detect and quantify phosphorylation events in complex mixtures, including cell lysates. There is significant interest in developing new screening methods to examine new potential drugs more accurately and faster than current methods.[38] These sensor peptides could potentially be exploited to detect kinase inhibitors in cell lysates as potential drug targets by ^{19}F NMR. Furthermore, these peptides could be introduced into live cells, via cell-penetrating peptides, to examine kinase inhibition in live cells. There is also potential for these probes to be used to study the mechanism of action of drugs in various systems. Having kinase-specific probes allows for the study of specific pathways which can be monitored for upregulation or suppression of signaling pathways upon which drugs may act.

3.4   Experimental

3.4.1   Peptide Synthesis, Purification, and Characterization

All peptides were synthesized using standard Fmoc solid phase peptide synthesis (SPPS) on Rink amide resin, using HATU, or COMU as a coupling reagent. Perfluorinated amino acids were coupled in four equivalent excess with COMU for 12 hours. All other amino acids were coupled in four equivalent excess for two hours with HBTU or HATU as a coupling reagent. All peptides were synthesized by hand or on a Rainin PS3 (Rainin Instruments, Woburn, MA). All peptides were acetylated on the N-terminus (5% acetic anhydride in pyridine, 3 mL, 3 × 5 min) and contained a C-terminal amide.

Peptides to be phosphorylated were synthesized with a trityl-protected serine residue. In a fritted tube the trityl group was removed after synthesis and acetylation with 2% trifluoroacetic acid (TFA) and 5% triethylsilane (TES) in CH_2Cl_2 (3 mL, 3 × 1 min). Following trityl deprotection the resin was washed with CH_2Cl_2 (3 mL, 3 × 1 min). The phosphitylation was performed by the addition of tetrazole (1.35 mmol; 3 mL of 3% tetrazole in acetonitrile) and O, O′-dibenzyl-N,N-diisopropylphosphoramitdite (1.52 mmol, 500 μL). The reaction was allowed to stir at room temperature for 4 to 5 hours. The solution was removed and the resin was washed with DMF (3 mL, 3 × 1 min) and CH_2Cl_2 (3 mL, 3 × 1 min). Oxidation was performed with t-butyl hydroperoxide (3 mL, 3 M solution in CH_2Cl_2) for 1 hour at room temperature. The solution was removed and the resin was washed with CH_2Cl_2 (3 mL, 3 × 1 min) and diethyl ether (3 mL, 3 × 1 min).

Peptides were cleaved from the resin and deprotected for 3 to 4 hours using 5% H_2O, 5% triisopropylsilane, and 90% TFA in 1 mL. TFA was removed by evaporation

and the peptides were precipitated with diethyl ether. Ether was removed by decantation, and the peptide was dissolved in water and filtered before HPLC. Peptides were purified by reverse phase HPLC (Vydac semipreparative C18, 10 × 250 mm, 5 μm particle size, 300 Å pore). Analytical data for peptides: 4R-FAbletide [retention time by HPLC (t_R, 37.1 min; expected mass, 1539.7, observed mass, 1562.3 (M + Na)$^+$], p4R-FAbletide [retention time by HPLC (t_R, 43.1 min; expected mass, 1619.7, observed mass, 1642.4 (M + Na)$^+$], 4R-FAbletideR [retention time by HPLC (t_R, 41.2 min; expected mass, 1548.7, observed mass, 1571.4 (M + Na)$^+$], p4R-FAbletideR [retention time by HPLC (t_R, 43.2 min; expected mass, 1628.7, observed mass, 1651.9 (M + Na)$^+$], 4R-FAbletideN [retention time by HPLC (t_R, 37.8 min; expected mass, 1507.4, observed mass, 1530.7 (M + Na)$^+$], p4R-FAbletideN [retention time by HPLC (t_R, 34.2 min; expected mass, 1597.4, observed mass, 1620.7 (M + Na)$^+$], 4R-FKemptide [retention time by HPLC (t_R, 47.7 min; expected mass, 1271.2, observed mass, 1272.7 (M + H)$^+$], p4R-FKemptide [retention time by HPLC (t_R, 43.4 min; expected mass, 1351.4, observed mass, 1351.7 (M + H)$^+$], 4S-FKemptide [retention time by HPLC (t_R, 45.5 min; expected mass, 1271.2, observed mass, 1272.7 (M + H)$^+$], p4S-FKemptide [retention time by HPLC (t_R, 42.8 min; expected mass, 1351.4, observed mass, 1351.7 (M + H)$^+$], 4R-FAktide [retention time by HPLC (t_R, 41.5 min; expected mass, 1925.6, observed mass, 963.3 (M + H)$^{2+}$], p4R-FAktide [retention time by HPLC (t_R, 38.8 min; expected mass, 2007.6, observed mass, 1004.3 (M + H)$^{2+}$], 4S-FAktide [retention time by HPLC (t_R, 38.3 min; expected mass, 1925.6, observed mass, 963.3 (M + H)$^{2+}$], and p4S-FAktide [retention time by HPLC (t_R, 42.6 min; expected mass, 2007.6, observed mass, 1004.3 (M + H)$^{2+}$],

3.4.2   4*R*-FKemptide phosphorylation by PKA

cAMP-dependent Protein Kinase catalytic subunit (PKA) was purchased from New England BioLabs (Cat. # P6000S). Reaction mixtures were prepared to a final volume of 25 μL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 2 mM DTT, 100 μM non-phosphorylated peptide, 400 μM ATP, and 1 μL enzyme solution (2,500 units). After incubation at 37 °C for 5 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 μL autoclaved water and 50 μL D$_2$O before transferring the resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.3   4*S*-FKemptide phosphorylation by PKA

cAMP-dependent Protein Kinase catalytic subunit (PKA) was purchased from New England BioLabs (Cat. # P6000S). Reaction mixtures were prepared to a final volume of 25 μL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 2 mM DTT, 100 μM non-phosphorylated peptide, 400 μM ATP, and 1 μL enzyme solution (2,500 units). After incubation at 37 °C for 5 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 μL autoclaved water and 50 μL D$_2$O before transferring the

resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.4   4*R*-FAktide phosphorylation by Akt

Protein Kinase B (Akt) was purchased from Signal Chem. Reaction mixtures were prepared to a final volume of 25 µL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 5 mM DTT, 200 µM β-glycerophosphate, 100 µM non-phosphorylated peptide, 400 µM ATP, and 2 µL enzyme solution (2 ng). After incubation at 37 °C for 60 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 µL autoclaved water and 50 µL D$_2$O before transferring the resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.5   4*S*-FAktide phosphorylation by Akt

Protein Kinase B (Akt) was purchased from Signal Chem. Reaction mixtures were prepared to a final volume of 25 µL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 5 mM DTT, 200 µM β-glycerophosphate, 100 µM non-phosphorylated peptide, 400 µM ATP, and 2 µL enzyme solution (2 ng). After incubation at 37 °C for 60 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 µL autoclaved water and 50 µL D$_2$O before transferring the resultant solution to an NMR

tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.6   4R-FAbltide phosphorylation by Abl Kinase

Abl Tyrosine Kinase (Abl) was purchased from New England BioLabs (Cat. # P6050S). Reaction mixtures were prepared to a final volume of 25 μL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT, 100 μM non-phosphorylated peptide, 400 μM ATP, and 2.5 μL enzyme solution (12.5 units). After incubation at 30 °C for 60 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 μL autoclaved water and 50 μL D$_2$O before transferring the resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.7   4R-FAbltideR phosphorylation by Abl Kinase

Abl Tyrosine Kinase (Abl) was purchased from New England BioLabs (Cat. # P6050S). Reaction mixtures were prepared to a final volume of 25 μL with 50 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 0.1 mM EDTA, 2 mM DTT, 100 μM non-phosphorylated peptide, 400 μM ATP, and 2.5 μL enzyme solution (12.5 units). After incubation at 30 °C for 60 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 μL autoclaved water and 50 μL D$_2$O before transferring the

resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.8 4*R*-FAbltideN phosphorylation by Abl Kinase

Abl Tyrosine Kinase (Abl) was purchased from New England BioLabs (Cat. # P6050S). Reaction mixtures were prepared to a final volume of 25 µL with 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 100 µM non-phosphorylated peptide, 400 µM ATP, and 2.5 µL enzyme solution (12.5 units). After incubation at 30 °C for 60 mins, the reaction mixture was heated to 100 °C for 15 minutes to inactivate the enzyme. The solution was then centrifuged for 30 seconds and diluted with 425 µL autoclaved water and 50 µL D_2O before transferring the resultant to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.9 Real Time NMR of 4*R*-FKemptide phosphorylation by PKA

PKA (1 µL, New England BioLabs) was diluted in buffer containing 50 mM Tris-HCl (pH 7.5) and 10 mM $MgCl_2$ (39 µL). Reaction mixtures were prepared to a final volume of 500 µL with 10 µM non-phosphorylated peptide, 80 µM ATP, 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM DTT, and 10% D_2O. Experiments were conducted at 310 K. One NMR experiment was taken before enzyme was added as $t = 0$. The tube was removed from the NMR and 5 µL (313 U) of the diluted enzyme was added. A total of six more experiments were completed without removing the sample from the NMR. Each experiment was 5 mins and 14 seconds for a total time of 31.4

mins. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.10 Cell Culture and cell extracts

HeLa cells were cultured at 37 °C humidified environment containing 5% CO_2 with Dulbecco's Modified Eagle Medium (DMEM) with 10% heat-inactivated fetal bovine serum (FBS), L-Glutamine (2 mM), penicillin (100 units/mL), and streptomycin (100 μg/mL). Twenty hours before lysate preparation, cells were starved with DMEM containing 0.5% FBS. The media was removed and the cells were washed with 4 mL 1× DPBS. The cells were trypsinized and centrifuged (3.5 rpm, 1 min). The pellet was resuspended in 2 mL 1× DPBS and centrifuged (3.5 rpm, 1 min). The pellet was resuspended in 1 mL Buffer A (0.4 M HEPES (pH 7.9), 60 mM $MgCl_2$, 400 mM KCl, 20 mM DTT, and 8 mM PMSF) and centrifuged (3.5 rpm, 1 min). The pellet was resuspended in Buffer A and incubated on ice for 10 minutes. The solution was then vortexed for 30 second, centrifuged (3.5 rpm, 1 min), and resuspended in Buffer B (100 mM HEPES (pH 7.9), 4.1 M NaCl, 14.7 mM $MgCl_2$, 200 uM EDTA, 5 mM DTT, 5 mM PMSF, and 2.5% gylcerol). The solution was incubated on ice for 15 minutes and then centrifuged (3.5 rpm, 5 min). The supernantant was divided into aliquots and frozen at -80 °C.

3.4.11 4*R*- FAbltideR phosphorylation by HeLa Cell Lysates

The reaction mixtures of peptide in HeLa cell extracts were prepared to a final volume of 100 μL with stock solutions were mixed to yield final concentrations of 160 μM ATP, 200 μM β-glycerophosphate, 200 μM sodium orthovanadate (Na_3VO_4), 50

mM Tris-HCl, 10 mM MgCl$_2$, 100 μM non-phosphorylated peptide, and 70 μL cell extracts. After incubation at 37 °C for two hours, the reaction mixture was diluted with 350 μL autoclaved water and 50 μL D$_2$O before transferring the resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.12 4*R*-FAbltideN phosphorylation by HeLa Cell Lysates

The reaction mixtures of peptide in HeLa cell extracts were prepared to a final volume of 100 μL with stock solutions were mixed to yield final concentrations of 160 μM ATP, 200 μM β-glycerophosphate, 200 μM sodium orthovanadate (Na$_3$VO$_4$), 50 mM Tris-HCl, 10 mM MgCl$_2$, 100 μM non-phosphorylated peptide, and 70 μL cell extracts. After incubation at 37 °C for two hours, the reaction mixture was diluted with 350 μL autoclaved water and 50 μL D$_2$O before transferring the resultant solution to an NMR tube. Upon completion of NMR experiments, the solution was injected on the HPLC to verify the extent of phosphorylation.

3.4.13 Real Time NMR of 2*S*, 4*R*-FKemptide phosphorylation by HeLa Cell Lysates

The reaction mixtures of peptide in HeLa cell extracts were prepared to a final volume of 500 μL as follows: stock solutions were mixed to yield final concentrations of 160 μM ATP, 200 μM β-glycerophosphate, 200 μM sodium orthovanadate (Na$_3$VO$_4$), 50 mM Tris-HCl, 10 mM MgCl$_2$, 20 μM non-phosphorylated peptide, 10% D$_2$O and 175 μL cell extracts. Experiments were conducted at 310 K. Peptide, ATP, and inhibitors, were dissolved in D$_2$O and buffer. The cell lysates were added and the

entire reaction mixture was transferred to an NMR tube. A total of twelve experiments were completed, removing the sample after the sixth to add supplemental ATP (80 µM, 10 µL). NMR experiments (128 scans, 5.14 mins) were run as per real time *in vitro* assays for a total of 62.8 mins. Reactions were then diluted with 500 µL of distilled water and injected on the HPLC to verify extent of phosphorylation.

Chapter 4

MECHANISTIC INTERROGATION OF POLYPROLINE HELICIES VIA 19F NMR UTLIZING 4,4-DIFLUOROPROLINE AS A SITE SPECIFIC PROBE

4.1 Introduction

Proline is a unique amino acid which, due to the rigid, cyclic side chain, is conformationally restricted, leading to unique structural elements within proteins. Proline residues are most commonly found in turns and at the beginning or ending of predominant secondary structures such as α-helices and β sheets.[53] Proline *cis-trans* isomerization is well documented in proteins, peptides, and even as monomers with protecting groups due to the decreased steric differences between the *cis* and *trans* isomers of proline versus other amino acids.[246-255] In all other canonical amino acids, *cis-trans* isomerization is unfavored due to steric clash in the *cis* conformation. The ω dihedral angle in the *cis* conformation is $0\,°$ which leads to the C_α of the two amino acids to occupy the same space, leading to unfavorable steric interactions.[256] In the case of Xaa-Pro sequences, the cyclic nature of proline does not have the same steric demands as other amino acids, allowing for the required $\omega = 0\,°C$ without the steric class observed in Xaa-Xaa sequences. The similar steric demands in both the *cis* and *trans* conformation for Xaa-Pro sequences, the free energy change between the two conformations is 1.8 kcal mol^{-1}, 3 kcal mol^{-1} less than that of Xaa-Xaa sequences (4.8 kcal mol^{-1}). This reduction in energy barriers between the *cis* and *trans* isomers leads

to higher populations of *cis* amide bond conformations in Xaa-Pro sequences than sequences without proline.

In some cases of Xaa-Pro sequences the *trans* amide bonds can, with certain torsion angles, be stabilized an n→π* interaction between the preceding carbonyl oxygen and the following carbonyl carbon.[254,257] The *trans* conformation is characterized by the ω = 180° dihedral angle.[258] The *cis* conformation is characterized by the ω = 0°, and is not capable of the n→π* interaction due to the orientation of the carbonyls.[258] Proline can undergo *cis-trans* isomerization, depending on local structure, on a time scale of seconds spontaneously. *Cis-trans* isomerization of a simple Ac-Pro-OMe in water has an energy barrier of 20 kcal mol^{-1} (Figure 4.1).[259]

<div align="center">

trans amide   *cis* amide
ω = 180º   ω = 0º

</div>
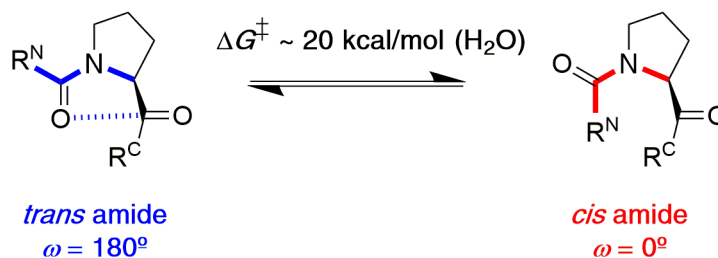
Figure 4.1   *cis-trans* isomerism of the proline amide bond. The energy barriers (activation free energies, $\Delta G^{\ddagger}$) for *trans*-to-*cis* and *cis*-to-*trans* amide bond rotation in water for Ac-Pro-OMe are 21.1 kcal/mol and 20.2 kcal/mol, respectively.[259] The magnitude of activation barrier is lower in more non-polar solvents, due to the reduced polarity of the transition state compared to the ground state.[246,255] The activation barrier is also reduced by proline isomerases.[260]

Approximately 5% of proline residues are in the *cis* amide bond conformation, compared to less than 0.03% of other residues, according to surveys of prolines in the PDB.[250,261] Proline *cis-trans* isomerization is often a key step in protein folding.

Despite the lower free energy difference to *cis-trans* isomerization relative to other amino acids, it is still a rate limiting step in protein folding due to the 20 kcal mol[-1] activation barrier. Small globular proteins can fold on the micro- to millisecond time scale.[262] *Cis-trans* isomerization can occur spontaneously (1 to 100 s. timescale), which is primarily dependent on local structure. All protein residues are incorporated into proteins as *trans* amide bonds, however, *cis-trans* isomerization event can be catalyzed by isomerases to increase the rate of isomerization. Isomerases do not alter the equilibrium of isomerization but simply alter the kinetics to increase the rate of isomerization. These isomerases, termed prolyl isomerases, are integral to increasing the rate of proper protein folding.[260,263,264]

Peptidyl prolyl isomerases are also implicated in cell signaling events, where the proline *cis-trans* isomerization is a key step in cell signaling. Fischer *et al*. demonstrated the conformation of a proline in a protease substrate can greatly influence reaction rates of enzymes.[247] In short, they demonstrated that the proline-dependent proteases, dipeptidyl peptidase IV, could only cleave an Ala-Pro amide bond in the *trans* conformation. When an additional proline was added to the substrate (Ala-Pro-Pro), approximately 10% *cis* amide bond was observed in the substrate, which led to much slower reaction rates for dipeptidyl peptidase IV cleavage. These data demonstrate *cis-trans* isomerization can be the rate-limiting step in proteolysis of proteins.

Proteolysis is not the only post-translational modification affected by proline conformation within a protein substrate. There is a class of kinases, known as proline-directed kinases, which utilize proline in the recognition sequence. A few examples of proline-directed kinases include the cyclin-dependent kinase (CDKs), the mitogen-

activated protein kinases (MAPKs), and glycogen synthase kinase 3 (GSK3).[249] The regulation of these kinases is directly linked to the conformation of the proline within the substrate. Peptidyl prolyl isomerases are responsible for interconversion of *cis* and *trans* amide bonds.[263] When proteins are synthesized within the cell, prolines are incorporated as *trans* amide bonds.[264] For proteins that require *cis* amide bonds for proper folding, peptidyl prolyl isomerases catalyze the conversion of the *trans* amide bond to the *cis* amide bond, increasing the rate of protein folding. More recently, however, *cis-trans* isomerization has been shown to be a key step in regulating phosphorylation events.[249] One example of *cis-trans* isomerization playing a key role in phosphorylation events is the activity of Pin1, a peptidyl prolyl isomerase.[249] Pin1 is known to catalyze isomerization events at phosphorylated Ser/Thr-Pro sequences. Pin1 catalyzes a change in conformation leading to dephosphorylation of the nearby phosphoserine or phosphotheronine. In the case of Cdc25c, a Cdc2-directed phosphatase, is dephosphorylated by Protein Phosphatase 2A (PP2A) after the key *cis* amide bond is converted to a *trans* amide bond by Pin1.[249] This dephosphorylation event alters Cdc25c activity, preventing dephosphorylation of downstream targets. Cdc25c is a key phosphatase in mitosis ultimately, at least in part, controlling cell proliferation and growth.[265] Furthermore, Pin1, while not entirely understood, is known to be overexpressed in certain cancers, including Her2+ breast cancer. The overexpression of Pin1 in cancer is unsurprising given its importance in helping to control phosphorylation events and control of cell signaling pathways.[249]

Given the apparent importance of *cis-trans* isomerization within proteins, new tools are needed to study proline isomerization and the conformation of prolines within proteins. As will be discussed, current methods in determining the

configuration of prolines within proteins are technically difficult. Developing new tools, which can be used not only to determine the conformation but monitor protein folding events, ideally in complex mixtures such as cell lysates or in live cells, is necessary to fully explore the impact of proline conformation on protein structure. This new tool should impart minimal effect on *cis-trans* isomerization in a peptide or protein context, it should be sterically similar in size to proline, and ideally, be expressible for straightforward integration into proteins.

4.1.1 Detection of *Cis-Trans* Isomerization and Polyproline Helices

Polyproline helices were originally identified in 1963 by powder extra diffraction.[266] These powder diffraction patterns were key in establishing the presence of the two unique structures of polyproline helices. Since these original powder diffraction, only PPII helical structures have been observed in single crystal x-ray diffraction structures. Notably, PPII helices are seen in protein crystal structures in the PDB in globular proteins as well as in natively disorder proteins.[250,267] There is no single crystal x-ray structure of a PPI helix structure. Crystal structures are important for characterizing the structures of PPI and PPII helices, but x-ray crystallography cannot be used to examine the kinetics of interconversion between PPI and PPII helix formation. Other methods detection of PPI and PPII helices are circular dichroism, ^{13}C chemical shift analysis, and more recently mass spectrometry.

Both PPI and PPII helices are often examined by circular dichroism.[55,56,254,267-274] PPI helices have a characteristic maximum at 215 nm, while PPII helices have a maximum at 228 nm.[275] CD is very effective for observation of polyproline helix structure and has provided key information about induction of PPI and PPII helix structure. The CD spectrum indicates the global structure of the peptide. This method

has been used to examine the formation of PPI and PPII helices as well as the interconversion from one structure to the other.[248,272,273,275,276] CD has been one of the primary methods to study the formation of PPI helices, including establishing the number of residues required to form a PPI helix and determining solvent effects on PPI formation (non-polar organic solvents more readily induce the PPI helix)[275]. CD has also been key in interrogating the mechanism of the polyproline interconversion including the kinetics of interconversion and potential cooperativity associated with PPI helix formation.[273] Despite the volume of work and the importance of CD in studying proline oligomers, new methods are needed to develop a full mechanistic understanding of polyproline helix interconversion. One of the primary drawbacks of utilizing CD is it only reports on global structure of the peptides in solution. The CD spectrum averages all of the various conformations in solution, preventing local, site-specific information from being obtained by this method. CD spectroscopy continues to be a valuable tool for studying polyproline helix structure; however, other supplementary methods are needed along with CD to develop an understanding of PPI and PPII helices and their interconversion.

PPI and PPII helices can also be identified by ^{13}C chemical shift analysis.[277,278] In *trans* conformations, ^{13}C chemical shift for the proline Cβ and Cγ are similar (1 to 4 ppm). In contrast the *cis* conformation exhibits a larger chemical shift dispersion for the proline Cβ and Cγ (5 to 10 ppm). This method is very effective for determining *cis-trans* isomerization in small systems, like single residues with protecting groups or dipeptides. However, it is problematic for peptides and proteins due to overlapping signals and requires a full sequence assignment for the entire protein or peptide. Furthermore, a large amount of material is required, or alternatively ^{13}C enrichment is

required, as was discussed in Chapter 1.[30] While this method can be effective and does allow for residue-specific information, it is time consuming, in both analysis and experiment time. Alternatively, NOE experiments can be used to gain better resolution and to make residue specific assignments, however, spectral overlap is still a problem, especially in proline-rich sequences.[279]

4.1.2   Polyproline I and Polyproline II Helices Have Different Organized Structures Which Interconvert From One Ordered Structure to the Other

In order to begin developing a new tool for studying *cis-trans* isomerization in proteins, potential probes must be studied in simplified systems as a proof of concept before introduction into more complex systems. One example of these model systems are proline oligomers which are known to undergo *cis-trans* isomerization.[276] Oligomers of proline are known to fold into two unique secondary structures based on proline's ability to more readily adopt *cis* amide bonds than other amino acids (Figure 4.2). Polyproline I (PPI) helices contain all *cis* amide bonds ($\phi = $ -83, $\psi = $ +158, $\omega = $ 0) in a compact right-handed helix. PPI helices exhibit 5.6 Å per turn (3.3 residues per turn) and are induced for oligomers of proline in organic solvents such as methanol, isopropanol, and acetonitrile.[266] Alternatively, polyproline II (PPII) helices contain all *trans* amide bonds ($\phi = -75$, $\psi = $ +145, $\omega = $ 180) in an extended left-handed helix.[276] PPII helices are induced in aqueous solutions and can be found in protein structures. Both PPI and PPII structures were solved by powder x-ray diffraction over 50 years ago.[266] However, the mechanism of interconversion between PPI and PPII helices and also what allows for the stabilization of PPI helices are not fully understood.

polyproline II helix (PPII)
$\phi = -75°$, $\psi = +145°$
$\omega = +180°$ (all *trans* amide bonds)

3.1 Å rise/residue
3 residues/turn
9.3 Å rise/turn
no helix dipole

H₂O ⇅ organic solvent

polyproline I helix (PPI)
$\phi = -83°$, $\psi = +158°$
$\omega = 0°$ (all *cis* amide bonds)

1.7 Å rise/residue
3.3 residues/turn
5.7 Å rise/turn
large helix dipole
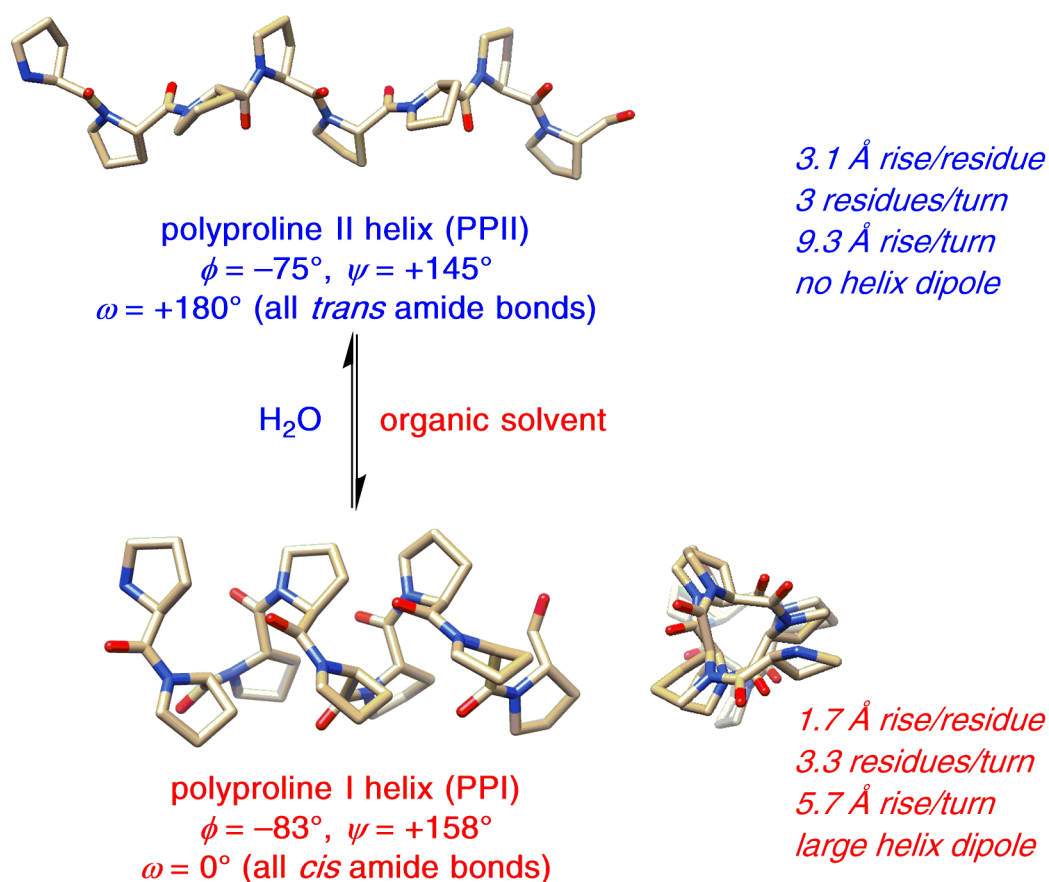
Figure 4.2   Polyproline I helix and polyproline II helix structure. (a) Proline oligomers can undergo an ordered-to-ordered transition between left-handed polyproline II helix (PPII) and right-handed polyproline I helix (PPI) as a function of solvent, with water favoring PPII and non-polar solvents favoring PPI. Structures are based on models based on the indicated ideal (canonical) torsion angles.

PPII helices are seen in structures in the PDB and are known to be stable in aqueous solutions.[280,281] Due to the stability and prevalence in protein structure, PPII helices are well-studied structures. PPII helices are, in part, stabilized by n→π* interactions between consecutive carbonyls. This interaction is due to the p-like orbitals of the lone pair on the first oxygen (n) overlapping with the π* molecular

orbital of the following carbonyl.[257] The geometry of this interaction results in the Bürgi-Dunitz trajectory of 109° from the donor carbonyl to the acceptor carbonyl.[257] These interactions demonstrate sub-van der Waals radii (less than 3.22 Å) as seen in Figure 4.3a. This favorable interaction restricts both ϕ and ψ as is visible in crystal structures of PPII sequences in the PDB.
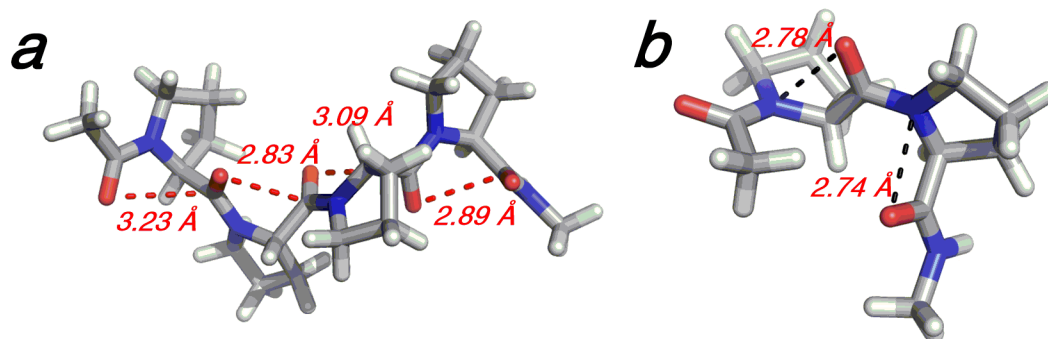
Figure 4.3   (a) Polyproline II helix from pdb 4qb3 (0.94 Å resolution, Brd4 bromodomain, residues 45-48). n →π* interactions (red dashed lines) and O...C=O distances are indicated. (b) Polyproline I helices form in organic solvents, but 2-residue PPI segments may be observed within proteins in the PBD (from pdb 3a72, residues Pro71 ($\phi, \psi = -72°, +166°$, *endo* ring pucker) and Pro72 ($\phi, \psi = -76°, +175°$; *endo* ring pucker), α-arabinase, 1.04 Å resolution). Notably, close intraresidue interactions (below the 3.07 Å sum of the van der Waals radii of oxygen and nitrogen, indicated by dashed lines) are observed between the carbonyl oxygen and the prior amide nitrogen (the nitrogen from the same residue, but from the previous amide bond).

Alternatively, very little is known about the mechanism of stabilization PPI helices despite the original powder diffraction structure being solved in 1956.[266] PPI helices are stabilized in organic solvent, but to date have not been found in a protein crystal structure. There are no single crystal structures of PPI helices, but they have

been studied in solution to determine key characteristics beyond the original powder diffraction pattern. There are also small, two reside-sequences in the PDB found in protein structures that occupy the same Ramachandran space as PPI helices (Figure 4.3b), indicating the PPI helices may have some biological relavence.

PPI helices have been examined in a number of organic solvents including methanol, 1-propanol, 1,4-dioxane, and n-butanol.[248,273,275] Lower solvent polarity is shown to promote PPI structure. Kakinoki *et al*. demonstrated that PPI more readily forms in 1-propanol (14 days to equilibrium at 5 °C) than in methanol (21 days to equilibrium at 5 °C) for a thirteen residue proline oligomer when monitored by circular dichroism.[275] These data are particularly interesting considering that PPI helices exhibit a helix dipole not observed in PPII helices. The dipole is opposite of that observed in a α-helix, with the positive charge on the C-terminus and the negative charge on the N-terminus.[282] Furthermore, PPI helices are promoted by increased length in oligomer.[275] Very little PPI structure is observed at equilibrium in in 1-propanol with a 6-residue proline oligomer compared to a 13-residue proline oligomer. No PPI structure is observed in 1-propanol with a 4-residue proline oligomer at equilibrium. These data are also inconsistent with increased PPI structure in non-polar solvents. The increased size of the oligomer also increases the strength of the helix dipole, which would be disfavored in non-polar solvents. These data demonstrate a need for new tools to study PPI structure as well as the PPI/PPII interconversion to understand the interactions that promote PPI structure.

Proline oligomers can reversibly interconvert between PPI and PPII structures with changes in solvent.[276] The interconversion between PPI and PPII was first described in 1953, but the exact mechanism of interconversion is not well understood.

This interconversion is particularly interesting considering it is an "ordered to ordered" transition state rather than an "ordered to disordered" transition such as protein unfolding as a result of heating.[251] The interconversion between PPI and PPII has been observed by CD and is completely reversible. The interconversion between the two structures requires the isomerization of each proline reside from *trans* to *cis* for PPII to PPI and from *cis* to *trans* for PPI to PPII. The isomerization of each proline residue has an energy barrier of 20 kcal mol^{-1} as seen in Figure 4.1.[259] The exact mechanism of the interconversion is not well understood, with a few studies examining the directionality and cooperativity of interconversion. Lin and Brandts examined the rate of hydrolysis of polyproline sequences by aminopeptidase D, which is a proline-directed protease which will only act on the *trans* conformation.[273,283] They examined both PPI and PPII helices interconverting in different solvents and measured the rate of proteolysis. Based on the rates of proteolysis, they proposed PPI helices form from the N-to-C terminus and PPII helices form from the C-to-N terminus. Early studies also examined cooperativity in the formation of PPI and PPII helix structures. Both Gornick *et al.* and Holzwarth and Chandrasekaren proposed cooperativity in the formation of polyproline helix structures. Gornick *et al.* examined the transition rates by specific rotation and x-ray crystallography, which led to sharp transition barriers rather than individual transition barriers for each residue.[284] Holzwarth and Chandrasekaren used calculations to propose cooperativity based on intramolecular electrostatics and intermolecular polymer-solvent interactions.[285] More recently, however, Beausoleil and Lubell have challenged the idea that the transition is cooperative.[286] They introduced 5-*tert*-butylproline, which has inherent steric effects which favor the *cis* conformation. They did not observe cooperativity with the

introduction of the 5-*tert*-butylproline, i.e. there was no increase in *cis* conformation, aside from the amide bond where the mutation was made. Given the work that has been done, there is still significant work to be done to determine the exact mechanism of interconversion.

4.1.3   4,4-Difluoroproline As A ^{19}F Probe for *cis trans* Isomerization

Fluorine is ideal as a site-specific probe for *cis-trans* isomerization. Prior chapters addressed the use of amino acids containing perfluoro-*tert*-butyl amino acids as probes for protein function. While very effective for detecting protein-protein interactions and post-translational modifications, the size and inherent structural preferences of the perfluoro-*tert*-butyl ether make it inappropriate to examine *cis-trans* isomerization in proline oligomers and protein structure. Single fluorine substitutions do not substantially increase the size of the proline, while still introducing fluorine atoms into the system for ^{19}F NMR. Renner *et al*. proposed the use of fluoroprolines in protein engineering based on the thermodynamics and equilibrium of *cis-trans* isomerization in the simple Ac-X-OMe model system.[259] 4*R*-fluoroproline, 4*S*-fluoroproline, and 4,4-difluoroproline were all examined in this context as well as unmodified proline. 4.4-Difluoroproline was found to have a lower activation barrier ($\Delta G^{\ddagger} = 19$ kcal mol^{-1}) to *cis-trans* isomerization than proline ($\Delta G^{\ddagger} = 20$ kcal mol^{-1}), leading to systems containing 4,4-difluoroproline equilibrating faster than systems containing proline. 4,4-Difluoroproline, however, did not alter the thermodynamics of *cis-trans* isomerization, unlike the 4*R*- and 4*S*-fluoroprolines, leading to similar populations of *cis* and *trans* as in proline. Others have found similar observations when examining fluoroprolines in other contexts. Single fluorine modification (4*R*-fluoroproline (Flp) or 4*S*-fluoroproline (flp)) contain inherent structural

preferences.[253,287] For example in the Ac-TYXN-NH$_2$ peptide model system, Flp and flp have a significant impact on the $K_{trans/cis}$ observed relative to Pro (Table 4.1).[253] These data are unsurprising due to the electronegative nature of fluorine and the inherent structural effects of 4R and 4S substitutions, as discussed in Chapter 1. Alternatively, the introduction of two fluorine atoms, despite the electronic changes to the system, lead to similar structure as that of Pro as observed in Table 4.10.[253] Similarly, Shoulders et $al.$ demonstrated 4,4-difluoroproline showed similar stability to proline in collagen model peptides.[288] Using the host-guest system Ac-(PXG)$_{10}$-NH$_2$ where X was Dfp, Pro, or Flp, which readily form the characteristic collagen triple helix, stability was measured by thermal denaturation via CD. All three peptides demonstrated a maxima at 225 nm; however, there was a stark contrast in the melting of the triple helix. Both the peptides containing proline and the 4,4-difluoroproline melted at similar temperatures (43 °C and 42 °C, respectively) compared to the 4R-fluoroproline (48 °C), indicating significant stabilization as a result of a single fluorine modification.

Ac-TYXN-NH$_2$ X =	$K_{trans/cis}$	$\Delta\Delta G$ kcal mol^{-1}
Flp	7.0	-0.56
Pro	2.7	0.00
Dfp	2.6	+0.02
flp	1.5	+0.35

Table 4.1 $K_{trans/cis}$ measurements and $\Delta\Delta G$ calculations for Flp, Pro, Dfp, and flp at the X position in the Ac-TYXN-NH$_2$ model peptide.[253] $K_{trans/cis}$ is defined as the ratio of the population of *trans* and *cis* as measured by ^{1}H NMR.

4,4-Difluoroproline has also been examined in synthetic proteins by Yu *et al.*, who introduced not only 4,4-difluoropline but also 4*R*-fluoroproline and 4*S*-fluoroproline into ß2-microglobulin.[254] ß2-microglobulin is a small soluble protein which is known to form insoluble plaques in patients on long-term hemodialysis. The solubility is, at least in part, dependent on the *cis-trans* isomerization of the proline at position 32 in the protein. The ß2-microglobulin was synthesized via solid-phase peptide synthesis with either proline or one of the fluoroproline derivatives, and native chemical ligation was used to synthesize the full length protein, which is a total of 99 residues long. *Cis-trans* isomerization is known to be the rate-limiting step of protein folding and denaturation in the wildtype protein. Upon the introduction of 4,4-difluoroproline at Pro$_{32}$, the *cis-trans* isomerization is no longer the rate limiting step for denaturation due to the decreased energy barrier associated with 4,4-difluoroproline. The mutant containing 4,4-difluoroproline showed reduced stability

under denaturing conditions, most notably a decreased melting temperature (Dfp T_m = 56 °C compared to Pro T_m = 61.5 °C).[254] The ^{19}F NMR spectra was also examined for the mutant containing the fluoroprolines. Interesting, even at high concentrations and long acquisition times, neither the 4*R* or the 4*S*-fluoroproline mutants were visible by ^{19}F NMR. The mutant containing 4,4-difluoroproline, however, was visible by ^{19}F NMR. Under non-denaturing conditions (pH 7.1), two sets of *cis* peaks were visible at -97 ppm and -112 ppm in the ^{19}F NMR spectra, indicating the 4,4-difluoroproline at residue 32 was in the *cis* conformation. Upon denaturation at pH 2.5, the *cis* peaks disappeared and a new peak appeared at -101 ppm indicated isomerization had occurred to the *trans* conformation. These data indicate 4,4-difluoroproline can be used to probe *cis-trans* isomerization in peptides and proteins.

4.2   Results

4.2.1   Synthesis and Characterization of 4,4-Difluoroproline

4,4-Difluoroproline has been synthesized previously both in solution as a monomer and on solid phase in a peptide context.[253,289] As demonstrated in Chapter 1, hydroxyproline is an inexpensive, readily available starting material from which 4-hydroxyproline derivatives can be synthesized. Demange *et al*. demonstrated the synthesis of 4,4-difluoroproline from 4*R*-hydroxyproline, as seen in Figure 4.4, in good yield in 1998 as the fully protected Boc amino acid.[289] This synthetic route requires Boc protection of the amine and methyl protection of the carbonyl as seen in Chapter 1. The alcohol is then oxidized using pyridinium dichromate (PDC) to form the ketone. The ketone is then converted into the geminal difluoride via diethylaminosulfur trifluoride. Having obtained the Boc and methyl protected 4,4-

difluoroproline, it can easily be deprotected and Fmoc protected, as seen in Figure 4.4. This same chemistry has also been applied by Thomas *et al*. on the Ac-TYXN-NH$_2$ model system on solid phase using "proline editing".[253] In short, Fmoc-hydroxyproline is incorporated into the peptide on solid phase. The alcohol on the 4-hydroxyproline is trityl protected on solid phase to complete the peptide synthesis. The trityl group can then be removed using 2% TES and 2% TFA in CH$_2$Cl$_2$, allowing for PDC oxidation, followed by conversion of the ketone to the difluoride via DAST.
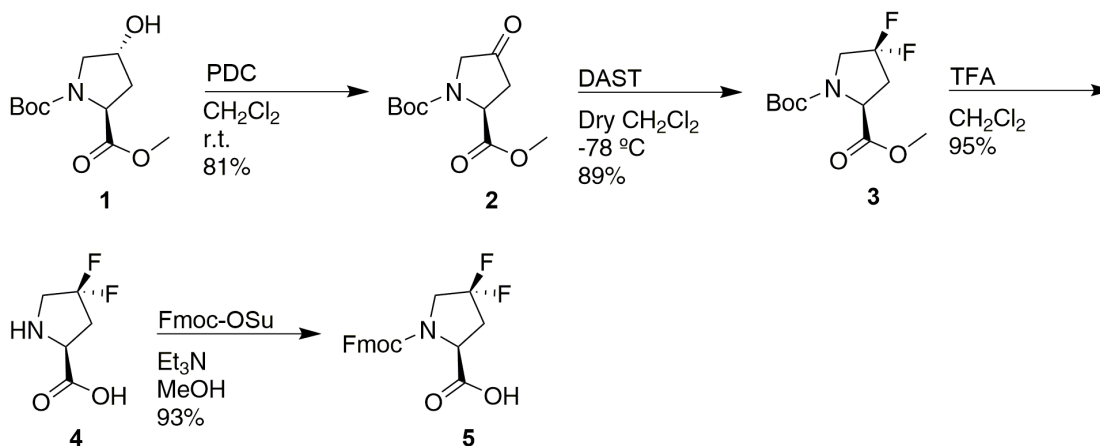
Figure 4.4   Solution phase synthesis of Fmoc-4,4-difluoroproline as described by Demange *et al*. This chemistry has also been applied on solid phase in a peptide by Pandey *et al*.

Both of these methods are useful synthetic routes. However, in the time between the publication of these papers and the work described herein, 4,4-difluoroproline became commercially available as the Boc amino acid. PDC oxidations, while high yielding and effective, are challenging to purify. The chromium metal oxidizes over the course of the reaction, resulting in a large amount of residual oxidized PDC which can be difficult to remove. Furthmore, PDC is toxic, introducing

unnecessary hazards to the synthetic route. Solid-phase chemistry was also considered

for this application; however, ultimately we would like to use this amino acid in

expressed proteins which requires the free monomer which can be dissolved in

expression media for incorporation into proteins. For these reasons, the Boc-4,4-

difluoroproline was purchased. The Boc-4,4-difluoroproline was deprotected using 2

M HCl to produce the free amino acid. The free amino acid was then Fmoc protected

using standard conditions described in Chapter 1 without purification, as seen in

Figure 4.5.[290] The Fmoc-4,4-difluoroproline can readily be incorporated into peptides.

This route minimized synthetic steps while readily accessing 4,4-difluoroproline as

both the free and Fmoc amino acid for either protein expression or peptide synthesis.
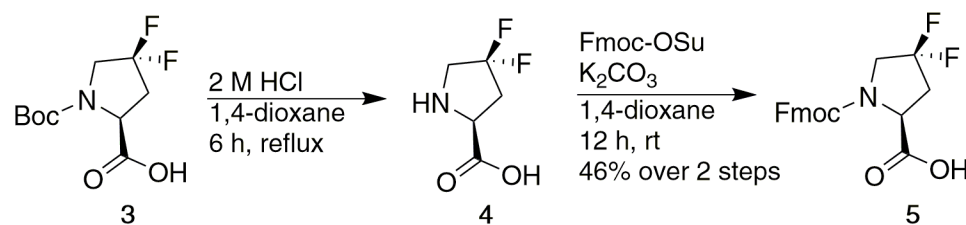
Figure 4.5   Fmoc-4,4-difluoroproline was synthesized from the commercially
available Boc-4,4-difluoroproline in two steps.

Fmoc-4,4-difluoroproline (**5**) readily crystalized from CDCl$_3$ leading to a

single-crystal structure which was solved by x-ray diffraction (Figure 4.6). The crystal

was examined to determine potential structural preferences as well as potential

interactions, such as ring pucker and n→π* interactions. We expect, based on solution

data in the Ac-TYXN-NH$_2$ peptide, for the 4,4-difluoroproline ($K_{trans/cis} = 2.6$, $\Delta\Delta G =$

+0.02 kcal mol^{-1}) to have similar conformations to that of proline ($K_{trans/cis} = 2.7$, $\Delta\Delta G$

= 0.00 kcal mol^{-1}).[253] In fact, Fmoc-4,4-difluoroproline crystalized with a *trans* amide bond ($\omega = +173$) and an *exo* ring pucker. The ϕ and ψ measured from the crystal structure were also consistent with a *trans* amide bond, with the ψ being in a slightly more extended confirmation than expected ($\phi = -61$, $\psi = +176$) (Figure 4.6). Furthermore, a weak n→π* interaction is visible between the consecutive carbonyls with 3.04 Å bond distance and a O…CO bond angle of 77°. The bond distance and angle are not ideal (> 3.00 Å bond distance, 109° bond angle) indicating a weakened n→π* interaction, potentially as a result of the fluorine substitutions which are known to significantly reduce the ammonium pKa of the free amino acid (pKa = 7.2) compared to Pro (pKa = 10.8) due to the inductive effects of fluorine.[43] The crystal structure data, along with solution data from prior work, indicate that 4,4-difluoroproline might make an ideal probe for *cis-trans* isomerization by ^{19}F NMR.
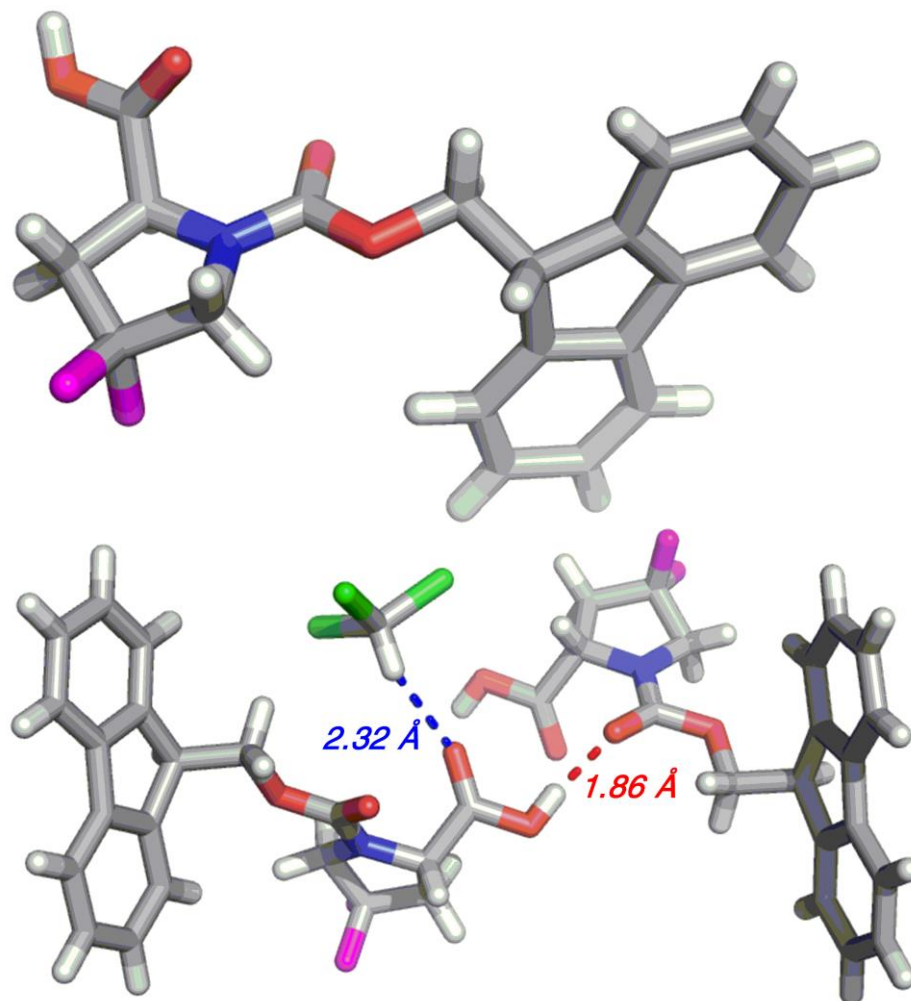
Figure 4.6   X-ray crystal structure (resolution 0.76 Å) of Fmoc-4,4-difluoroproline
(**5**), which crystallized from CDCl₃. **5** exhibited a compact value of ϕ but
an extended value of ψ ($\varphi = -61$, $\psi = +176$), a *trans* amide bond ($\omega =$
+173˚), and a Cγ–*exo* ring pucker. In addition, a close interaction was
observed between the carboxylic acid carbonyl oxygen and the carbamate
nitrogen (O...N distance 2.78 Å). A weak n→π* interaction was observed
between consecutive carbonyls (O...C=O distance 3.04 Å, < OCO 77˚
(ideal 109˚)). Crystal assembly was mediated by a close hydrogen bond
between the carboxylic acid OH and the carbamate carbonyl (O–H...O=C
distance 1.86 Å, indicated in red). In addition, a close C–H...O interaction
was observed between the C–D of chloroform and the acid carbonyl
(D...O=C 2.32 Å (2.22 Å with normalized C–D bond lengths), indicated
in blue).

4,4-Difluoroproline had previously been incorporated into the Ac-TYXN-NH$_2$ and the Ac-TAXN-NH$_2$ peptide model systems by Pandey *et al*. and the peptide was examined by ^{1}H and ^{19}F NMR to determine the *cis* amide bond population; however, no other characterization was reported.[43] These data were the first reported ^{19}F NMR spectra of 4,4-difluoroproline in a peptide or protein context, so in order to verify that the same species was synthesized and to verify previous results, the synthesized Fmoc-4,4-difluoroproline was incorporated into the Ac-TYXN-NH$_2$ peptide model system and was examined by ^{19}F NMR (Figure 4.7). These data are consistent with the previously observed ^{19}F NMR spectra. The decoupled ^{19}F NMR spectrum appears as a *trans* peak at -101 ppm with two sets of peaks at least 2 ppm away which are representative of the pro-4*S* and the pro-4*R* fluorines (Figure 4.7a). The *trans* conformation appears as one peak, indicating that the pro-4*R* and pro-4*S* are magnetically equivalent, meaning the chemical shift of the two fluorine atoms is less than the $^2J_{FF}$ coupling constant. The pro-4*S* and the pro-4*R* *cis* peaks, however, are not magnetically equivalent and therefore appear as two separate peaks in both cases where $^2J_{FF} = 252$ Hz. The ^{19}F NMR was also measured without decoupling. Due to the proximity of the fluorine atoms to each other as well as the protons on the prolyl ring, the fluorine atoms can couple to each other leading to the more complex splitting pattern observed in Figure 4.7b. These data suggest the potential application of 4,4-difluoroproline as a probe of *cis-trans* isomerization, with a large chemical shift change for the *cis* population, and a small chemical shift change for the *trans* population.
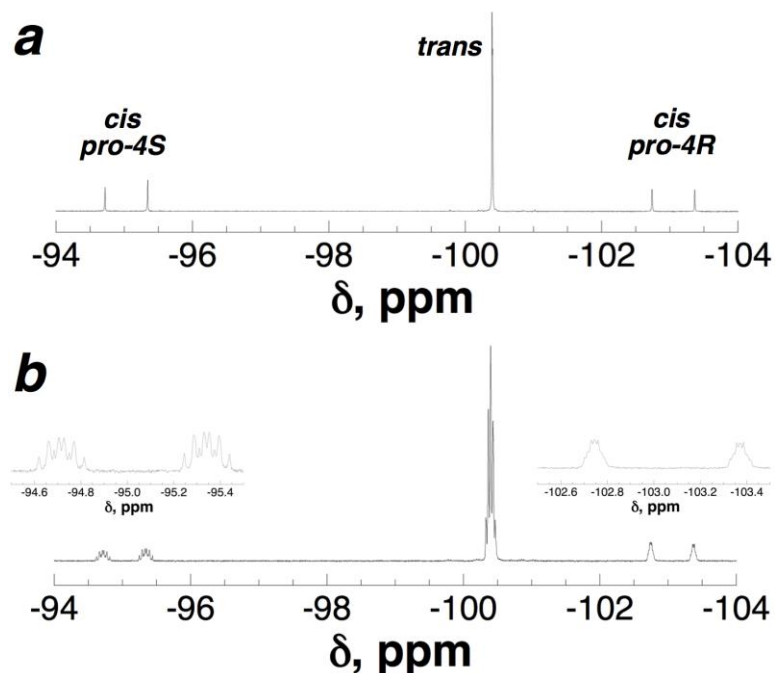
Figure 4.7 ^{19}F NMR spectra (376 MHz) of the peptide Ac-TYDfpN-NH$_2$ at 25 ˚C in D$_2$O with 5 mM phosphate pH 4 and 25 mM NaCl. (a) ^{1}H-decoupled NMR spectrum. (b) ^{1}H-coupled NMR spectrum. Peaks associated with *cis* and *trans* amide bonds are indicated. Only 1 fluorine peak is observed in the *trans* conformation due to magnetic equivalence of the fluorines ($\Delta\delta < {^2J_{FF}}$).

While the NMR characterization of the Ac-TAXN-NH$_2$ peptide was not repeated, it demonstrates the potential of the 4,4-difluoroproline to be utilized as a *cis-trans* isomerization probe. The Ac-TAXN-NH$_2$ model peptide system is utilized like the Ac-TYXN-NH$_2$ system but shows a significant reduction in the amount of *cis* amide bond formation due to the lack of an aromatic-Pro sequence. The decoupled ^{19}F NMR spectrum of the Ac-TA(4,4-Dfp)N-NH$_2$ showed a similar pattern as the Ac-TY(4,4-Dfp)N-NH$_2$ peptide, with a large chemical shift dispersion of the *cis* peaks ($\Delta\delta$

266

= 8 to 10 ppm) and $^2J_{FF}$ = 252 Hz (Figure 4.8) .[43] The *trans* peaks, however, exhibited strong second order effects, unlike the Ac-TY(4,4-Dfp)N-NH$_2$ peptide, due to similar chemical shifts ($\Delta\delta \cong {}^2J_{FF}$).
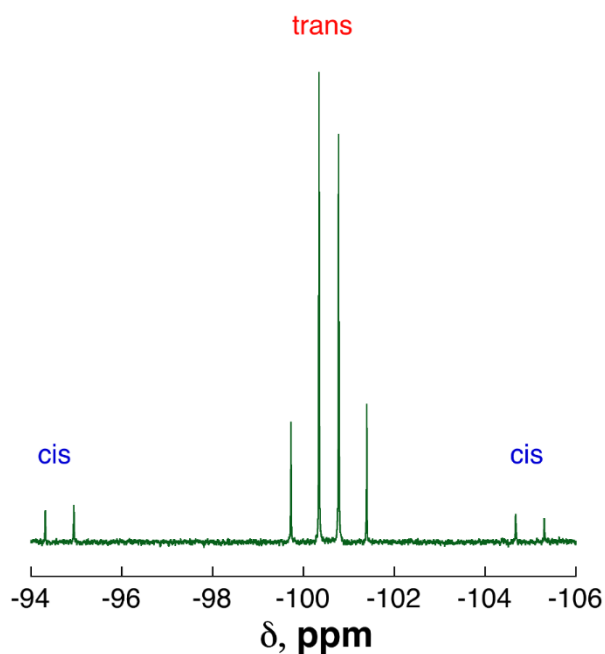
Figure 4.8   Decoupled ^{19}F NMR (376 MHz) spectrum of the Ac-TA(4,4-Dfp)N-NH$_2$ model peptide at 25 °C in 10% D$_2$O/90% H$_2$O with 5 mM phosphate buffer (pH 4) and 25 mM NaCl.[43]

4.2.2   4,4-Difluoroproline Promotes PPII Helices Similarly to Proline

Fmoc-4,4-difluoroproline was incorporated into the Ac-GPPXPPGY-NH$_2$ model system to examine the impact of 4,4-difluoroproline on PPII structure. As described in Chapter 1, the Ac-GPPXPPGY-NH$_2$ model system is used to determine PPII propensity. Fmoc-4,4-difluoroproline was introduced into the peptide at the X position using standard amide bonding coupling (2 equiv amino acid with HATU in

8% DIPEA in DMF). The peptide was purified via reverse phase HPLC before CD

experiments were performed. The peptide was subjected to CD under the previously

described conditions (5 mM phosphate buffer, pH 7, with 25 mM KF at 25 °C) (Figure

4.9). 4,4-Difluoroproline exhibits a strong PPII propensity ($[\theta]_{229} = 2030$

deg●cm^2●dmol^{-1}), but somewhat less than Pro ($[\theta]_{228} = 2950$ deg●cm^2●dmol^{-1}). These

data might be indicative of a modestly greater preference for the extended

conformation, which was observed in the ψ torsion angle in the x-ray crystal

structure.[56]
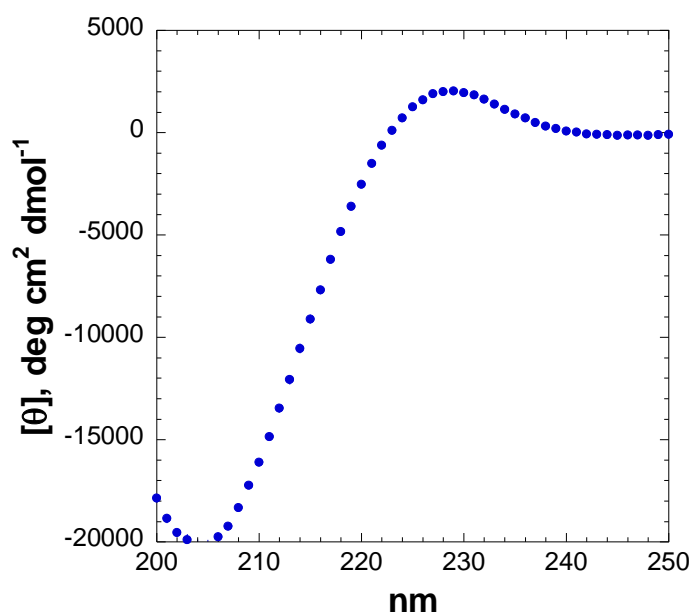
Figure 4.9   CD spectrum of Ac-GPP(Dfp)PPGY-NH$_2$ in 5 mM phosphate buffer (pH
7) with 25 mM KF at 25 °C. PPII is indicated by the magnitude of the
positive band at 228 nm. Data is representative of three independent trials
and error bars indicate standard error.

Despite only having five consecutive proline residues, and thus unlikely to adopt a PPI helix, the Ac-GPP(Dfp)PPGY-NH$_2$ was also examined by CD in acetonitrile (Figure 4.9). The purpose of this experiment was to determine if the organic solvent inherently disfavors PPII structure and to verify the modification to 4,4-difluoroproline does not promote PPI structure in short proline oligomers. When subjected to CD in deuterated acetonitrile, a modest blue shift was observed (λ_{max} = 227 compared to λ_{max} = 229 in buffer). While there is no comparison of this peptide in organic solvent, we expect a possible decrease in PPII structure, but little to no PPI structure to be visible in this peptide. In fact, we do observe a reduction in PPII structure ($[\theta]_{227}$ = 1522 deg•cm^2•dmol^{-1}) indicating a potential increase in the amount of *cis* amide bond formation at equilibrium. However, these data broadly indicate that PPII is not inherently greatly destabilized by organic solvent.
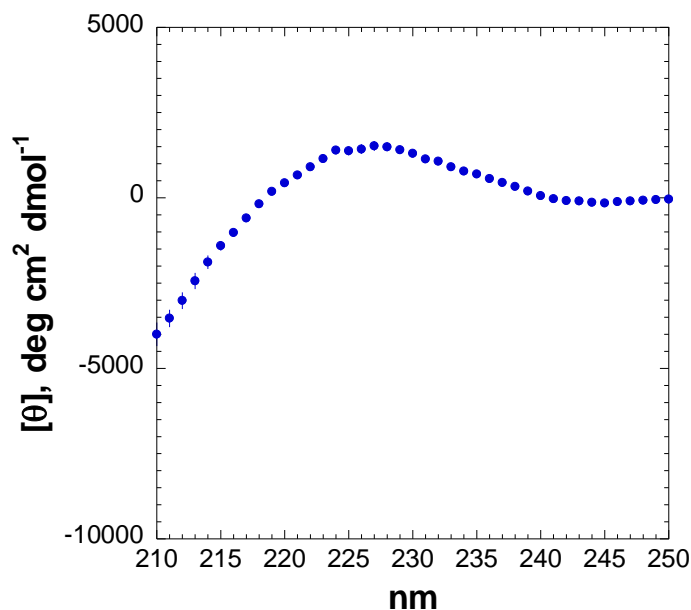
Figure 4.10 CD spectrum of Ac-GPP(Dfp)PPGY-NH$_2$ in deuterated acetonitrile at 25
°C. PPII is indicated by the magnitude of the positive band at 228 nm.
Data is representative of three independent trials and error bars indicate
standard error.

As was demonstrated in the peptide Ac-TY(Dfp)N-NH$_2$, the potential increase

in population of species with *cis* amide bonds should be visible by ^{19}F NMR in the

peptide Ac-GPP(Dfp)PPGY-NH$_2$. As seen in Figure 4.10, the ^{19}F NMR spectrum of

the peptide Ac-TY(Dfp)N-NH$_2$ in 5 mM phosphate buffer (pH 4) with 25 mM NaCl

showed distinct *cis* peaks at -94 and -104 ppm and a *trans* peak at -100 ppm. A similar

pattern was expected in the Ac-GPP(Dfp)PPGY-NH$_2$ peptide ^{19}F NMR spectra. The

peptide Ac-GPP(Dfp)PPGY-NH$_2$ was subjected to ^{19}F NMR in both buffered water (5

mM phosphate, pH 4 with 25 mM NaCl) and deuterated acetonitrile. The ^{19}F NMR

spectrum, shown in Figure 4.11, in buffered water shows a similar pattern to the Ac-TA(Dfp)N-NH$_2$ peptide, with four visible peaks as a result of second order effects.[254] Minor species due to minor conformational heterogeneity are visible by ^{19}F NMR in buffer, which is expected as the peptide promotes PPII structure which contains all *trans* amide bonds, but some conformation heterogeneity is observed. In the same system, Brown and Zondlo observed as much as 5 to 10% *cis* amide bond formation in all cases.[56] In contrast, the ^{19}F NMR spectrum, shown in Figure 4.12, in deuterated acetonitrile shows a similar pattern, indicating second order effects for the *trans* conformation as well as the previously observed *cis* peaks at -96 and -101 ppm. The *trans* conformation displays a larger chemical shift dispersion as a result of the change in solvent. Fluorine is well documented, as demonstrated in previous chapters, to display chemical shift changes as a result of a change in the chemical environment, including solvent. It is not unexpected that a change in chemical shift is observed with the change from buffered water to acetonitrile. There are a significant number of other species clearly visible in the ^{19}F NMR spectrum in Figure 4.12. This peptide contains 5 proline residues, each of which is capable of *cis-trans* isomerization leading to 32 different combinations of *cis* and *trans* amide bonds at each X-Pro amide. These smaller species visible are most likely a result of combinations *cis-trans* isomerization events at all proline sites within the peptide. These data are consistent with the observed modest decrease in PPII structure observed by CD in acetonitrile. Together, the CD and ^{19}F NMR data demonstrate that 4,4-difluoroproline promotes PPII structure, similar to but somewhat less than unmodified proline, but is capable of detection of *cis-trans* isomerization in both water and organic solvent.
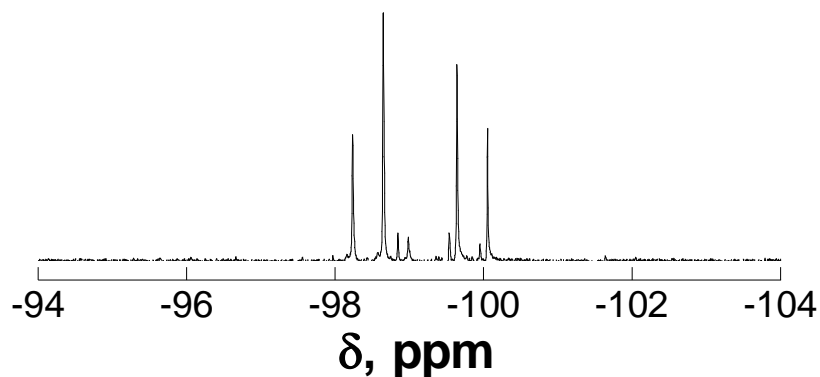
Figure 4.11 ^{19}F NMR spectrum of Ac-GPP(Dfp)PPGY-NH$_2$ in 5 mM phosphate buffer (pH 4) with 25 mM NaCl and 10% D$_2$O.
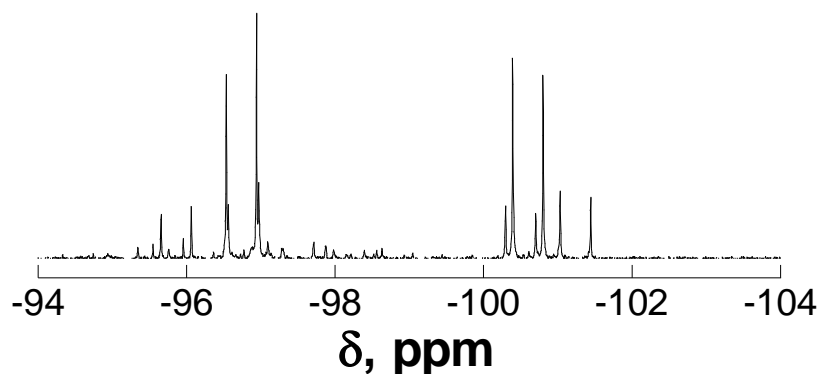
Figure 4.12 ^{19}F NMR spectrum of Ac-GPP(Dfp)PPGY-NH$_2$ in deuterated acetonitrile.

4.2.3   Design of Peptides Containing 4,4-Difluoroproline to Interrogate Polyproline Helix Structure and Interconversion

In order to probe polyproline helix structure and the interconversion between PPI and PPII, we designed a series of peptides to probe the local structure within proline oligomers. As previously described in Section 4.1.2, at least six consecutive proline residues are required to form a stable PPI helix in solution.[248,275,276] Conversely, long proline oligomers can be difficult to synthesize and purify due to

272

inefficient coupling and *cis-trans* isomerization leading to multiple peaks by reverse phase HPLC. For these reasons, we chose to synthesize and examine a 9 residue proline oligomer as a model system to explore PPI, PPII, and polyproline helix interconversion (Figure 4.13). A control peptide containing nine proline residues was synthesized. All peptides were synthesized with an N-terminal acetylation and a C-terminal amide. A series of peptides containing 4,4-difluoroproline was synthesized. 4,4-Difluoroproline was introduced at the 2, 5, or 8 positions within the 9 residue proline oligomer. The purpose of these modifications was to examine the local structure at the beginning, middle, and end of the proline oligomer. For the examination of local structure at the beginning and end of the peptide, we did not examine the first and last positions, to examine 4,4-difluoroproline always in a Pro-Dfp-Pro context, minimizing terminus effects on the ^{19}F NMR. Furthermore, as was already demonstrated, the energy barrier for *cis-trans* isomerization is slightly reduced from that of proline due to the inductive effects of the fluorine modification. Increased *cis-trans* isomerization at the termini of the proline oligomer may potentially influence the kinetics of interconversion of the polyproline helices. All of the peptides containing 4,4-difluoroproline were synthesized using standard solid-phase peptide synthesis and purified by reverse phase HPLC. Peptide identities were verified by mass spectrometry.

Pro₉   Ac-PPPPPPPPP-NH₂
Pro₉-P2Dfp   Ac-PXPPPPPPP-NH₂
Pro₉-P5Dfp   Ac-PPPPXPPPP-NH₂
Pro₉-P8Dfp   Ac-PPPPPPPXP-NH₂

X =

Dfp
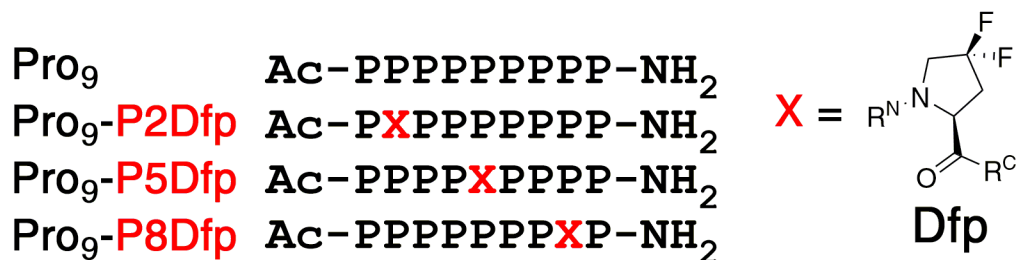
Figure 4.13 Polyproline helix model peptides containing 4,4-difluoroproline (Dfp) at the 2, 5, and 8 position within the 9 residue proline oligomer. All peptides were acetylated on the N-terminus and contained a C-terminal amine.

4.2.4 Polyproline II to Polyproline I Interconversion

The most common method for observing the interconversion from PPII to PPI is following the interconversion by CD. PPII helices have a characteristic CD signature indicated by a maximum at 228 nm and PPI helices have a characteristic CD signature indicated by a maximum at 215 nm.[248,275,282] For this reason, we chose to examine each peptide by CD to determine the ideal solvent, which would be both CD and NMR compatible, and the time and temperature required. Furthermore, CD experiments allow us to obtain global information as to how the 4,4-difluoroproline impacts structure. Previous experiments show PPI effectively forms in methanol, 1-propanol, 1-butanl, 1,4-dixane, chloroform and acetonitrile[248,275,276]. While this list of solvents is not exhaustive, they are well documented to form a PPI helix. Due to the nature of experiments, utilizing both CD and NMR to examine the interconversion of PPI and PPII helices, deuterated acetonitrile was chosen. Deuterated acetonitrile is compatible with both CD and NMR while promoting PPI helix formation. Furthermore, acetonitrile is miscible with water, allowing mixtures of solvents to be used to examine the kinetics of interconversion. All peptides were equilibrated to PPII

in water for at least 12 hours before the water was removed. The dried peptide was redissolved in deuterated acetonitrile (CD_3CN) and examined by CD. Data were recorded at 5 mins, 15 mins, 30 mins, 1 hour, 2, hours, 4 hours, and 6 hours, and the sample was held at 25 °C for the entire experiment.

Figure 4.14 Circular dichroism spectra indicating the transition from PPII to PPI (a) in CD$_3$CN for Pro$_9$ (b), Pro$_9$-P2Dfp (c), Pro$_9$-P5Dfp (d), and Pro$_9$-P8Dfp (e). All peptides were incubated in CD$_3$CN for six hours. Data were recorded at 5 mins (purple), 15 mins (blue), 30 mins (cyan), 1 hour (green), 2 hours (yellow), 4 hours (orange), and 6 hours (red) at 25 °C.

The control peptide (Pro$_9$) was fully equilibrated to PPI in 6 hours (Figure 4.14b). The Pro$_9$ experiment was used as a basis to examine the effect of modifications in peptides containing 4,4-difluoroproline. Within the first 5 mins, there was already a significant reduction in the amount of PPII helix formation. The interconversion proceeds as is visible with the decrease in the magnitude of the mean residue ellipticity at 228 nm and the increase in the magnitude of the mean residue ellipticity at 215 nm.

The three peptides containing 4,4-difluoroproline at the 2, 5, or 8 position were also examined in the same manner as described above. Peptides were incubated in water to equilibrate them to PPII before the water was removed. The peptides were then re-dissolved in deuterated acetonitrile and monitored by CD over 6 hours to observe the interconversion from PPII to PPI. In all three cases, the peptides reached equilibrium in 6 hours, as with the control Pro$_9$ peptide. The similar rate of interconversion indicates that 4,4-difluoroproline does not have a dramatic effect on the formation of PPI structure, demonstrating that it can be used in place of proline in proline oligomers without dramatic structural effects. The rate of interconversion is similar to that of the control Pro$_9$ peptide. However, there is an important trend within the data. Pro$_9$-P8Dfp equilibrates to PPI more rapidly than the other two peptides containing 4,4-difluoroproline (Figure 4.14e). Pro$_9$-P2Dfp equilibrates to PPI the slowest of the three peptides containing 4,4-difluoroproline (Figure 4.14c).

In order to quantify the rate at which the peptides converted from PPII to PPI, the ratio of $[\Theta]_{215}/[\Theta]_{223}$, where $[\Theta]_{215}$ is the maximum indicative of PPI structure and $[\Theta]_{223}$ is the isodichroic point, by circular dichroism, was plotted versus time (Figure 4.15). This ratio was chosen because it measures the appearance of PPI structure,

independent of concentration. It is important to note that all four peptides are starting from approximately the same amount of PPI structure. Since the first time point was not measured until five minutes, the starting point for these data are not necessarily all PPII structure; however, all of the peptides are starting from approximately the same percentage of PPI helical structure.   All four peptides were plotted and the data were fit to a linear regression, where the slope of the line is indicative of the rate of appearance of PPI structure over time assuming a pseudo-first order rate law. Comparison of the slope of each fit reveals the trend of interconversion from fastest to slowest as follows: Pro_9-P2Dfp (slope = 2.90 s^{-1}), Pro_9 (slope = 2.69 s^{-1} ± 0.98), Pro_9-P5Dfp (slope = 2.31 s^{-1}), Pro_9-P8Dfp (slope = 1.92 s^{-1}). These data are indicative of the directionality of the formation of PPI helix. As previously mentioned, 4,4-difluoroproline has a reduced free energy barrier compared to proline. If all four peptides, or all three of the peptides containing 4,4-difluoroproline, equilibrated at the same rate, these data would not indicate directionality of the propagation of the helix. However, if the Pro_9-P2Dfp equilibrates faster than the Pro_9-P5Dfp and Pro_9-P8Dfp, it would indicate the helix is forming at the N-terminus and propagating towards the C terminus. In fact, the trend observed is Pro_9-P2Dfp equilibrates the fastest followed by Pro_9-P5Dfp and Pro_9-P8Dfp, indicating the polyproline I helix is forming from the N terminus and propagating towards the C terminus. These data are consistent with the observations made by Lin and Brandts.[283]
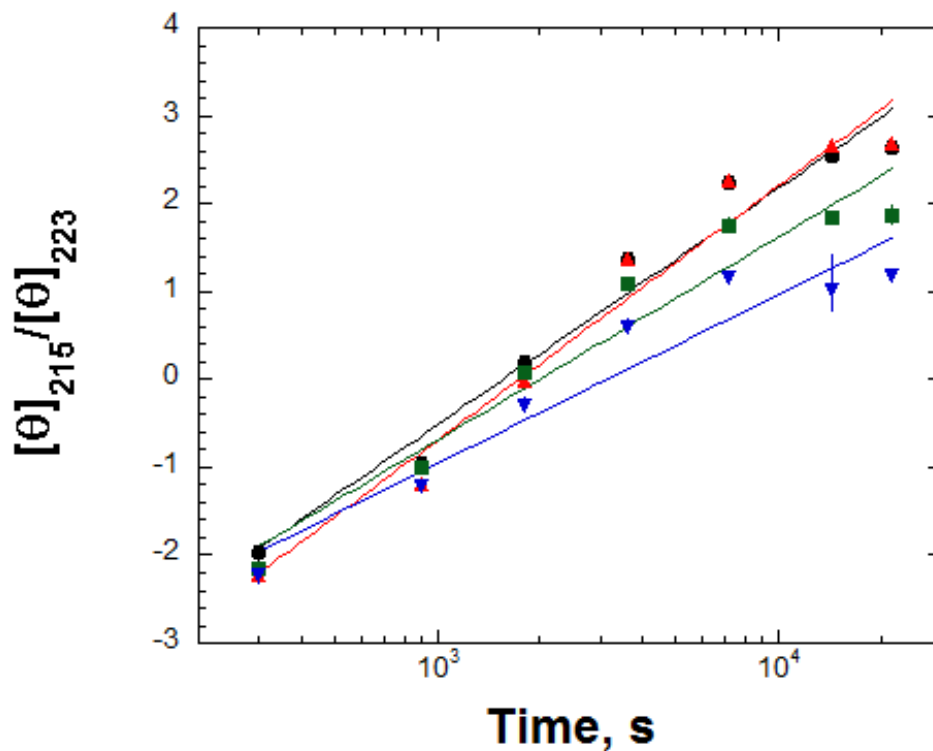
Figure 4.15 Analysis of the appearance of PPI over time, at $[\Theta]_{215}/[\Theta]_{223}$, where $[\Theta]_{215}$ is the maximum indicative of PPI structure and $[\Theta]_{223}$ is the isodichroic point, by circular dichroism for Pro_9 (black circles) Pro-P2Dfp (red triangles), Pro-P5Dfp (green squares), and Pro-P8Dfp (blue triangles). All curve fits were completed using KaleidaGraph using a logarithmic regression $(y = mx + b)$.

The CD data report on overall global structure; however, the introduction of the 4,4-difluoroproline allows for the observation of site-specific structural information, via ^{19}F NMR. Due to the proximity of the two fluorines and the neighboring protons on the prolyl ring, all NMR spectrum were measured with proton

decoupling. Peptides were again fully equilibrated to PPII structure in water before removing the water and re-dissolving in deuterated acetonitrile for NMR experiments. Peptides were dissolved in deuterated acetonitrile and data were collected at 5 mins, 15 mins, 30 mins, 1 hour, 2 hours, 4 hours, and 6 hours at 25 °C. NMR experiments. All NMR experiments were performed with a 10 ppm sweep width, 128 scans, and an acquisition time of 0.8 seconds. The NMR parameters were optimized to minimize the amount of time required to acquire the NMR spectrum for each experiment.

Polyproline II

Ac-PXPPPPPPP-NH₂ **Ac-PPPPXPPPP-NH₂** **Ac-PPPPPPPXP-NH₂**

Polyproline I
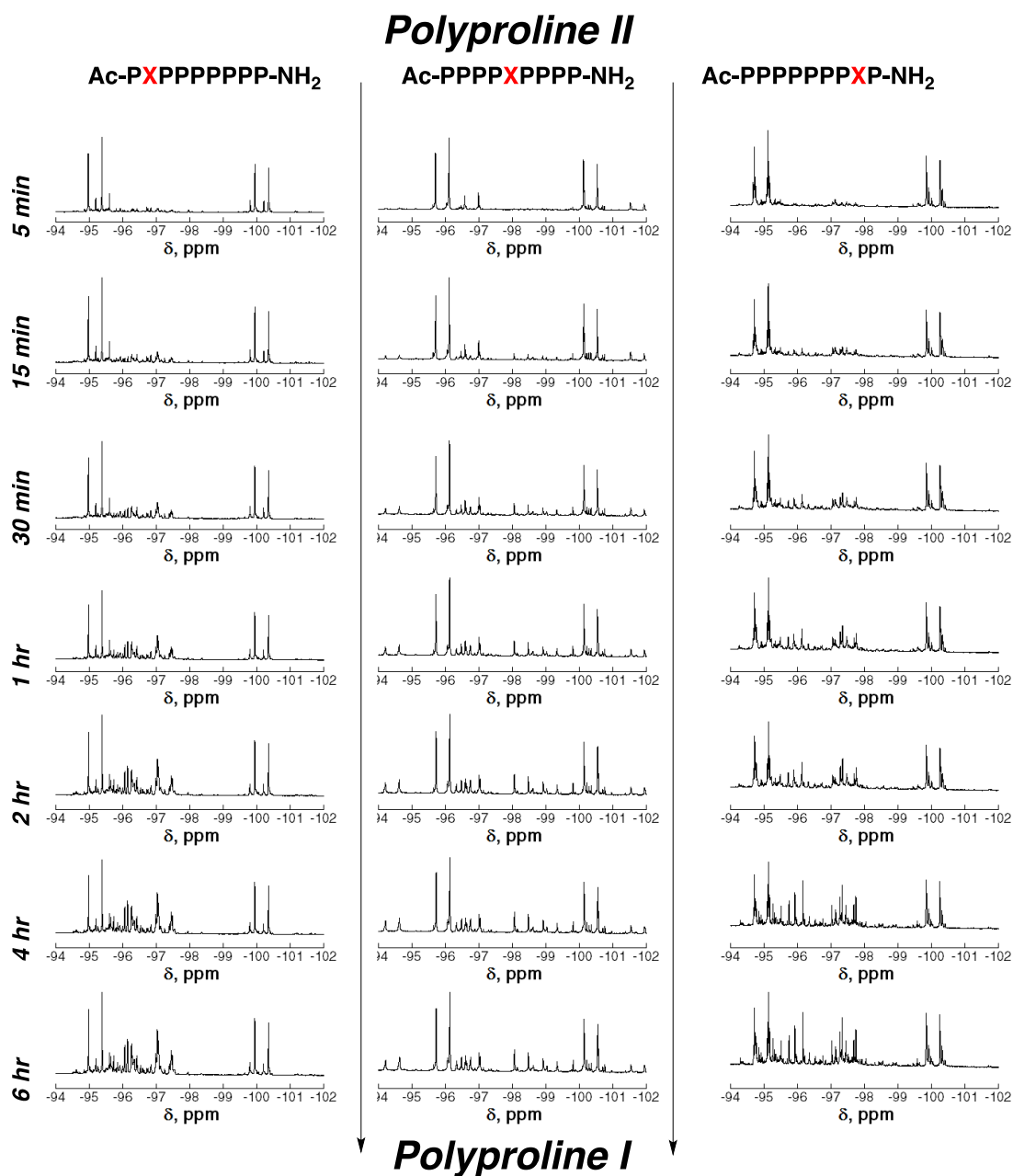
Figure 4.16 ^{19}F NMR spectra indicating the transition from PPII to PPI in CD$_3$CN for Pro$_9$-P2Dfp (left), Pro$_9$-P5Dfp (middle), and Pro$_9$-P8Dfp (right). All peptides were incubated in CD$_3$CN for six hours with data recorded at 5 mins, 15 mins, 30 mins, 1 hour, 2 hours, 4 hours, and 6 hours at 25 °C.

To gain site-specific information, all three peptides containing 4,4-difluoroproline were subjected to ^{19}F NMR to examine the interconversion from PPII to PPI in deuterated acetonitrile. These data demonstrate a different pattern by ^{19}F NMR than was previously observed in water. Notably, the *trans* amide peaks have a large chemical shift change ($\Delta\delta$ = 4 to 5 ppm), while the *cis* amide peaks have a smaller chemical shift change ($\Delta\delta$ < 5 ppm) in deuterated acetonitrile (Figure 4.16). This pattern was seen for all peptides containing 4,4-difluoroproline in all of the examined organic solvents (CH_2Cl_2 Figure 4.16).

The interconversion of Pro_9-P2Dfp was followed over the course of 6 hours by ^{19}F NMR (Figure 4.16, right column). At 5 mins, a small amount of *cis* conformation is visible by ^{19}F NMR, indicating that even at 5 mins, there is some heterogeneity in local structure within the proline oligomer. While there appear to be multiple species containing some quantity of *cis* amide bonds, there is one major species and one minor species. The heterogeneity of the proline oligomer is not visible by CD, because CD only reports on the overall global structure. As the interconversion proceeds over six hours, an increasing amount of *cis* amide bonds are visible. Due to the nine proline residues in the peptide, there are multiple different *cis* amide bonds that appear over the course of the six hours. Within one hour, a significant amount of *cis* amide bonds are visible by ^{19}F NMR, leading to as many as four to five different substantial minor species. Interestingly, despite the large amount of PPI visible by CD at the end of six hours, a significant amount of *trans* amide bonds are still visible by ^{19}F NMR at six hours. The amount of heterogeneity observed at six hours gives valuable, previously unattainable information about the local structure. Noteably, there is one major

species, corresponding with the assigned *trans* peaks in the 5 min time point. There are at least four to five significant minor species visible in the six hour time point, indicating a substantial population of *cis* amide bonds within the peptide after six hours.

The ^{19}F NMR data, when compared to the CD data over the same time-course, support the same trends observed globally. In both cases, the structure is not completely PPII at 5 minutes. There is a small amount of PPI character (*cis* amide bond) visible at the first time point. Also, both the CD spectrum and the ^{19}F NMR show little change after two hours, indicating the peptide is mostly PPI structure at two hours. Finally, the six hour time point demonstrates a significant amount of heterogeneity even when the peptide has fully equilibrated to PPI as observed by CD. These data taken together demonstrate that, while both global and local structure indicate a significant population of *cis* amide bond formation, the global structure is not necessarily indicative of local structure at all positions.

The Pro$_9$-P5Dfp interconversion from PPII to PPI was also examined by ^{19}F NMR (Figure 4.16, middle column). At five minutes, there is one major species and two to three minor species, indicating primarily PPII structure. More *cis* amide bond formation is visible by 30 mins, with greater than five different minor species present, relative to Pro$_9$-P2Dfp, which is consistent with the CD results. Again, a significant amount of *trans* amide bond is visible even at six hours, despite the full equilibration to PPI observed by CD, however there are at least five minor species present, indicating a substantial quantity of *cis* amide bond formation. The other notable difference between the Pro$_9$-P5Dfp and the Pro$_9$-P2Dfp is the different in the peak pattern for the *cis* amide peaks. This change is, at least in part, due to the different

position of the 4,4-difluoroproline. We speculate that the chemical environment in the center of the helix is significantly different than that of the termini. *Cis-trans* isomerization is possible at each amide bond, leading to 512 combinations of *cis* and *trans* amides, within the peptide. At the center of the helix (Pro$_9$P5Dfp), there are four amide bonds in either direction, each of which will have populations of *cis* and *trans* peaks, while at the termini (Pro$_9$P2Dfp), only two amide bonds are on one side of the 4,4-difluoroproline. The 4,4-difluoroproline, based on distance, will not sense all conformations at all positions equally.

Finally, Pro$_9$-P8Dfp interconversion from PPII to PPI was examined by ^{19}F NMR. As with Pro$_9$-P2Dfp and Pro$_9$-P5Dfp, a small percentage of *cis* amide bond is visible even at 5 mins with one major species and two to three minor species visible. More species containing *cis* amide bonds appear to be visible at 30 mins than in Pro$_9$-P2Dfp or Pro$_9$-P5Dfp. These data are consistent with the CD indicating that the PPI helix propagates from the C-terminus to the N-terminus. The large population of *cis* amide bonds belonging to multiple different species visible at six hours is indicative of "fraying" at the C-terminus of the peptide. These data are further indicative of the PPI helix propagating from the C-terminus to the N-terminus. If the PPI helix were forming in the opposite direction, we would expect to see a lower *trans* amide bond population in the fluorine NMR. These data also confirm the potential of 4,4-difluoroproline can be used to monitor local structure in proline oligomers.

Finally, all three peptides containing 4,4-difluoroproline were examined by ^{19}F NMR in CDCl$_3$ (Figure 4.17). PPI helices, as previously mentioned, are more favored in non-polar organic solvents. While CDCl$_3$ is not compatible with CD at 210-230 nm due to high dynode voltage values due to absorbance of CDCl$_3$, the solvent is

compatible with NMR. The NMR spectra of the peptides containing 4,4-difluoroproline should fully equilibrate to PPI in CDCl$_3$, allowing us to observe the PPI endpoint. In fact, we observe a similar trend in these ^{19}F NMR data as was observed in the 6 hour time points in deuterated acetonitrile, with some notable differences. The Pro$_9$-P2Dfp peptide shows primarily a single species in CDCl$_3$, with only the *cis* conformation observed demonstrating that the local structure at the N-terminus of the peptide is entirely PPI. In contrast, a mixture of *cis* and *trans* peaks, with two major species and one to three minor species, are visible in the Pro$_9$-P5Dfp peptide in CDCl$_3$, demonstrating that the local structure is more flexible in the middle of the peptide than at the N-terminus. Finally, a higher percentage of *trans* amide bonds, with two to three major species visible, is visible in Pro$_9$-P8Dfp in CDCl$_3$, demonstrating a larger amount of interconversion at the C-terminus of the peptide. These data are consistent with the observations made in examining the ^{19}F NMR spectra of these peptides in deuterated acetonitrile, demonstrating the PPI helix is formed initially from the C-terminus and propagated to the N-terminus.
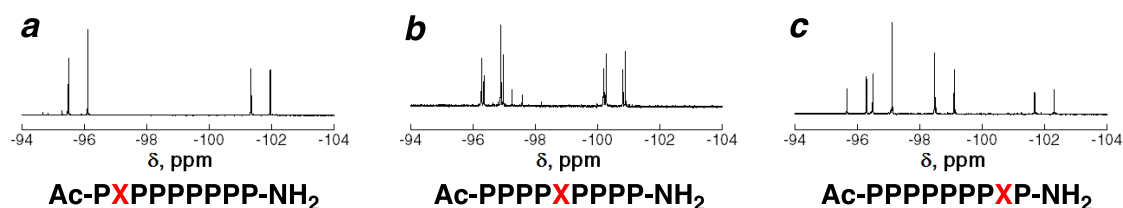
Figure 4.17 ^{19}F NMR spectra of the peptides Pro$_9$-P2Dfp (a), Pro$_9$-P5Dfp (b), and Pro$_9$-P8Dfp (c) incubated in CDCl$_3$ overnight at room temperature.

4.2.5 Polyproline I to Polyproline II Interconversion

Interconversion from PPI to PPII helices was also examined by CD to help develop a mechanism of interconversion, not only from PPII to PPI but also in the reverse direction of the "ordered to ordered" transition of proline oligomers.[251] Again, due to the extensive number of previous examples of monitoring the interconversion from PPI to PPII, we began examining the system by CD. As previously demonstrated in the Ac-GPPXPPGY-NH$_2$ peptide model system, 4,4-difluoroproline does not impede the formation of PPII helices. To verify these results in the Pro$_9$ system and also to explore the directionality of PPII helix formation the Pro$_9$, Pro$_9$-P2Dfp, Pro$_9$-P5Dfp, Pro$_9$-P6Dfp were examined by CD and ^{19}F NMR to monitor the interconversion from PPI to PPII helix.

The Pro$_9$ peptide was subjected to CD experiments to monitor the interconversion from PPI to PPII (Figure 4.18a). The Pro$_9$ peptide was equilibrated for a minimum of six hours in 50 µL deuterated acetonitrile. Prior to CD measurements, 50 µL D$_2$O and 400 µL H$_2$O were added to a final volume of 500 µL. These conditions were chosen, in contrast to the alternative approach of removal of solvent by evaporation and redilution in H$_2$O, due to an observed reduction in PPI helix formation with the removal of the deuterated acetonitrile prior to redissolving in 10% D$_2$O in water for CD measurements, presumably due to small amounts of residual water that evaporate last. The presence of 10% deuterated acetonitrile does not substantially inhibit PPII formation as reported previously. The Pro$_9$ peptide was fully equilibrated to PPII in 2 hours at 15 °C. CD spectra were recorded at the following time points: 5 mins, 15 mins, 30 mins, 45 mins, 1 hour, 1.5 hours, and 2 hours. The Pro$_9$ peptide model system shows a significant amount of PPI helix formation at 5 mins and complete equilibration to PPII helix in 2 hours (Figure 4.18b). Other

temperatures were examined, including 0.5 °C, 5 °C, and 10 °C, and 25 °C. The 15 °C

temperature was chosen because the CD instrument could maintain 15 °C and the

interconversion of the Pro_9 peptide could be measured on a reasonable time scale (i.e.

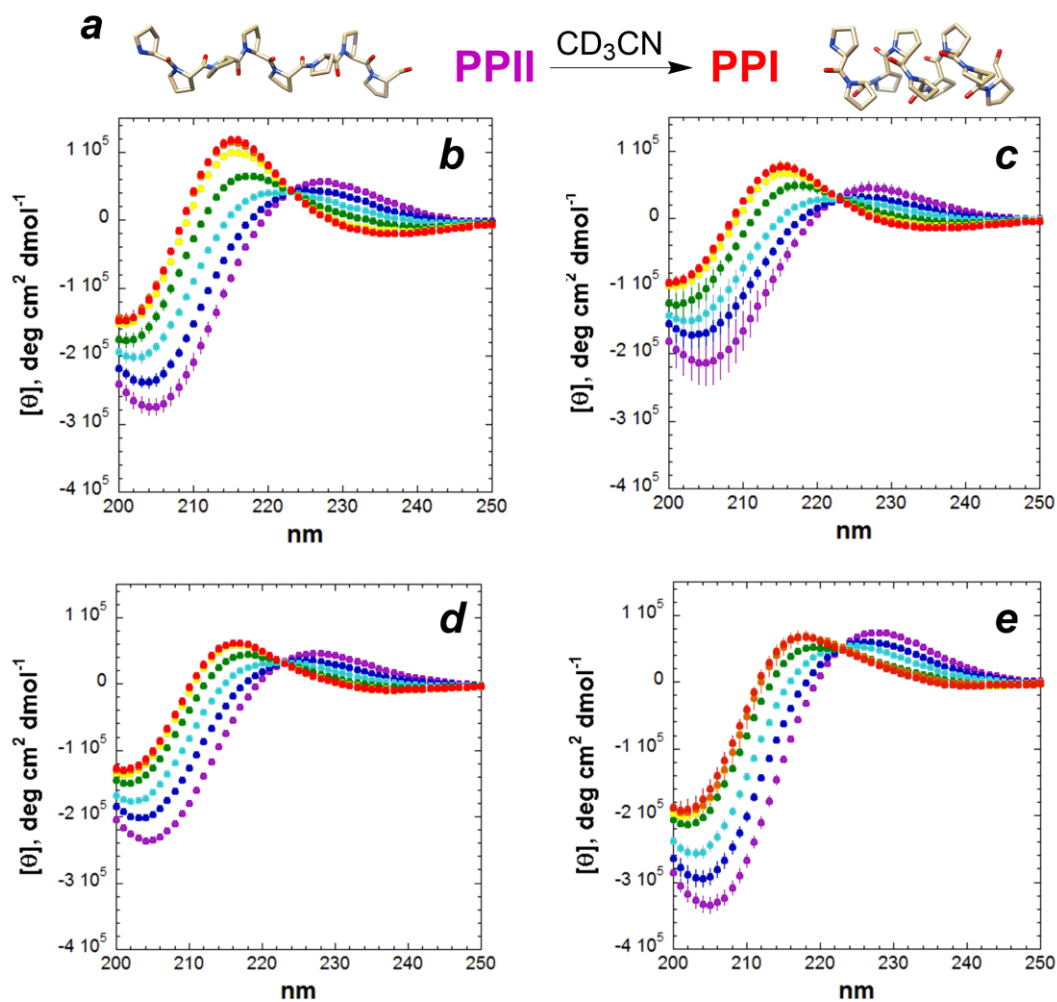hours rather than days at lower temperatures or mins in the case of 25 °C).
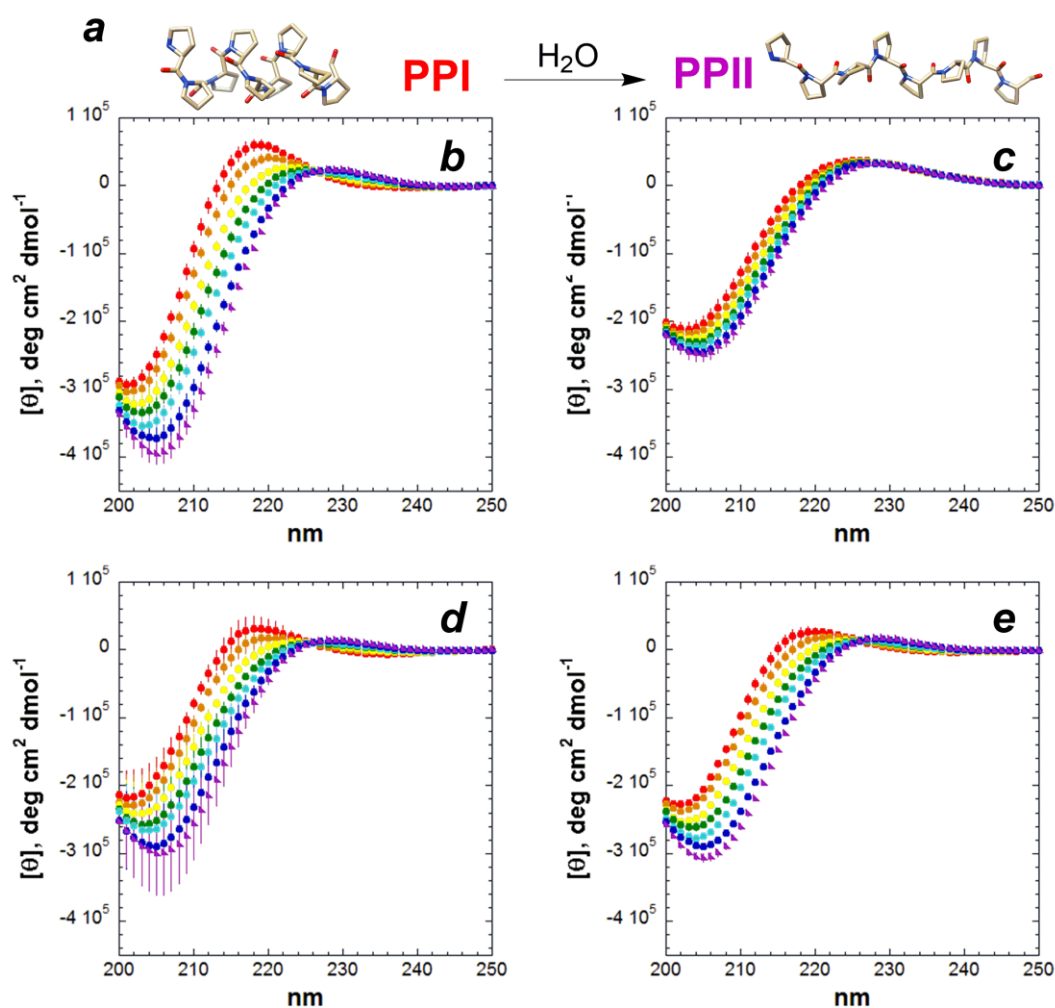
Figure 4.18 Circular dichroism spectra indicating the transition from PPII to PPI (a) in CD$_3$CN for Pro$_9$ (b), Pro$_9$-P2Dfp (c), Pro$_9$-P5Dfp (d), and Pro$_9$-P8Dfp (e). All peptides were incubated in 10% CD$_3$CN, 10% D$_2$O, and 80% H$_2$O for two hours. CD spectra were recorded at the following time points: 5 mins (red), 15 mins (orange), 30 mins (yellow), 45 mins (green), 60 mins (cyan), 90 mins (blue), 120 mins (purple) at 15 °C. Data represents three independent trials and error bars indicate standard error.

The peptides containing 4,4-difluoroproline were also examined in the same manner as the Pro_9 peptide. All three peptides were independently equilibrated in 50 µL deuterated acetonitrile at room temperature (Figure 4.18a). After a minimum of six hours, 50 µL D_2O and 400 µL H_2O were added prior to CD measurements. The peptide Pro_9-P2Dfp equilibrated to entirely PPII structure much faster than any of the other peptides examined (Figure 4.18c). Even at 5 mins, almost no PPI structure is observable globally by CD. Within 30 mins, the peptide Pro_9-P2Dfp had fully equilibrated to PPII, as observed by CD. This observation is particularly noteworthy because the Pro_9-P2Dfp had the highest PPI population of the fluorinated peptides. In contrast, Pro_9-P5Dfp demonstrated equilibration to PPII helix structure more slowly than Pro_9-P2Dfp, with a substantial amount of PPI helix visible at both 5 and 15 mins (Figure 4.18d). A full hour was required before the equilibrium shifted towards visible quantities of PPII structure visible by CD, and the Pro_9-P5Dfp peptide is fully equilibrated in 2 hours, similar to the Pro_9 peptide. Similarly, the Pro_9-P8Dfp peptide exhibits a substantial population of PPI helix structure at 5 and 15 mins. The Pro_9-P8Dfp peptide also demonstrates a similar rate of equilibration to the Pro_9 and the Pro_9-P5Dfp, which is significantly slower than the Pro_9-P2Dfp. These data indicate the PPII helix is forming in the opposite direction of PPI helices, beginning at the N-terminus and propagating to the C-terminus. The directionality of PPII helix formation is the same directionality observed in α-helices.

The CD data for each peptide was again analyzed for the rate of interconversion by plotting the ratio of $[\Theta]_{228}/[\Theta]_{226}$, where $[\Theta]_{228}$ is the maximum observed for PPII helices and $[\Theta]_{226}$ is the isodichroic point, versus time. (Figure 4.19) Unlike the time course discussed in the previous section, the interconversion from PPI

to PPII happens much fast and at five minutes, each peptide demonstrates very different amounts of PPII structure. Even at five minutes, the Pro$_9$-P2Dfp (red triangles) is almost entirely PPII. The Pro$_9$-P5Dfp (green squares) starts from the lowest amount of PPII structure of the four peptides but the rate (m = 0.53 s^{-1}) is the second fastest, leading to more PPII structure at the end of two hours. The Pro$_9$ and Pro$_9$-P8Dfp have similar both similar starting points and similar rates (Pro$_9$ m = 0.35 s^{-1}, Pro$_9$-P8Dfp m = 0.39 s^{-1}). As with the previous section, we expect the rate of interconversion, due to the decreased free energy barrier associated with 4,4-difluoroproline to give an indication of directionality. The Pro$_9$-P2Dfp is interconverting much faster than the other peptides, followed by Pro$_9$-P5Dfp, and then Pro$_9$-P8Dfp, clearly indicated the PPII helix is forming from the N terminus and propagating towards the C terminus. These data are again consistent with the observations of Lin and Brandts.[283]
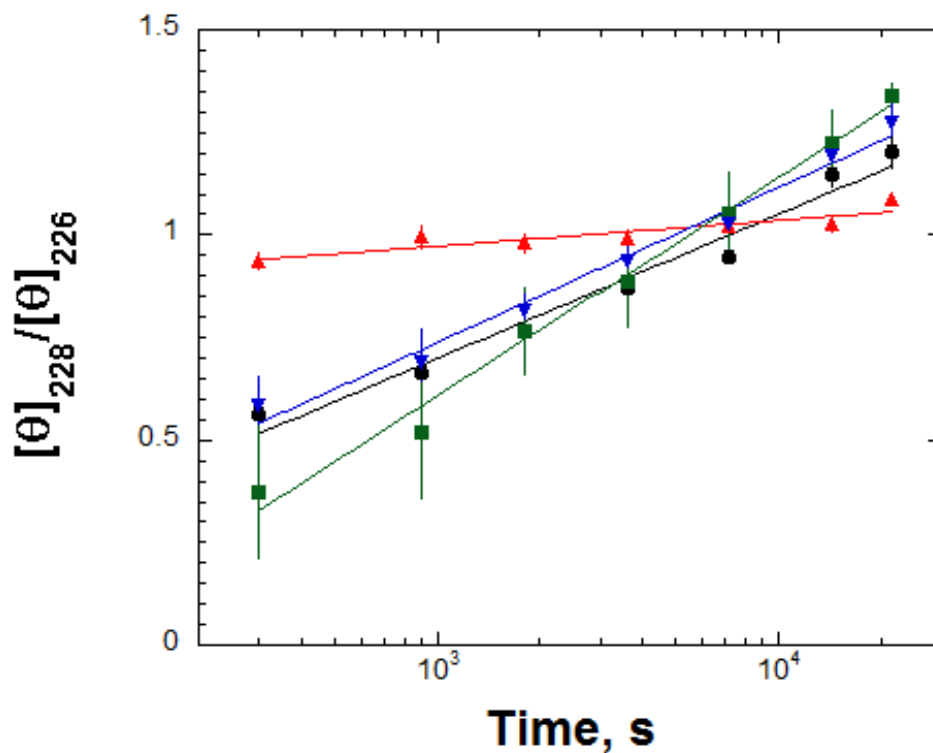
Figure 4.19 Analysis of the appearance of PPI over time, as a ratio of $[\Theta]_{228}/[\Theta]_{226}$, where $[\Theta]_{228}$ is the maximum indicative of PPII structure and $[\Theta]_{226}$ is the isodichroic point, by circular dichroism for Pro_9 (black circles) Pro-P2Dfp (red triangles), Pro-P5Dfp (green squares), and Pro-P8Dfp (blue triangles). All curve fits were completed using KaleidaGraph using a linear regression (y = mx + b).

Given the compelling CD data indicating the directionality of PPII helix formation, all the peptides containing 4,4-difluoroproline were also examined by ^{19}F NMR. The peptides were again equilibrated in 50 µL deuterated acetonitrile for a minimum of six hours prior to NMR experiments (Figure 4.20). After equilibration,

cold (approximately 0 °C) 50 μL D_2O and 400 μL H_2O were added to the peptide before the NMR experiments. The Pro$_9$-P2Dfp peptide showed a substantial amount of *cis* amide bond conformation at 5 mins by ^{19}F NMR, despite the CD data showing a substantial amount of PPII helix formation (Figure 4.20, first column). These data indicate that despite the rapid interconversion from PPI to PPII observed in the global structure, the local structure at the N-terminus of the peptide shows a substantial amount of *cis-trans* isomerization. Over the course of two hours, the majority of species containing *cis* amide bonds disappear from the ^{19}F NMR. However, even after the full two hours, a significant amount of species with *cis* amide bonds is still visible. Much like the Pro$_9$-P8Dfp interconverting from PPII to PPI, a significant amount of "fraying" is visible due to continued interconversion between *cis* and *trans* is visible at the terminus of the peptide. Alternatively, the Pro$_9$-P5Dfp peptide showed significantly fewer species, 1 major and 3 minor species, containing *cis* amide bonds at 5 mins relative to the Pro$_9$-P5Dfp peptide at the same timepoint (Figure 4.20 middle column). The data demonstrate the 4,4-difluoroproline at the 5 position within the peptide locally experiences fewer species containing *cis* amide bonds than the N-termini. Nearly no species containing *cis* amide bonds are visible by ^{19}F NMR by 45 mins in the Pro$_9$-P5Dfp peptide under these conditions which is consistent with the CD data. Similarly, the Pro$_9$-P8Dfp ^{19}F NMR spectrum shows a significant amount of *trans* amide bond conformation even at 5 mins (Figure 4.20 left column). Much like the Pro$_9$-P5Dfp peptide, nearly all *cis* peaks have disappeared from the ^{19}F NMR spectrum by 45 mins, showing one major species and only trace minor species. These NMR data demonstrate the local structure at the termini of the peptide as well as the center of the peptide. Given the amount of heterogeneity visible at the N-terminus of

292

the peptide and the minimal amount of heterogeneity visible at the center and C-terminus of the peptide, these data are consistent with the CD data. These data, taken together with the CD data, demonstrate the PPII helix forms from N to C terminus of the peptide, opposite of the PPI helix.
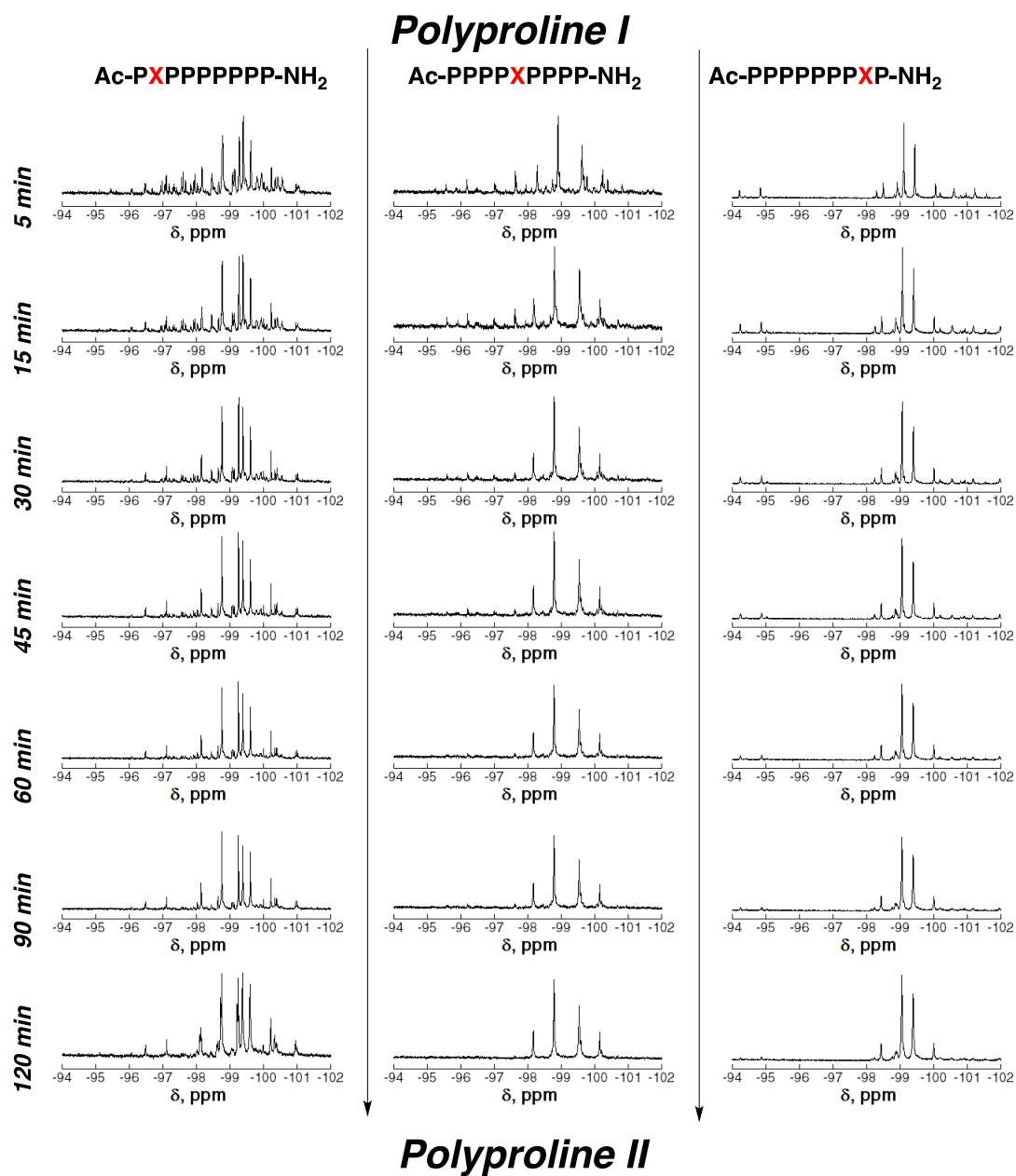
Figure 4.20 ^{19}F NMR spectra indicating the transition from PPII to PPI in CD_3CN for Pro_9-P2Dfp (left), Pro_9-P5Dfp (middle), and Pro_9-P8Dfp (right). All peptides were incubated in CD_3CN for six hours with data recorded at 5 mins, 15 mins, 30 mins, 1 hour, 2 hours, 4 hours, and 6 hours at 25 °C.

The CD and ^{19}F data are consistent in indicating the PPII helix forms from the N-terminus to the C-terminus. Both the rate of interconversion, as a result of the reduced free energy of interconversion, as measured by CD and the amount of heterogeneity observed by ^{19}F NMR, indicate the Pro$_9$-P2Dfp peptide forms a PPII helix much faster than any of the other peptides. Proline residues can be stabilized, in the *trans* conformation via n to π^* interactions, where the oxygen carbonyl can interact with the following carbonyl, stabilizing the PPII helix and indicating potential cooperativity in the formation of the PPII helix (Figure 4.21).
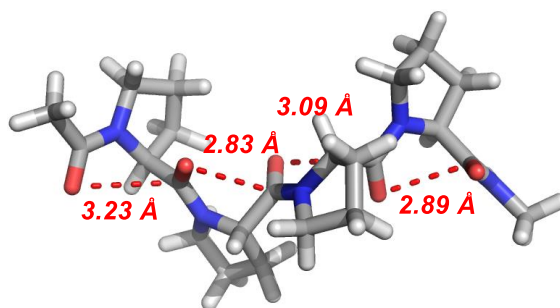
Figure 4.21 Crystal structure of a PPII helix (PDB: 4QB3), where potential n to π^* interactions are shown by the red dotted lines.

4.3   Summary and Discussion

In this work, we have demonstrated that 4,4-difluoroproline can be used to detect *cis-trans* isomerization within a peptide context by examining polyproline helix interconversion in Pro$_9$ oligomer peptides. Fmoc-4,4-difluoroproline was synthesized from Boc-4,4-difluoroproline and crystalized. 4,4-Difluoroproline was also examined in a series of peptide model systems including Ac-TYXN-NH$_2$ for re-analysis of the

^{19}F NMR spectra and to determine absolute stereochemistry of the fluorine atoms. 4,4-Difluoroproline was also examined in the PPII model system Ac-GPPXPPGY-NH$_2$ in buffer (5 mM phosphate buffer, pH 7, with 25 mM KF at 25 °C) demonstrating slightly less PPII helix propensity than Pro, most likely due to the enhanced steric effects of fluorine versus hydrogen, promoting a somewhat more extended conformation. The peptide Ac-GPP(Dfp)PPGY-NH$_2$ was also examined in deuterated acetonitrile, which showed further reduction in PPII structure relative to buffer; however, no PPI helix structure was observed by CD, indicating that organic solvents do not inherently disfavor PPII helix structure. These data are also consistent with previous observations that at least six proline residues are required for PPI helix formation in peptides. Furthermore, the peptide Ac-GPP(Dfp)PPGY-NH$_2$ was examined by ^{19}F NMR in both buffer and deuterated acetonitrile. The major species observed in both ^{19}F NMR spectra represented the *trans* conformation, which is expected based on the observed CD spectra, although, a higher percentage of *cis* conformation was observed by ^{19}F NMR in deuterated acetonitrile. These NMR data are consistent with the reduced PPII helix observed by CD in deuterated acetonitrile.

4,4-Difluoroproline was then introduced into proline oligomers, containing nine proline residues, at the 2, 5, or 8 position in order to examine the interconversion from PPII helix to PPI helix. Peptides were equilibrated in water to adopt PPII prior to re-dissolving in deuterated acetonitrile and monitoring the interconversion from PPII to PPI over the course of six hours at room temperature by CD and ^{19}F NMR. The Pro$_9$-P8Dfp peptide equilibrated to PPI more rapidly than the Pro$_9$-P5Dfp or the Pro$_9$-P2Dfp based on CD analysis. The Pro$_9$-P2Dfp equilibrated the slowest of the three peptides. ^{19}F NMR analysis of the three peptides containing 4,4-difluoroproline

interconverting from PPII to PPI in deuterated acetonitrile demonstrated a significant amount of heterogeneity in the Pro$_9$-P8Dfp peptide when compared to Pro$_9$-P5Dfp and Pro$_9$-P2Dfp. Both the CD and ^{19}F NMR experiments demonstrate that the PPI helix forms from the C-terminus of the peptide and propagates to the N-terminus, in the opposite direction of α-helix and PPII helix. The increased heterogeneity is observed by ^{19}F NMR due to the decreased energy barrier associated with 4,4-difluoroproline, allowing for isomerization at the terminus of the peptide. In the cases of the Pro$_9$-P5Dfp and Pro$_9$-P2Dfp peptides, the other proline residues subsequent to the 2 and 5 position are initiating the PPI helix, preventing heterogeneity at later residues. All three peptides were also examined in CDCl$_3$ which should induce more PPI helix than acetonitrile, as PPI helices are most favored in nonpolar organic solvents. The same trend of increased heterogeneity was observed, indicating that the PPI helix is being generated from the C-terminus. Based on both crystallographic data (Figure 4.3b) and experimental data, we believe PPI helices are propagated from the C-terminus to the N-terminus through a series of reverse n→π* interactions where the carbonyls interact with the preceding amide nitrogen. The reverse n→π* interaction is currently being explored via computational experiments.

The Pro$_9$-P2Dfp, Pro$_9$-P5Dfp, and Pro$_9$-P8Dfp peptides were also examined to explore the interconversion from PPI to PPII over the course of two hours at 15 °C. Peptides were equilibrated in deuterated acetonitrile and then diluted with 10% D$_2$O and 80% H$_2$O and monitored by both CD and ^{19}F NMR. The Pro$_9$-P2Dfp equilibrated to PPII faster than Pro$_9$-P5Dfp and Pro$_9$-P8Dfp. The Pro$_9$-P8Dfp equilibrated to PPII the slowest of the three peptides containing 4,4-difluoroproline, indicating the PPII helix forms from the N-terminus and propagates to the C-terminus which is the same

directionality as an α-helix and the opposite of a PPI helix. The peptides were also minored by ^{19}F NMR to explore the local structure at the 2, 5, and 8 positions in the peptide while equilibrating from PPI to PPII. The Pro$_9$-P2Dfp peptide showed the most heterogeneity of the three peptides, indicating more *cis-trans* isomerization at the 2 position than the 5 and 8 position. After two hours, almost no heterogeneity is observed in either the Pro$_9$-P5Dfp or Pro$_9$-P8Dfp peptides. These data are consistent with the observations by CD indicating the PPII helix forming at the N-terminus and propagating to the C-terminus. We believe the directionality of the formation of a PPII helix is a result of the n→π* interactions where the nitrogen amide interacts with the carbonyl of the following residue. The n→π* interactions are known in both PPII and α-helices so it is unsurprising that this interaction would be a major driving force for the formation of PPII helices.

The experiments presented here, utilizing 4,4-difluoroproline to monitor local structure in proline oligomers, demonstrates the potential of this amino acid to act as a probe for *cis-trans* isomerization in peptides and proteins. There are a number of different future directions which can be explored utilizing 4,4-difluoroproline. There are a number of experiments which could potentially add to our understanding of PPI and PPII helices. Not currently addressed in this body of work is the length of the proline oligomer required to form PPI helices. Based on the Ac-GPP(Dfp)PPGY-NH$_2$ data as well as data from other labs, we know that at least six residues are required to form a stable PPI helix, despite the PPI helix-like structure observed in a few PDB structures consisting of only two residues.[248,275] The length requirement could potentially be addressed by introducing 4,4-difluoroproline into proline oligomers of increasing length and monitoring how local structure is impacted within increasing

lengths as well as how multiple positions within longer helices Furthermore, we did not explore the termini, i.e. the first and last position within the peptide, due to potential termini effects on ^{19}F NMR. Having developed a potential mechanism for PPI stabilization and polyproline helix, we could begin to address any potential termini effects by introducing 4,4-difluoroproline at the first and last position within a proline oligomer.

We also expect that 4,4-difluoroproline can be used as a probe of *cis-trans* isomerization in proteins. 4,4-Difluoroproline has been incorporated into proteins, via expression with proline auxotrophic bacteria.[291,292] The elastin-1 and barstar proteins have both been successfully expressed with 4,4-difluoroproline incorporated via the proline auxotrophic cell lines. Both of these proteins contain proline residues which are known to undergo *cis-trans* isomerization. Interestingly, neither report of expressed proteins containing 4,4-difluoroproline provide a ^{19}F NMR spectrum of the protein. As mentioned in the introduction, Yu and Hilvert synthetically produced a small 99-residue protein containing 4,4-difluoroproline and did indeed successfully demonstrate detection of *cis-trans* isomerization in their protein.[254] We expect this work could be expanded to expressed proteins. This would be especially beneficial in determining the conformation of prolines in the absence of protein crystal structures. Many protein crystal structures do not have good enough resolution (sub-2.0 Å) to determine the conformation of proline residues. Without improved resolution, these proline residues are often identified as *trans*, due to the higher population of *trans* amide bonds within proteins. Incorporation of 4,4-difluoroproline would allow for identification of the proline conformation via quick and simple ^{19}F NMR experiments. As seen in Yu and Hilvert's ^{19}F NMR spectrum as well as peptide models presented

here, the *cis* conformation has a larger $\Delta\delta_{FF}$ while the *trans* conformation has a much smaller $\Delta\delta_{FF}$ which allows for identification via a simple 1-dimentional ^{19}F NMR spectrum.[254] Furthermore, the lack of fluorine atoms in proteins means there are no competing signals like ^{1}H and ^{13}C NMR experiments.

Expressed proteins containing 4,4-difluoroproline also have the ability to report on *cis-trans* isomerization in real time via ^{19}F NMR. Even in the highest resolution crystal structures, no kinetic information can be obtained via x-ray crystallography. As demonstrated in the ^{19}F experiments reported in the Pro$_9$ oligomers containing 4,4-difluoroproline, we can obtain kinetic information as *cis-trans* isomerization occurs within peptides and proteins. Similar to Yu and Hilvert's experiment, denaturation of proteins as well as protein folding could be monitored by ^{19}F NMR. While Yu and Hilvert measured pH-dependent denaturation by ^{19}F NMR, a number of other denaturing conditions could be explored including increasing concentrations of denaturing agents (such as urea or surfactants) as well as parameters such as melting[254]. Proteins which can spontaneously fold, such as barstar which has previously been expressed with 4,4-difluoroproline, could be observed not only denaturing but also refolding after denaturation. Barstar is a well characterized protein which has been used extensively to study protein folding and is expressible in gram quantities.[259,293] Furthermore, barstar has a known thermostability, approximately 70 °C.[293] Utilizing barstar expressed with 4,4-difluoroproline, thermal denaturation can be monitored by ^{19}F NMR simply by increasing the temperature on the NMR instrument. The spontaneous refolding can also be monitored by decreasing the temperature on the NMR instrument. These experiments would be a good proof-of-principle of monitoring protein denaturation and folding by ^{19}F NMR in real time.

Another potential application of this work is the ability to monitor *cis-trans* isomerization of prolyl isomerases by ^{19}F NMR. As mentioned previously, prolyl isomerases catalyze *cis-trans* isomerization in proteins to allow proteins to fold faster, as *cis-trans* isomerization is often the rate-limiting step in protein folding.[264] Prolyl isomerases are known to be activated in cellular growth and proliferation pathways and certain cancer phenotype including Her2+ breast cancers.[249] Using simple and straightforward ^{19}F NMR experiments, over-expression of prolyl isomerases in cancer cell lines could be identified. 4,4-Difluoroproline could potentially also be used to measure the kinetics of prolyl isomerase enzymes by ^{19}F NMR. Furthermore, 4,4-difluoroproline could potentially be used to image prolyl isomerases in live cells by ^{19}F NMR. Utilizing expressed proteins containing 4,4-difluoroproline or synthetic substrates containing 4,4-difluoroproline which can be introduced into cells via cell penetrating peptides, prolyl isomerase activity could be measured within cells via ^{19}F NMR.

4.4 Experimental

4.4.1 Synthesis of Fmoc-4,4-Difluoroproline

L-4,4-difluoroproline (2) Boc-4,4-difluoroproline (0.50 g, 2.00 mmol) was dissolved in 4 HCl (10 mL) and 1,4-dixoane (10 mL). The solution was heated to reflux and allowed to stir for six hours. Upon completion of the reaction, the solution was neutralized with 1 M NaOH to pH 7 and the solvent was removed under reduced pressure. Crude product was used in the next reaction without further purification. [1]H (400 MHz, MeOD) δ 4.82-4.78 (t, *J* = 8.7, 8.6 Hz, 1H), 3.92-3.77 (m, 2H), 3.07-2.95

(m, 1H), 2.86-2.73 (m, 1H). ^{13}C NMR (150.8 MHz, MeOD) δ 167.9, 128.7, 126.2, 123.8, 62.9. ^{19}F NMR (376.3 MHz, MeOD) δ −98.80 (dd, J = 240, 240 Hz).

Fmoc-L-4,4-difluoroproline (3) Crude compound **2** (0.17 g, 2.00 mmol) was dissolved in 1,4-dioxane (13 mL). Fmoc-OSu (0.81 g, 2.39 mmol) and K_2CO_3 (0.33 g, 2.39 mmol) were added, and the resultant solution was stirred for 14 hours at room temperature. The 1,4-dioxane was removed under reduced pressure, and the crude product was acidified with 2 M HCl to pH 2. The crude product was extracted with ethyl acetate (2 × 50 mL). The solvent was removed, and the crude product was redissolved in CH_2Cl_2. The crude mixture was purified via column chromatography (0−4% methanol in CH_2Cl_2 v/v) to obtain compound **3** (0.34 g, 0.91 mmol) as a white solid in 46% yield over two steps. ^{1}H (400 MHz, MeOD) δ 7.81−7.79 (d, J = 7.6 Hz, 2H, major), 7.77-7.75 (d, J = 8.7 Hz, 1H, minor), 7.71-7.63 (m, 0.6H, minor), 7.61-7.53 (m, 3H), 7.48-7.40 (m, 3H), 7.37-7.32 (m, 3H), 4.67-4.64 (dd, J = 5.6, 5.5 Hz, 1H), 4.59-4.50 (m, 2H), 4.49-4.44 (m, 1H, minor) 4.40-4.37 (m, 0.6H, minor), 4.30-4.27 (dd, J = 6.9, 6.9 Hz, 1H, major), 4.30-4.27 (dd, J = 5.8, 5.6 Hz, 0.6H, minor) 4.10-3.91 (m, 1H, minor), 3.90-3.77 (m, 3H, major), 2.81-2.67 (m, 3H, major), 2.59-2.48 (m, 0.6H, minor). ^{13}C NMR (150.8 MHz, MeOD) δ 174.1, 173.1, 155.2, 143.7, 143.4, 143.3, 141.4, 127.9, 127.8, 127.2, 127.1, 124.9, 124.9, 124.8, 124.7, 120.1, 120.0, 68.4, 67.7, 57.0, 56.4, 53.1, 47.0, 37.3, 37.0. ^{19}F NMR (376.3 MHz, MeOD) δ -97.9, -98.0, -98.3, -98.5, -98.6, -98.7, -98.9, -98.9, -99.4 (J = 24 Hz), -99.9, -100.0, -100.1, -100.5, -100.7.

4.4.2 Peptide Synthesis and Purification

All peptides were synthesized using standard Fmoc solid phase peptide synthesis with MBHA Rink amide resin (0.1 or 0.25 mmol). Before the first amino acid coupling, resin was swelled in DMF (15 min). Each amino acid coupling was performed using Fmoc amino acids (0.4 or 1.0 mmol, 4 equiv) and HATU (0.4 or 0.8 mmol, 4 equiv). Each coupling cycle consisted of the removal of the Fmoc group (20% piperidine in DMF, 3 × 15 min) and the amide coupling (Fmoc amino acid, HBTU, 8% DIPEA in DMF, 90 min) except Fmoc-4,4-dilfluoroproline which was coupled for 14 hours. After addition of the last residue, the N-terminal Fmoc group was removed (20% piperidine in DMF, 3 × 15 min) and the amino terminus was acetylated (5% acetic anhydride in pyridine, 3 × 5 min). The resin was washed with DMF and CH2Cl2 and dried with ethyl ether. Peptides were cleaved and deprotected from the resin in a solution of 5% H2O, 5% triisopropylsilane (TIS), and 90% TFA. Cleavage reactions were allowed to stir for 90 minutes at room temperature. TFA was removed by evaporation and the peptides were precipitated with ether. Ether was removed and the peptide was dissolved in water and filtered before HPLC. All peptides were purified using reverse phase HPLC on a Vydac 300 Å pore C18 semi-preparative column (10 × 250 nm, 5 μm particle size) or on a Varian Microsorb MV 100 Å pore C18 analytical column (10 × 250 nm, 5 μm particle size. Peptides were purified by a linear gradient of 0 – 50% buffer B (20% water, 80% acetonitrile, 0.05% TFA) in buffer A (98% water, 2% acetonitrile, 0.06% TFA) over 60 minutes. Purity was demonstrated by reinjection of peptide on the HPLC. HPLCs were conducted under the parameters described earlier on a Varian Microsorb MV 100 Å pore C18 analytical column. Ac-PPPPPPPPP-NH2: t R = 40.4 min, expected 933.1, found 955.7 (M + Na). Ac-P(Dfp)PPPPPP-NH2: t_R = 44.3, expected 968.1, found 991.8 (M +

Na$^+$). Ac-PPPP(Dfp)PPPP-NH$_2$: t$_R$ = 46.7 min, expected 968.1, found 991.8 (M + Na$^+$). Ac-PPPPPPP(Dfp)P-NH$_2$: t$_R$ = 45.1 min, expected 968.1, found 991.7 (M + Na$^+$). Ac-GPP(Dfp)PPGY-NH$_2$: t$_R$ = 31.0 min, expected 857.9, found 880.6 (M + Na$^+$).

4.4.3   Circular Dichromism (CD)

Circular dichroism (CD) experiments were completed on a Jasco model J-810 spectropolarimeter. All data were collected using a 0.1 cm cell (Starna Cells, Atascadero, CA) Data in Figure 3 were collected in either deuterated acetonitrile or 5 mM phosphate buffer pH 7.4 and 25 mM KF and collected at 25 °C. Peptides were dissolved in solvent and allowed to equilibrate for a minimum of six hours at room temperature prior to data collection. Data in Figure 4 were collected in deuterated acetonitrile. Peptides were equilibrated to PPII in water for at least two hours at room temperature. The water was lyophilized prior to use. Peptides were dissolved in deuterated acetonitrile at room temperature. Data was collected at 25 °C over six hours. Data in Figure 6 were collected in 10% deuterated acetonitrile, 10% D$_2$O, and 80% water. Peptides were equilibrated for a minimum of six hours in 50 μL deuterated acetonitrile. Peptides, D$_2$O and water stocks were chilled to zero degrees in ice prior to the start of the experiment. D$_2$O (50 μL) and water (400 μL) were added to peptide stocks and CD experiments were taken at 15 °C. Data represent the average of at least three independent trials. Error bars indicate standard error. Data were background-corrected but were not smoothed.

4.4.4 ¹H NMR Spectroscopy of Peptides

All peptide NMR experiments were performed on a a Brüker 600 MHz (¹⁹F 564.5 MHz) NMR spectrometer equipped with a 5-mm Brüker SMART probe in water with 5 mM phosphate buffer (pH 4) with 25 mM NaCl and 10% D_2O. TSP was used as a standard for ¹H NMR.

4.4.5 ¹⁹F NMR Spectroscopy of Peptides

Peptides were equilibrated to PPII in water for at least two hours at room temperature. The water was lyophilized prior to use. Peptides were dissolved in deuterated acetonitrile at room temperature. Data in Figure 5 was collected at 25 °C over six hours on a 600 mHz NMR equipped with an autosampler. Data in Figure 7 were collected in 10% deuterated acetonitrile, 10% D_2O, and 80% water. Peptides were equilibrated for a minimum of six hours in 50 µL deuterated acetonitrile. Peptides, D_2O and water stocks were chilled to zero degrees in ice prior to the start of the experiment. D_2O (50 µL) and water (400 µL) were added to peptide stocks and NMR experiments were taken at 15 °C on a 400 mHz NMR. All ¹⁹F experiments were completed with a 10 ppm sweep width, 128 scans, and an acquisition time of 0.8 seconds.

4.4.6   X-ray Crystallography

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 ($\lambda = 0.71073$ 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 consistent for P2 1 and P2 1 /m. The absence of a molecular mirror, occupancy of two, and the synthetic method are consistent with the noncentrosymmetric space group option. 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 U eq of the attached atom. Atomic scattering factors are contained in the SHELXTL program library (Sheldrick, G., op. cit.).

Empirical formula	C21 H18 Cl3 F2 N O4
Formula weight	492.71
Temperature	200(2) K
Wavelength	0.71073 A
Crystal system, space group	Monoclinic, P2(1)
Unit cell dimensions	a = 11.1277(16) A alpha = 90 deg.
	b = 5.8944(8) A beta = 100.464(7) deg.
	c = 16.870(2) A gamma = 90 deg.
Volume	1088.1(3) A^3
Z, Calculated density	2, 1.504 Mg/m^3
Absorption coefficient	0.467 mm^-1
F(000)	504
Crystal size	0.489 x 0.159 x 0.094 mm
Theta range for data collection	1.861 to 27.713 deg.
Limiting indices	-14<=h<=13, -7<=k<=7, -21<=l<=22
Reflections collected / unique	16808 / 5042 [R(int) = 0.0583]
Completeness to theta = 25.242	99.9 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7456 and 0.5611
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5042 / 1 / 281
Goodness-of-fit on F^2	1.015
Final R indices [I>2sigma(I)]	R1 = 0.0511, wR2 = 0.1227
R indices (all data)	R1 = 0.0731, wR2 = 0.1349
Absolute structure parameter	-0.10(5)
Extinction coefficient	n/a
Largest diff. peak and hole	0.331 and -0.311 e.A^-3

Table 4.1.   Crystal data and structure refinement for Fmoc-4,4-difluoroproline

	x	y	z	U(eq)
C(1)	3620(4)	1438(8)	6527(2)	48(1)
C(2)	4115(4)	486(10)	5824(3)	55(1)
C(3)	3112(4)	713(10)	5099(2)	54(1)
C(4)	2404(4)	2838(8)	5277(2)	43(1)
C(5)	1028(4)	2417(7)	5081(2)	38(1)
C(6)	2381(4)	4939(7)	6517(2)	41(1)
C(7)	2240(4)	6435(8)	7802(2)	44(1)
C(8)	3137(3)	6712(7)	8594(2)	37(1)
C(9)	4263(3)	8082(7)	8511(2)	34(1)
C(10)	5162(4)	7621(8)	8070(2)	42(1)
C(11)	6110(4)	9175(8)	8082(3)	48(1)
C(12)	6134(4)	11180(9)	8521(3)	49(1)
C(13)	5246(4)	11647(8)	8968(2)	43(1)
C(14)	4294(4)	10102(7)	8963(2)	36(1)
C(15)	3253(4)	10118(7)	9383(2)	36(1)
C(16)	2917(4)	11653(8)	9933(2)	43(1)
C(17)	1892(4)	11177(9)	10262(2)	50(1)
C(18)	1222(4)	9216(9)	10052(3)	52(1)
C(19)	1558(4)	7685(8)	9507(2)	47(1)
C(20)	2566(4)	8139(7)	9176(2)	37(1)
C(21)	9131(4)	2988(8)	7116(3)	48(1)
Cl(1)	7804(1)	1601(3)	6652(1)	69(1)
Cl(2)	8776(1)	5513(2)	7563(1)	73(1)
Cl(3)	10012(2)	1278(3)	7851(1)	90(1)
F(1)	4511(4)	-1696(7)	5967(2)	84(1)
F(2)	5117(3)	1664(9)	5717(2)	89(1)
N(1)	2817(3)	3214(6)	6144(2)	42(1)
O(1)	1736(3)	6459(6)	6181(2)	50(1)
O(2)	2778(3)	4831(5)	7319(2)	46(1)
O(3)	377(3)	2233(6)	5561(2)	52(1)
O(4)	680(3)	2236(6)	4288(2)	49(1)

Table 4.2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (A^2 x 10^3) for Fmoc-4,4-difluoroproline. U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

<u>Bond Lengths</u>

C(1)-N(1)	1.450(5)	C(9)-C(14)	1.411(6)
C(1)-C(2)	1.505(6)	C(10)-C(11)	1.394(6)
C(1)-H(1A)	0.99	C(10)-H(10)	0.95
C(1)-H(1B)	0.99	C(11)-C(12)	1.392(7)
C(2)-F(2)	1.353(6)	C(11)-H(11)	0.95
C(2)-F(1)	1.367(7)	C(12)-C(13)	1.376(6)
C(2)-C(3)	1.504(6)	C(12)-H(12)	0.95
C(3)-C(4)	1.538(7)	C(13)-C(14)	1.396(6)
C(3)-H(3A)	0.99	C(13)-H(13)	0.95
C(3)-H(3B)	0.99	C(14)-C(15)	1.464(6)
C(4)-N(1)	1.469(5)	C(15)-C(16)	1.394(6)
C(4)-C(5)	1.527(6)	C(15)-C(20)	1.403(6)
C(4)-H(4)	1	C(16)-C(17)	1.385(6)
C(5)-O(3)	1.186(5)	C(16)-H(16)	0.95
C(5)-O(4)	1.328(4)	C(17)-C(18)	1.387(7)
C(6)-O(1)	1.221(5)	C(17)-H(17)	0.95
C(6)-N(1)	1.333(6)	C(18)-C(19)	1.387(7)
C(6)-O(2)	1.346(4)	C(18)-H(18)	0.95
C(7)-O(2)	1.447(5)	C(19)-C(20)	1.368(6)
C(7)-C(8)	1.525(5)	C(19)-H(19)	0.95
C(7)-H(7A)	0.99	C(21)-Cl(1)	1.744(5)
C(7)-H(7B)	0.99	C(21)-Cl(2)	1.746(5)
C(8)-C(20)	1.517(5)	C(21)-Cl(3)	1.753(5)
C(8)-C(9)	1.518(6)	C(21)-H(21)	1
C(8)-H(8)	1	O(4)-H(4A)	0.84
C(9)-C(10)	1.378(6)		

<u>Bond Angles</u>

N(1)-C(1)-C(2)	101.7(3)	C(9)-C(10)-C(11)	119.0(4)
N(1)-C(1)-H(1A)	111.4	C(9)-C(10)-H(10)	120.5
C(2)-C(1)-H(1A)	111.4	C(11)-C(10)-H(10)	120.5
N(1)-C(1)-H(1B)	111.4	C(12)-C(11)-C(10)	120.5(4)
C(2)-C(1)-H(1B)	111.4	C(12)-C(11)-H(11)	119.8

H(1A)-C(1)-H(1B)	109.3	C(10)-C(11)-H(11)	119.8
F(2)-C(2)-F(1)	105.1(4)	C(13)-C(12)-C(11)	121.0(4)
F(2)-C(2)-C(3)	110.7(4)	C(13)-C(12)-H(12)	119.5
F(1)-C(2)-C(3)	113.3(4)	C(11)-C(12)-H(12)	119.5
F(2)-C(2)-C(1)	110.0(4)	C(12)-C(13)-C(14)	119.1(4)
F(1)-C(2)-C(1)	111.3(4)	C(12)-C(13)-H(13)	120.5
C(3)-C(2)-C(1)	106.4(4)	C(14)-C(13)-H(13)	120.5
C(2)-C(3)-C(4)	104.2(4)	C(13)-C(14)-C(9)	119.9(4)
C(2)-C(3)-H(3A)	110.9	C(13)-C(14)-C(15)	131.3(4)
C(4)-C(3)-H(3A)	110.9	C(9)-C(14)-C(15)	108.8(3)
C(2)-C(3)-H(3B)	110.9	C(16)-C(15)-C(20)	120.4(4)
C(4)-C(3)-H(3B)	110.9	C(16)-C(15)-C(14)	131.1(4)
H(3A)-C(3)-H(3B)	108.9	C(20)-C(15)-C(14)	108.5(3)
N(1)-C(4)-C(5)	111.1(3)	C(17)-C(16)-C(15)	118.4(4)
N(1)-C(4)-C(3)	103.8(3)	C(17)-C(16)-H(16)	120.8
C(5)-C(4)-C(3)	110.7(4)	C(15)-C(16)-H(16)	120.8
N(1)-C(4)-H(4)	110.4	C(16)-C(17)-C(18)	120.7(4)
C(5)-C(4)-H(4)	110.4	C(16)-C(17)-H(17)	119.6
C(3)-C(4)-H(4)	110.4	C(18)-C(17)-H(17)	119.6
O(3)-C(5)-O(4)	125.3(4)	C(19)-C(18)-C(17)	120.9(4)
O(3)-C(5)-C(4)	125.5(3)	C(19)-C(18)-H(18)	119.5
O(4)-C(5)-C(4)	109.2(3)	C(17)-C(18)-H(18)	119.5
O(1)-C(6)-N(1)	125.1(4)	C(20)-C(19)-C(18)	118.9(4)
O(1)-C(6)-O(2)	124.0(4)	C(20)-C(19)-H(19)	120.5
N(1)-C(6)-O(2)	110.9(3)	C(18)-C(19)-H(19)	120.5
O(2)-C(7)-C(8)	106.9(3)	C(19)-C(20)-C(15)	120.7(4)
O(2)-C(7)-H(7A)	110.3	C(19)-C(20)-C(8)	128.8(4)
C(8)-C(7)-H(7A)	110.3	C(15)-C(20)-C(8)	110.4(3)
O(2)-C(7)-H(7B)	110.3	Cl(1)-C(21)-Cl(2)	110.7(3)
C(8)-C(7)-H(7B)	110.3	Cl(1)-C(21)-Cl(3)	111.5(3)
H(7A)-C(7)-H(7B)	108.6	Cl(2)-C(21)-Cl(3)	109.0(2)
C(20)-C(8)-C(9)	102.4(3)	Cl(1)-C(21)-H(21)	108.5
C(20)-C(8)-C(7)	110.0(3)	Cl(2)-C(21)-H(21)	108.5
C(9)-C(8)-C(7)	113.3(3)	Cl(3)-C(21)-H(21)	108.5
C(20)-C(8)-H(8)	110.3	C(6)-N(1)-C(1)	126.1(3)
C(9)-C(8)-H(8)	110.3	C(6)-N(1)-C(4)	120.7(3)
C(7)-C(8)-H(8)	110.3	C(1)-N(1)-C(4)	113.0(3)

C(10)-C(9)-C(14)	120.6(4)	C(6)-O(2)-C(7)	116.2(3)
C(10)-C(9)-C(8)	129.6(4)	C(5)-O(4)-H(4A)	109.5
C(14)-C(9)-C(8)	109.8(3)		

Symmetry transformations used to generate equivalent atoms:

Table 4.3.   Bond lengths [A] and angles [deg] for Fmoc-4,4-difluoroproline.

Table S4. Anisotropic displacement parameters (A^2 x 10^3) for neal026. The anisotropic displacement factor exponent takes the form: -2 pi^2 [h^2 a*^2 U11 + ... + 2 h k a* b* U12]

	U11	U22	U33	U23	U13
C(1)	60(3)	48(2)	33(2)	1(2)	0(2)
C(2)	52(3)	73(3)	38(2)	-4(2)	2(2)
C(3)	51(3)	75(4)	35(2)	-11(2)	4(2)
C(4)	46(2)	52(2)	27(2)	1(2)	-4(2)
C(5)	44(2)	37(2)	32(2)	-7(2)	1(2)
C(6)	39(2)	40(2)	37(2)	-2(2)	-7(2)
C(7)	43(2)	44(2)	39(2)	-7(2)	-2(2)
C(8)	40(2)	39(2)	30(2)	-3(2)	1(1)
C(9)	35(2)	37(2)	29(2)	4(2)	-1(1)
C(10)	46(2)	46(2)	33(2)	4(2)	3(2)
C(11)	40(2)	59(3)	44(2)	12(2)	9(2)
C(12)	39(2)	54(3)	51(2)	16(2)	1(2)
C(13)	43(2)	42(2)	38(2)	4(2)	-4(2)
C(14)	39(2)	38(2)	26(2)	2(2)	-6(1)
C(15)	43(2)	37(2)	24(2)	2(1)	-7(1)
C(16)	53(2)	41(2)	31(2)	-2(2)	-4(2)
C(17)	54(2)	57(3)	37(2)	-4(2)	5(2)
C(18)	45(3)	66(3)	46(2)	2(2)	14(2)
C(19)	46(2)	51(3)	44(2)	-2(2)	8(2)
C(20)	41(2)	39(2)	29(2)	3(2)	0(2)
C(21)	51(2)	49(3)	44(2)	4(2)	10(2)
Cl(1)	68(1)	85(1)	51(1)	-1(1)	3(1)
Cl(2)	84(1)	58(1)	76(1)	-11(1)	10(1)
Cl(3)	96(1)	64(1)	95(1)	10(1)	-26(1)
F(1)	100(2)	91(3)	57(2)	-11(2)	1(2)
F(2)	49(2)	149(4)	66(2)	0(2)	4(1)
N(1)	49(2)	40(2)	30(2)	-4(1)	-9(1)
O(1)	49(2)	47(2)	44(2)	-2(1)	-13(1)
O(2)	57(2)	42(2)	34(1)	-7(1)	-7(1)
O(3)	50(2)	67(2)	40(2)	-12(2)	12(1)

O(4)   44(2)   68(2)   30(1)   -8(1)   -3(1)

_____

Table S5. Hydrogen coordinates (x 10^4) and isotropic displacement parameters (A^2 x 10^3) for Fmoc-4,4-difluoroproline.

_____

	x	y	z	U(eq)
H(1A)	3163	276	6775	57
H(1B)	4281	2062	6943	57
H(3A)	2576	-640	5038	65
H(3B)	3453	914	4601	65
H(4)	2627	4163	4963	52
H(7A)	1446	5860	7903	52
H(7B)	2102	7911	7519	52
H(8)	3381	5193	8835	44
H(10)	5137	6266	7763	51
H(11)	6744	8864	7789	57
H(12)	6773	12242	8511	59
H(13)	5280	13002	9276	51
H(16)	3380	12993	10077	52
H(17)	1646	12205	10636	59
H(18)	522	8916	10285	62
H(19)	1095	6343	9367	56
H(21)	9628	3360	6695	58
H(4A)	-71	1947	4179	73

_____

Torsion angles [deg] for Fmoc-4,4-difluoroproline.

N(1)-C(1)-C(2)-F(2)	87.3(4)	C(13)-C(14)-C(15)-C(20)	-179.4(4)
N(1)-C(1)-C(2)-F(1)	-156.6(4)	C(9)-C(14)-C(15)-C(20)	-1.3(4)
N(1)-C(1)-C(2)-C(3)	-32.7(5)	C(20)-C(15)-C(16)-C(17)	-0.3(5)
F(2)-C(2)-C(3)-C(4)	-89.0(5)	C(14)-C(15)-C(16)-C(17)	-177.4(4)
F(1)-C(2)-C(3)-C(4)	153.3(4)	C(15)-C(16)-C(17)-C(18)	0.3(6)
C(1)-C(2)-C(3)-C(4)	30.6(6)	C(16)-C(17)-C(18)-C(19)	-0.1(7)
C(2)-C(3)-C(4)-N(1)	-15.8(5)	C(17)-C(18)-C(19)-C(20)	-0.1(7)
C(2)-C(3)-C(4)-C(5)	-135.1(4)	C(18)-C(19)-C(20)-C(15)	0.1(6)
N(1)-C(4)-C(5)-O(3)	-5.5(7)	C(18)-C(19)-C(20)-C(8)	177.8(4)
C(3)-C(4)-C(5)-O(3)	109.2(5)	C(16)-C(15)-C(20)-C(19)	0.1(6)
N(1)-C(4)-C(5)-O(4)	175.9(4)	C(14)-C(15)-C(20)-C(19)	177.8(4)
C(3)-C(4)-C(5)-O(4)	-69.3(4)	C(16)-C(15)-C(20)-C(8)	-178.0(3)
O(2)-C(7)-C(8)-C(20)	172.8(3)	C(14)-C(15)-C(20)-C(8)	-0.3(4)
O(2)-C(7)-C(8)-C(9)	73.3(4)	C(9)-C(8)-C(20)-C(19)	-176.3(4)
C(20)-C(8)-C(9)-C(10)	178.7(4)	C(7)-C(8)-C(20)-C(19)	63.0(6)
C(7)-C(8)-C(9)-C(10)	-62.9(5)	C(9)-C(8)-C(20)-C(15)	1.7(4)
C(20)-C(8)-C(9)-C(14)	-2.5(4)	C(7)-C(8)-C(20)-C(15)	-119.1(4)
C(7)-C(8)-C(9)-C(14)	115.9(4)	O(1)-C(6)-N(1)-C(1)	177.6(4)
C(14)-C(9)-C(10)-C(11)	0.4(5)	O(2)-C(6)-N(1)-C(1)	-0.5(6)
C(8)-C(9)-C(10)-C(11)	179.2(4)	O(1)-C(6)-N(1)-C(4)	-8.5(7)
C(9)-C(10)-C(11)-C(12)	-1.2(6)	O(2)-C(6)-N(1)-C(4)	173.5(4)
C(10)-C(11)-C(12)-C(13)	1.7(6)	C(2)-C(1)-N(1)-C(6)	-162.4(4)
C(11)-C(12)-C(13)-C(14)	-1.5(6)	C(2)-C(1)-N(1)-C(4)	23.3(5)
C(12)-C(13)-C(14)-C(9)	0.8(5)	C(5)-C(4)-N(1)-C(6)	-60.6(5)
C(12)-C(13)-C(14)-C(15)	178.7(4)	C(3)-C(4)-N(1)-C(6)	-179.6(4)
C(10)-C(9)-C(14)-C(13)	-0.2(5)	C(5)-C(4)-N(1)-C(1)	114.1(4)
C(8)-C(9)-C(14)-C(13)	-179.2(3)	C(3)-C(4)-N(1)-C(1)	-4.9(5)
C(10)-C(9)-C(14)-C(15)	-178.6(3)	O(1)-C(6)-O(2)-C(7)	9.9(6)
C(8)-C(9)-C(14)-C(15)	2.4(4)	N(1)-C(6)-O(2)-C(7)	-172.1(4)
C(13)-C(14)-C(15)-	-2.1(7)	C(8)-C(7)-O(2)-C(6)	-157.0(4)

C(16)

C(9)-C(14)-C(15)-C(16)   176.1(4)

Table 4.4.   Torsion angles [deg] for Fmoc-4,4-difluoroproline.

REFERENCES

1. Cobb SL, Murphy CD. 19F NMR applications in chemical biology. *J Fluorine Chem*. 2009;130(2):132 143.

2. Ruiz-Cabello J, Barnett BP, Bottomley PA, Bulte JW. Fluorine (19F) MRS and MRI in biomedicine. *NMR Biomed*. 2011;24(2):114-129.

3. Ojima I, ed. *Fluorine in medicinal chemistry and chemical biology.* Oxford United Kingdom: Blackwell Publishing Ltd; 2009. 10.1002/9781444312096.

4. Wang, J., Sánchez-Roselló, M., Aceña, J. L., del Pozo, C., Sorochinsky, A. E., Fustero, S., Soloshonok, V., Liu, H. Fluorine in pharmaceutical industry: Fluorine-containing drugs introduced to the market in the last decade (2001–2011). *Chem. Rev*. 2014;114(4):2432-2506.

5. Marsh, E. N. G., Suzuki, Y. Using 19F NMR to probe biological interactions of proteins and peptides. *ACS Chem. Biol.* 2014;9(6):1242-1250.

6. Janjic JM, Ahrens ET. Fluorine-containing nanoemulsions for MRI cell tracking. *Wiley Interdiscip Rev Nanomed Nanobiotechnol*. 2009;1(5):492-501.

7. Laverman P, Boerman OC, Corstens FH, Oyen WJ. Fluorinated amino acids for tumour imaging with positron emission tomography. *Eur J Nucl Med Mol Imaging*. 2002;29(5):681-690.

8. Merkel L, Budisa N. Organic fluorine as a polypeptide building element: In vivo expression of fluorinated peptides, proteins and proteomes. *Org Biomol Chem*. 2012;10(36):7241-7261.

9. Mykhailiuk PK, Afonin S, Palamarchuk GV, Shishkin OV, Ulrich AS, Komarov IV. Synthesis of trifluoromethyl-substituted proline analogues as 19F NMR labels for peptides in the polyproline II conformation. *Angew Chem Int Ed Engl*. 2008;47(31):5765-5767.

10. O'Hagan D. Fluorine in health care: Organofluorine containing blockbuster drugs. *J Fluorine Chem*. 2010;131(11):1071-1081.

11. Yoder NC, Kumar K. Fluorinated amino acids in protein design and engineering. *Chem Soc Rev*. 2002;31(6):335-341.

12. Muller K, Faeh C, Diederich F. Fluorine in pharmaceuticals: Looking beyond intuition. *Science*. 2007;317(5846):1881-1886.

13. Park, B. K., Kitteringham, N. R., O'Neill, P. M. Metabolism of fluorine-containing drugs. *Annual Review of Pharmacology and Toxicology*. 2001;41:1-906.

14. Yu S. Review of 18F-FDG synthesis and quality control. *Biomed Imaging Interv J*. 2006;2(4):1-11.

15. Caravan P, Ellison JJ, McMurry TJ, Lauffer RB. Gadolinium(III) chelates as MRI contrast agents: Structure, dynamics, and applications. *Chem Rev*. 1999;99(9):2293-2352.

16. Salwiczek, M., Nyakatura, E. K., Gerling, U. I. M., Ye, S., Koksch, B. Fluorinated amino acids: Compatibility with native protein structures and effects on protein–protein interactions. *Chemical Society Reviews*. 2021;41:2135-2171.

17. Bretscher LE, Jenkins CL, Taylor KM, DeRider ML, Raines RT. Conformational stability of collagen relies on a stereoelectronic effect. *J Am Chem Soc*. 2001;123(4):777-778.

18. Chiu, H.-P., Kokona, B., Fairman, R., Cheng, R. P. Effect of highly fluorinated amino acids on protein stability at a solvent-exposed position on an internal strand of protein G B1 domain. *J. Am. Chem. Soc.* 2009;131(37):13192-13193.

19. Wang, P., Tang, Y., Tirrell, D. A. Incorporation of trifluoroisoleucine into proteins in vivo. *J. Am. Chem. Soc.* 2003;125(23):6900-6906.

20. Kawahara, K., Nemoto, N., Motooka, D., Nishi, Y., Doi, M., Uchiyama, S., Nakazawa, T., Nishiuchi, Y., Yoshida, T., Ohkubo, T., & Kobayashi, Y. Polymorphism of collagen triple helix revealed by 19F NMR of model peptide [pro-

4(R)-hydroxyprolyl-gly]3-[pro-4(R)-fluoroprolyl-gly]-[pro-4(R)-hydroxyprolyl-gly]3. *J. Phys. Chem. B*. 2012;116(32):6908-6915.

21. Bilgiçer B, Xing X, Kumar K. Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores. *J Am Chem Soc*. 2001;123(47):11815-11816.

22. Tang Y, Tirrell DA. Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J Am Chem Soc*. 2001;123(44):11089-11090.

23. Buer, B. C., & Marsh, E. N. G. Fluorine: A new element in protein design. *Protein Sci*. 2012;21(4):453-462.

24. Horng, J.-C., Raleigh, D. P. Phi-values beyond the ribosomally encoded amino acids: Kinetic and thermodynamic consequences of incorporating trifluoromethyl amino acids in a globular protein. *J Am Chem Soc*. 2003;125(31):9286-9287.

25. Woll MG, Hadley EB, Mecozzi S, Gellman SH. Stabilizing and destabilizing effects of phenylalanine --> F5-phenylalanine mutations on the folding of a small protein. *J Am Chem Soc*. 2006;128(50):15932-15933.

26. Zheng H, Gao J. Highly specific heterodimerization mediated by quadrupole interactions. *Angew Chem Int Ed Engl*. 2010;49(46):8635-8639.

27. Chiu, H.-P., Suzuki, Y., Gullickson, D., Ahmad, R., Kokona, B., Fairman, R., & Cheng, R. P. Helix propensity of highly fluorinated amino acids. *J. Am. Chem. Soc.* 2006;128(49):15556-15557.

28. Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., Vidal, M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*. 2005;437:1173-1178.

29. Salwiczek M, Nyakatura EK, Gerling UI, Ye S, Koksch B. Fluorinated amino acids: Compatibility with native protein structures and effects on protein-protein interactions. *Chem Soc Rev*. 2012;41(6):2135-2171.

30. Gardner, K. H., Kay, L. E. The use of 2h, 13c, 15n multidimensional nmr gto study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure*. 1998;27:1-528.

31. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J Am Chem Soc*. 1994;116(26):11655-11666.

32. Månsson, S., Johansson, E., Magnusson, P., Chai, C.-M., Hansson, G., Petersson, J. S., Petersson, J. S., Ståhlberg, F., Golman, K. 13C imaging—a new diagnostic platform. *European Radiology*. 2006;16(1):57-67.

33. Lu X, Fu Wang R. A concise review of current radiopharmaceuticals in tumor angiogenesis imaging. *Curr Pharm Des*. 2012;18(8):1032-1040.

34. Jaroniec, C. P., Filip, C., Griffin, R. G. 3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly 13C,15N-labeled solids. *J. Am. Chem. Soc.* 2002;124(36):10728-10742.

35. Shulman R, Brown T, Ugurbil K, Ogawa S, Cohen S, den Hollander J. Cellular applications of 31P and 13C nuclear magnetic resonance. *Science*. 1979;205(4402):160-166.

36. Takaoka, Y., Kiminami, K., Mizusawa, K., Matsuo, K., Narazaki, M., Matsuda, T., Hamachi, I. A general NMR method for rapid, efficient, and reliable biochemical screening. *J. Am. Chem. Soc.* 2003;125(47):14620-14625.

37. Yue, X., Taraban, M. B., Hyland, L. L., Yu, Y. B. Avoiding steric congestion in dendrimer growth through proportionate branching: A twist on da vinci's rule of tree branching. *J. Org. Chem.* 2012;77(20):8879-8887.

38. Dalvit C. Ligand- and substrate-based 19F NMR screening: Principles and applications to drug discovery. *Prog Nucl Magn Reson Spectrosc*. 2007;51(4):243-271.

39. Qiu X, Qing F. Recent advances in the synthesis of fluorinated amino acids. *European Journal of Organic Chemistry*. 2011;2011(18):3261-3278.

40. Dalvit C, Mongelli N, Papeo G, et al. Sensitivity improvement in 19F NMR-based screening experiments: Theoretical considerations and experimental applications. *J. Am. Chem. Soc.* 2005;127:13380-13385.

41. Dalvit, C., Ardini, E., Flocco, M., Fogliatto, G. P., Mongelli, N., Veronesi, M. A general NMR method for rapid, efficient, and reliable biochemical screening. *J. Am. Chem. Soc.* 2003;125(47):14620-14625.

42. Jackson, J. C., Hammill, J. T., Mehl, R. A. Site-specific incorporation of a 19F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* 2007;129(5):1160-1166.

43. 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. *J. Am. Chem. Soc.* 2013;135(11):4333-4363.

44. Tressler, C. M., Zondlo, N. J. (2S,4R)- and (2S,4S)-perfluoro-tert-butyl 4-hydroxyproline: Two conformationally distinct proline amino acids for sensitive application in 19F NMR. *J. Org. Chem.* 2014;79(12):5880-5886.

45. Sebesta DP, O'Rourk SS, Pieken WA. Facile preparation of perfluoro-tert-butyl ethers by the mitsunobu reaction. *J Org Chem*. 1996;61(1):361-362.

46. Nabuurs RJ, Kapoerchan VV, Metaxas A, et al. Polyfluorinated bis-styrylbenzenes as amyloid-beta plaque binding ligands. *Bioorg Med Chem*. 2014;22(8):2469-2481.

47. Jiang, Z.-X., Yu, Y. B. The synthesis of a geminally perfluoro-tert-butylated ß-amino acid and its protected forms as a potential pharmacokinetic modulator and reporter for peptide-based pharmaceuticals. *J. Org. Chem.* 2007;72(4):1464-1467.

48. Weng Z, Rickles RJ, Feng S, et al. Structure-function analysis of SH3 domains: SH3 binding specificity altered by single amino acid substitutions. *Mol Cell Biol*. 1995;15(10):5627-5634.

49. Hutti JE, Jarrell ET, Chang JD, et al. A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods*. 2004;1(1):27-29.

50. Carraz M, Zwart W, Phan T, Michalides R, Brunsveld L. Perturbation of estrogen receptor α localization with synthetic nona-arginine LXXLL-peptide coactivator binding inhibitors. *Chem Biol*. 2009;16(7):702-711.

51. Doig AJ, Baldwin RL. N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci*. 1995;4(7):1325-1336.

52. Elbaum, M. B., Zondlo, N. J. OGlcNAcylation and phosphorylation have similar structural effects in a-helices: Post-translational modifications as inducible start and stop signals in a-helices, with greater structural effects on threonine modification. *Biochemistry*. 2014;53(14):2242-2260.

53. 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 aromatici+1–prolyli+2 amide bond: Implications for type VI ß-turn formation. *Peptide Science*. 2006;84(2):192-204.

54. Cochran, D. A. E., Doig, A. J. Effect of the N2 residue on the stability of the a-helix for all 20 amino acids. *Protein Sci*. 2001;10(7):1305-1311.

55. Pandey AK, Thomas KM, Forbes CR, Zondlo NJ. Tunable control of polyproline helix (PPII) structure via aromatic electronic effects: An electronic switch of polyproline helix. *Biochemistry*. 2014;53(32):5307-5314.

56. Brown, A. M., Zondlo, N. J. A propensity scale for type II polyproline helices (PPII): Aromatic amino acids in proline-rich sequences strongly disfavor PPII due to Proline–Aromatic interactions. *Biochemistry*. 2012;51(25):5041-5051.

57. Trapp N, Scherer H, Hayes SA, et al. The perfluorinated alcohols (F5C6)(F3C)2COH and (F5C6)(F10C5)COH: Synthesis, theoretical and acidity

studies, spectroscopy and structures in the solid state and the gas phase. *Phys Chem Chem Phys*. 2011;13(13):6184-6191.

58. Swamy, K. C. K., Kumar, N. N. B., Balaraman, E., Kumar, K. V. P. P. Mitsunobu and related reactions: Advances and applications. *Chem. Rev.* 2009;109(6):2551-2651.

59. Gómez-Vidal, J. A., Forrester, M. T., Silverman, R. B. Mild and selective sodium azide mediated cleavage of p-nitrobenzoic esters. *Org. Lett.* 2001;3(16):2477-2479.

60. Buer BC, Levin BJ, Marsh EN. Perfluoro-tert-butyl-homoserine as a sensitive 19F NMR reporter for peptide-membrane interactions in solution. *J Pept Sci*. 2013;19(5):308-314.

61. Wolfe S, Wilson M, Cheng M, Shustov GV, Akuche CI. Cyclic hydroxamates, especially multiply substituted [1,2]oxazinan-3-ones. *Canadian Journal of Chemistry*. 2003;81(8):937-960.

62. Herr RJ. A whirlwind tour of current mitsunobu chemistry. *Albany Molecular Research, Inc. Technical Reports*. 1999;3(19):1-36.

63. Martin SF, Dodge JA. Efficacious modification of the mitsunobu reaction for inversions of sterically hindered secondary alcohols. *Tetrahedron Lett*. 1991;32(26):3017-3020.

64. Cherney, R. J., Wang, L. Efficient mitsunobu reactions with N-phenylfluorenyl or N-trityl serine esters. *J. Org. Chem*. 1996;61(7):2544-2546.

65. Gram, H. F., Mosher, C. W., Baker, B. R. Potential anticancer agents.1 XVII. alkylating agents to phenylalanine mustard. I. *J. Am. Chem. Soc*. 1959;81(12):3103-3108.

66. Wang, Y., Lin, Q. Synthesis and evaluation of photoreactive tetrazole amino acids. *Org. Lett*. 2009;11(16):3570-3573.

67. Colescott RL, Herr RR, Dailey JP. Analogs of phenylalanine containing sulfur. *J Am Chem Soc*. 1957;79(15):4232-4235.

68. Ganther HE. Selenotyrosine and related phenylalanine derivatives. *Bioorg Med Chem*. 2001;9(6):1459-1466.

69. Hobbs DW, Still WC. Synthesis of diphenyl thioether derivatives of peptides and aminoacids. *Tetrahedron Lett*. 1987;28(25):2805-2808.

70. Hobbs DW, Clark Still W. Synthesis of a thioether analog of the macrocyclic tripeptide K-13. *Tetrahedron Lett*. 1989;30(40):5405-5408.

71. Houghten RA, Rapoport H. Synthesis of pure p-chlorophenyl-L-alanine form L-phenylalanine. *J Med Chem*. 1974;17(5):556-558.

72. Sengupta S, Bhattacharyya S. Aqueous heck reaction of amino acid derived arenediazonium salts: Semisynthetic modification of phenylalanine and tyrosine. *Tetrahedron Lett*. 1995;36(25):4475-4478.

73. Zhu B, Ge J, Yao SQ. Developing new chemical tools for DNA methyltransferase 1 (DNMT 1): A small-molecule activity-based probe and novel tetrazole-containing inhibitors. *Bioorg Med Chem*. 2015;23(12):2917-2927.

74. Shirakawa, E., Itoh, K., Higashino, T., Hayashi, T. Tert-butoxide-mediated arylation of benzene with aryl halides in the presence of a catalytic 1,10-phenanthroline derivative. *J. Am. Chem. Soc.* 2010;132(44):15537-15539.

75. Mann G, Hartwig JF. Nickel- vs palladium-catalyzed synthesis of protected phenols from aryl halides. *J Org Chem*. 1997;62(16):5413-5418.

76. Tomori H, Fox JM, Buchwald SL. An improved synthesis of functionalized biphenyl-based phosphine ligands. *J Org Chem*. 2000;65(17):5334-5341.

77. Wolfe JP, Singer RA, Yang BH, Buchwald SL. Highly active palladium catalysts for suzuki coupling reactions. *J Am Chem Soc*. 1999;121(41):9550-9561.

78. Olmstead WN, Margolin Z, Bordwell FG. Acidities of water and simple alcohols in dimethyl sulfoxide solution. *J Org Chem*. 1980;45(16):3295-3299.

79. El-Faham A, Subiros Funosas R, Prohens R, Albericio F. COMU: A safer and more effective replacement for benzotriazole-based uronium coupling reagents. *Chemistry*. 2009;15(37):9404-9416.

80. Zalipsky S, Chang JL, Albericio F, Barany G. Preparation and applications of polyethylene glycol-polystyrene graft resin supports for solid-phase peptide synthesis. *Reactive Polymers*. 1994;22(3):243-258.

81. Yanagisawa D, Amatsubo T, Morikawa S, et al. In vivo detection of amyloid beta deposition using (1)(9)F magnetic resonance imaging with a (1)(9)F-containing curcumin derivative in a mouse model of alzheimer's disease. *Neuroscience*. 2011;184:120-127.

82. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J*. 1998;75(1):422-427.

83. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7(1):95-99.

84. Pawson T, Schlessingert J. SH2 and SH3 domains. *Current Biology*. 1993;3(7):434 442.

85. Brister, M. A., Pandey, A. K., Bielska, A. A., Zondlo, N. J. OGlcNAcylation and phosphorylation have opposing structural effects in tau: Phosphothreonine induces particular conformational order. *J. Am. Chem. Soc.* 2014;136(10):3803-3816.

86. Zondlo NJ. Aromatic–Proline interactions: Electronically tunable CH/pi interactions. *Acc. Chem. Res.* 2014;46(4):1039-1049.

87. Hansch C, Leo A, Taft RW. A survey of hammett substituent constants and resonance and field parameters. *Chem Rev*. 1991;91(2):165-195.

88. Thomas, K. M., Naduthambi, D., Zondlo, N. J. Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* 128;7:2216-2217.

89. Forbes CR. *Aromatic amino acids in peptides and proteins: Novel syntheses, influences on structure, and the nature of C-H/pi and S-H/pi interactions.* [PhD]. University of Delaware: University of Delaware; 2016.

90. 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. *J. Am. Chem. Soc.* 2012;134(6):2950-2953.

91. Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G., Schultz, P. G. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry*. 2011;50(11):1894-1900.

92. Dumas, A., Lercher, L., Spicer, C. D., Davis, B. G. Designing logical codon reassignment – expanding the chemistry in biology. *Chem. Sci.* 2015;6:50-69.

93. Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S. M., Brinker, A., Caldwell, J., Bursulaya, B., Spraggon, G., Brock, A., Ryu, Y., Uno, T., Schultz, P. G., Geierstanger, B. H. In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* 2008;130(29):9268-9281.

94. Chen, W.-N., Kuppan, K. V., Lee, M. D., Jaudzems, K., Huber, T., Otting, G. O-tert-butyltyrosine, an NMR tag for high-molecular-weight systems and measurements of submicromolar ligand binding affinities. *J. Am. Chem. Soc.* 2015;137(13):4581-4586.

95. Tenbaum S, Baniahmad A. Nuclear receptors: Structure, function and involvement in disease. *Int J Biochem Cell Biol*. 1997;29(12):1325-1341.

96. Hudis CA. Trastuzumab — mechanism of action and use in clinical practice. *N Engl J Med*. 2007;357(1):39-51.

97. Heinlein, C. A., Chang, C. Androgen receptor in prostate cancer. *Endocrine Reviews*. 2004;25(2):276-308.

98. Harries M, Smith I. The development and clinical use of trastuzumab (herceptin). *Endocr Relat Cancer*. 2002;9(2):75-85.

99. Vanneman M, Dranoff G. Combining immunotherapy and targeted therapies in cancer treatment. *Nat Rev Cancer*. 2012;12(4):237-251.

100. Birkedal-Hansen H, Moore WG, Bodden MK, et al. Matrix metalloproteinases: A review. *Crit Rev Oral Biol Med*. 1993;4(2):197-250.

101. Wells, J. A., McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*. 2007;450:1001-1009.

102. Davey NE, Van Roey K, Weatheritt RJ, et al. Attributes of short linear motifs. *Mol Biosyst*. 2012;8(1):268-281.

103. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition*. 2005;18(5):343 384.

104. Van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J., Davey, N. E. Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 2014;114(13):6733-6778.

105. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. *Nature Reviews Molecular Cell Biology*. 2013;15(1):49-63.

106. Sattler M, Liang H, Nettesheim D, et al. Structure of bcl-xL-bak peptide complex: Recognition between regulators of apoptosis. *Science*. 1997;275(5302):983-986.

107. van Delft MF, Huang DC. How the bcl-2 family of proteins interact to regulate apoptosis. *Cell Res*. 2006;16(2):203-213.

108. Baell JB, Huang DC. Prospects for targeting the bcl-2 family of proteins to develop novel cytotoxic drugs. *Biochem Pharmacol*. 2002;64(5-6):851-863.

109. Wang JL, Liu D, Zhang ZJ, et al. Structure-based discovery of an organic compound that binds bcl-2 protein and induces apoptosis of tumor cells. *Proc Natl Acad Sci U S A*. 2000;97(13):7124-7129.

110. Phizicky EM, Fields S. Protein-protein interactions: Methods for detection and analysis. *Microbiol Rev*. 1995;59(1):94-123.

111. Xing S, Wallmeroth N, Berendzen KW, Grefen C. Techniques for the analysis of protein-protein interactions in vivo. *Plant Physiol*. 2016;171(2):727-758.

112. Blikstad C, Ivarsson Y. High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Commun Signal*. 2015;13:38.

113. Zolghadr, K., Mortusewicz, O., Rothbauer, U., Kleinhans, R., Goehler, H., Wanker, E. E., Cardoso, M. C., Leonhardt, H. A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. *Molecular & Cellular Proteomics*. 2008;7(11):2279-2287.

114. Stanisz GJ, Odrobina EE, Pun J, et al. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine*. 2005;54(3):507-512.

115. Mizukami, S., Takikawa, R., Sugihara, F., Hori, Y., Tochio, H., Wälchli, M., Shirakawa, M., Kikuchi, K. Paramagnetic relaxation-based 19F MRI probe to detect protease activity. *J. Am. Chem. Soc.* 2008;130(3):794-795.

116. Jiang, Z.-X., Liu, X., Jeong, E.-K., Yu, Y. B. Symmetry-guided design and fluorous synthesis of a stable and rapidly excreted imaging tracer for 19F MRI. *Angewandte Chemie International Edition*. 2009;48(26):4755-4758.

117. Acharya, A., Das, I., Chandhok, D., Saha, T. Redox regulation in cancer. *Oxid Med Cell Longev.* 2010;3(1):23-34.

118. VanderHart, D. L., Gutowsky, H. S. Rigid-lattice NMR moments and line shapes with chemical-shift anisotropy. *Journal of Chemical Physics*. 1968;49:261-271.

119. Hunt E, Meyer H. Anisotropy of the 19F chemical shift in trapped CH3F molecules. *J Chem Phys*. 1964;41(2):353.

120. Danielson, M. A., Biemann, H.-P., Koshland, D. E., Falke, J. J. Attractant- and disulfide-induced conformational changes in the ligand binding domain of the chemotaxis aspartate receptor: A 19F NMR study. *Biochemistry*. 1994;33(20):6100-6109.

121. Janjic, J. M., Srinivas, M., Kadayakkara, D. K. K., Ahrens, E. T. Self-delivering nanoemulsions for dual fluorine-19 MRI and fluorescence detection. *J. Am. Chem. Soc*. 2008;130(9):2832-2841.

122. Srinivas M, Heerschap A, Ahrens ET, Figdor CG, de Vries IJ. (19)F MRI for quantitative in vivo cell tracking. *Trends Biotechnol*. 2010;28(7):363-370.

123. Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., Dunker, A. K. Coupled folding and binding with a-helix-forming molecular recognition elements. *Biochemistry*. 2005;44(37):12454-12470.

124. Ball LJ, Kuhne R, Schneider-Mergener J, Oschkinat H. Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew Chem Int Ed Engl*. 2005;44(19):2852-2869.

125. Galande, A. K., Bramlett, K. S., Trent, J. O., Burris, T. P., Wittliff, J. L., Spatola, A. F. Potent inhibitors of LXXLL-based Protein–Protein interactions. *ChemBioChem*. 2005;6(11):1991-1998.

126. Lorenz S, Deng P, Hantschel O, Superti-Furga G, Kuriyan J. Crystal structure of an SH2–kinase construct of c-abl and effect of the SH2 domain on kinase activity. *Biochem J*. 2015;468(2):283 291.

127. Ren S, Uversky VN, Chen Z, Dunker AK, Obradovic Z. Short linear motifs recognized by SH2, SH3 and ser/thr kinase domains are conserved in disordered protein regions. *BMC Genomics*. 2008;9(Suppl 2):S26.

128. Frame MC. Src in cancer: Deregulation and consequences for cell behaviour. *Biochim Biophys Acta*. 2002;1602(2):114-130.

129. Simon JA, Schreiber SL. Grb2 SH3 binding to peptides from sos: Evaluation of a general model for SH3-ligand interactions. *Chem Biol*. 1995;2(1):53-60.

130. Glass CK, Rose DW, Rosenfeld MG. Nuclear receptor coactivators. *Curr Opin Cell Biol*. 1997;9(2):222-232.

131. Burley S, Petsko G. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science*. 1985;229(4708):23-28.

132. Kuiper GG, Carlsson B, Grandien K, et al. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*. 1997;138(3):863-870.

133. Shiau AK, Barstad D, Loria PM, et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell*. 1998;95(7):927-937.

134. Dubbink HJ, Hersmus R, Pike ACW, et al. Androgen receptor ligand-binding domain interaction and nuclear receptor specificity of FXXLF and LXXLL motifs as determined by L/F swapping. *Molecular Endocrinology*. 2006;20(8):1742-1755.

135. Dahlman-Wright K, Cavailles V, Fuqua SA, et al. International union of pharmacology. LXIV. estrogen receptors. *Pharmacol Rev*. 2006;58(4):773-781.

136. Klinge CM, Jernigan SC, Mattingly KA, Risinger KE, Zhang J. Estrogen response element-dependent regulation of transcriptional activation of estrogen receptors alpha and beta by coactivators and corepressors. *J Mol Endocrinol*. 2004;33(2):387-410.

137. Rodriguez, A. L., Tamrazi, A., Collins, M. L., Katzenellenbogen, J. A. Design, synthesis, and in vitro biological evaluation of small molecule inhibitors of estrogen receptor a coactivator binding. *J. Med. Chem.* 2004;47(3):600-611.

138. Dutertre M, Smith CL. Molecular mechanisms of selective estrogen receptor modulator (SERM) action. *J Pharmacol Exp Ther*. 2000;295(2):431-437.

139. Ring A. Mechanisms of tamoxifen resistance. *Endocrine Related Cancer*. 2004;11(4):643-658.

140. Wu X, Bayle JH, Olson D, Levine AJ. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev*. 1993;7(7A):1126-1132.

141. Freedman DA, Wu L, Levine AJ. Functions of the MDM2 oncoprotein. *Cell Mol Life Sci*. 1999;55(1):96-107.

142. Juven-Gershon T, Oren M. Mdm2: The ups and downs. *Mol Med*. 1999;5(2):71-83.

143. Kussie PH, Gorina S, Marechal V, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*. 1996;274(5289):948-953.

144. Zondlo SC, Lee AE, Zondlo NJ. Determinants of specificity of MDM2 for the activation domains of p53 and p65: proline27 disrupts the MDM2-binding motif of p53. *Biochemistry*. 2006;45(39):11945-11957.

145. Haridas V. From peptides to non-peptide alpha-helix inducers and mimetics. *European Journal of Organic Chemistry*. 2009;2009(30):5112-5128.

146. Becerril J, Hamilton AD. Helix mimetics as inhibitors of the interaction of the estrogen receptor with coactivator peptides. *Angewandte Chemie International Edition*. 2007;46(24):4471-4473.

147. Vassilev, L., Vu, B., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., Liu, E. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. 204;303:844-848.

148. Zhao, Y., Aguilar, A., Bernard, D., Wang, S. Small-molecule inhibitors of the MDM2–p53 Protein–Protein interaction (MDM2 inhibitors) in clinical trials for cancer treatment. *J. Med. Chem.* 2015;58(3):1038-1052.

149. Stoll R, Renner C, Hansen S, et al. Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53. *Biochemistry*. 2001;40(2):336-344.

150. Vassilev LT. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. 2004;303(5659):844-848.

151. Yin H, Lee G, Park HS, et al. Terphenyl-based helical mimetics that disrupt the p53/HDM2 interaction. *Angew.Chem.Int.Ed Engl.* 2005;44(18):2704-2707.

152. Ferreira ME, Hermann S, Prochasson P, Workman JL, Berndt KD, Wright APH. Mechanism of transcription factor recruitment by acidic activators. *J Biol Chem*. 2005;280(23):21779-21784.

153. Presta LG, Rose GD. Helix signals in proteins. *Science*. 1988;240(4859):1632-1641.

154. Pollard T. A guide to simple and informative binding assays. *Mol Biol Cell*. 2010;21(23):4061-4067.

155. Lea, W. A., Simeonov, A. Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov*. 2011;6(1):17-32.

156. Myers JK, Pace CN, Scholtz JM. Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci*. 1998;7(2):383-388.

157. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci*. 1994;3(5):843-852.

158. Wishart, D. S., Sykes, B. D., Richards, F. M. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*. 1992;31(6):1647-1651.

159. Balaguer P, Francois F, Comunale F, et al. Reporter cell lines to study the estrogenic effects of xenoestrogens. *Sci Total Environ*. 1999;233(1-3):47-56.

160. Lindgren M, Hällbrink M, Prochiantz A, Langel Ü. Cell-penetrating peptides. *Trends Pharmacol Sci*. 2000;21(3):99-103.

161. Naduthambi, D., Bhor, S., Elbaum, M. B., Zondlo, N. J. Synthesis of a tetrasubstituted tetrahydronaphthalene scaffold for a-helix mimicry via a MgBr2-catalyzed Friedel–Crafts epoxide cycloalkylation. *Org. Lett.* 2013;15(18):4892-4895.

162. Chen, X., Zaro, J. L., Shen, W.-C. Fusion protein linkers: Property, design and functionality. *Adv Drug Deliv Rev*. 2013;65(10):1357-1369.

163. Balakrishnan S, Scheuermann MJ, Zondlo NJ. Arginine mimetics using a-guanidino acids: Introduction of functional groups and stereochemistry adjacent to recognition guanidiniums in peptides. *ChemBioChem*. 2012;13(2):259-270.

164. Wittekind M, Mapelli C, Lee V, et al. Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: Direct refinement against NOEs, J-couplings and 1H and 13C chemical shifts. *J. Mol. Bio.* 1997;264(4):933-952.

165. Charlot, C., Dubois-Pot, H., Serchov, T., Tourrette, Y., Wasylyk, B. A review of post-translational modifications and subcellular localization of ets transcription factors: Possible connection with cancer and involvement in the hypoxic response. *Methods Mol Biol*. 2010;647:3-30.

166. Humphrey SJ, James DE, Mann M. Protein phosphorylation: A major switch mechanism for metabolic regulation. *Trends Endocrinol Metab*. 2015;26(12):676-687.

167. Hunter T. Protein kinases and phosphatases: The yin and yang of protein phosphorylation and signaling. *Cell*. 1995;80(2):225-236.

168. Neurath H, Walsh KA. Role of proteolytic enzymes in biological regulation (a review). *Proc Natl Acad Sci U S A*. 1976;73(11):3825-3832.

169. Spiro RG. Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12(4):43R-56R.

170. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;294(5):1351-1362.

171. Giepmans BN, Adams SR, Ellisman MH, Tsien RY. The fluorescent toolbox for assessing protein location and function. *Science*. 2006;312(5771):217-224.

172. Lawrence DS. Chemical probes of signal-transducing proteins. *Acc Chem Res*. 2003;36(6):401-409.

173. Rothman DM, Shults MD, Imperiali B. Chemical approaches for investigating phosphorylation in signal transduction networks. *Trends Cell Biol*. 2005;15(9):502-510.

174. Ting AY, Kain KH, Klemke RL, Tsien RY. Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *Proc Natl Acad Sci U S A*. 2001;98(26):15003-15008.

175. Vieth M, Higgs RE, Robertson DH, Shapiro M, Gragg EA, Hemmerle H. Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim Biophys Acta*. 2004;1697(1-2):243-257.

176. Wu D, Sylvester JE, Parker LL, Zhou G, Kron SJ. Peptide reporters of kinase activity in whole cell lysates. *Biopolymers*. 2010;94(4):475-486.

177. Ozes ON, Mayo LD, Gustin JA, Pfeffer SR, Pfeffer LM, Donner DB. NF-kappaB activation by tumour necrosis factor requires the akt serine-threonine kinase. *Nature*. 1999;401(6748):82-85.

178. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002;298(5600):1912-1934.

179. Hart GW, Chou T-, Dang CV. C-myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J Biol Chem*. 1995;270(32):18961-18965.

180. Persad, S., Attwell, S., Gray, V., Mawji, N., Deng, J. T., Leung, D., Yan, J., Sanghera, J., Walsh, M. P., Dedhar, S. Regulation of protein kinase B/akt-serine 473 phosphorylation by integrin-linked kinase. *J.Biol.Chem.* 2001;276:27462-27469.

181. Yang, P.-C., Mahmood, T. Western blot: Technique, theory, and trouble shooting. *N Am J Med Sci.* 2012;4(9):429-434.

182. Dephoure N, Gould KL, Gygi SP, Kellogg DR. Mapping and analysis of phosphorylation sites: A quick guide for cell biologists. *Mol Biol Cell*. 2013;24(5):535-542.

183. Espina V, Woodhouse EC, Wulfkuhle J, Asmussen HD, Petricoin EF,3rd, Liotta LA. Protein microarray detection strategies: Focus on direct detection technologies. *J Immunol Methods*. 2004;290(1-2):121-133.

184. Roda A, Guardigli M, Michelini E, Mirasoli M, Pasini P. Peer reviewed: Analytical bioluminescence and chemiluminescence. *Anal Chem*. 2003;75(21):462 A-470 A.

185. Rychlewski L, Kschischo M, Dong L, Schutkowski M, Reimer U. Target specificity analysis of the abl kinase using peptide microarray data. *J Mol Biol*. 2004;336(2):307-311.

186. Xue L, Tao WA. Current technologies to identify protein kinase substrates in high throughput. *Front Biol (Beijing)*. 2013;8(2):216-227.

187. Picotti P. Phosphoproteomics takes it easy. *Nature Biotechnology*. 2015;33:929-930.

188. Engholm-Keller, K., Larsen, M. R. Technologies and challenges in large-scale phosphoproteomics. *Proteomics*. 2013;13(6):910-931.

189. Zondlo, S. C., Gao, F., Zondlo, N. J. Design of an encodable tyrosine kinase-inducible domain: Detection of tyrosine kinase activity by terbium luminescence. *J. Am. Chem. Soc.* 2010;132(16):5619-5621.

190. Devkota, A. K., Kaoud, T. S., Warthaka, M., Dalby, K. N. Fluorescent peptide assays for protein kinases. *Curr Protoc Mol Biol*. 2010;91:18.17.1-18.17.7.

191. Zhang J, Ma Y, Taylor SS, Tsien RY. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc Natl Acad Sci U S A*. 2001;98(26):14997-15002.

192. Kunkel, M. T., Ni, Q., Tsien, R. Y., Zhang, J., Newton, A. C. Spatio-temporal dynamics of protein kinase B/akt signaling revealed by a genetically encoded fluorescent reporter. *Journal of Biological Chemistry*. 2005;280:5581-5587.

193. Balakrishnan, S., Zondlo, N. J. Design of a protein kinase-inducible domain. *J. Am. Chem. Soc.* 2006;128(17):5590-5591.

194. Lewit-Bentley A, Rety S. EF-hand calcium-binding proteins. *Curr Opin Struct Biol*. 2000;10(6):637-643.

195. Franz KJ, Nitz M, Imperiali B. Lanthanide-binding tags as versatile protein coexpression probes. *ChemBioChem*. 2003;4(4):265-271.

196. Selvin PR, Hearst JE. Luminescence energy transfer using a terbium chelate: Improvements on fluorescence energy transfer. *Proc Natl Acad Sci U S A*. 1994;91(21):10024-10028.

197. Stockman BJ. 2-fluoro-ATP as a versatile tool for 19F NMR-based activity screening. *J Am Chem Soc*. 2008;130(18):5870-5871.

198. Knowles JR. Enzyme-catalyzed phosphoryl transfer reactions | annual review of biochemistry. *Ann. Rev. Biochem.* 1980;49:877-919.

199. Baldo JH, Hansen PE, Shriver JW, Sykes BD. 2-fluoro-ATP, a fluorinated ATP analog. 19F nuclear magnetic resonance studies of the 2-fluoro-ADP . myosin subfragment-1 complex. *Can J Biochem Cell Biol*. 1983;61(2-3):115-119.

200. Goldman, J. M., Melo, J. V. Chronic myeloid leukemia. *New England Journal of Medicine*. 2003;349:1451-1464.

201. Bridges AJ. Chemical inhibitors of protein kinases. *Chem Rev*. 2001;101(8):2541-2572.

202. Cohen P. Protein kinases--the major drug targets of the twenty-first century? *Nat Rev Drug Discov*. 2002;1(4):309-315.

203. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive philadelphia chromosome. *Blood*. 1975;45(3):321-334.

204. Salah, E., Ugochukwu, E., Barr, A. J., von Delft, F., Knapp, S., Elkins, J. M. Crystal structures of ABL-related gene (ABL2) in complex with imatinib, tozasertib (VX-680), and a type I inhibitor of the triazole carbothioamide class. *J Med Chem*. 201;54(7):2359-2367.

205. Wilson, C., Agafonov, R. V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L., Kern, D. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science*. 2015;347(6224):882-886.

206. Kim C, Xuong NH, Taylor SS. Crystal structure of a complex between the catalytic and regulatory (RIalpha) subunits of PKA. *Science*. 2005;307(5710):690-696.

207. Taylor SS, Yang J, Wu J, Haste NM, Radzio-Andzelm E, Anand G. PKA: A portrait of protein kinase dynamics. *Biochim Biophys Acta*. 2004;1697(1-2):259-269.

208. Maller JL, Kemp BE, Krebs EG. In vivo phosphorylation of a synthetic peptide substrate of cyclic AMP-dependent protein kinase. *Proc Natl Acad Sci U S A*. 1978;75(1):248-251.

209. Chan, T. O., Rittenhouse, S. E., Tsichlis, P. N. AKT/PKB and other D3 phosphoinositide-regulated kinases: Kinase activation by phosphoinositide-dependent phosphorylation. *Annual Review of Biochemistry*. 1999;68:965-1014.

210. Hers I, Vincent EE, Tavare JM. Akt signalling in health and disease. *Cell Signal*. 2011;23(10):1515-1527.

211. Franke TF. PI3K/akt: Getting it right matters. *Oncogene*. 2008;27(50):6473-6488.

212. Fresno Vara JA, Casado E, de Castro J, Cejas P, Belda-Iniesta C, Gonzalez-Baron M. PI3K/akt signalling pathway and cancer. *Cancer Treat Rev*. 2004;30(2):193-204.

213. Carracedo, A., Pandolfi, P. P. Oncogene - the PTEN-PI3K pathway: Of feedbacks and cross-talks. *Oncogene*. 2008;27:5527-5541.

214. Mackenzie RW, Elliott BT. Akt/PKB activation and insulin signaling: A novel insulin signaling pathway in the treatment of type 2 diabetes. *Diabetes Metab Syndr Obes*. 2014;7:55-64.

215. Xu, J., Wang, P., Yang, H., Zhou, J., Li, Y., Li, X., Xue, W., Yu, C., Tian, Y., Zhu, F. Comparison of FDA approved kinase targets to clinical trial ones: Insights from their system profiles and drug-target interaction networks. *Biomed Res Int*. 2016;2016:1-9.

216. Reddy EP, Smith MJ, and Srinivasan A. Nucleotide sequence of abelson murine leukemia virus genome: Structural similarity of its transforming gene product to other onc gene products with tyrosine-specific kinase activity. *Proc.Natl.Acad.Sci.U.S.A*. 1983;80(12):3623-3627.

217. Rychlewski L, Kschischo M, Dong L, Schutkowski M, Reimer U. Target specificity analysis of the abl kinase using peptide microarray data. *J Mol Biol*. 2004;336(2):307-311.

218. Zondlo NJ. Phosphothreonine is uniquely ordered: Strong induction of structure upon threonine phosphorylation. *J Pept Sci*. 2014;20:S69-S70.

219. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein identification and analysis tools on the ExPASy server - springer. In: Walker J.M., ed. *The proteomics protocols handbook*. Vol 2017. Totowa, NJ: Humana Press; 2005:571-607. Accessed 1/29/2017.

220. Chakrabarti P, Pal D. The interrelationships of side-chain and main-chain conformations in proteins. *Progress in Biophysics and Molecular Biology*. 2001;76(1):1-102.

221. Ho BK, Coutsias EA, Seok C, Dill KA. The flexibility in the proline ring couples to the protein backbone. *Protein Sci*. 2005;14(4):1011-1018.

222. Obata T, Yaffe MB, Leparc GG, et al. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J Biol Chem*. 2000;275(46):36108-36115.

223. Pearce LR, Komander D, Alessi DR. The nuts and bolts of AGC protein kinases. *Nature Reviews Molecular Cell Biology*. 2010;11:9-22.

224. Madhusudan, Trafny, E. A., Xuong, N.-H., Adams, J. A., Eyck, L. F.Ten., Taylor, S. S. and Sowadski, J. M. cAMP-dependent protein kinase: Crystallographic

insights into substrate recognition and phosphotransfer. *Protein Science*. 1994;3(2):176-187.

225. Lippa B, Pan G, Corbett M, et al. Synthesis and structure based optimization of novel akt inhibitors. *Bioorg Med Chem Lett*. 2008;18(11):3359-3363.

226. Tenenbaum HC. Levamisole and inorganic pyrophosphate inhibit beta-glycerophosphate induced mineralization of bone formed in vitro. *Bone Miner*. 1987;3(1):13-26.

227. Morita A, Zhu J, Suzuki N, et al. Sodium orthovanadate suppresses DNA damage-induced caspase activation and apoptosis by inactivating p53. *Cell Death Differ*. 2006;13(3):499-511.

228. Lochner A, Moolman JA. The many faces of H89: A review. *Cardiovasc Drug Rev*. 2006;24(3-4):261-274.

229. Engh RA, Girod A, Kinzel V, Huber R, and Bossemeyer D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89. *J.Biol.Chem.* 1996;271(42):26157-26164.

230. Rust, H. L., Thompson, P. R. Kinase consensus sequences – A breeding ground for crosstalk. *Chemical Biology*. 2011;6(9):881-892.

231. James TL. Phosphorus-31 NMR as a probe for phosphoproteins. *CRC Crit Rev Biochem*. 1985;18(1):1-30.

232. Mak A, Smillie LB, Barany M. Specific phosphorylation at serine-283 of alpha tropomyosin from frog skeletal and rabbit skeletal and cardiac muscle. *Proc Natl Acad Sci U S A*. 1978;75(8):3588-3592.

233. Matheis G, Whitaker JR. 31P NMR chemical shifts of phosphate covalently bound to proteins. *Int J Biochem*. 1984;16(8):867-873.

234. Freedberg DI, Selenko P. Live cell NMR. *Annual Review of Biophysics*. 2014;43:177-192.

235. Theillet FX, Smet-Nocca C, Liokatis S, et al. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR*. 2012;54(3):217-236.

236. Srinivas M, Heerschap A, Ahrens ET, Figdor CG, de Vries IJ. (19)F MRI for quantitative in vivo cell tracking. *Trends Biotechnol*. 2010;28(7):363-370.

237. Dürr UHN, Grage SL, Witter R, Ulrich AS. Solid state 19F NMR parameters of fluorine-labeled amino acids. part I: Aromatic substituents. *Journal of Magnetic Resonance*. 2008;191(1):7-15.

238. Takaoka, Y., Kiminami, K., Mizusawa, K., Matsuo, K., Narazaki, M., Matsuda, T., Hamachi, I. Systematic study of protein detection mechanism of self-assembling

19F NMR/MRI nanoprobes toward rational design and improved sensitivity. *J. Am. Chem. Soc.* 2011;133(30):11725-11731.

239. Takaoka, Y., Sun, Y., Tsukiji, S., Hamachi, I. Mechanisms of chemical protein 19F-labeling and NMR-based biosensor construction in vitro and in cells using self-assembling ligand-directed tosylate compounds. *Chem. Sci.* 2011;2:511-520.

240. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S., & Smith, J. W. Substrate hydrolysis by matrix metalloproteinase-9. *Journal of Biological Chemistry*. 2001;276:20578-20578.

241. Andersen J. Neurodegeneration: Oxidative stress in neurodegeneration: Cause or consequence? *Nature Reviews Neuroscience*. 2004;5:18-25.

242. Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. *Nature*. 2000;408(6809):239-247.

243. Maritim AC, Sanders RA, Watkins JB. Diabetes, oxidative stress, and antioxidants: A review. *J Biochem Mol Toxicol*. 2003;17(1):24 38.

244. Radi R. Protein tyrosine nitration: Biochemical mechanisms and structural basis of functional effects. *Acc Chem Res*. 2013;46(2):550-559.

245. Ischiropoulosm H. Biological tyrosine nitration: A pathophysiological function of nitric oxide and reactive oxygen species. *Archives of Biochemistry and Biophysics*. 1998;356(1):1-11.

246. Dugave C, Demange L. Cis-trans isomerization of organic molecules and biomolecules: Implications and applications. *Chem Rev*. 2003;103(7):2475-2532.

247. Fischer G, Heins J, Barth A. The conformation around the peptide bond between the P1- and P2-positions is important for catalytic activity of some proline-specific proteases. *Biochim Biophys Acta*. 1983;742(3):452-462.

248. Horng J-. Stereoelectronic effects on polyproline conformation. *Protein Science*. 2006;15(1):74 83.

249. Lu Z, Hunter T. Prolyl isomerase Pin1 in cancer. *Cell Res*. 2014;24(9):1033-1049.

250. MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. *J Mol Biol*. 1991;218(2):397-412.

251. Moradi M, Babin V, Roland C, Darden TA, Sagui C. Conformations and free energy landscapes of polyproline peptides. *Proc Natl Acad Sci U S A*. 2009;106(49):20746-20751.

252. Shen Y, Bax A. Prediction of xaa-pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR*. 2010;46(3):199-204.

253. Thomas KM, Naduthambi D, Tririya G, Zondlo NJ. Proline editing: A divergent strategy for the synthesis of conformationally diverse peptides. *Org Lett*. 2005;7(12):2397-2400.

254. Torbeev V.Y. HD. Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate ß2-microglobulin amyloid assembly. *Proc Natl Acad Sci*. 2013;110(50):20051-20056.

255. Wedemeyer WJ, Welker E, Scheraga HA. Proline cis-trans isomerization and protein folding. *Biochemistry*. 2002;41(50):14637-14644.

256. Ramachandran GN, Mitra AK. An explanation for the rare occurrence of cis peptide units in proteins and polypeptides. *Journal of Molecular Biology*. 1976;107(1):85-92.

257. Bartlett, G. J., Choudhary, A., Raines, R. T., Woolfson, D. N. N-->pi* interactions in proteins. *Nat Chem Biol*. 2010;6(8):615-620.

258. Pal D, Chakrabarti P. Cis peptide bonds in proteins: Residues involved, their conformations, interactions and locations. *J Mol Biol*. 1999;294(1):271-288.

259. Renner C, Alefelder S, Bae JH, Budisa N, Huber R, Moroder L. Fluoroprolines as tools for protein design and engineering. *Angew Chem Int Ed Engl*. 2001;40(5):923-925.

260. Schmid FX. Prolyl isomerase: Enzymatic catalysis of slow protein-folding reactions. *Annual Review of Biophysics and Biomolecular Structure*. 1993;22:123-143.

261. Stewart DE, Sarkar A, Wampler JE. Occurrence and role of cis peptide bonds in protein structures. *J Mol Biol*. 1990;214(1):253-260.

262. Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. *Current Opinion in Structural Biology*. 2004;14(1):76-88.

263. Fischer G. Peptidyl-prolyl cis/trans isomerases and their effectors. *Angewandte Chemie International Edition*. 1994;33(14):1415-1436.

264. Shaw PE. Peptidyl-prolyl isomerases: A new twist to transcription. *EMBO Rep*. 2002;3(6):521-526.

265. Boutros R, Lobjois V, Ducommun B. CDC25 phosphatases in cancer cells: Key players? good targets? *Nature Reviews Cancer*. 2007;7(7):495-507.

266. Cowan PM, McGavin S. Structure of poly-L-proline. *Nature*. 1955;176(4480):501-503.

267. Rucker, A. L., Creamer, T. P. Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Sci*. 2002;11(4):980-985.

268. Shoulders, M. D., Kamer, K. J., Raines, R. T. Origin of the stability conferred upon collagen by fluorination. *Bioorg Med Chem Lett.* 2009;19(14):3859-3862.

269. Blout ER, Carver JP, Gross J. Intrinsic cotton effects in collagen and poly-L-proline. *J Am Chem Soc*. 1963;85(5):644 646.

270. Burger K, Rudolph M, Fehn S, Sewald N. Synthesis of (2S)-4,4-difluoroproline, (2S,4R)-4-fluoroproline and their derivatives from (S)-aspartic acid [1]. *J Fluorine Chem*. 1994;66(1):87 90.

271. Engel-Andreasen J, Wich K, Laursen JS, Harris P, Olsen CA. Effects of thionation and fluorination on cis-trans isomerization in tertiary amides: An investigation of N-alkylglycine (peptoid) rotamers. *J Org Chem*. 2015;80(11):5415-5427.

272. Kim W, Hardcastle KI, Conticello VP. Fluoroproline flip-flop: Regiochemical reversal of a stereoelectronic effect on peptide and protein structures. *Angewandte Chemie International Edition*. 2006;45(48):8141-8145.

273. Lin L, Brandts JF. Kinetic mechanism for conformational transitions between poly-L-prolines I and II: A study utilizing the cis-trans specificity of a proline-specific protease. *Biochemistry*. 1980;19(13):3055-3059.

274. Woody R. Circular dichroism spectrum of peptides in the poly(pro)II conformation. *J. Am. Chem. Soc.* 2009;131(23):8234-8245.

275. 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.

276. Harrington WF, Sela M. Studies on the structure of poly-L-proline in solution. *Biochim Biophys Acta*. 1958;27(1):24-41.

277. Dorman DE, Torchia DA, Bovey FA. Carbon-13 and proton nuclear magnetic resonance observations of the conformation of poly(L-proline) in aqueous salt solutions. *Macromolecules*. 1973;6(1):80-82.

278. Dorman DE, Bovey FA. Carbon-13 magnetic resonance spectroscopy. spectrum of proline in oligopeptides. *J Org Chem*. 1973;38(13):2379-2383.

279. Troganis A., Gerothanassis I.P., Athanassiou Z., Mavromoustakos T., Hawkes G.E., Sakarellos C. Thermodynamic origin of cis/trans isomers of a proline-containing beta-turn model dipeptide in aqueous solution: A combined variable temperature 1H-NMR, two-dimensional 1H,1H gradient enhanced nuclear overhauser effect spectroscopy (NOESY), one-dimensional steady-state intermolecular 13C,1H NOE, and molecular dynamics study. *Biopolymers*. 2000;53(1):72-83.

280. Adzhubei AA, Sternberg MJ, Makarov AA. Polyproline-II helix in proteins: Structure and function. *J Mol Biol*. 2013;425(12):2100-2132.

281. Chebrek R, Leonard S, de Brevern AG, Gelly J-. PolyprOnline: Polyproline helix II and secondary structure assignment database. *Database*. 2014;2014(0):bau102 bau102.

282. Applequist J. Theoretical p-p* absorption and circular dichroic spectra of helical poly(L-proline) forms I and II. *Biopolymers*. 1981;20(11):2311-2322.

283. Lin LN, Brandts JF. Evidence suggesting that some proteolytic enzymes may cleave only the trans form of the peptide bond. *Biochemistry*. 1979;18(1):43-47.

284. Gornick F, Mandelkern L, Diorio AF, Roberts DE. Evidence for a cooperative intramolecular transition in poly-L-proline. *J Am Chem Soc*. 1964;86(13):2549-2555.

285. Holzwarth G, Chandrasekaran R. Cooperativity in poly-L-proline 1-11 transitions. *Macromolecules*. 1969;2(3):245-250.

286. Beausoleil E. LWD. An examination of the steric effects of 5-tert-butylproline on the conformation of polyproline and the cooperative nature of type II to type I helical interconversion. *Biopolymers*. 2000;53(3):249-256.

287. Renner, C., Alefelder, S., Bae, J. H., Budisa, N., Huber, R., Moroder, L. Fluoroprolines as tools for protein design and engineering. *Angew.Chem.Int.Ed Engl.* 2001;40(5):923-925.

288. Shoulders, M. D., Kamer, K. J., Raines, R. T. Origin of the stability conferred upon collagen by fluorination. *Bioorg Med Chem Lett.* 2009;19(14):3859-3862.

289. Demange L, Ménez A, Dugave C. Practical synthesis of boc and fmoc protected 4-fluoro and 4-difluoroprolines from trans-4-hydroxyproline. *Tetrahedron Lett*. 1998;39(10):1169 1172.

290. Tressler CM, Zondlo NJ. (2S,4R)- and (2S,4S)-perfluoro-tert-butyl 4-hydroxyproline: Two conformationally distinct proline amino acids for sensitive application in 19F NMR. *J Org Chem*. 2014;79(12):5880-5886.

291. Kim W, George A, Evans M, Conticello VP. Cotranslational incorporation of a structurally diverse series of proline analogues in an escherichia coli expression system. *Chembiochem*. 2004;5(7):928-936.

292. Kim, W., McMillan, R. A., Snyder, J. P., Conticello, V. P. A stereoelectronic effect on turn formation due to proline substitution in elastin-mimetic polypeptides. *J. Am. Chem. Soc.* 2005;127(51):18121-18132.

293. Makarov AA, Protasevich II, Lobachov VM, et al. Thermostability of the barnase-barstar complex. *FEBS Lett*. 1994;354(3):251 254.

Appendix A

RIGHTS AND PERMISSIONS

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Synthesis of Perfluoro-tert-butyl Tyrosine, for Application in 19F NMR, via a Diazonium-Coupling Reaction

Author: Caitlin M. Tressler, Neal J. Zondlo

Publication: Organic Letters

Publisher: American Chemical Society

Date: Dec 1, 2016

Copyright © 2016, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

BACK | CLOSE WINDOW

This ACS article is provided to You under the terms of this Standard ACS *AuthorChoice/Editors' Choice* usage agreement between You and the American Chemical Society ("ACS"), a federally-chartered nonprofit located at 1155 16th Street NW, Washington DC 20036. Your access and use of this ACS article means that you have accepted and agreed to the Terms and Conditions of this Agreement. ACS and You are collectively referred to in this Agreement as "the Parties").

5.1.1 1. SCOPE OF GRANT
ACS grants You non-exclusive and nontransferable permission to access and use this ACS article subject to the terms and conditions set forth in this Agreement.

5.1.2 2. PERMITTED USES
a. For non-commercial research and education purposes only, You may access, download, copy, display and redistribute articles as well as adapt, translate, text and data mine content contained in articles, subject to the following conditions:

i. The authors' moral right to the integrity of their work under the Berne Convention (Article 6bis) is not compromised.

ii. Where content in the article is identified as belonging to a third party, it is your responsibility to ensure that any reuse complies with copyright policies of the owner.

iii. Copyright notices or the display of unique Digital Object Identifiers (DOI's), ACS or journal logos, bibliographic (e.g. authors, journal, article title, volume, issue, page numbers) or other references to ACS journal titles, web links, and any other journal-specific "branding" or notices that are included in the article or that are provided by the ACS with instructions that such should accompany its display, should not be removed or tampered with in any way. The display of ACS AuthorChoice or *ACS Editors' Choice* articles on non-ACS websites must be accompanied by prominently displayed links to the definitive published versions of those articles on the ACS website.

iv. Any adaptations for non-commercial purposes must prominently link to the definitive published version on the ACS website and prominently display the statement: "This is an unofficial adaptation of an article that appeared in an ACS publication. ACS has not endorsed the content of this adaptation or the context of its use."

v. Any translations for non-commercial purposes, for which a prior translation agreement with ACS has not been established, must prominently link to the definitive published version on the

ACS website and prominently display the statement: "This is an unofficial translation of an article that appeared in an ACS publication. ACS has not endorsed the content of this translation or the context of its use."

b. Each time You distribute this ACS article or an adaptation, ACS offers to the recipient a license to this ACS article on the same terms and conditions as the license granted to You under this License.

c. For permission to use ACS copyrighted articles beyond that permitted here, visit: http://pubs.acs.org/copyright/permissions.html

5.1.3 3. PROHIBITED USES

a. Use of this ACS article for commercial purposes is prohibited. Examples of such prohibited commercial purposes include but are not limited to:

i. Copying or downloading of articles, or linking to such postings, for further distribution, sale or licensing, for a fee;

ii. Copying, downloading or posting by a site or service that incorporates advertising with such content;

iii. The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee;

iv. Use of articles or article content (other than normal quotations with appropriate citation) by a for-profit organizations for promotional purposes, whether for a fee or otherwise;

v. Sale of translated versions of the article that have not been authorized by license or other permission from the ACS

5.1.4 4. TERMINATION

ACS reserves the right to limit, suspend, or terminate your access to and use of the ACS Publications Division website and/or all ACS articles immediately upon detecting a breach of this License.

5.1.5 5. COPYRIGHTS; OTHER INTELLECTUAL PROPERTY RIGHTS

Except as otherwise specifically noted, ACS is the owner of all right, title and interest in the content of this ACS article, including, without limitations, graphs, charts, tables illustrations, and copyrightable supporting information. This ACS article is protected under the Copyright Laws of the United States Codified in Title 17 of the U.S. Code and subject to the Universal Copyright Convention and the Berne Copyright Convention. You agree not to remove or obscure copyright notices. You acknowledge that You have no claim to ownership of any part of this ACS article or other proprietary information accessed under this Agreement.

The names "American Chemical Society," "ACS" and the titles of the journals and other ACS products are trademarks of ACS.

5.1.6 6. DISCLAIMER OF WARRANTIES; LIMITATION OF LIABILITY

ACS warrants that it is entitled to grant this Agreement.

EXCEPT AS SET FORTH IN THE PRECEDING SENTENCE, ACS MAKES NO WARRANTY OR REPRESENTATION OF ANY KIND, EXPRESS OR IMPLIED, WITH RESPECT TO THIS ACS ARTICLE INCLUDING, BUT NOT LIMITED TO WARRANTIES AS TO THE ACCURACY OR COMPLETENESS OF THE ACS ARTICLE, ITS QUALITY, ORIGINALITY, SUITABILITY, SEARCHABILITY, OPERATION, PERFORMANCE, COMPLIANCE WITH ANY COMPUTATIONAL PROCESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

ACS SHALL NOT BE LIABLE FOR: EXEMPLARY, SPECIAL, INDIRECT, INCIDENTAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF OR IN CONNECTION WITH THE AGREEMENT GRANTED HEREUNDER, THE USE OR INABILITY TO USE ANY ACS PRODUCT, ACS'S PERFORMANCE UNDER THIS AGREEMENT, TERMINATION OF THIS AGREEMENT BY ACS OR THE LOSS OF DATA, BUSINESS OR GOODWILL EVEN IF ACS IS ADVISED OR AWARE OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL THE TOTAL AGGREGATE LIABILITY OF ACS OUT OF ANY BREACH OR TERMINATION OF THIS AGREEMENT EXCEED THE TOTAL AMOUNT PAID BY YOU TO ACS FOR ACCESS TO THIS ACS ARTICLE FOR THE CURRENT YEAR IN WHICH SUCH CLAIM, LOSS OR DAMAGE OCCURRED, WHETHER IN CONTRACT, TORT OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, DUE TO NEGLIGENCE.

The foregoing limitations and exclusions of certain damages shall apply regardless of the success or effectiveness of other remedies. No claim may be made against ACS unless suit is filed within one (1) year after the event giving rise to the claim.

5.1.7 7. GENERAL

This Agreement sets forth the entire understanding of the Parties. The validity, construction and performance of this Agreement shall be governed by and construed in accordance with the laws of the District of Columbia, USA without reference to its conflicts of laws principles. You acknowledge that the delivery of the ACS article will occur in the District of Columbia, USA. You shall pay any taxes lawfully due from it, other than taxes on ACS's net income, arising out of your use of this ACS article and/or other rights granted under this Agreement. You may not assign or transfer its rights under this Agreement without the express written consent of ACS.

5.1.8 8. ACCEPTANCE

You warrant that You have read, understand, and accept the terms and conditions of this Agreement. ACS reserves the right to modify this Agreement at any time by posting the modified terms and conditions on the ACS Publications Web site. Any use of this ACS article after such posting shall constitute acceptance of the terms and conditions as modified.