PROGRAMMING STRAND DISPLACEMENT CIRCUITS FOR DYNAMIC PROTEIN ASSEMBLY IN CANCER CELLS

by

Rebecca P. Chen

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 Chemical Engineering

Winter 2019

PROGRAMMING STRAND DISPLACEMENT CIRCUITS FOR DYNAMIC PROTEIN ASSEMBLY IN CANCER CELLS

by

Rebecca P. Chen

Approved: _____
Eric Furst, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Levi T. Thompson, Ph.D.
Dean of the College of Engineering

Approved: _____
Douglas J. Doren, Ph.D.
Interim 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: _____

Wilfred Chen, 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: _____

Millicent O. Sullivan, 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: _____

Maciek R. Antoniewicz, 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: _____

Brian J. Bahnson, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would first and foremost like to acknowledge my advisor, Professor Wilfred Chen. During this journey, he has always been available to offer support and guidance. His boundless enthusiasm for his work is contagious and has spurred my own passion for this field. Despite his odd motivational speeches, they were truly effective in helping me power through hurdles. Through his teaching, I have grown tremendously as a scientist and as a person. I am truly grateful to have found a mentor and friend that I will value for life.

I would like to thank Professor Millicent Sullivan, Professor Maciek Antoniewicz, and Professor Brian Bahnson for serving on my dissertation committee and providing constructive feedback.

Next, I would like to express my deepest gratitude to former and current members of the Chen group. We have a great group culture that is maintained and contributed to by every single member. Dr. Qing Sun, Dr. Heejae Kim, Dr. Daniel Blackstock, Dr. Andrew Swartz, and Dr. Miso Park are past members who have all given me great advice and guidance. I would also like to thank my current group members, Kay Siu, Emily Hartzell, Andrew Gaynor, Emily Berckman, Rachel Lieser, Alex Mitkas, Victoria Hunt, Daniel Yur, and Hopen Yang. I have valued getting to know each of you and forming great friendships. It has been my pleasure blowing up your phones with memes. I am especially thankful for my officemates Andrew, Emily, and Kay for all the insightful conversations about science and life over the years. Thank you, Andrew, for philosophical discussions and free magnets. Thank you,

Emily, for baking me those great tobin dog cupcakes. Lastly, I would most like to thank Kay Siu, who joined the lab at the same time six years ago. We have gotten to know each other so well, and I deeply appreciate your support and friendship all these years.

Finally, I would like to acknowledge my family for their constant love and backing. I want to thank my sister, Samantha, for all the inside jokes, Fitbit travel adventures, and winter binge fests. To my parents, Dennis and Jennifer, I want to express my deepest gratitude for always finding ways to support and take care of me, even from the other side of the country. Thank you for all the care packages full of food and goodies. Family visits and vacations have given me great joy over the years.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Nature has evolved biological systems to exist as highly intricate and dynamic networks, which is exemplified in complex diseases like cancer. To tackle this multifaceted disease, there has been a strong push to develop "smart therapeutics" that can match these complexities. Despite new targeted approaches, there are still challenges in developing multi-input responsive therapeutics. This dissertation addresses these challenges by building towards a therapeutic computing device that utilizes programmable nucleic acid circuits to control protein-based therapeutic action. A flexible platform technology was established to harness toehold-mediated strand displacement for dynamic protein assembly. Key aspects for realizing this platform as a novel class of smart therapeutics are explored in this thesis.

First, the foundation was laid by synthesizing protein-DNA conjugates to be tested within strand displacement circuits. We showed that DNA strand displacement can be used to dynamically control the spatial proximity and corresponding fluorescence resonance energy transfer (FRET) between two fluorescent proteins with multi-input, reversible, and amplification architectures. Next, the power and utility of this technology as a synthetic computing platform was demonstrated by driving the dynamic reconstitution of a split enzyme for targeted prodrug activation based on the sensing of cancer-specific miRNAs.

Subsequently, we addressed the major bottleneck that lays in using sequence constrained biological inputs to run *de novo* circuits. A novel strategy called associative strand displacement was developed to elegantly interface miRNA inputs

with synthetic components. This sequence decoupling allows any miRNA sequence to be targeted without compromising function or efficiency of circuits that have been optimized *de novo*. We applied the design principles of our strategy towards integrating Boolean logic and amplification architectures, as well as creating a four-input miRNA classifier.

Lastly, to further prove the feasibility of our technology as a therapeutic, we implemented a genetically-encoded hybrid device inside live HeLa cells. Our strategy uniquely utilizes Cas6 endoribonucleases for their picomolar binding affinity and cleavage activity to drive self-assembly of protein and RNA device components. This Cas6-guided approach allowed protein assembly and disassembly to be controlled by RNA hybridization and strand displacement, respectively. These promising results are important stepping stones that support the future execution of more complex architectures with therapeutic outputs.

Ultimately, our technology shows the powerful utility of combing nucleic acids and proteins into hybrid devices, especially when toehold-mediated strand displacement is used to generate computing power for dynamic behavior. Beyond disease therapeutics, this technology has widespread applicability and can be expanded to generate synthetic programmable protein switches for any biological system of interest.

Chapter 1

INTRODUCTION

1.1 Programming biological switches for smart therapeutics

Nature has evolved biological systems to exist as highly complex and dynamic networks[1, 2]. Through biomolecular sensors and switches, cellular organisms can respond quickly to changing environmental cues[3-5]. In the 20th century, scientists mainly took a reductionist approach to understanding these biological systems, breaking down the cell into its individual parts and studying each entity one at a time. More recently, technological advances in the areas genomics and proteomics have led to a significant paradigm shift and the emergence of new fields such a systems biology and bioinformatics[6, 7]. The ability to collect large wealth of biological data has made abundantly clear the enormous complexity and intricate connectivity of biological networks that make up living beings.

One prevalent human disease that exemplifies this level of complexity is cancer. To tackle this multifaceted disease, there has been a strong push to develop "smart therapeutics" that can match these complexities towards the ultimate goal of personalized medicine[8]. This began by moving beyond small molecule-based chemotherapeutics and developing targeted biologics. Presently, monoclonal antibodies have attained huge commercial success and have become a standard treatment option. Even "smarter" therapies lay on the horizon, such as T-cell therapy[9], oncolytic viruses[10], and cancer vaccines[11, 12]. However, these targeted therapies generally look to a single overexpressed surface marker, and while they offer great

improvements over traditional chemotherapeutics, they still do not utilize the full wealth of information we now have for cancer. This includes the knowledge that multiple biomarkers must often be considered to properly identify a cell as being cancerous. Moving forward there is a need to develop multi-input responsive "smart therapeutics" to further reduce off target effects.

The development of smart therapeutics is a challenge well suited to synthetic biology. Lying at the intersection of reductionist and systems approaches, the powerful toolbox of biological components is harnessed and adapted to design artificial biological networks and devices with increasing complexity[3, 13]. Towards creating a smart therapeutic, we envision a biological computing device in which multiple endogenous inputs are processed for decision-making towards a therapeutic output[14, 15]. Only when multiple disease indicators are met is the therapeutic device turned on for effective cell killing.

1.2 Nucleic acid-protein hybrid computing device

A biological computing device is comprised of three functional components: input sensing, computation and decision making, and output actuation (**Figure 1.1**)[15]. This dissertation sets out to demonstrate that a nucleic acid-protein hybrid device allows each functional component of the smart therapeutic to be maximized. In our device, nucleic acids are utilized for sensing and computation, and proteins execute the therapeutic output. Nucleic acids have a high degree of programmability, while proteins have diverse and useful functionality; each cannot be matched by the other in their respective advantages. As such, their combined power offers advantages non-hybrid strategies cannot.

The simple and predictable nature of base pairing makes nucleic acids a highly programmable class of biomolecules. By changing their sequence, they can be easily designed to target endogenous nucleic acid biomarkers of interest[16] (mRNA, microRNA, small nuclear RNA, and other types of non-coding RNA) with high specificity and modularity. Furthermore, the high affinity of hybridization (binding) allows for high detection sensitivity. In fact, this simple premise has led DNA microarrays to become the gold standard for high-throughput global expression profiling[17]. Furthermore, aptamer technology allows nucleic acids to be utilized for sensing other types of analytes as well, such as proteins and small molecules[18]. Even more importantly, nucleic acids have been emerging as a powerful medium for biological computation[19-21]. Since first pioneered in 2000 for DNA nanotechnology by Yurke et al.[22], the process of toehold-mediated strand displacement has revolutionized how nucleic acids can be programmed for reaction pathways exhibiting complex logic behavior (detailed in **Section 1.3**). However, nucleic acids do not readily serve as useful output modules, because as standalone entities they do not possess any native catalytic functionality. One exception is a class of nucleic acids called deoxyribozymes[23, 24] and ribozymes[24-26]. Still, the reactions these catalytic nucleic acids carry out are limited in substrate (primarily ligating and cleaving oligonucleotides), thus hindering the scope of their application as output modules.

On the other hand, the same 20 amino acid code that hinders their programmability enables proteins to fold into three-dimensional structures that exhibit a vast array of behavior and catalytic activities. More importantly, there are many techniques for engineering proteins to exhibit switch-like behavior. A particularly successful approach has been to assemble proteins onto a scaffold as a proximity-

based toggle to activate proteins or to redirect protein function[27, 28]. Through a combination of naturally existing parts and engineering techniques to elevate these parts, proteins serve as a treasure trove for output modules.

Overall, this is not too different than how nature has chosen to organize itself. In general, nucleic acids are responsible for information storage and transfer, while proteins are the primary actors of cellular processes. Furthermore, the existence of ribonucleoproteins, such as ribosomes and spliceosomes, demonstrate nature's evolved coordination between RNA and protein to form complex nanomachinery for accomplishing cellular tasks[29]. In this thesis, we explore how to design and build artificial nucleic acid-protein hybrid devices to achieve novel therapeutic functionality. Towards this goal, we utilized both approaches of synthetically producing and genetically-encoding such a hybrid computing device.

1.3 Toehold-mediated strand displacement

Traditionally, nucleic acid hybridization is thought to result in a static complex, whose interaction is disrupted through enzyme-dependent unwinding or thermal denaturation. However, enzyme-free unwinding is possible through a process called toehold-mediated strand displacement, in which two nucleic acid strands hybridize to each other to displace a previously hybridized strand (**Figure 1.2**). It is initiated at complementary single-stranded domains called toeholds. Upon toehold binding, a random walk process called branch migration results in the displacement of the incumbent strand by the invader strand[30]. Ultimately, there is the formation of a new duplex and a new single-stranded output. Strand displacement is thermodynamically driven by free energy enthalpic gains of base-pair hybridization and entropic gains of strand release. By varying toehold length and composition,

strand displacement rate constants can be modulated by over a factor of 10^6. Longer toeholds and higher GC composition of toeholds favor stronger toehold binding, and thus faster strand displacement[30]. The sequence-specific nature of hybridization also means that orthogonal strand displacement reactions can be run in parallel.

Strand displacement cascades can be built by using the output of a strand displacement reaction as the input of a downstream reaction. By layering and linking reactions, autonomous computation circuits can be rationally designed to exhibit a variety of behaviors that resemble digital circuitry, such as Boolean-logic, amplification, thresholding, recycling, and fan-in/out[19, 21, 31]. Strand displacement circuits have been successfully scaled up to perform highly complex computation. For example, Qian et al. built a four-bit square root circuit comprised of 130 unique strands[32]. The same group also went on to build an artificial neural network capable of associative memory[33]. Besides species-based reaction pathways, strand displacement can also be integrated into previously static nucleic acid nanostructures (scaffolds[34], tweezers[22], origami[35], etc.) to give them dynamic behavior.

Throughout this thesis, strand displacement reactions and cascades are schematically depicted with the following conventions. Nucleic acid strands are represented by lines with arrows on the 3' end. Functional domains are grouped by color and denoted with numbers or letters. Starred domains are used to identify complementarity. For example, domain a* is complementary to domain a. Hybridization is indicated by grey shading or black dots between the strands.

1.3.1 Catalytic Hairpin Assembly (CHA)

One of the most innovative strategies for building strand displacement architecture is the use of kinetically trapped hairpins in place of multi-strand

complexes[21, 36, 37]. Toeholds are sequestered and inactivated in the stem or loop regions until strand displacement is used to open the hairpin. In particular, a scheme called catalytic hairpin assembly (CHA) will be utilized throughout this thesis (**Figure 1.3**). This circuit is comprised of two complementary DNA hairpins (H1 and H2) which initially cannot interact due to their hairpin structures. The addition of an input catalyst strand opens H1 through toehold-mediated strand displacement, exposing a previously sequestered toehold, which can go on to open H2. When H2 is opened, not only does it hybridize to H1, but it simultaneously displaces the catalyst strand. The regenerated catalyst strand goes on to initiate further hairpin assembly.

First developed by Yin et al.[36] and then further optimized by Jung et al.[38], CHA was shown to achieve 50- to 100-fold signal amplification. Furthermore, by layering two CHA circuits signal amplification was pushed to 7000-fold[39]. The amplification properties of CHA are particularly attractive for the purposes of disease application where endogenous target inputs may be present at low concentrations[21]. In Chapter 2, a strategy is developed to use protein-DNA conjugates within CHA. In Chapter 3, a novel method is established to expand CHA with multi-input logic behavior and to interface optimized *de novo* CHA with sequence constrained miRNA targets.

1.4 Nucleic acid to protein attachment

The efficiency of nucleic acid to protein attachment is crucial for maximizing their combined functionality in a hybrid device. This linkage serves as the decisive transduction juncture between nucleic acid-based input processing and computation to protein-dependent output actuation. The general strategies used to link nucleic acids to protein are covalent attachment or affinity interaction, and both approaches are taken in this thesis for their respective benefits (**Figure 1.4**).

Covalent attachment offers the stability of a permanent connection between nucleic acid and protein. Our group has previously developed a method to produce and purify protein-DNA conjugates through the use of HaloTag technology[40, 41]. HaloTag is a mutant dehalogenase that has been engineered to bind irreversibly to chloroalkane ligands[42]. By chemically modifying DNA oligos with chloroalkane ligands, any DNA strand of interest can be conjugated to any HaloTag fusion protein of interest (**Figure 1.4**). This highly specific reaction occurs rapidly under mild physiological conditions with high efficiency. Compared to traditional methods of reacting with native residues[43, 44], our strategy avoids chemical modifications to the proteins themselves and allows consistent single-site labeling, thus preserving protein activity and providing greater control and uniformity of protein-DNA labeling orientation. Furthermore, the inclusion of an elastin-like polypeptide (ELP) purification tag allows for highly pure protein-DNA conjugates to be obtained through inverse transition cycling (ITC)[45]. In Chapter 2, HaloTag conjugation is utilized to covalently attach DNA to fluorescent reporter proteins and a split prodrug activating enzyme.

It is fundamentally impossible for affinity interactions to match the stability of a covalent bond. However, current covalent attachment strategies involve synthetic conjugation that requires *in vitro* processing steps that are not feasible for genetically-encoded systems. Instead, affinity interactions between nucleic acids and proteins must be exploited. To best approach the stability of covalent attachment, high binding affinity between the protein and nucleic acid is key. A variety of DNA and RNA binding proteins have been utilized for protein assembly on nucleic acid scaffolds[28]. Zinc finger proteins and PUF family proteins can be engineered for binding to different DNA and RNA sequences, respectively, with micromolar to nanomolar range

affinity[46, 47]. Meanwhile, other binding proteins recognize sequence-specific substrates. For example, coat proteins from RNA bacteriophages have been a serviceable source of RNA binders of short hairpin sequences. The most commonly used MS2 and PP7 coat proteins bind their respective RNA substrates with low nanomolar affinity[48, 49]. However, this thesis chooses to shift towards an emerging family of endoribonucleases called Cas6 that are part of Type I CRISPR/Cas systems[50]. Cas6 proteins bind specific RNA loop sequences with picomolar dissociation constants, making them one of the strongest RNA binders discovered to date[51]. This allows for highly efficient Cas6-RNA binding to best mimic covalent attachment (**Figure 1.4**). Cas6 proteins also exhibit single-turnover cleavage of their RNA substrate. Meaning, after cleaving their RNA substrate at a specific 3' location, Cas6 remain bound to the cleavage product and do not go on to cleave other substrates. We explore this cleavage function as way to control the generation of discrete RNA strands. In Chapter 4, a novel demonstration of Cas6-guided protein assembly on a RNA scaffold is executed inside mammalian cells.

1.5 Nucleic acid-based scaffolds for protein assembly and actuation

Nucleic acids lend themselves towards being attractive building materials for synthetic protein assembly scaffolds, since they take on predictable structures that allow for the precise spatial organization of proteins to be easily programmed and modulated[28]. Whether through the use of DNA conjugation for hybridization-based docking or the use of protein affinity-based binding to the scaffold, multiple demonstrations have proven that these scaffolds offer proximity-based enhancements in a variety of protein systems[28], including FRET between fluorescent proteins[52], product formation in enzyme cascades[49, 53], and reconstitution of split proteins[49].

For building a therapeutic device, the proteins to be assembled on the scaffold must be chosen carefully. Minimal leakiness is crucial to preventing nonspecific activation of the therapy. However, upon activation, the output must be strong enough for effective treatment. Prodrug activating enzymes are promising therapeutic outputs that convert non-toxic prodrugs into their toxic form[54, 55]. The catalytic turnover of prodrug conversion makes for strong output amplification. Efforts are being made to engineer these enzymes for conditional activity towards cancer targeting. For example, Ear et al. created a split version of yeast cytosine deaminase (yCD), a prodrug activating enzyme that converts non-toxic 5-fluorocytosine (5-FC) into the toxic chemotherapeutic, 5-fluorouracil (5-FU)[56]. In Chapter 2, split yCD is conjugated to DNA, and the assembly of the split fragments is controlled through DNA strand displacement. We demonstrate that a dynamic DNA scaffold can be utilized to control split yCD assembly for effective prodrug activation in the presence of specific miRNA inputs.

Besides the therapeutic split yCD output, different protein reporters are also used within this thesis for detailed characterization purposes. In Chapter 2, a fluorescent protein pair is used to measure fluorescence resonance energy transfer (FRET), which serves as a spectroscopic ruler in determining the distances between proteins[57]. FRET between CFP (cyan) and YFP (yellow) conjugated with DNA allowed real-time tracking of protein assembly within our DNA circuit. In Chapter 4, a split nanoluciferase reporter system was used for the characterization of genetically-encoded Cas6 protein-RNA designs. The split enzyme has been engineered for low background affinity (K_d = 190 μM)[58], and the catalytic turnover of its substrate into

light allows for highly sensitive detection even at the low expression levels that were used in this study.

1.6 Dissertation Overview

This dissertation focuses on the development of a therapeutic device that utilizes programmable nucleic acid circuits to control protein function. Each chapter focuses on different aspects and strategies towards this goal.

Chapter 2 lays out the foundation for dynamic protein assembly controlled by DNA strand displacement. Through *in vitro* synthesis and characterization, protein-DNA conjugates are shown to be effectively utilized within DNA circuits without adverse consequences to protein or circuit behavior. Furthermore, the use of miRNA inputs to trigger split yCD activity is a powerful demonstration of the therapeutic potential for this technology.

In Chapter 3, the focus is switched towards an innovative strategy for interfacing input processing of biological targets with optimized *de novo* circuit components. A novel scheme is developed to achieve CHA with multi-input logic-gated behavior. Most importantly, the design decouples sequence constraints between input miRNA and CHA circuit components, allowing for uncompromised input detection and circuit computation.

Chapter 4 takes on the challenge of executing strand displacement controlled protein assembly in live mammalian cells. Cas6 proteins are used innovatively for their picomolar binding affinity and RNA cleavage activity to achieve efficient RNA-protein attachment and discretization of RNA strand displacement components, respectively. Protein assembly and disassembly based on simple strand displacement architectures reveal the potential for more complex designs in the future.

Finally, in Chapter 5, the conclusions of each chapter will be discussed in relation to each other and the overall goal of programmable nucleic acid-protein hybrid therapeutics. It will provide an overview of what has been accomplished in this work, as well as exciting implications and future directions for these technologies.

FIGURES

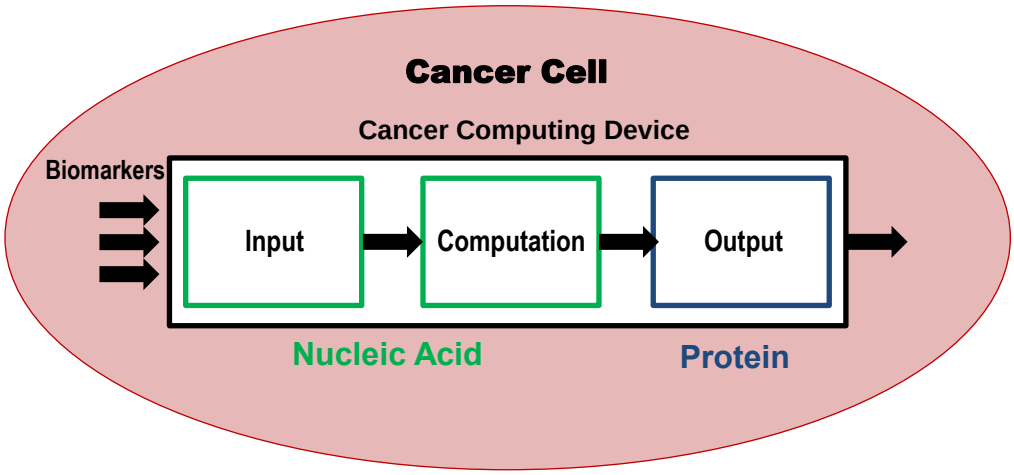

Figure 1.1 Therapeutic computing device based on a nucleic acid-protein hybrid system. Nucleic acids are utilized for complex input processing and computational decision making. Meanwhile proteins are harnessed to allow for diverse and useful output functionalities.

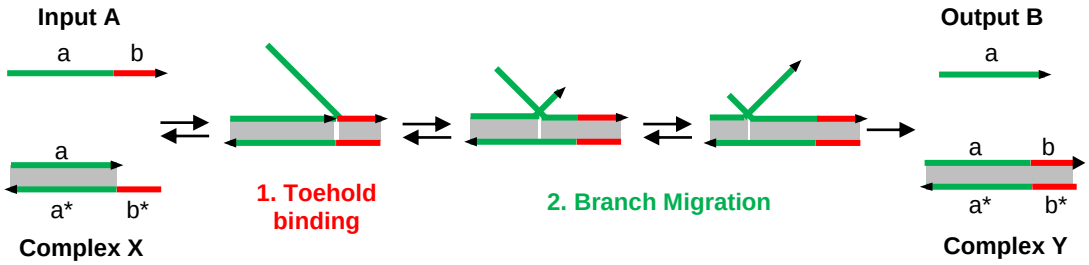

Figure 1.2 Toehold-mediated strand displacement. Input A can initiate binding to complex X through complementary toehold domains (red). Next, branch migration of the invading strand (input A) displaces the incumbent strand resulting in a newly formed complex Y and single stranded output B.

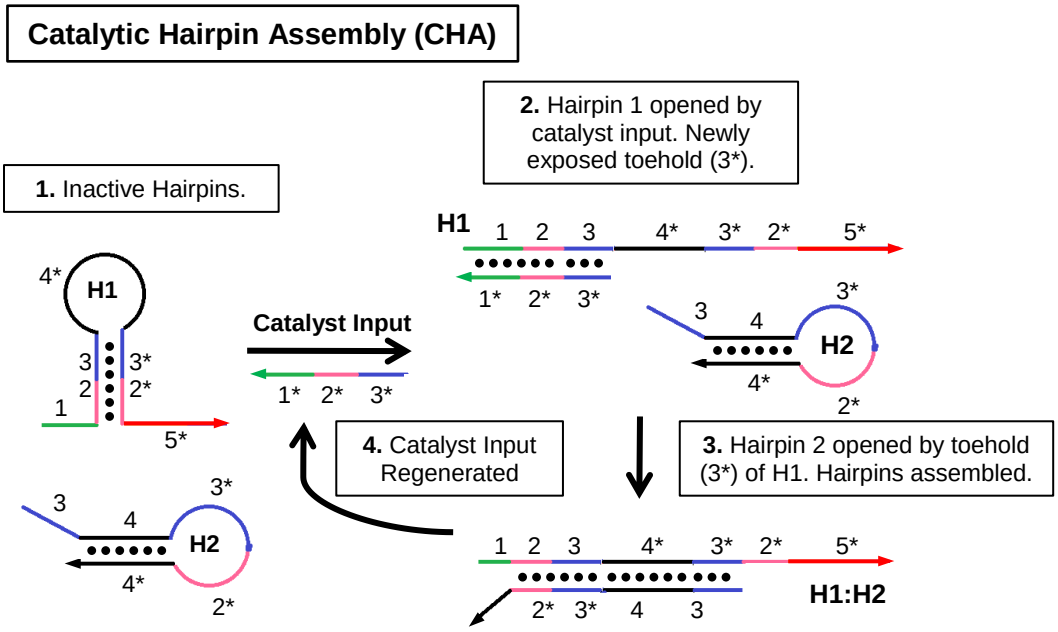

Figure 1.3 Catalytic hairpin assembly reaction pathway. Step 1.) CHA hairpins H1 and H2 initially cannot interact because toeholds are sequestered within stem and loop regions. **Step 2.)** The addition of the catalyst input opens H1 through a strand displacement reaction initiated at toehold domain 1 (green). This results in a newly exposed toehold domain 3* (blue). **Step 3.)** Hairpin H2 is opened through strand displacement by toehold domain 3* of H1, resulting in hairpin assembly. **Step 4.)** As H1 and H2 assemble, H2 strand displaces the catalyst input and regenerates it for further cycles of hairpin assembly.

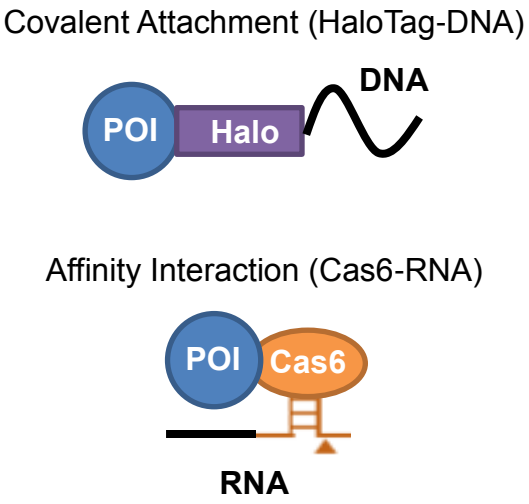

Figure 1.4 Protein to nucleic acid attachment strategies. Chloroalkane-modified DNA oligos can be covalently conjugated to HaloTag fusions to proteins of interest (POI). Cas6 fusion proteins allow for high affinity interaction between protein and RNA that contains Cas6 binding site sequences (orange). Cas6 also cleave their RNA substrate at a specific location (orange triangle) in a single-turnover manner.

REFERENCES

1. Alm, E. & Arkin, A. P. Biological networks. *Current Opinion in Structural Biology* **13**, 193-202 (2003).

2. Bhalla, U. S. & Iyengar, R. Emergent Properties of Networks of Biological Signaling Pathways. *Science* **283**, 381-387 (1999).

3. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray A. W. From molecular to modular cell biology. *Nature* **402**, C52 (1999).

4. Ptashne, M. Principles of a switch. *Nature Chemical Biology* **7**, 484-487 (2011).

5. Govern, C. C. & Rein ten Wold, P. Optimal resource allocation in cellular sensing systems. *Proceedings of the National Academy of Sciences* **111**, 17486 (2014).

6. Vidal, M. A unifying view of 21st century systems biology. *FEBS Letters* **583**, 3891-3894 (2009).

7. Wolkenhauer, O. Systems biology: the reincarnation of systems theory applied in biology? *Briefings in bioinformatics* **2**, 258-270 (2001).

8. Gonzalez-Angulo, A. M., Hennessy, B. T. & Mills, G. B. Future of Personalized Medicine in Oncology: A Systems Biology Approach. *Journal of Clinical Oncology* **28**, 2777-2783 (2010).

9. June Carl, H., O'Connor Roddy, S., Kawalekar Omkar, U., Ghassemi Saba & Milone Michael, C. CAR T cell immunotherapy for human cancer. *Science* **359**, 1361-1365 (2018).

10. Twumasi-Boaten, K., Pettigrew, J. L., Kwok, E., Bell, J. & Nelson, B. H. Oncolytic viruses as engineering platforms for combination immunotherapy. *Nature reviews. Cancer* **18**, 419-432 (2018).

11. Ugur, S. & Özlem, T. Personalized vaccines for cancer immunotherapy. *Science* **359**, 1355-1360 (2018).

12. Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nature Reviews. Immunology* **18**, 168-182 (2018).

13. Lemke, E. A. & Schultz, C. Principles for designing fluorescent sensors and reporters. *Nature Chemical Biology* **7**, 480-483 (2011).

14. Condon, A. Automata make antisense. *Nature* **429**, 351-352 (2004).

15. Shapiro, E., Benenson, Y., Adar, R., Gil, B. & Ben-Dor, U. An autonomous molecular computer for logical control of gene expression. *Nature* **429**, 423-429 (2004).

16. Xi, X. *et al*. RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Non-coding RNA* **3** (2017).

17. Botstein, D., Dunham, M. J. & Gresham, D. Comparing whole genomes using DNA microarrays. *Nature Reviews Genetics* **9**, 291-302 (2008).

18. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381-403 (2007).

19. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature chemistry* **3**, 103-113 (2011).

20. Chen, Y., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nature nanotechnology* **10**, 748-760 (2015).

21. Jung, C. & Ellington, A. D. Diagnostic applications of nucleic acid circuits. *Accounts of chemical research* **47**, 1825-1835 (2014).

22. Yurke, B., Turberfield, A. J., Mills, A. P., Jr, Simmel, F. C. & Neumann, J. L. A DNA-fueled molecular machine made of DNA. *Nature* **406**, 605 (2000).

23. Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends in Biochemical Sciences* **41**, 595-609 (2016).

24. Jäschke, A. Artificial ribozymes and deoxyribozymes. *Current Opinion in Structural Biology* **11**, 321-326 (2001).

25. Walter, N. G. & Engelke, D. R. Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London, England)* **49**, 199 (2002).

26. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends in Biochemical Sciences* **40**, 648-661 (2015).

27. Siu, K. *et al*. Synthetic scaffolds for pathway enhancement. *Current Opinion in Biotechnology* **36**, 98-106 (2015).

28. Chen, R. *et al*. Biomolecular scaffolds for enhanced signaling and catalytic efficiency. *Current Opinion in Biotechnology* **28**, 59-68 (2013).

29. Staley, Jonathan P|Woolford, John L. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Current Opinion in Cell Biology* **21**, 109-118 (2009).

30. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. (2009).

31. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science* **314**, 1585-1588 (2006).

32. Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196-1201 (2011).

33. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372 (2011).

34. Zhang, D. Y., Hariadi, R. F., Choi, H. M. & Winfree, E. Integrating DNA strand-displacement circuitry with DNA tile self-assembly. *Nature Communications* **4**, 1965 (2013).

35. Douglas, S. M., Bachelet, I. & Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831-834 (2012).

36. Yin, P., Choi, H. M. T., Calvert, C. R. & Pierce, N. A. Programming biomolecular self-assembly pathways. (2008).

37. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15275-15278 (2004).

38. Li, B., Ellington, A. D. & Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic acids research* **39**, e110 (2011).

39. Chen, X., Briggs, N., McLain, J. R. & Ellington, A. D. Stacking nonenzymatic circuits for high signal gain. *Proceedings of the National Academy of Sciences* **110**, 5386 (2013).

40. Blackstock, D. & Chen, W. Halo-tag mediated self-labeling of fluorescent proteins to molecular beacons for nucleic acid detection. *Chemical communications (Cambridge, England)* **50**, 13735-13738 (2014).

41. Sun, Q. & Chen, W. HaloTag mediated artificial cellulosome assembly on a rolling circle amplification DNA template for efficient cellulose hydrolysis. *Chemical communications (Cambridge, England)* **52**, 6701-6704 (2016).

42. Los, G. V. *et al*. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS chemical biology* **3**, 373-382 (2008).

43. Lapiene, V., Kukolka, F., Kiko, K., Arndt, A. & Niemeyer, C. M. Conjugation of fluorescent proteins with DNA oligonucleotides. *Bioconjugate chemistry* **21**, 921-927 (2010).

44. Lovrinovic, M. & Niemeyer, C. M. Rapid synthesis of DNA–cysteine conjugates for expressed protein ligation. *Biochemical and Biophysical Research Communications* **335**, 943-948 (2005).

45. Meyer, D. E., Trabbic-Carlson, K. & Chilkoti, A. Protein purification by fusion with an environmentally responsive elastin-like polypeptide: effect of polypeptide length on the purification of thioredoxin. *Biotechnology progress* **17**, 720-728 (2001).

46. Moore, M. & Ullman, C. Recent developments in the engineering of zinc finger proteins. *Briefings in functional genomics & proteomics* **1**, 342-355 (2003).

47. Abil, Z., Denard, C. A. & Zhao, H. Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *Journal of biological engineering* **8**, 7 (2014).

48. Lim, F. & Peabody, D. S. RNA recognition site of PP7 coat protein. *Nucleic Acids Research* **30**, 4138-4144 (2002).

49. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **333**, 470-474 (2011).

50. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Brouns, S. J. J. & Oost, V. D., John. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology* **13**, 722-736 (2015).

51. Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences* **40**, 58-66 (2015).

52. Blackstock, D., Sun, Q. & Chen, W. Fluorescent protein-based molecular beacons by zinc finger protein-guided assembly. *Biotechnology and Bioengineering* **112**, 236-241 (2015).

53. Sun, Q., Madan, B., Tsai, S., DeLisa, M. P. & Chen, W. Creation of artificial cellulosomes on DNA scaffolds by zinc finger protein-guided assembly for efficient cellulose hydrolysis. *Chemical communications (Cambridge, England)* **50**, 1423-1425 (2014).

54. Altaner, C. Prodrug cancer gene therapy. *Cancer Letters* **270**, 191-201 (2008).

55. Zawilska, J. B., Wojcieszak, J. & Olejniczak, A. B. Prodrugs: a challenge for the drug development. *Pharmacological reports : PR* **65**, 1-14 (2013).

56. Michnick, S. W. & Ear, P. H. A general life-death selection strategy for dissecting protein functions. *Nature Methods* **6**, 813-816 (2009).

57. Miyawaki, A. Visualization of the Spatial and Temporal Dynamics of Intracellular Signaling. *Developmental Cell* **4**, 295-305 (2003).

58. Dixon, A. S. *et al*. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS chemical biology* **11**, 400-408 (2016).

Chapter 2

DYNAMIC PROTEIN ASSEMBLY BY PROGRAMMABLE DNA STRAND DISPLACEMENT

Abstract

Inspired by the remarkable ability of natural protein switches to sense and respond to a wide range of environmental queues, here we report a strategy to engineer synthetic protein switches by using DNA strand displacement to dynamically organize proteins with highly diverse and complex logic gate architectures. We show that DNA strand displacement can be used to dynamically control the spatial proximity and the corresponding fluorescence resonance energy transfer (FRET) between two fluorescent proteins. Performing Boolean logic operations enabled the explicit control of protein proximity using multi-input, reversible, and amplification architectures. We further demonstrate the power and utility of this technology beyond sensing as a synthetic computing platform by driving the dynamic reconstitution of a split enzyme for targeted prodrug activation based on the sensing of cancer-specific miRNAs.

Chapter 2 is adapted and reproduced with permission from Springer Nature. Source:

Chen, R. P., Blackstock, D., Sun, Q. & Chen, W. Dynamic protein assembly by programmable DNA strand displacement. *Nature chemistry* **10**, 474-481 (2018). (DOI: 10.1038/s41557-018-0016-9)

2.1 Introduction

Living cells exhibit a remarkable ability to sense and respond to a wide range of signals[1], and dynamic control of protein assembly is one of the key mechanisms in actuating many cellular reactions in response to these evolving stimuli[2]. A key challenge in synthetic biology is designing more sophisticated sensors and actuators that exert dynamic control over protein assembly to mimic their biological counterparts in a much more tunable and predictable manner[3–5]. The highly programmable nature of DNA hybridization provides a simple way to spatially and temporally control the formation of DNA nanostructures with precise geometries[6–8]. By positioning proteins onto these DNA nanostructures, protein complexes can be spatially organized in a topologically calculated manner[9–12]. The use of dynamic DNA devices whose operation is based on toehold-mediated strand displacement can further produce reconfigurable and autonomous DNA nanostructures to allow for dynamic protein assembly[13–16].

Headway has been made toward building DNA-protein hybrid structures with dynamic mechanical properties that can be controlled through simple strand displacement. DNA tweezers were designed to regulate the efficiency of several enzyme cascades based on proximity control by modulating the switchable distance between the open and closed conformation[17,18]. Others have taken advantage of the rigidity of double stranded DNA to control enzyme-inhibitor complex formation[19]. A simple NOR logic gate based on strand displacement was constructed by the displacement of two zinc-finger recognition motifs from a template strand for split luciferase complementation[20]. While these examples are dynamic in nature and demonstrate the execution of simple logic gate architectures, their structural restrictions preclude the ability to achieve higher complexity and fully exploit the

power of strand displacement for molecular computing. Here, we demonstrate the use of strand displacement as a generalizable principle of constructing multi-input, reversible, and amplifiable dynamic protein-DNA nanoassemblies suitable for modulating protein proximity and activities. The modularity of the design enables the creation of a universal set of input-guided logic devices for regulated dynamic protein assembly.

2.2 Materials and Methods

2.2.1 Construction of expression vectors

The HaloTag vector was purchased from Promega. The HaloTag sequence was PCR amplified and then inserted into pET24a-ELP[KV8F]$_{40}$ using the SacII and EcoRI sites to form pET24a-ELP-HaloTag. The CFP and YFP sequences were PCR amplified, then double digested with NotI and BlpI and inserted into separate constructs to form pET24a-ELP-HaloTag-CFP and pET24a-ELP-HaloTag-YFP. F[1]yCD and F[2]yCD were PCR amplified and digested with SacI and SacII sites. The HaloTag sequence was PCR amplified, double digested with SacII and XhoI. A three piece ligation with pET24a-ELP[KV8F]$_{80}$ digested with SacI and XhoI sites was performed to construct pET24a-ELP-F[1]yCD-HaloTag and pET24a-ELP-F[2]yCD-HaloTag.

2.2.2 Protein expression and purification

All constructs were transformed into *E. coli* BLR [F- ompT hsdSB (r-B m-B) gal dcm(DE3) Δ (srl-recA)306::Tn10(TetR); Novagen, Madison,WI] cells for protein expression. Starting cultures for CFP/YFP and split yCD were grown to OD 0.7 at 37°C, induced with 250μM isopropyl-thiogalactopyranoside (IPTG), and grown at

20°C overnight. Cells were gathered by centrifugation, resuspended in Ni-NTA column buffer, and then sonicated. The soluble fraction was isolated and transferred to a Ni-NTA His Bind (Novagen, Madison,WI) column for purification via the C-terminal his6 tag. Residual imidazole was removed from purified proteins by ELP precipitation of the proteins with 1M Na_2SO_4 at 37°C. The pellet was recovered at 15,000xg for 15 min and washed with 37°C phosphate buffered saline to remove any residual salt. The proteins were then re-solubilized in cold DNA hybridization buffer (20mM Tris, 150mM NaCl, 5mM $MgCl_2$, pH7.4) and centrifuged again to remove any insoluble matter. For split yCD conjugates no Histag purification was performed. Proteins were purified through two cycles of ELP inverse transition cycling.

2.2.3 Protein-DNA labeling

The specified oligos with 5' and 3' amine modifications were reacted with the HaloTag Succinimidyl (O4) Chlorohexane (CH) ligand (Promega P6751) for HaloTag attachment. The CH ligand was mixed with the DNA at a molar ratio of 30:1 and incubated at room temperature for 24 h. Excess CH ligand was removed using a 3,000 Da ultrafiltration column (Vivaspin 500, Sartorius Stedim Biotech). The purified CH-oligos were then reacted with the purified HaloTag fusion protein, using a 3x molar excess of CH-oligo, and incubated overnight at 4°C. Utilizing the ELP tag, the excess/unattached oligo was removed by pelleting the protein-DNA and washing away any residual DNA through inverse transition cycling. The protein only, protein plus excess DNA, and purified protein-DNA samples were loaded onto a 10% SDS-PAGE and stained with Coomassie blue for analysis of labeling efficiency by densitometry. More details can be found in previous literature using this method[25].

2.2.4 Electromobility shift assay (EMSA)

Mobility shift experiments were carried out in a 4.5% or 6% non-denaturing acrylamide gel. The protein only, protein plus excess DNA, and purified protein-DNA samples were loaded and run for 45 min at a constant 90V with TBE(0.5x). The gel was then stained with ethidium bromide for 5 min before imaging.

2.2.5 FRET strand displacement reactions

All oligo sequences and final concentrations used within circuit architectures can be found in **Table 2.1** and **Table 2.2**, respectively.

Single input. The CFP-A strand was prehybridized with the scaffold blocking strand B and scaffold strand C at a molar ratio of 3:3:2μM, respectively. The YFP-D strand was prehybridized with its sequestering strand at a molar ratio of 3:4μM, respectively. These mixtures were incubated overnight at 4°C in DNA hybridization buffer (20mM Tris, 150mM NaCl, 5mM $MgCl_2$, pH7.4). The prehybridized complexes (5x concentrated) were then mixed and diluted in hybridization buffer. The input F was added at 1.2x (960nM) the concentration of the CFP sequestering strand (800nM). Fluorescence data was taken every 30 sec for 45 min immediately after adding the input.

Multi-input. CFP and YFP strands were prehybridized with their sequestering strands at a molar ratio of 3:4μM, respectively. The scaffold was prehybridized with its blocking strands at a ratio of 2:3:3μM (scaffold:block(CFP):block(YFP)). After overnight incubation at 4°C in DNA hybridization buffer, the prehybridized complexes (5x concentrated) were mixed and diluted to their working concentrations. The inputs F and I were added at 1.2x (960nM) the concentration of the CFP/YFP sequestering strand (800nM). Fluorescence data was taken every minute for 60min

immediately after adding the strands. Then, 1.2x (1.15μM) of the "NOT" input was added to the sample containing both inputs F and I, and data was taken for an additional 30 min.

Catalytic hairpin assembly. The YFP strand was prehybridized with the scaffold blocking strand G and scaffold strand C at a molar ratio of 1.5: 1.5:1.25μM, respectively. The CFP strand was prehybridized with its sequestering strand at a molar ratio of 1.5:1.8μM, respectively. These mixtures were incubated overnight at 4°C in hybridization buffer. The CHA hairpins (H1 PAGE purified, H2 standard desalted) were prepared in hybridization buffer at 4uM by separately heating to 95°C for 5 min before cooling to 4°C at a rate of 0.1°C/sec. The prehybridized complexes (5x concentrated) and hairpins (10x concentrated) were then mixed and diluted in hybridization buffer. The final concentration of the hairpins was 400nM and various concentrations of miR-122, from 1x (400nM) down to 0.05x (20nM) that of the hairpin concentration, were added for immediate fluorescence measurement.

2.2.6 Fluorescence measurements

All fluorescence measurements were taken using a Synergy H4 (BioTek) plate reader. The CFP and YFP samples were excited at a wavelength of 434 nm and the fluorescence emission spectra were taken from 460 nm to 600 nm. FRET ratio data was gathered by measuring the emission at 478 nm and 524 nm using a 434 nm excitation. For the split yCD strand displacement kinetics experiment, strand B and strand C were modified with 3'BHQ1 and 5'FAM, respectively. Fluorescence data was taken using 520 nm emission and 470 nm excitation.

2.2.7 yCD activity assay

For split yCD reconstitution, DNA strand displacement reactions were carried out in the same manner as the CFP/YFP system. The 5x concentrated prehybridized complexes were prepared in DNA hybridization buffer and incubated at 4°C for at least 24 h. For the one-input and two-input circuits, the 1x concentration of the split yCD with their sequestration strands was 500nM and 750nM, respectively. The central scaffold and blocking strands were hybridized for a final 1x concentration of 500nM and 600nM, respectively. The prepared concentrated complexes were diluted and mixed together in the presence of freshly prepared HEK293T cell lysate by glass bead lysis. Each 50µl reaction contains the lysate of 10^5 cells. The appropriate miRNA inputs (RNA synthesized from IDT) were then added at 1.2X (900nM) the sequestration strand concentration (750nM). For the amplification circuit, the same strand concentrations were used as the one-input circuit, except now the input strand I is replaced with hairpins H1 and H2 (900nM of each). With the same HEK293T cell lysate preparation, various catalyst input concentrations, miR-122, corresponding to 1x (900nM), 0.25x (225nM), and 0.1x (90nM) of the hairpin concentration, were added to the reaction. All circuits were allowed to proceed for 2 h at room temperature before 5-FC was added at a final concentration of 1mg/ml. Samples were incubated at room temperature for 12 h, and then assayed for 5-FU production. GIA39(ΔbCD) *E. coli* cells were precultured in a 2ml of LB media for 30 h before dilution to OD=0.05 in LB media. In a 96-well plate, 150µl of cells were added to 50µl of split yCD circuit reaction. The cell mixtures were grown at 37°C with shaking for 14 h, and OD measurements were recorded by measuring absorbance at 600nm on a plate reader.

2.2.8 MTT cell viability assay

HeLa cells were cultured in MEM (Corning) supplemented with 10% fetal bovine serum (Sigma), 100 U ml–1 penicillin, and 100 µg ml–1 streptomycin (GE Healthcare). 16 h after seeding in 24-well plates, at 10,000 cells per well, 500µl fresh media containing various concentrations of 5-FU (0-500µg ml–1) were added to cells. Two days later, cells were incubated with MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (Life Technologies), 1.2mM in media) for 2 h at 37°C. Next, 500µl of stop solution (10% SDS, 0.01M HCl) were added to each well. After thorough mixing, plate was incubated at 37°C for 4h. Lastly, absorbance was measured at 570nm on plate reader.

2.3 Results and Discussion

2.3.1 Dynamic assembly of fluorescent proteins by DNA strand displacement

To demonstrate the feasibility of dynamic protein organization using toehold-mediated logic circuits, we first used the fluorescent protein pair, cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), as model proteins. The assembly and disassembly of the two proteins can be easily monitored in real-time using the well-known fluorescent resonance energy transfer (FRET) property of the CFP-YFP pair[21–23]. The basic design of the one-input protein-DNA device is shown in **Figure 2.1A**. Both the CFP-DNA conjugate (strand A) and the YFP-DNA conjugate (strand D) are initially sequestered onto two non-interacting DNA complexes. Since all toeholds are initially blocked, the mixture of the two sequestered protein-DNA complexes remains inactive and does not yield a signal unless acted upon by an input (strand F). Once the input is introduced, the YFP-D strand is displaced, freeing toehold (d) to bind and displace blocking strand B on the CFP complex. This results in

co-localization of CFP and YFP onto the same scaffold strand C for induced protein proximity.

To attach CFP and YFP to the DNA strands, CFP and YFP fusions with the HaloTag protein[24] were used to provide site-specific decoration of DNA strands[25] (**Figure 2.2**). Each fusion protein was designed to contain either CFP or YFP at the C-terminus and an elastin-like-polypeptide (ELP) tag[26,27] at the N-terminus for simplified purification. A chlorohexane (CH)-modified DNA strand was covalently linked to each fusion protein by utilizing the suicide ligand recognition properties of HaloTag. Both fusion proteins were purified using two cycles of thermally triggered precipitation and solubilization before labeling with CH-modified DNA strands. The resulting protein-DNA conjugates were purified again by an additional ELP inverse transition cycle to remove any excess DNA. For both proteins, the labeling reaction was highly effective with more than 95% labeling efficiency (**Figure 2.2**).

To test the one-input design, the CFP-A conjugate was first hybridized with strands B and C overnight. Similarly, the YFP-D conjugate was prehybridized with strand E. When the two prehybridized complexes were mixed together, no significant increase in FRET was observed even after 45 min incubation as all the toeholds were sequestered (**Figure 2.1B**). However, upon addition of input strand F, the sequestering strand E on the YFP conjugate was rapidly displaced. The released YFP-D conjugate contains an exposed single stranded DNA toehold (d) domain that is able to hybridize with toehold (d*) on strand C and trigger the displacement of strand B. This resulted in the co-assembly of CFP-A and YFP-D onto a single scaffold strand C as reflected by a rapid increase in FRET reaching 90% of the full response within 30 min (**Figure**

2.1B). Even an RNA input strand is capable of eliciting a similar FRET response (**Figure 2.1B**).

2.3.2 Dynamic FRET control using multi-input, reversible, and amplification architectures.

2.3.2.1 Multi-input circuit

We next investigated whether dynamic protein assembly can be driven by a more complex two-input "AND" gate design. This new circuit includes an extra sequestering strand H for the CFP-A conjugate and a blocking strand G, such that the scaffold strand C is initially not loaded with any protein (**Figure 2.3A**). Assembly of the two fluorescent proteins is triggered only in the presence of two distinct inputs, F and I. As in the case of the single-input circuit, the background signal was minimal in the absence of inputs (**Figure 2.3B**). A small increase in the signal was observed upon introducing either one of the inputs, suggesting a low level of leak in the strand displacement process. In contrast, a significant increase in FRET was observed upon the addition of both inputs reaching saturation within 40 min (**Figure 2.3B**). Not surprisingly, the two-input circuit is kinetically slower than the single input design since four different strand displacement reactions rather than two are required before FRET can occur.

2.3.2.2 NOT strand disassembly

To further demonstrate that DNA strand displacement is capable of triggering not only protein assembly but also disassembly, we utilize an exposed toehold (e) on the YFP strand to turn off the FRET response (**Figure 2.4**). Addition of a "NOT" strand containing the (e*) domain rapidly displaced YFP-D from the scaffold strand C,

and effectively turned off the FRET signal within 10 min (**Figure 2.3B**). This result opens up the exciting possibility of regulated cycling of protein assembly and disassembly.

2.3.2.3 Amplification through catalytic hairpin assembly (CHA)

One of the most useful features of toehold-mediated strand displacement circuits is the ability to amplify even very small biological signals, a feature that is essential for RNA-driven intracellular circuits, as the typical miRNA concentrations are relatively low even when highly expressed in disease cells. Catalytic hairpin assembly (CHA)[28, 29,30], in which two kinetically trapped hairpins (H1 and H2) are used to amplify the circuit response in an autocatalytic and cascade fashion **(Figure 2.5**), is used for signal amplification. Spontaneous hybridization between H1 and H2 is hindered because the complementary regions are blocked by the stems, but upon the addition of the catalyst input, the stem portion of H1 is opened by toehold-mediated strand displacement. A newly exposed domain (3*) within H1 can then hybridize to toehold (3) on H2 and trigger branch migration, ultimately displacing the catalyst strand for additional rounds of hairpin assembly. Ideally, direct labeling of fluorescent proteins to H1 and H2 would be perfect for protein assembly as they will be adjacent to each other after hybridization (**Figure 2.5**). However, the high-temperature annealing conditions crucial for preventing incomplete and/or improper hairpin formation[30,31] that leads to system leakage preclude direct protein conjugation to the hairpins. Instead, we designed a CHA circuit that trigged the release of input I in the previous two-input design (**Figure 2.5**). The stem of H1 sequesters toehold (j), rendering input I inactive until H1 is opened by the catalyst strand, which is designed to match the sequence of miR-122, a cancer-specific target[32,33]. In the absence of the

catalyst strand, there is no FRET detected, however upon miR-122 addition, drastic FRET increase was observed (**Figure 2.6**). Even an input:H1 ratio of 1:10 is adequate to achieve full FRET activation, indicative of signal amplification. At even lower target concentrations FRET activation was observed, but the slower rate prevented our ability to measure full activation before extensive sample evaporation. This level of amplification is on par with previous demonstrations of CHA[29–31]. The ability to provide amplification is a unique advantage of our design over other strand displacement devices currently available to control protein proximity. The overreliance of other methods on non-dynamic structural elements severely limits the ability to build or expand to include amplification characteristics. By cascading orthogonal CHA reactions, signal amplification can be controlled over several orders of magnitude, thus opening up the landscape of biological inputs and protein assembly outputs that can be practically paired together for useful applications.

2.3.3 DNA-logic devices for prodrug activation.

Perhaps the most intriguing utility of protein-DNA logic devices is the ability to perform multi-input computation. By exploiting biomolecules as inputs and biologically active responses as outputs, biomolecular computing devices could be produced for "logical" control of biological processes. This type of synthetic computing device is ideally suited for sensing multiple cancer-specific biomarkers toward the creation of "smart therapeutics" that detect the disease state and actuate an appropriate therapeutic response for cell killing. Yeast cytosine deaminase (yCD) possesses the ability to deaminate 5-fluorocytosine (5-FC), a non-toxic prodrug, into 5-fluorouracil (5-FU), a chemotherapeutic agent that can be used to inhibit DNA synthesis[32,33]. The recent discovery of split yCD fragments (referred to as F1 and F2)[34]

that are conditionally reassembled in the presence of flanking interacting domains provides the framework for a new mechanism of prodrug activation based on induced split yCD assembly by toehold-mediated strand displacement.

Although the circuits described above were effective in bringing CFP and YFP into close proximity, the reconstitution of split yCD is more stringent as the two split fragments must be side-by-side for reconstitution to occur. To investigate this feasibility, the F1 and F2 fragments were fused to the ELP-HaloTag protein. After conjugating with the desired DNA strands, the F1-A and F2-D conjugates were mixed with different amounts of the scaffold strand C (**Figure 2.7**). Dual binding of the two conjugates onto strand C was first confirmed by native PAGE analysis (**Figure 2.7B**). The resulting yCD activity was evaluated by 5-FC activation into 5-FU. After incubation with 5-FC, the resulting reaction mixtures were added to *E. coli* (Δbcd) cultures, and the effect on cell growth was used to indicate the level of 5-FU production (**Figure 2.7C**). In the absence of strand C, no growth inhibition was detected indicating a negligible amount of yCD reconstitution and 5-FU production. Upon the addition of strand C, there was a direct correlation between the amount of strand C added (yCD reconstitution) and the level of growth inhibition, confirming the binding-induced reconstitution of split yCD fragments and prodrug activation.

To achieve biomarker-specific prodrug activation, similar one-input and two-input circuits were designed to reconstitute the split yCD fragments using two cancer-specific miRNAs, miR-21 and miR-122, as inputs[35,36]. In the absence of CFP and YFP, a fluorophore dye and a quencher were added to the 5' end of strand C and 3' end of strand B, respectively, in order to probe the kinetics of the one-input circuit (**Figure 2.8A**). No increase in fluorescence was detected when the two sequestered

yCD fragments were mixed, indicating virtually no background leakage for the newly designed circuit (**Figure 2.8B**). Although the increase in fluorescence upon the addition of miR-21 was slightly slower reaching a plateau only after 40 min, full activation of split yCD was observed as indicated by a significant reduction in cell growth (**Figure 2.9A**) after 14 h incubation. Time course data demonstrate that cell growth stays relatively flat for the sample in which the split yCD activity has been turned "ON", while the "OFF" sample exhibits the typical logarithmic growth curve (**Figure 2.10**).

Similar responses were observed using the two-input circuit with full yCD activation detected only in the presence of both miR-21 and miR-122. A small level of growth inhibition was detected with the additional of only one input (**Figure 2.9B**), consistent with the low level of leak from the CFP-YFP experiments. Time course data again show logarithmic growths in the absence of either one or both inputs, but virtually no growth in the presence of both inputs (**Figure 2.10**). All yCD activity circuit experiments were done with RNA inputs and in the presence of HEK293T cell lysate to emulate the noise of a cellular environment and further support the robustness of our device inside a cell.

Finally, we sought to demonstrate that the amplification circuit design can be similarly employed for yCD activation and growth inhibition (**Figure 2.9C).** Compared to the one-input circuit (**Figure 2.9A**), there is slightly more background growth inhibition due to minimal leak for the CHA hairpins. However, amplification is evident as even 0.1X miR-122 (90nM) is sufficient to elicit a significant growth inhibition (34% drop in the growth when compared to no input). The level of amplification is even more impressive considering the fact that the extent of inhibition

is less than maximum since 5-FC was added 2 h into the CHA reaction, far before the amplification can reach completion. This was done to better discriminate differences in the amplification levels. For prodrug activation, amplification is particularly useful to maximize cell death even while dealing with low miRNA concentrations. While our split yCD results are *in vitro*, recent studies demonstrating successful execution of strand displacement circuits inside living mammalian cells[37,38] suggest that our approach for controlled split yCD reconstitution is highly achievable intracellularly. In an *in vivo* context, changes such as a backbone modification to 2′-O-methyl ribonucleotides for increased stability against nuclease degradation will be important38. While the use of *E. coli* allows fast and easy characterization of a large number of split yCD circuit designs, it is important that similar growth inhibitions would translate well to cancer HeLa cells. As depicted in **Figure 2.11**, both *E. coli* GIA39 and HeLa cells exhibit similar sensitivity to 5-FU, confirming the effectiveness of our split yCD reconstitution approach even for cancer cells.

2.4 Conclusion

In this study, we have introduced a new dynamic protein assembly framework based on DNA logic circuits driven by toehold-mediated strand displacement. The modularity of the circuit design enables the construction of a wide range of input-guided DNA logic devices as a powerful information carrier for dynamic protein assembly. We demonstrated that protein proximity control is useful not only for modulating enzyme activity but is perhaps most ideally suited for creating Boolean logic computing, sensing, and responsive devices for highly amplifiable and programmable prodrug activation. Unlike previous demonstrations of DNA-guided protein assembly, which relied either on simple hybridization scaffolds or structural

elements (such as DNA tweezers) that severely limit the complexity of the logic architecture achievable, our design is flexible and can be easily combined into highly modular multiple-input logical devices. Since strand displacement reactions are sequence dependent[39], different dynamic complexes can be designed to operate orthogonally and integrated into programmable reaction networks to perform complex computations[16] of endogenous biological inputs that are aberrantly expressed in disease cells. More importantly, the current CHA design is capable of eliciting split yCD reconstitution and growth inhibition even with 90nM miR-122, consistent with the typical miR-122 level detected in liver tissues[40,41]. This sensitivity further suggests the potential utility of our design for cell-based applications. In Chapter 3, miRNA-gated CHA will be expanded upon to explore more efficient interfacing between miRNA and optimized *de novo* CHA.

Moving forward, the ultimate realization of this technology is the execution of such a programmable device inside live cells. Development of such a "cell classifier" would open a new realm of possibilities in smart diagnostics and therapeutics. It has been previously shown that toehold-mediated strand displacement can be programmed to assess endogenous signals inside live cells[37,38]. However, output signals were solely based on changing fluorescence from fluorophore/quencher labeled oligos, which ultimately limits therapeutic capabilities. On the other hand, there have been demonstrations of "cell classifiers" using protein as outputs. In this example, a synthetic device that senses the high/low state of six cancer-related miRNAs in parallel was built to initiate transcription of the apoptosis gene hBax in HeLa cells[42]. This multi-input logic circuit selectively identified HeLa cells and triggered apoptosis without affecting other non-HeLa cell types. However, implementing such synthetic

circuits within a patient is far from trivial, requiring delivery and expression of multiple foreign genes in cancer cells efficiently and safely. In contrast, our cell classifiers are easy to design and all optimization can first be done *in vitro* before cellular deployment, making this one of the most modular approaches reported to date.

The efficient delivery of protein-DNA conjugates/complexes remains an obstacle to our technology. While there has been a wealth of work on the delivery of nucleic acids[43–45] and proteins independently[46,47], reports of hybrid protein-DNA conjugate delivery are quite limited. However, it is plausible that a combination of these previous strategies can be used to successfully deliver our protein-DNA device. It is likely that further modification of the conjugates (such as nucleic acid backbone changes[38], protein PEGylation[47], etc.) will also be necessary for our device to reach its full therapeutic potential. For more complex computation architectures, our design relies heavily on the delivery of different DNA-protein components in the required ratios. In anticipation of delivery challenges, the development of strategies to minimize the number of oligos involved such as the use of more hairpin designs and to make our device more robust to varying complex ratios would be beneficial.

TABLES

Table 2.1 Oligo sequences for protein assembly circuits. All oligos were purified through standard desalting, unless otherwise specified. Domains are separated by underscores. Abbreviations: /3AmMC6T/ - 3' Amino Modifier C6 dT, /5AmMC12/ - 5' Amino Modifier C12, /56/FAM – 5’ 6-FAM (fluorescein), /3BHQ1/ - 3’ Black Hole Quencher 1.

Name	Design	Sequence (5'→3')
Multi-input CFP/YFP		
CFP strand **A**	kab	CTTTCCTACA_CCTACG_TCTCCAACTAACTTACGG/3AmMC6T/
Scaffold blocking strand **B** (YFP site)	c	CCCTCATTCAATACCCTACG
Scaffold strand **C**	d*c*7Tb*a*	TGGAGA_CGTAGGGTATTGAATGAGGG_TTTTTTT_CCGTAAGTTAGTTGGAGA_CGTAGG
YFP strand **D**	cde	/5AmMC12/CCCTCATTCAATACCCTACG_TCTCCA_AGCAACCTCAAACAGACAC
YFP sequestration strand **E**	f*e*d*	GGTAC_GTGTCTGTTTGAGGTTGCT_TGGAGA
Input strand **F** (YFP release strand)	def	TCTCCA_AGCAACCTCAAACAGACAC_GTACC
Scaffold blocking strand **G** (CFP site)	b	TCTCCAACTAACTTACGG
CFP sequestration strand **H**	a*k*j*	CGTAGG_TGTAGGAAAG_TATGGT
Input strand **I** (CFP release strand)	jka	ACCATA_CTTTCCTACA_CCTACG
Catalytic Hairpin Assembly (FRET and Split yCD)		
H1 (miR-122 CHA), PAGE purified	1234*3*2*(j)ka	CAAACAC_CATTGTC_ACACTCCA_CCATGTGTAGA_TGGAGTGT_GACAATG_CTTTCCTACA_CCTACG
H2 (miR-122 CHA)	343*2*(j)4*	ACACTCCA_TCTACACATGG_GACAATG_GTGTTTG_CCATGTGTAGA_CAAACAC
YFP sequestration strand H	a*k*j*	CGTAGG_TGTAGGAAAG_CATTGTC
Input strand I (non CHA input)	jka	GACAATG_CTTTCCTACA_CCTACG
Catalyst CHA input (miR-122)	3*2*(j)1*	UGGAGUGU_GACAAUG_GUGUUUG
Multi-input Split yCD		
F1 strand	kab	GTGACAATGG_TGTTTG_TCTCCAACTAACTTACGG/3AmMC6T/
F2 strand	cde	/5AmMC12/CCCTCATTCAATACCCTACG_TAGCTT_ATCAGACTGA
Scaffold strand	d*c*7Tb*a*	/56FAM/AAGCTA_CGTAGGGTATTGAATGAGGG_TTTTTTT_CCGTAAGTTAGTTGGAGA_CAAACA
Scaffold blocking strand (F1 site)	b	TCTCCAACTAACTTACGG/3BHQ1/ (HPLC purified)
Scaffold blocking strand (F2 site)	c	CCCTCATTCAATACCCTACG
F1 sequestration strand	a*k*j*	CAAACA_CCATTGTCAC_ACTCCA
F2 sequestration strand	f*e*d*	TCAACA_TCAGTCTGAT_AAGCTA
miR-122 input (F1 release strand)	jka	UGGAGU_GUGACAAUGG_UGUUUG
miR-21 input (F2 release strand)	def	UAGCUU_AUCAGACUGA_UGUUGA

Table 2.2 Final oligo concentrations in strand displacement circuits.

Name	Final Concentration (nM)	
One-input and Multi-input CFP/YFP		
CFP strand **A**	600	
Scaffold blocking strand **B** (YFP site)	600	
Scaffold strand **C**	500	
YFP strand **D**	600	
YFP sequestration strand **E**	800	
Input strand **F** (YFP release strand)	960	
Scaffold blocking strand **G** (CFP site)	600	
CFP sequestration strand **H**	800	
Input strand **I** (CFP release strand)	960	
Catalytic Hairpin Assembly (FRET and Split yCD)		
CFP/F1 strand **A**	300	600
CFP/F1 sequestration strand **H**	360	750
YFP/F2 strand **D**	300	600
Scaffold strand **C**	250	500
Scaffold blocking strand **G** (CFP/F1 site)	300	600
H1 (miR-122 CHA)	400	900
H2 (miR-122 CHA)	400	900
Input strand **I** (non CHA input)	400	900
Catalyst CHA input (miR-122)	400 (1X), 100 (0.25X), 40 (0.1X), 20 (0.05X)	900 (1X), 225 (0.25X), 90 (0.1X)
Multi-input Split yCD		
F1 strand	500	
F2 strand	500	
Scaffold strand	500	
Scaffold blocking strand (F1 site)	600	
Scaffold blocking strand (F2 site)	600	
F1 sequestration strand	750	
F2 sequestration strand	750	
miR-122 input (F1 release strand)	900	
miR-21 input (F2 release strand)	900	

FIGURES

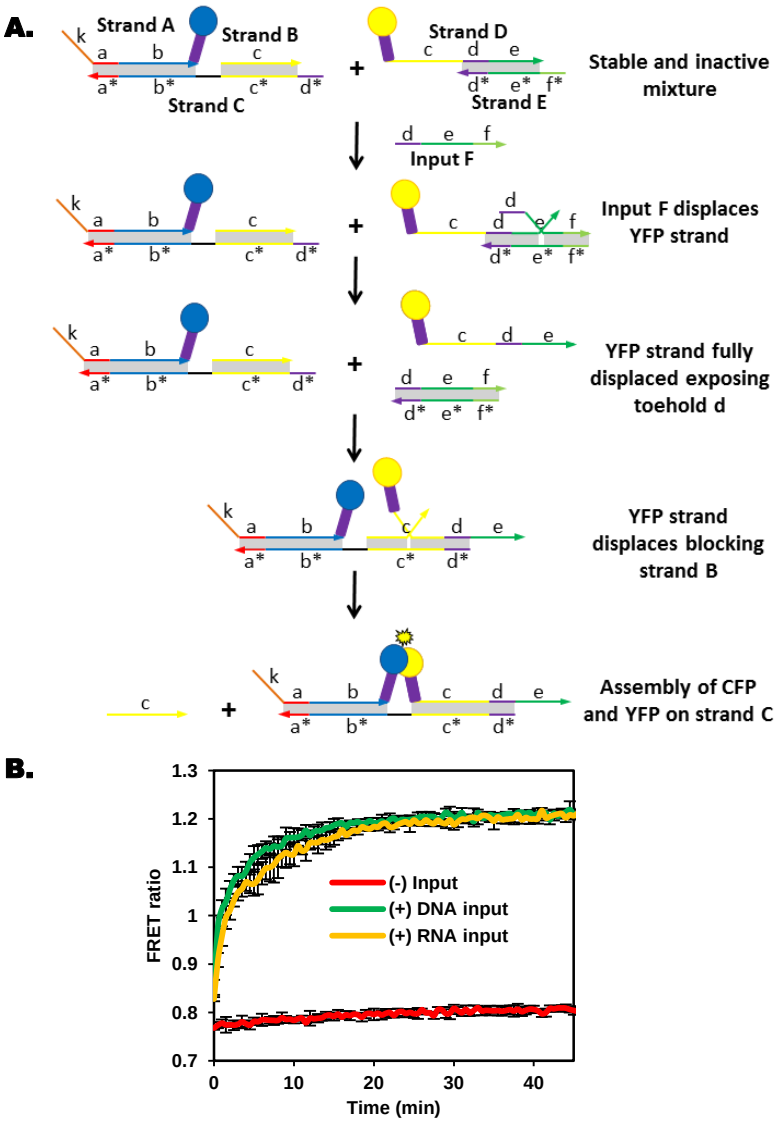

Figure 2.1 One-input protein-DNA device mechanism and response. A.) YFP and CFP are attached to the 5' and 3' end of Strand D and Strand A, respectively. Strand displacement is initiated when complementary toeholds hybridize and branch migration allows for the formation of new complexes and output strands. **B.)** Without a sequence specific input no FRET response is detected over time (red). When the 1X (960 nM) input is added, a rapid rise in FRET ratio is detected and reaches a maximum within 10 min. Both DNA (green) and RNA (yellow) result in the same fast kinetic response. Error bars represent standard deviation (n=3).

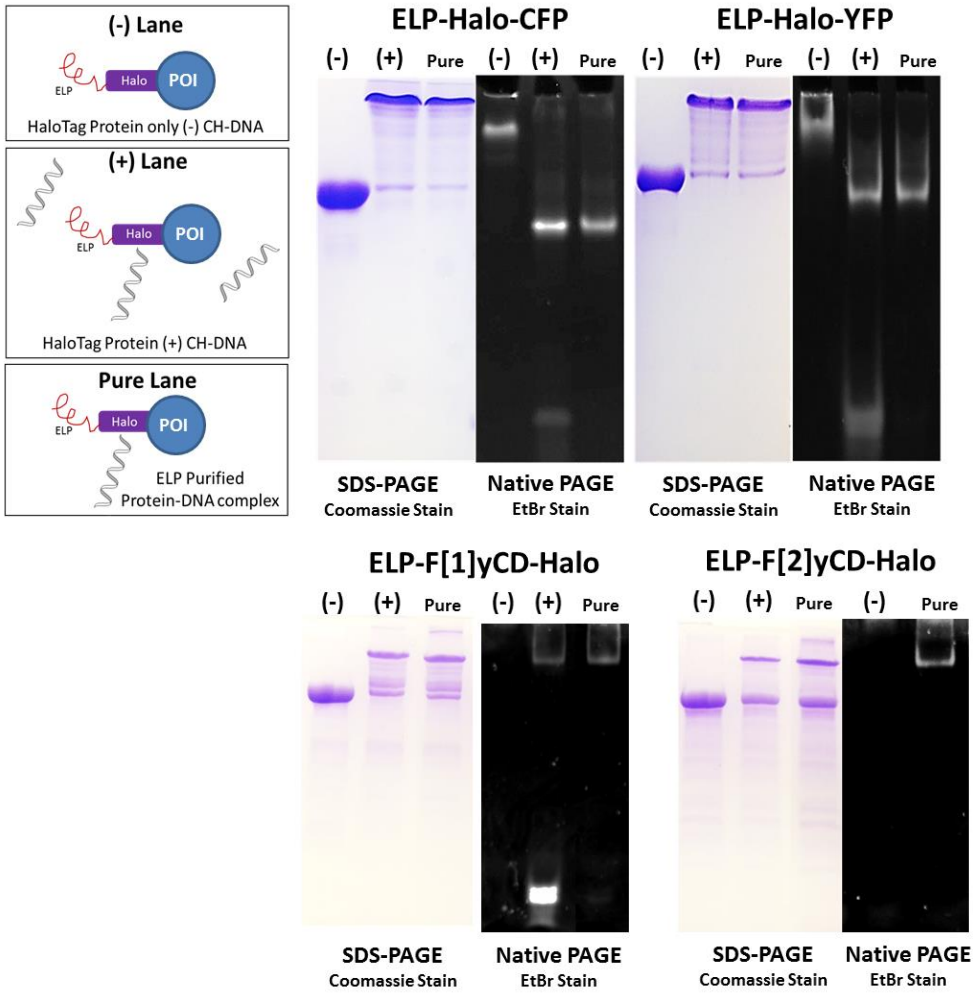

Figure 2.2 HaloTag-mediated DNA labeling of proteins. Purified proteins of interest were incubated with a 3-fold excess of the respective chlorohexane (CH)-modified DNA oligo. Both unconjugated (-) and conjugated (+) proteins were analyzed by SDS-PAGE and native PAGE gel. Reaction mixtures were purified by one cycle of ELP-triggered ITC to remove the excess CH-DNA (Pure). Labeling efficiencies were determined by densitometry comparison of the unlabeled bands since proteins labeled with DNA migrated and stained on the gel compactly and were unreliable for comparison. The labeling efficiencies were close to 95% or above for CFP and YFP. For split yCD (85% for F[1]yCD and 60% for F[2]yCD), we suspect that some fraction of the split yCD is misfolded and remained unlabeled. These unlabeled split yCD proteins cannot participate in the reconstitution experiments and will have minimal effects to our experiments. Purification yield after labeling is 90% or above with <5% free CH-DNA remaining as determined by densitometry of SDS-PAGE and native PAGE, respectively.

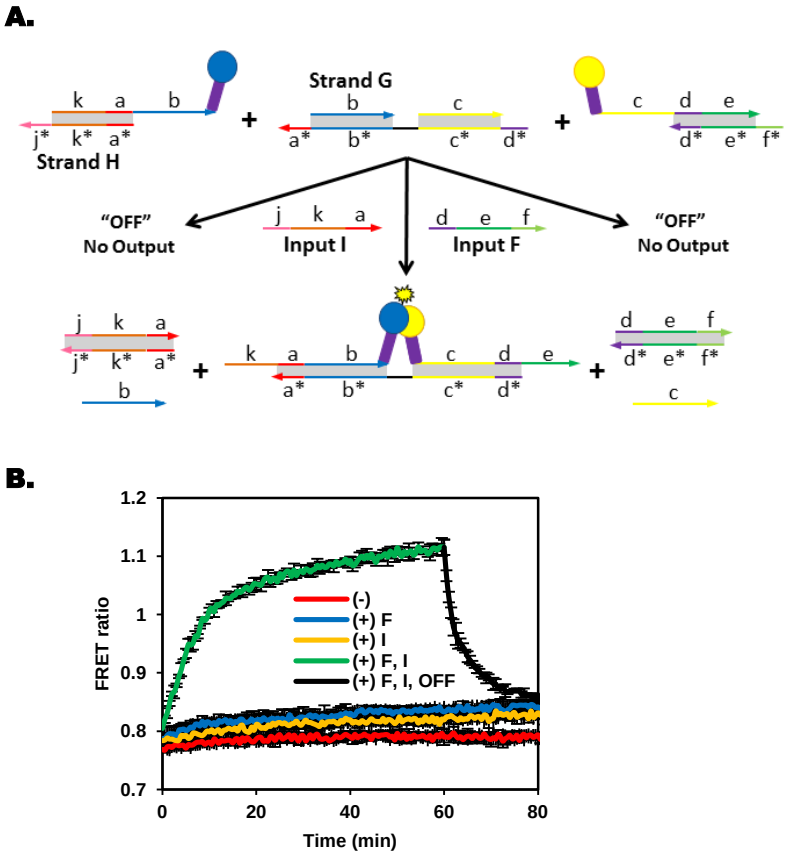

Figure 2.3 Multi-input and reversible device architectures. A.) Two-input protein-DNA circuit mechanism. Similar to the one-input circuit design except now both proteins are blocked from interacting with the central DNA scaffold. Only in the presence of both inputs will both proteins be brought into close proximity. **B.)** Two-input circuit FRET response. Without either input, no increase in FRET ratio is detected (red). With only 1.2X (960 nM) input F (blue) or input I (yellow) there is a slight increase in FRET ratio. This 10% leak is attributed to incomplete sequestration of toeholds by strand E and strand H. With both input F AND input I, the maximum FRET response is reached in approximately 60 min (green). The slower kinetics of the two-input compared with the one-input system is explained by the greater amount of strand displacement reactions needed for both CFP and YFP assembly. Once the "OFF" strand (1.15μM) is added, the FRET signal is rapidly turned off (black). Error bars represent standard deviation (n=3).

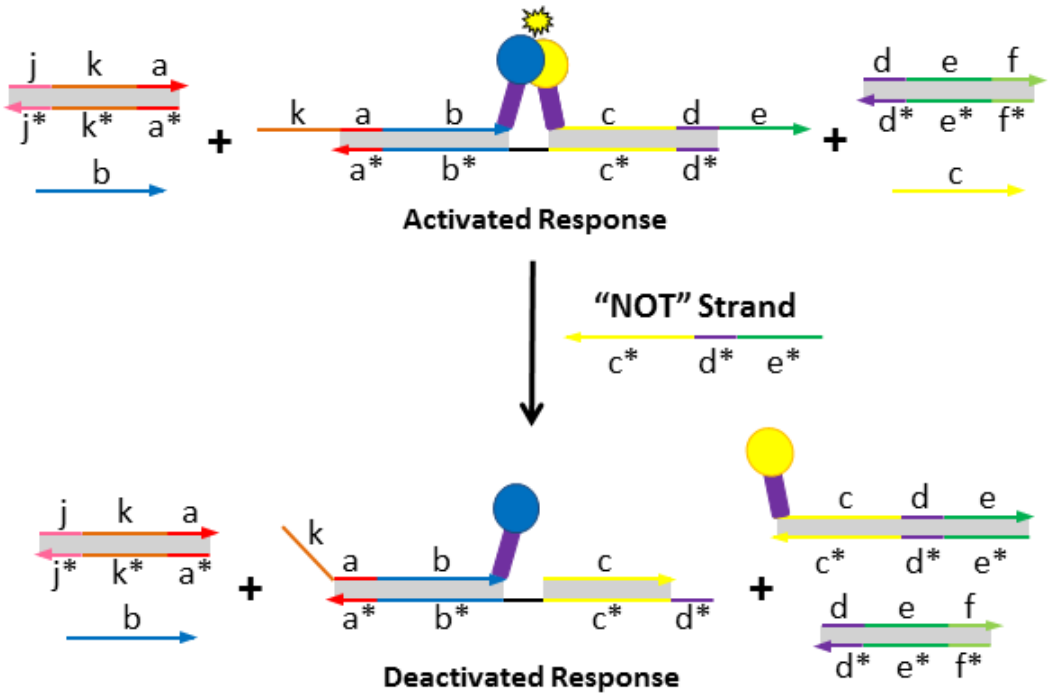

Figure 2.4 NOT reaction mechanism. Domain e (green) on the CFP-labeled strand D serves as a toehold for the displacement of CFP from the central scaffold strand C with the "NOT" strand. Adding the "NOT" strand results in a rapid disappearance of FRET as shown in **Figure 2.3B**.

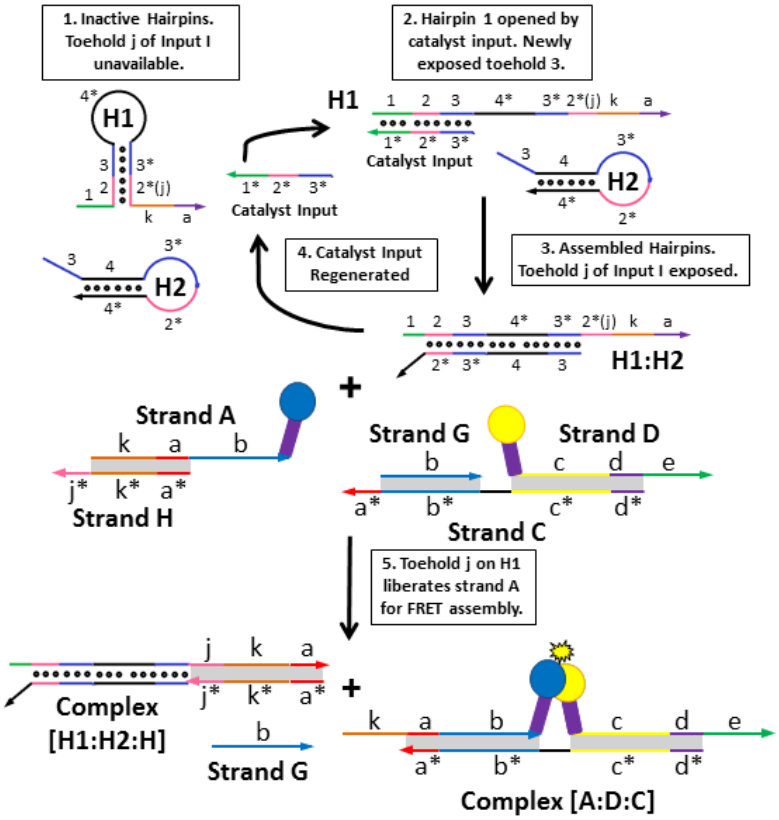

Figure 2.5 Signal amplification protein-DNA device. CHA mechanism for FRET amplification. H1 and H2 have no exposed complementary regions and thus cannot interact. However, in the presence of a specific input strand, toehold 1 (green) opens up H1 and exposes toehold 3* (blue) for the formation of the H1:H2 duplex and exposes toehold 2*(j), shown in pink. More importantly the catalytic input strand is regenerated and recycled for more duplex formation. The exposed toehold 2*(j) allows H1 to act as Input I and activate protein assembly as depicted in **Figure 2.3A**.

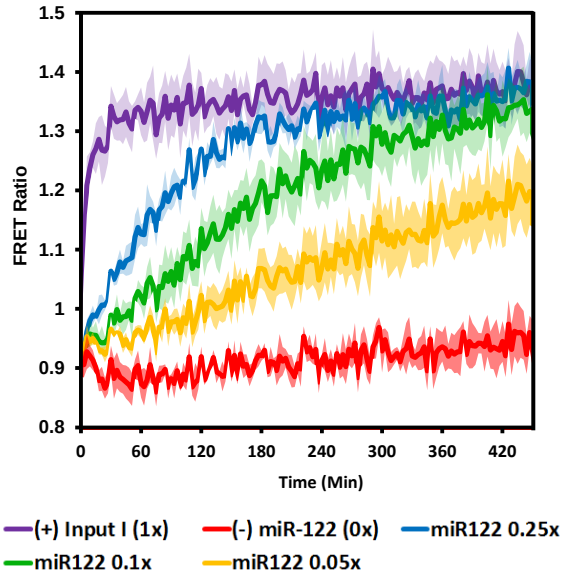

Figure 2.6 FRET response for the one-input miR-122 CHA amplification circuit. Addition of the 1X (400 nM) miRNA-122 results in a rapid FRET signal (purple). Even 0.1X (40 nM) input achieves a full FRET response (green). The shaded regions represent standard deviations for n=3.

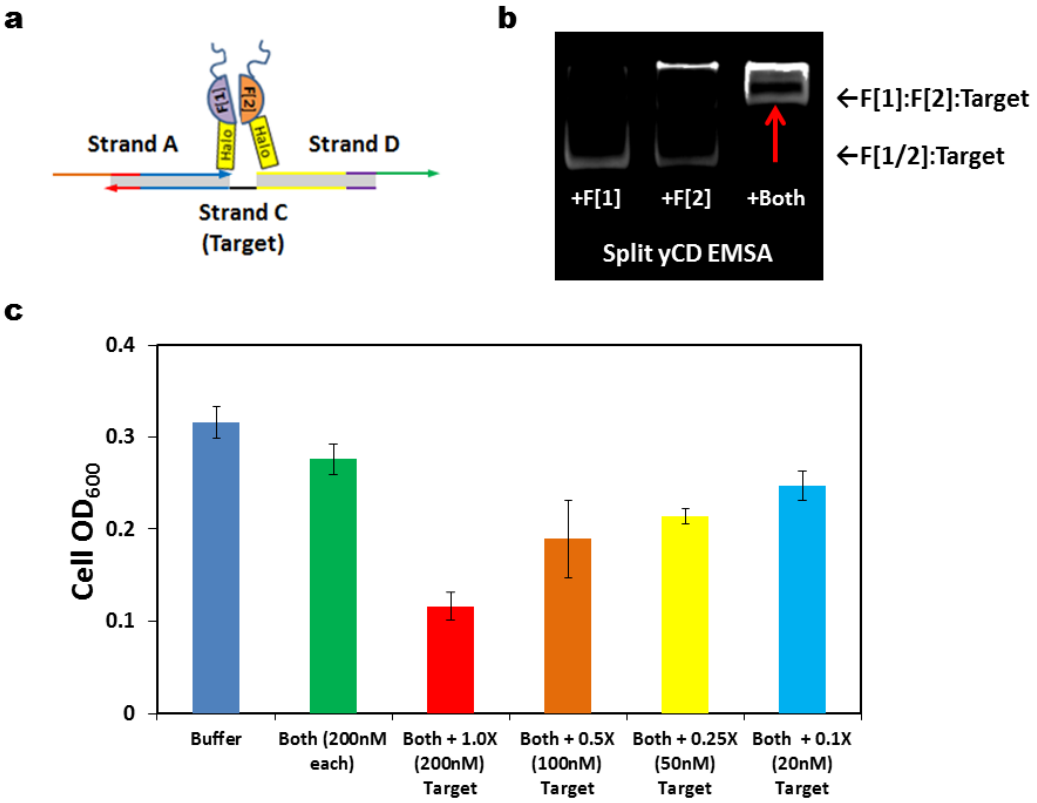

Figure 2.7 DNA binding induced split yCD reconstitution. A.) Schematic of split yCD reconstitution. F[1]yCD and F[2]yCD are labeled with strand A and D, respectively, which hybridize to strand C. **B.)** Electromobility shift assay on an ethidium bromide stained 6% native PAGE gel. Lane 1 and 2 are mixtures of F[1]yCD-A with strand C and F[2]yCD-D with strand C, respectively. The third lane is a mixture of all three components. In this lane, we observe a band shifting upwards due to the slower mobility of the dual bound split yCD complex. Because high concentrations of protein-DNA conjugates (500 nM) are needed to visualize the fluorescent bands, a fraction of the conjugates are retained inside the well and appear as a bright fluorescent band on top of the well. This situation is amplified more when both proteins are presented. It appears that F[2]yCD is more prone to aggregation, and thus we see more F[2]yCD stuck in the well relative to F[1]yCD **C.)** *In vitro* split yCD activity assay. Both F[1]yCD-A and F[2]yCD-D (200 nM) were mixed with various amounts of target strand C and the prodrug 5-FC (1mg/ml). After 12 h these reaction mixtures are incubated with *E. coli* (Δbcd) cells, and OD_{600} was measured after 14 h. In the absence of the scaffolding DNA (green) there is no yCD activity. However, with increasing amounts of target scaffolding DNA added, greater growth retardation is observed, indicating higher yCD conversion of 5-FC into toxic 5-FU. Error bars represent standard deviations (n=3).

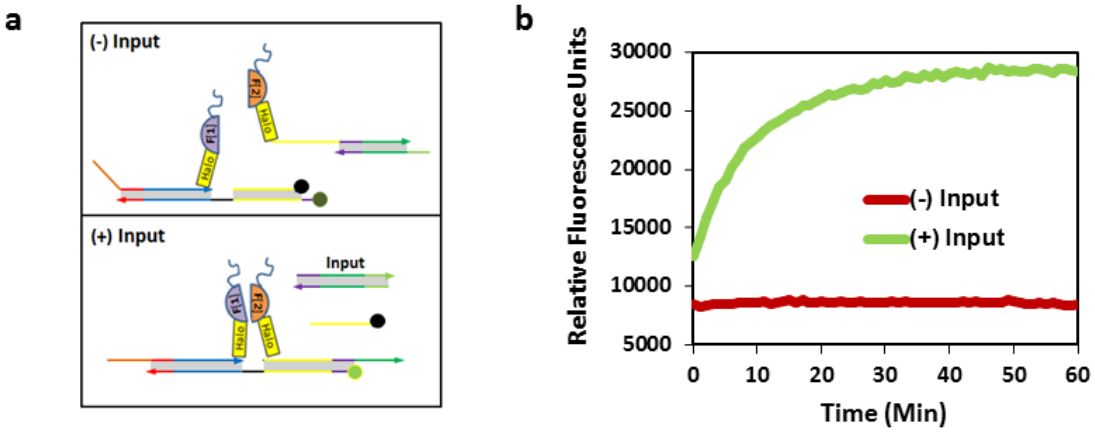

Figure 2.8 Kinetics of split yCD circuits. A.) Schematic of strand displacement guided split yCD reconstitution. 5' end of strand C and 3' end of strand B were labeled with FAM and BHQ1, respectively, so that the kinetics of strand displacement with labeled split yCD could be determined. **B.)** Without the input, no fluorescence leak was observed over time (red), but upon addition of the input the maximum fluorescence was reached within 40 min (green).

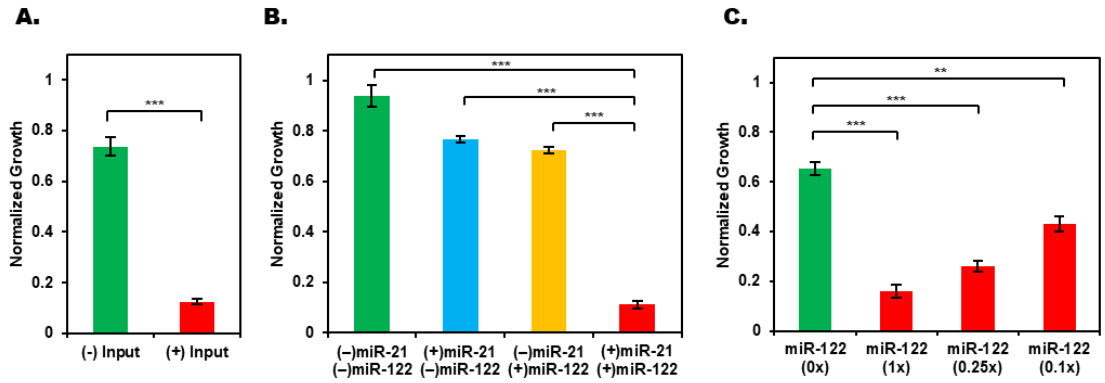

Figure 2.9 miRNA input-guided prodrug activation device. A.) Reconstitution of split yCD activity by a one-input (miR-122) circuit. The one-input split yCD circuit was allowed to proceed for 2 h before 5-FC was added. Production of 5-FU was allowed to proceed for 12 h, and the inhibition effect on the growth of GIA39(ΔbCD) *E. coli* cells (initial OD=0.05) was compared. Without the input (green), there was minimal yCD activity as observed by a small reduction in the cell growth compared to control cells with F1-A and F2-D. In the presence of 1X (900 nM) miR-122 input (red), a significant growth reduction is observed, indicating the assembly and reconstitution of split yCD. **B.)** Reconstitution of split yCD activity using a two-input (miR-21 and miR-122) circuit. In the absence of either input (green) or the presence of only one input (blue and yellow), there is minimal split yCD assembly and inhibition on cell growth. When 1X (900 nM) of both miR-21 and miR-122 are present (red), significant growth reduction is observed. **C.)** The use of a one-input (miR-122) amplification circuit for split yCD reconstitution. Slightly more background activity (green) compared to A.) due to hairpin leak. Even 0.1x (90 nM) miR-122 is sufficient to elicit a detectable growth inhibition. Full activation in 0.25x and 0.1x samples not observed due to adding 5-FC 2 h into reaction. In all experiments, error bars represent standard deviation of n=3 independent reactions. Significance was determined by unpaired Student's t-test with *** for $p \leq 0.001$ and ** for $p \leq 0.01$. All circuits were run in the presence of HEK293T cell lysate. Normalized growth is defined by the controls of F1-A and F2-D by itself (set as 1) and F1-A and F2-D directly docked to strand C (set as 0).

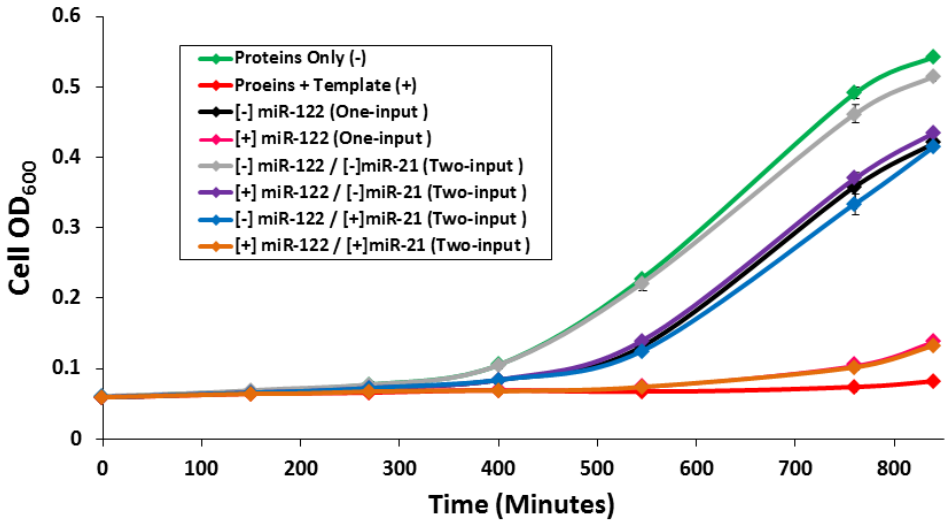

Figure 2.10 Time course of split yCD activity cell growth inhibition. All time points for GIA39 growth assay. DNA labeled split yCD by itself served as the negative control growth curve (green), while split yCD directly docked onto strand C served as the positive control of full growth inhibition (red). In the absence of the proper miRNA inputs logarithmic growth similar to the negative control was observed. However, when the target miRNAs were present almost complete growth inhibition is observed (yellow and orange). The miRNA input split yCD circuits were setup and run at room temperature for 2 h in the presence of HEK293T cell lysate. Next, the prodrug, 5-FC (1mg/ml), was added incubated for another 12 h at room temperature. These samples were then added to GIA39 cells, which were pre-cultured for 30 h before being diluted to OD=0.05 for t=0 of the growth assay. Error bars indicate standard deviation for n=3.

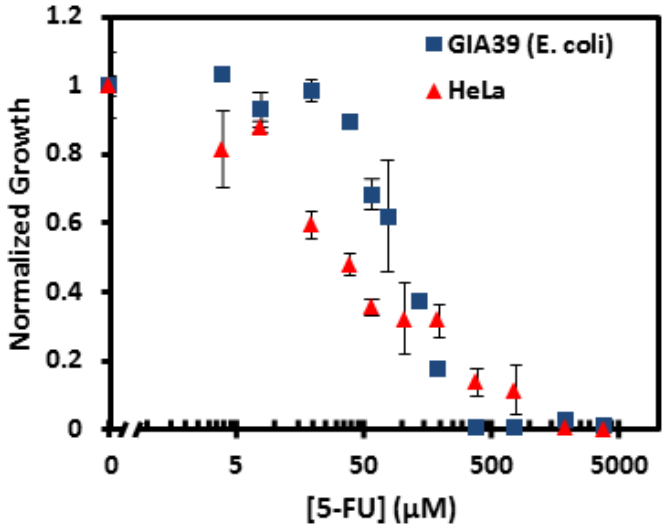

Figure 2.11 Cell growth effects of 5-FU in *E. coli* and HeLa. Growth in GIA39 was measured by OD_{600} at 14 h while in HeLa it was determined by absorbance at 570nm from a MTT viability assay. Growth is normalized between no addition of 5-FU and maximal growth inhibition observed. The growth inhibition in GIA39 correlates very closely to that observed in HeLa suggesting that *in vitro* split yCD circuit results would translate very well in mammalian cells.

REFERENCES

1. Govern, C. C. & ten Wolde, P. R. Optimal resource allocation in cellular sensing systems. *Proc. Natl. Acad. Sci.* **111,** 17486–17491 (2014).

2. Lalonde, S. *et al.* Molecular and cellular approaches for the detection of protein-protein interactions: Latest techniques and current limitations. *Plant Journal* **53,** 610–635 (2008).

3. Papapostolou, D. & Howorka, S. Engineering and exploiting protein assemblies in synthetic biology. *Mol. Biosyst.* **5,** 723 (2009).

4. Grünberg, R. & Serrano, L. Strategies for protein synthetic biology. *Nucleic Acids Res.* **38,** 2663–2675 (2010).

5. Kelwick, R., MacDonald, J. T., Webb, A. J. & Freemont, P. Developments in the Tools and Methodologies of Synthetic Biology. *Front. Bioeng. Biotechnol.* **2,** (2014).

6. Sadowski, J. P., Calvert, C. R., Zhang, D. Y., Pierce, N. A. & Yin, P. Developmental self-assembly of a DNA tetrahedron. *ACS Nano* **8,** 3251–3259 (2014).

7. Modi, S. *et al.* A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* **4,** 325–330 (2009).

8. Krishnan, Y. & Simmel, F. C. Nucleic acid based molecular devices. *Angewandte Chemie - International Edition* **50,** 3124–3156 (2011).

9. Fu, J. *et al.* Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* **9,** 531–536 (2014).

10. Sun, Q., Madan, B., Tsai, S.-L., DeLisa, M. P. & Chen, W. Creation of artificial cellulosomes on DNA scaffolds by zinc finger protein-guided assembly for efficient cellulose hydrolysis. *Chem. Commun.* **50,** 1423–1425 (2014).

11. Wilner, O. I. *et al.* Enzyme cascades activated on topologically programmed DNA scaffolds. *Nat. Nanotechnol.* **4,** 249–254 (2009).

12. Piperberg, G., Wilner, O. I., Yehezkeli, O., Tel-Vered, R. & Willner, I. Control of bioelectrocatalytic transformations on DNA scaffolds. *J. Am. Chem. Soc.*

131, 8724–8725 (2009).

13. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3,** 103–113 (2011).

14. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science (80-.).* **314,** 1585–1588 (2006).

15. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proc. Natl. Acad. Sci.* **101,** 15275–15278 (2004).

16. Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science (80-.).* **332,** 1196–1201 (2011).

17. Xin, L., Zhou, C., Yang, Z. & Liu, D. Regulation of an enzyme cascade reaction by a DNA machine. *Small* **9,** 3088–3091 (2013).

18. Liu, M. *et al.* A DNA tweezer-actuated enzyme nanoreactor. *Nat. Commun.* **4,** (2013).

19. Janssen, B. M. G., Engelen, W. & Merkx, M. DNA-directed control of enzyme-inhibitor complex formation: A modular approach to reversibly switch enzyme activity. *ACS Synth. Biol.* **4,** 547–553 (2015).

20. Prokup, A. & Deiters, A. Interfacing synthetic DNA logic operations with protein outputs. *Angew. Chemie - Int. Ed.* **53,** 13192–13195 (2014).

21. Hwang, Y.-C., Chen, W. & Yates, M. V. Use of fluorescence resonance energy transfer for rapid detection of enteroviral infection in vivo. *Appl. Environ. Microbiol.* **72,** 3710–5 (2006).

22. Blackstock, D., Sun, Q. & Chen, W. Fluorescent protein-based molecular beacons by zinc finger protein-guided assembly. *Biotechnol. Bioeng.* **112,** 236–241 (2015).

23. Cantera, J. L., Chen, W. & Yates, M. V. Detection of infective poliovirus by a simple, rapid, and sensitive flow cytometry method based on fluorescence resonance energy transfer technology. *Appl. Environ. Microbiol.* **76,** 584–588 (2010).

24. Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3,** 373–382 (2008).

25. Blackstock, D. & Chen, W. Halo-tag mediated self-labeling of fluorescent proteins to molecular beacons for nucleic acid detection. *Chem. Commun. Chem. Commun* **50,** 13735–13738 (2014).

26. Kostal, J., Mulchandani, A., Gropp, K. E. & Chen, W. A temperature responsive biopolymer for mercury remediation. *Environ. Sci. Technol.* **37,** 4457–4462 (2003).

27. Liu, F., Tsai, S. L., Madan, B. & Chen, W. Engineering a high-affinity scaffold for non-chromatographic protein purification via intein-mediated cleavage. *Biotechnol. Bioeng.* **109,** 2829–2835 (2012).

28. Li, B., Ellington, A. D. & Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic acids research* **39**, e110 (2011)..

29. Chen, X., Briggs, N., McLain, J. R. & Ellington, A. D. Stacking nonenzymatic circuits for high signal gain. *Proc. Natl. Acad. Sci.* **110,** 5386–5391 (2013).

30. Yin, P., Choi, H. M. T., Calvert, C. R. & Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451,** 318–322 (2008).

31. Jiang, Y. S., Bhadra, S., Li, B. & Ellington, A. D. Mismatches improve the performance of strand-displacement nucleic acid circuits. *Angew. Chemie - Int. Ed.* **53,** 1845–1848 (2014).

32. Erbs, P. *et al.* In vivo cancer gene therapy by adenovirus-mediated transfer of a bifunctional yeast cytosine deaminase/uracil phosphoribosyltransferase fusion gene. *Cancer Res.* **60,** 3813–3822 (2000).

33. Austin, E. A & Huber, B. E. A first step in the development of gene therapy for colorectal carcinoma: cloning, sequencing, and expression of Escherichia coli cytosine deaminase. *Mol. Pharmacol.* **43,** 380–387 (1993).

34. Ear, P. H. & Michnick, S. W. A general life-death selection strategy for dissecting protein functions. *Nat. Methods* **6,** 813–816 (2009).

35. Bandiera, S., Pfeffer, S., Baumert, T. F. & Zeisel, M. B. miR-122 – A key factor and therapeutic target in liver disease. *J. Hepatol.* **62,** 448–457 (2015).

36. Coulouarn, C., Factor, V. M., Andersen, J. B., Durkin, M. E. & Thorgeirsson, S. S. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* **28,** 3526–36

(2009).

37. Hemphill, J. & Deiters, A. DNA computation in mammalian cells: MicroRNA logic operations. *J. Am. Chem. Soc.* **135,** 10512–10518 (2013).

38. Groves, B. *et al.* Computing in mammalian cells with nucleic acid strand exchange. *Nat. Nanotechnol.* **11,** 287–294 (2016).

39. Zhang, D. Y. & Winfree, E. Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Res.* **38,** 4182–4197 (2010).

40. Chang, J. *et al.* miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol.* **1,** 106–113 (2004).

41. Laterza, O. F. *et al.* Plasma MicroRNAs as Sensitive and Specific Biomarkers of Tissue Injury. *Clin. Chem.* **55,** 1977 LP-1983 (2009).

42. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science (80-.).* **333,** 1307–1311 (2011).

43. Draghici, B. & Ilies, M. A. Synthetic nucleic acid delivery systems: Present and perspectives. *Journal of Medicinal Chemistry* **58,** 4091–4130 (2015).

44. Lächelt, U. & Wagner, E. Nucleic Acid Therapeutics Using Polyplexes: A Journey of 50 Years (and Beyond). *Chemical Reviews* **115,** 11043–11078 (2015).

45. Lehto, T., Ezzat, K., Wood, M. J. A. & El Andaloussi, S. Peptides for nucleic acid delivery. *Advanced Drug Delivery Reviews* **106,** 172–182 (2016).

46. Fu, A., Tang, R., Hardie, J., Farkas, M. E. & Rotello, V. M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjugate Chemistry* **25,** 1602–1608 (2014).

47. Pisal, D. S., Kosloski, M. P. & Balu-Iyer, S. V. Delivery of therapeutic proteins. *Journal of Pharmaceutical Sciences* **99,** 2557–2575 (2010).

Chapter 3

TUNABLE AND MODULAR MIRNA CLASSIFIER THROUGH SPLIT ASSOCIATIVE STRAND DISPLACEMENT

Abstract

The programmability of nucleic acids allows nanodevices with complex behaviors and features to be designed and engineered *de novo*. These devices are conceptually sound, but suffer major inefficiencies when adapted towards biologically relevant inputs. Often times, dynamic and structural elements are combined to build higher-order devices. These elements must be formed precisely in order to guarantee function and therefore do not tolerate sequence constraints well. Here, we devise a novel strategy called associative strand displacement to overcome this obstacle by decoupling sequence constraints between biological inputs and *de novo* strand displacement circuits. By splitting circuit inputs into their toehold and branch migration regions, and controlling their association through logic gate architectures, we demonstrate how any miRNA sequence of interest can be interfaced with *de novo* DNA circuits, including catalytic hairpin assembly and a four-input classifier. We established simple design principles that can be followed to optimize kinetics and apply thresholds.

3.1 Introduction

DNA nanotechnology serves as a promising field for developing biomolecular "computers" that process information with Boolean logic behavior[1-3]. Due to the simple but powerful concept of toehold mediated strand displacement, DNA can be "programed" to behave dynamically both with high sequence specificity and tunable kinetics. Layering and linking different strand displacement reactions with creative structural components allow diverse "circuits" to be built in which desired multi-input combinations can be used to trigger an output result.

Chapter 2 focused on expanding output functionality of DNA devices through conjugation of protein to DNA. In this chapter, the focus is shifted to input processing and transduction in DNA circuits, particularly in the context of microRNA (miRNA) inputs. These short (~22nt), non-coding, and single-stranded RNAs play a crucial role in the down-regulation of gene expression by hybridizing to the 3' untranslated region of target mRNA[4, 5]. Advances in high-throughput genomics and bioinformatics have spurred extensive efforts to characterize miRNAs and their expression profiles in different cell types[6, 7], especially in disease contexts[8-11]. These studies have revealed miRNA to be a promising biomarker and therapeutic target, and multi-input miRNA profiles can be utilized to build highly specific cell classifiers[12-14]. Mature miRNAs are especially attractive for strand displacement because they are typically more accessible than longer RNA targets whose secondary confirmations are less predictable.

Most strand displacement circuits developed so far have been rationally and computationally designed *de novo*. Nucleic acid modeling packages, such as NUPACK developed by the Pierce group, offer computational methods to model and design optimal sequences for desired reaction pathway behaviors[15, 16]. While these

advances offer powerful methods for reverse engineering synthetic DNA circuitry with ideal behavior, the challenge remains in adapting them for biologically relevant contexts. Incorporating sequence constraints from biological targets into DNA devices often compromises device function and efficiency. Linear strand displacement cascades, such as the two-input miRNA circuit from Chapter 2, have seen the most success in being adapted to biological targets[12, 17]. However, DNA devices with more complex elements, such as structural hairpin features, have seen limited success. The best approach is to find a way to interface these *de novo* circuits with biological targets without implementing sequence constraints. In this chapter, we develop a strategy in which miRNA sequences are processed and transduced into running *de novo* DNA circuits.

Previously in Chapter 2, a miR-122-gated CHA circuit was executed. The original synthetic catalyst strand was replaced with miR-122, and the hairpin sequences were changed accordingly. Even though the CHA circuit was functional and had a detectable signal in the presence of miR-122, significantly more hairpin leak was observed over the original CHA circuit. By sequence constraining to the target, the fidelity and integrity of hairpin formation was compromised. Unstable or improper hairpin folding can lead to non-functional hairpins and non-specific hairpin assembly, resulting in slower kinetics and increased background leak[18, 19]. To improve our miRNA-gated CHA design, we sought not only to relax these sequence constraints, but also to introduce multi-input functionality. Since the hairpins are the leakiest component, it would be ideal to condense the number of hairpins necessary to compute multiple inputs.

To achieve multi-input CHA, we introduced the concept of split associative strand displacement (**Figure 3.1**). This idea was inspired by split protein reporters, which similarly rely on binding interactions for proximity-based protein reconstitution[20]. In this strategy, the DNA strand of interest (in our case the CHA catalyst strand) is split into two components. Each component by itself is incapable of strand displacement. However, when brought into close proximity by a docking strand, the association of the split components allows for subsequent activation of strand displacement. Strand displacement can be split elegantly between the toehold and branch migration regions, since both are required to complete the process.

Previously, proximity-based strand displacement has been demonstrated in T-junction formations, in which the split components are designed to hybridize to each other instead of to a docking strand[19, 21, 22]. In one example, the split DNA strands were conjugated with antibodies, and only upon binding to the same antigen did the T-junction form to allow strand displacement[21]. In another study, a similar T-junction complex was applied to form an initiator strand used to run hybridization chain reaction[19]. These previous T-junction demonstrations strongly supported that a split associative strand displacement design using a docking strand would also be successful. Advantageously, replacing the T-junction with a docking strand allows for sequence independence between the two split strand components.

Here, we demonstrate that a split associative catalyst strand can be used to run CHA conditionally. The split catalyst strand was optimized for docking length and split location. Next, we developed a strategy for translating miRNA sequences into activation of the split associative catalyst through an indirect associative strand displacement scheme (**Figure 3.5**). The split catalyst components are pre-hybridized

with blocking strands to prevent association with the docking strand. The presence of target miRNA triggers the release of the split catalyst strands from the blocking strand, allowing split catalyst activation on the docking strand. We demonstrate how the facile design principles of toehold length and clamping allow for fine tuning of circuit behavior. Multi-input CHA is demonstrated for the first time with AND, NOT, and NOTAND logic behaviors. Furthermore, the associative strand displacement strategy is generalized beyond CHA to run a four-input miRNA classifier. Ultimately, this novel strategy offers a modular framework in which any miRNA of interest can be adopted to run logic-gated strand displacement. This powerful method allows for the elegant interfacing of endogenous targets with *de novo* DNA devices.

3.2 Materials and Methods

3.2.1 Preparation of DNA complexes

DNA oligos were purchased from Integrated DNA Technologies with standard desalting unless otherwise indicated. Oligo sequences along with specific strand modifications are listed in **Table 3.1**. All DNA was reconstituted in hybridization buffer (20mM Tris, 150mM NaCl, 5mM $MgCl_2$, pH 7.4). DNA mimics of miRNA targets were used in all experiments in this study.

Pre-hybridized complexes were prepared by heating followed by slow cooling in a thermocycler (BioRad) as follows. Complexes were prepared in hybridization buffer and then heated to 95°C for 10 min. Afterwards the solution was slowly cooled to 4°C at a ramp rate of 0.1°C/sec. DNA complexes were then stored at 4°C overnight before use. H1, H2, and the reporter complex were prepared each separately at 30µM. Blocked staples and/or docking strand complexes were prepared at 10x the final

concentration. For the four-input miRNA classifier, all complexes were prepared at 10x concentrated.

3.2.2 Circuit running conditions

All CHA experiments were run at 37°C. A ratio of 1:4 was used for catalyst components to hairpins. The fluorophore/quencher reporter duplex, H1, and H2 were all used at a final concentration of 1μM. The one-strand catalyst, various associative catalyst staples, and docking strand were used at 250nM. The blocking strands were added in 1.2x (300nM) excess of the staples, and the inputs were added in 1.2x (360nM for AND, 300nM for NOT) excess of the blocking strands. The final concentration of miRNA input was denoted with the Boolean logic value of 1. For the NOT-gate a logic value of 4 was also tested and is equivalent to 1200nM. In general, a master mix of reporter duplex, H1, and H2 was first prepared and added to wells. Next, various logic gate components were added. The miRNA inputs were added last and then immediately assayed on the plate reader.

The four-input miRNA classifier was run at 25°C. The blocked central scaffold complex was prepared with 400nM of scaffold strand and 800nM of blocking strands. The blocked fluorophore and blocked quencher duplexes were prepared at 750nM of blocking strand and 500nM of fluorophore or quencher strand. NOTAND circuit components were prepared at 900nM for docking and staple strands. For staple blocking strands, 1080nM was used. The miRNA inputs were added at 2x (1800nM for NOT and 2160nM for AND) and these concentrations correspond to the Boolean value of 1.

3.2.3 Fluorescence measurements of DNA circuits

For the characterization of associative catalyst designs, samples were set up at 100µl per well in black 96-well plates. Fluorescence measurements were taken using the Synergy H4 (BioTek) plate reader at a sensitivity setting of 85. Excitation and emission wavelengths of 470nm and 520nm, respectively, were used to measure FAM fluorescence. Fluorescence readings were taken every 30 sec for the first 2 h and afterwards every 2 min. For the four-input miRNA classifier, 25µl samples were set up in clear PCR tubes and caps for optical reactions. Fluorescence measurements were taken on the FAM channel using a qPCR machine (BioRad, CFX96). Measurements were taken every 5 min.

3.3 Results and Discussion

3.3.1 Conditional behavior of split associative catalyst

The behavior of the split associative catalyst was assayed by measuring fluorescence of a fluorophore/quencher reporter duplex. Upon hairpin assembly, the newly exposed toehold (2*) domain on H1 strand displaces the fluorophore strand from the quencher strand (**Figure 3.2**). Fluorescence measurements were normalized by subtracting the signal of the CHA hairpins alone. Hairpin leak was observed over time, and this phenomenon has been studied in great detail previously[18, 19]. While significant to overall CHA behavior, it does not affect the associative catalyst design itself, and thus it is not a focus here.

In the earliest design, we envisioned using miRNA directly as the staples for our split catalyst. This idea was abandoned quickly not only because of the sequence constraints it placed on the CHA circuit, but also because miRNAs (~22nt) are too short. When a 22nt length constraint was used, only a 10nt hybridization domain

remained after 12nt was used for the catalyst component (one-strand catalyst is 24nt). The 10nt docked split catalyst (red) was found to turn on the CHA circuit faster than when only a single staple (purple and orange) was present (**Figure 3.3A**). However, the CHA activation of the split catalyst was extremely slow, and staple 1 (purple) also resulted in hairpin assembly. It was obvious that a 22nt length constraint was not feasible, and longer docking regions were necessary for efficient hybridization and formation of the split catalyst complex. The split location also needed to be closer to the toehold and branch migration juncture to improve conditionality and eliminate the leak observed in staple 1.

3.3.2 Optimization of docking region length and split location

The kinetic defects observed in the initial design were attributed to inefficient hybridization and effectively low concentration of split catalyst formation. Increasing the docking length improved hybridization, and CHA kinetics greatly increased (**Figure 3.3**). A maximum split catalyst initial rate was achieved with a 16nt docking region, and the 20nt version offered no further improvement (**Figure 3.3B-C**). 50% GC content was used, so a 16nt length serves as a good heuristic, but for specific sequences of interest, melting temperature provides a more quantitative metric for determining docking length.

The leak observed from staple 1 in the initial design was unsurprising since it contained a significant portion of the catalyst branch migration region. Staple 1 opens the stem of H1 halfway, destabilizing it enough for hairpin assembly to occur slowly. To eliminate this leak, the split location was shifted closer to the junction of the toehold and branch migration domains. This junction was denoted as location (0), while split locations invading into the branch migration region are referred to as (+1),

(+2), etc. The initial split location was (+4), and moving towards location (0), decreasing staple 1 leak was observed (**Figure 3.4A**). At the (+1) location no leak was observed even over long time periods (>12h), and this split location was used for future experiments. Overall, CHA completion took ~90min with the optimized split catalyst, three times slower than the one-strand catalyst (**Figure 3.4B**).

3.3.3 Design of indirect associative strand displacement scheme (AND-Gate)

With a two-input conditional split catalyst developed and optimized, we moved forward to interface the split catalyst with miRNA sequences. The catalyst part of the staple was kept constant while the docking region was adapted for miRNA compatibility. Because the docking region does not have structural constraints, its sequence is more easily exchanged without consequence. Using a novel scheme called indirect associative strand displacement (**Figure 3.5**), miRNA inputs served as inputs that trigger the strand displacement of a blocking strand from the staple through toehold exchange[23]. The unblocked staple was then able to hybridize to the docking strand for associative catalyst formation. An AND-gated CHA circuit was created by blocking both staples.

In this scheme, there is non-trivial competition between the miRNA and the staples for the docking strand since they share the same complementarity to the docking strand (**Figure 3.6**). To combat this competition, clamps[24] (made up of Gs and Cs) were added to the ends of the design components. By doing so, the staples now have greater complementarity than the miRNA for the docking strand. Secondly, the clamp can serve as a toehold, and the staple can strand displace competing miRNA off the docking strand (**Figure 3.6**). Thirdly, the clamp serves to stabilize the blocked staple complex and prevent leak.

In this study, DNA mimics of miRNA targets were used for initial proof of concept since DNA is less prone to degradation from user handling. From previous studies of our own and others, qualitative behavior of strand displacement between RNA and DNA is consistent[25, 26]. However, using RNA will result in slower kinetics of the indirect associative strand displacement circuit, since RNA-DNA interactions are stronger than that of DNA-DNA[27]. A greater contribution to competitive hybridization of the docking strand can be expected for RNA.

3.3.4 Optimization of clamp length for AND-gated CHA

The inclusion of a clamp allowed for miRNA competition for the docking strand to be combatted (**Figure 3.6**), but it also introduced a toehold exchange reaction for the release of the staple from the blocking strand (**Figure 3.7**). These two processes are at odds had to be balanced to find the optimal clamp length for fastest CHA kinetics. For minimizing competition, a longer clamp is desired because it makes staple hybridization more thermodynamically favorable than miRNA hybridization. Meanwhile, in a toehold exchange reaction, the existence of an incumbent toehold (clamp) and an invading toehold (Th1) on the blocking strand creates a reversible seesaw reaction (**Figure 3.7**). Both the staple and the miRNA have unique toeholds that can displace the other from the blocking strand. For equilibrium to favor the release of the staple, the invading toehold from the miRNA must be stronger than the incumbent toehold of the clamp[23]. Thus, for toehold exchange a shorter clamp is favorable.

Clamp length was varied for the indirect associative catalyst AND-gate. The AND-gate was designed for the same two-input combination as Chapter 2, miR-122 and miR-21. A 6nt toehold and 16nt branch migration region were utilized for each

miRNA. When the docking strand and staples were mixed, clamp length had little effect on CHA kinetics (**Figure 3.8A**). This was expected since a 16nt docking domain was already sufficient without the clamp. For the same reason, when the docking strand and blocked staples were combined, no leak was observed regardless of clamp length (**Figure 3.8B**). However, in both these cases, if a hypothetical target sequence is particularly AT-rich having a clamp would supplement weak hybridization.

Next, the docking strand was mixed with the staples (without blocking strands) and miRNA targets to look at competitive hybridization (**Figure 3.8C**). For all clamp lengths, CHA was significantly slower than in the absence of miRNA because of competition for the docking strand. Without a clamp (grey) very minimal CHA was observed, but as clamp length was increased, CHA kinetics also increased, as staples outcompeted the miRNA for the docking strand. Lastly, the docking strand and blocked staples were combined with the two miRNA targets (**Figure 3.8D**). Clamp length had a bell-shaped trend with the 5nt clamp being optimal. Despite the benefits of the longer clamp (6nt and 7nt) observed in **Figure 3.8C**, the clamp is too long for toehold exchange equilibrium to favor efficient release of the staple from the blocking strand. For the shorter clamp designs (≤5nt), CHA was faster than in **Figure 3.8C**, since blocking strands are present to sequester miRNA for reduced competitive hybridization with the staples for the blocking strand. The 5nt clamp being optimal corroborates nicely the principles of toehold exchange and completive hybridization. It is the longest clamp length without exceeding the miRNA toehold length (6nt). As a general heuristic, we recommend a clamp length 1nt shorter than the miRNA toehold length to maximize its kinetic benefits. Using the 5nt clamp, the AND-gated CHA was

tested against the full truth table, and CHA was only observed when both miR-122 and miR-21were present (**Figure 3.9**). Hairpin assembly reached completion in ~60min, only twice as long as the docked staples positive control.

3.3.5 NOT-gated CHA with threshold tuning

For a NOT-gated associative catalyst, the staple is initially hybridized to the docking strand. A toehold on the miRNA initiates strand displacement of the staple from the docking strand, thus disassembling the associative catalyst (**Figure 3.10A**). No clamp is necessary here since there are no competing hybridization events. Because the associative catalyst is initially active, how fast strand displacement occurs will determine how stringent the NOT-gate is. Speed of strand displacement is increased by longer toeholds and higher strand concentrations[25]. To demonstrate how the threshold of the NOT-gate can be tuned, a 6nt and 8nt toehold version of a NOT miR-141 gated associative catalyst was tested at various concentrations of the input (**Figure 3.10B**). With the 6nt toehold, CHA is not turned off effectively with the lower concentration of miR-141 (dark red). However, in the 8nt toehold design, the lower concentration is able to turn off CHA after a small initial increase in fluorescence (light green). In both designs, adding 4-fold more miR-141 caused strand displacement and disassembly of all initial associative catalysts almost instantaneously (bright red and dark green). This provides a framework for tuning the threshold of NOT-gated behavior by varying toehold length.

3.3.6 NOTAND-gated CHA

The AND-gate and NOT-gate designs for the associative catalyst were combined to create a NOTAND-gated CHA circuit (**Figure 3.11A**). In this scheme,

staple 1 is pre-hybridized to the docking strand, and staple 2 is pre-hybridized to a blocking strand. Only if miRNA 1 is absent (preserving staple 1 on the docking strand) and miRNA 2 is present (unblocking staple 2 and allowing hybridization to the docking strand) will the associative catalyst form. A NOT miR-141 AND miR-21 design was constructed and tested against all input combinations. Only in the absence of miR-141 and presence of miR-21 did CHA occur (**Figure 3.11B**). When both miR-141 and miR-21 were added there was no initial leak, unlike the NOT miR-141 design in **Figure 3.10B**. The time delay for the strand displacement of staple 2 from the blocking strand eliminates initial catalyst activity before the NOT displacement is completed.

3.3.7 Four-input miRNA classifier

Thus far, the focus has been on utilizing the associative strand displacement strategy for creating a conditional catalyst strand for running CHA. However, the strategy and design principles outlined so far are generalizable to any strand displacement reaction of interest. To demonstrate this, we revisited our CFP/YFP two-input circuit (**Figure 2.3A**) from Chapter 2. The circuit was easily expanded to four inputs by splitting each of the original inputs (Strand I and F) at their toehold and branch migration junctures to form staples that can be logic gated by the indirect associative strand displacement strategy (**Figure 3.12**). Each associative input was NOT-AND gated by two miRNA targets for a total of four unique targets. For easy characterization purposes, a fluorophore and quencher were placed at the original protein attachment locations (Strand D and Strand A). Only when the correct four-input condition is fulfilled can the fluorophore and quencher strand assemble on the

scaffold strand for fluorescence quenching. This is unlike the previous CHA circuits shown in this chapter; here, activation results in an observed decrease in fluorescence.

A four-input miRNA classifier was designed to identify a NOT mir-141, miR-21, NOT miR-146, AND miR-17 condition. This miRNA combination has been previously used to successfully distinguish between several cancer cell lines[13]. All possible Boolean combinations of the four miRNA were challenged to the four-input circuit, and an optimal reaction end point of 6 hours was chosen for fluorescence measurement (**Figure 3.13**). A room temperature condition was chosen to mimic the use of the miRNA classifier as a simple diagnostic assay. Only the desired miRNA combination resulted in a significant 70% drop in fluorescence, while other conditions had minimal leak.

The circuit behaved slower than the gated-CHA designs for a number of reasons including the greater number of strand displacement reactions, the lack of an amplification step, and most importantly a lower running temperature (25°C versus 37°C). We envision that a patient RNA sample can be mixed with the circuit in a one pot reaction. After reaction, measured fluorescence is converted to a simple binary diagnostic output without extensive data processing. Furthermore, the panel of miRNA assayed can be easily expanded in the same pot by using orthogonal fluorophores to run orthogonal circuits in parallel.

3.4 Conclusion

Computational advances have enabled complex DNA strand displacement circuits and devices to be designed *de novo* for optimal kinetics, high efficiency, and robustness. However, these synthetic devices cannot readily accept biologically relevant targets as device inputs. Often, when DNA nanotechnology is adapted from

proof-of-concept to biological contexts, device performance is compromised due to sequence constraints. In order to overcome this, we have developed a novel translational strategy to connect biological targets to *de novo* devices. Our associative strand displacement approach was used achieve miRNA logic-gated CHA. Docking length and split location were optimized for an associative catalyst that was only 3-fold slower than the one-strand catalyst. Next, various indirect associative catalyst designs were executed with AND, NOT, and NOTAND behavior. Toehold length and clamp length serve as tunable parameters for adapting any miRNA of interest. Above all, the indirect associative strand displacement strategy completely sequence decoupled CHA circuit components from miRNA inputs. This powerful strategy is generalizable to any strand displacement reaction, and we used it to expand a two-input circuit to a four-input miRNA classifier. The miRNA classifier successfully identified the correct miRNA condition, and provides an initial basis for creating simple diagnostic assays.

Overall, this study lays out the framework and design principles for associative strand displacement. Future characterization and optimization with actual miRNA instead of DNA mimics will be necessary to better determine the kinetics and sensitivity of our circuits, particularly for diagnostic applications. For the four-input circuit, integrating CHA for amplification would help achieve greater sensitivity to detect low target concentrations. One possible improvement for future iterations is to scale-down the number of DNA strands involved. In the current design, going from the two-input to four-input miRNA classifier requires 8-10 additional strands. If the staple and blocking strand duplex is converted into a hairpin or molecular beacon design, the two strands can be collapsed into one.

TABLES

Table 3.1 Oligo sequences used in associative strand displacement. Regions corresponding to the split input are underlined. For the 4-input circuit, strand names used for the two-input circuit in Chapter 2 are labeled in parentheses.

Oligo	Sequence (5' to 3')
CHA	
H1	GT CAG TGA GCT AGG TTA GAT GTC GCC ATG TGT AGA CGA CAT CTA ACC TAG CAC TTG TCA TAG AGC AC
H2	AGA TGT CGT CTA CAC ATG GCG ACA TCT AAC CTA GCC CAT GTG TAG A
F	5' FAM AGTGCTCTATGACAAGT GCTAGGTTA
Q	ACTTGTCATAGAGCAC T 3'BHQ1
1-strand Cat	CGACATC_TAACCTAGC_TCACTGAC
Docking Length Optimization	
Staple 1	CGCACTACTGGAGTGTGACA TAGCTCACTGAC
Staple 2	CGACATCTAACC GACTGATGTTGATGGCATCC
Dock 10	AACATCAGTC TGTCACACTC
Dock 12	TCAACATCAGTC TGTCACACTCCA
Dock 16	GCCATCAACATCAGTC TGTCACACTCCAGTAG
Dock 20	GGATGCCATCAACATCAGTC TGTCACACTCCAGTAGTGCG
Split Location (used with Dock 20)	
Staple 1 (+3)	CGCACTACTGGAGTGTGACA AGCTCACTGAC
Staple 2 (+3)	CGACATCTAACCT GACTGATGTTGATGGCATCC
Staple 1 (+2)	CGCACTACTGGAGTGTGACA GCTCACTGAC
Staple 2 (+2)	CGACATCTAACCTA GACTGATGTTGATGGCATCC
Staple 1 (+1)	CGCACTACTGGAGTGTGACA CTCACTGAC
Staple 2 (+1)	CGACATCTAACCTAG GACTGATGTTGATGGCATCC
miR-122 AND miR-21 (5nt clamp)	
Docking Strand	GCGGCTCAACATCAGTCTGACATTGTCACACTCCACCCGC
Staple 1	GCGGGTGGAGTGTGACAATGCTCACTGAC
Block 1	CAAACACCATTGTCACACTCCACCCGC
Staple 2	CGACATCTAACCTAGTCAGACTGATGTTGAGCCGC
Block 2	GCGGCTCAACATCAGTCTGATAAGCTA
NOT miR-141/ NOT miR-141 AND mir-21	
Staple 1 (6nt toehold)	TGTCTGGTAAAGATGGCTCACTGAC
Staple 1 (8nt	TCTGGTAAAGATGGCTCACTGAC

toehold)	
Staple 2	CGACATCTAACCTAGTCAGACTGATGTTGAGCCGC
Block 2	GCGGCTCAACATCAGTCTGATAAGCTA
Docking Strand	GCGGCGGCTCAACATCAGTCTGACCATCTTTACCAGACAGTGTTA
4-input	
NOT miR-141 AND mir-21	
Staple 1	TGTCTGGTAAAGATGGCTTTCCTACACCTACG
Staple 2	ACCATATCAGACTGATGTTGAGCCGC
Block 2	GCGGCTCAACATCAGTCTGATAAGCTA
Docking Strand	GCGGCGGCTCAACATCAGTCTGACCATCTTTACCAGACAGTGTTA
NOT miR-146 AND mir-17	
Staple 1	TCTCCAAGCAACCTCA TGAGAACTGAATTC
Staple 2	CGCCGCAAAGTGCTTACAGTACTGACAC
Block 2	CTACCTGCACTGTAAGCACTTTGCGGCG
Docking Strand	AACCCATGGAATTCAGTTCTCAACTGTAAGCACTTTGCGGCG
Detection Scheme	
Fluorophore Strand (Strand D)	5'FAM CCCTCATTCAATACCCTACGTCTCCAAGCAACCTCA
F. Blocking Strand (Strand E)	GTGTCAGTTGAGGTTGCTTGGAGA
Quencher Strand (Strand A)	CTTTCCTACACCTACGTCTCCAACTAACTTACGG 3'BHQ1
Q. Blocking Strand (Strand H)	CGTAGGTGTAGGAAAGTATGGT
Scaffold Strand (Strand C)	TGGAGACGTAGGGTATTGAATGATTTCCGTAAGTTAGTTGGAGACG TAGG
Scaffold Blocking Strand 1 (Strand G)	TCTCCAACTAACTTACGG
Scaffold Blocking Strand 2 (Strand B)	CCCT CATTCAATACCCTACG

FIGURES

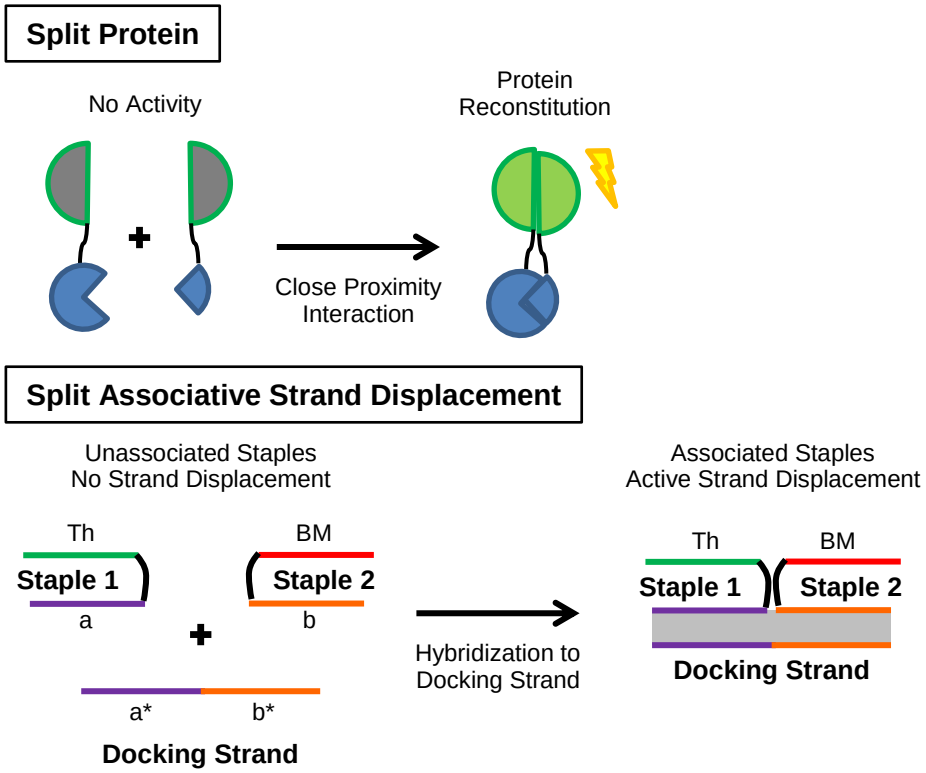

Figure 3.1 Associative strand displacement scheme. Analogous to split protein systems, in which protein fragments have no activity until brought into close proximity for protein reconstitution, the split associative strand displacement strategy splits the DNA strand of interest at its toehold (Th) and branch migration (BM) junction. The split strands are linked to domains that allow the staples to reassemble upon hybridization to the docking strand. Only when the split strand is associated can strand displacement occur.

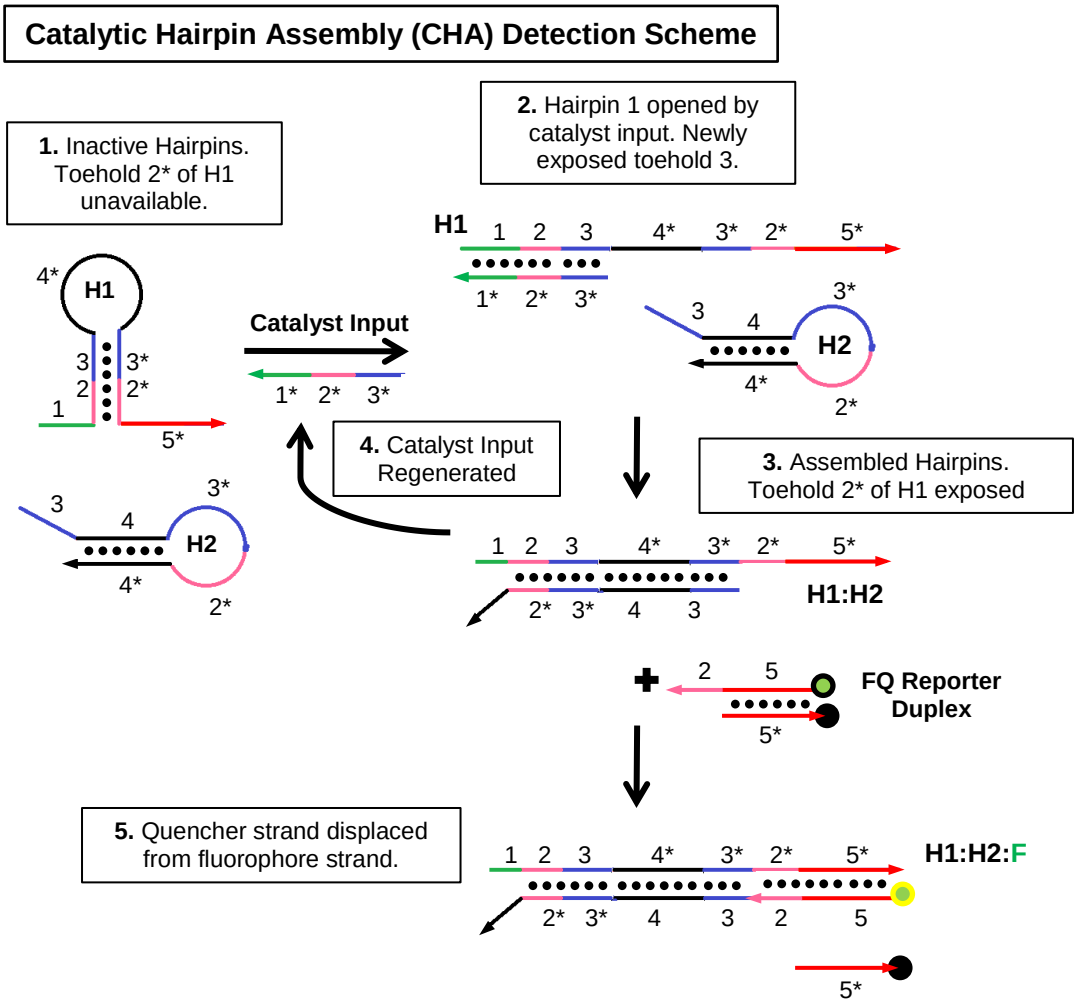

Figure 3.2 Detection scheme for catalytic hairpin assembly (CHA). H1 and H2 hairpins have sequestered toeholds until the catalyst input binds and opens H1. A newly exposed toehold allows H2 and H1 to assemble, while simultaneously displacing and regenerating the catalyst input. When H1 is opened, the exposed toehold (2*) is used to initiate strand displacement of the fluorophore strand from the quencher strand in a reporter duplex. Increase in florescence is used to track and quantify CHA.

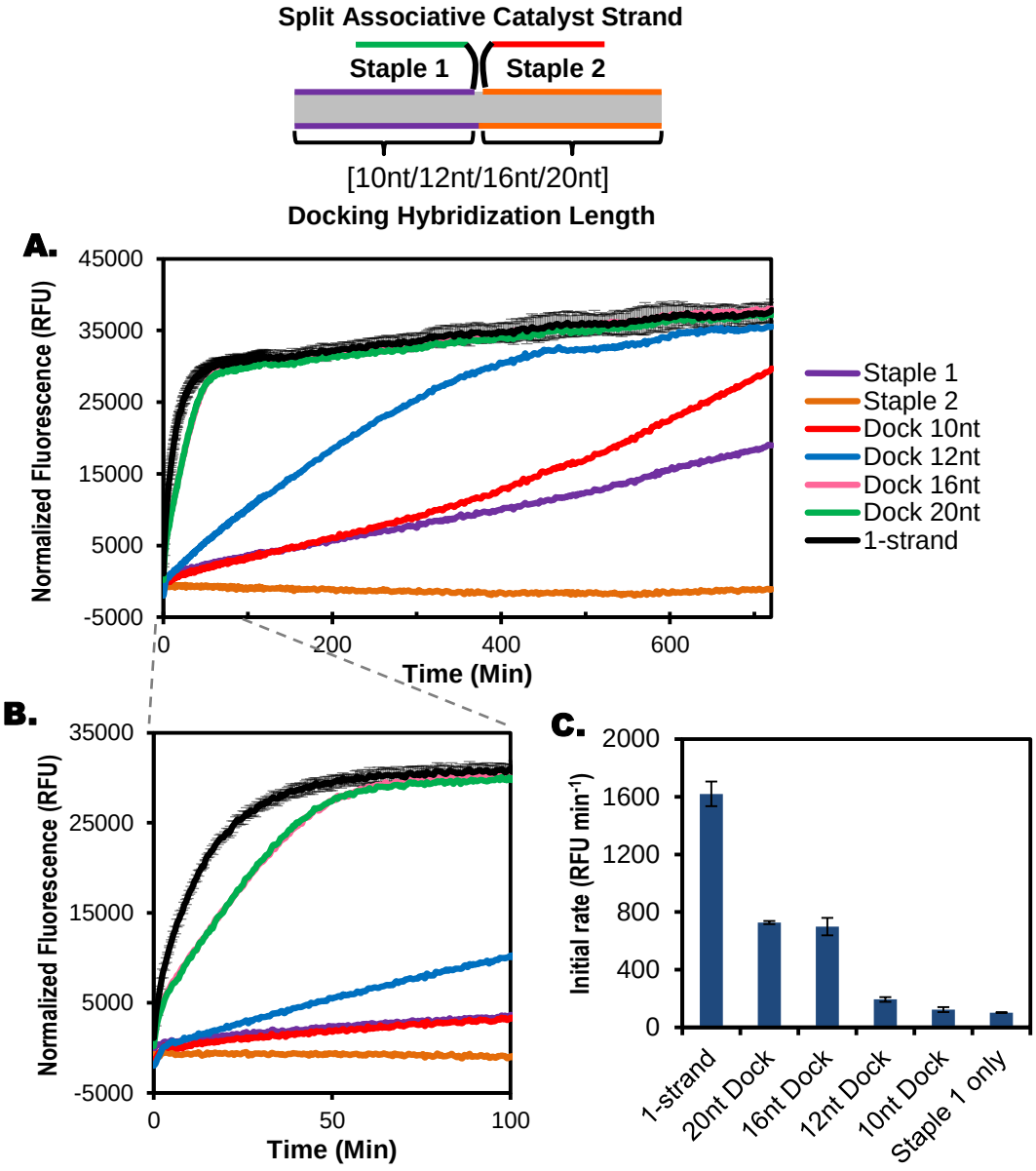

Figure 3.3 Effect of docking hybridization length. A.) Various hybridization domain lengths (10nt, 12nt, 16nt, and 20nt) were tested for split associative catalyst activity. The shorter the hybridization domain the slower CHA occurred. The one-strand catalyst (black) was used as a positive control for comparison. Negative controls of staple 1(purple) or staple 2(orange) alone indicated that staple 1 by itself was leaky. For clarity, error bars representing standard deviation (n=3) were only shown for the one-strand catalyst but are representative of other samples. Other traces are the average values obtained from triplicate samples. **B.)** Fluorescence plotted on a shorter time scale. **C.)** Initial rate of various catalyst configurations.

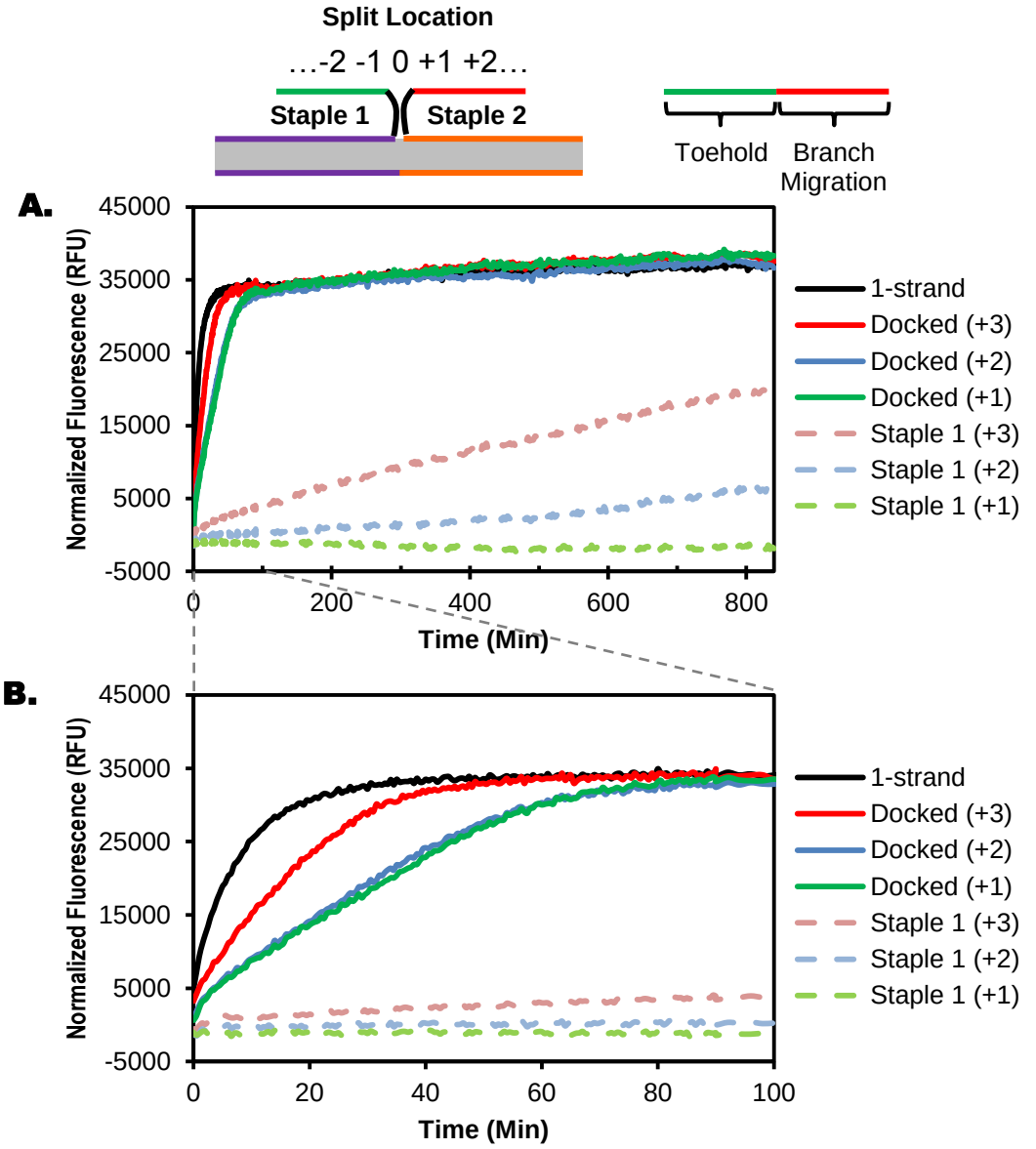

Figure 3.4 Optimization of split location. A.) The location at which the catalyst strand was split was varied at the (+1), (+2), and (+3) locations. Location (0) is the junction between the toehold and branch migration domains). Staple 1 by itself was still capable of initiating hairpin assembly in the (+3) and (+2) locations (dashed pink and dashed blue). At location (+1), no staple 1 leak is observed (dashed green). **B.)** Fluorescence plotted on a shorter time scale for better visualization of the kinetics of the various docked split catalysts.

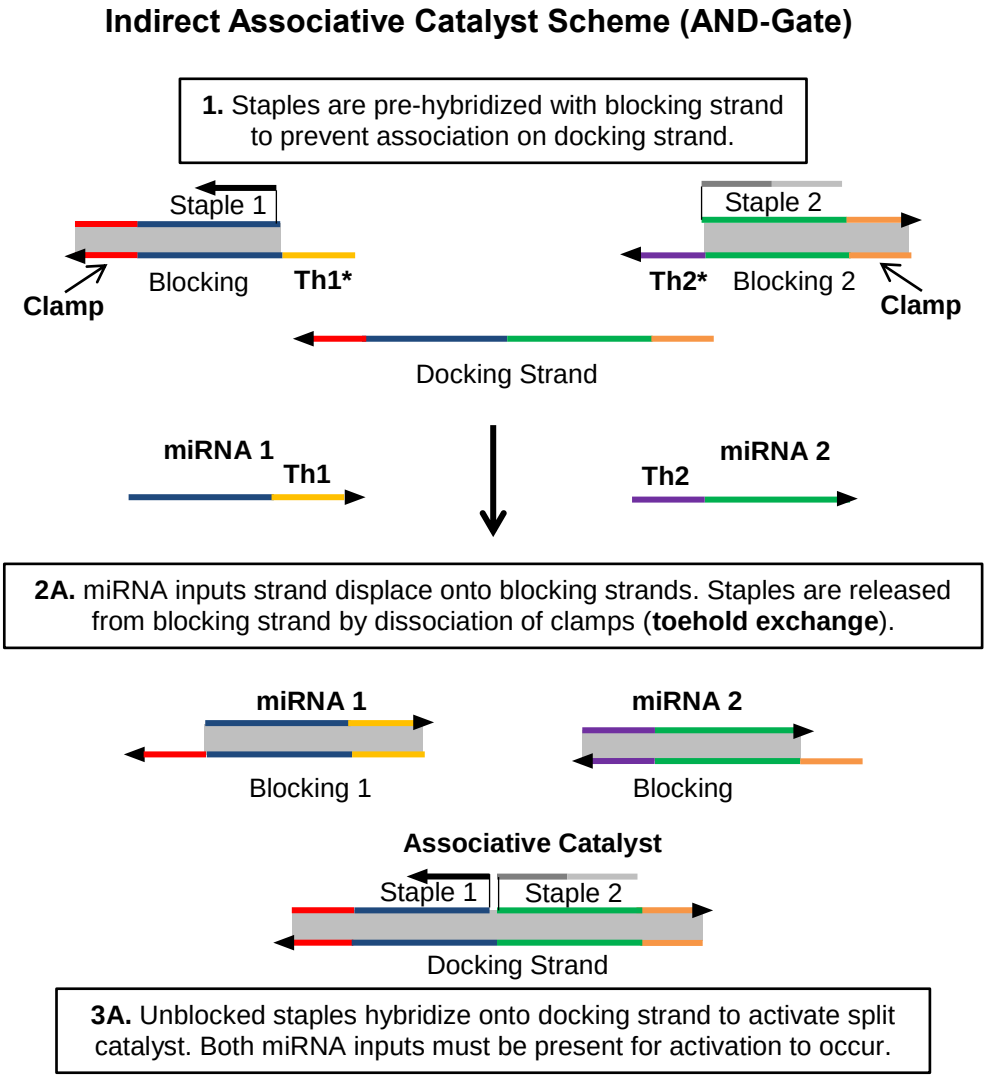

Figure 3.5 Indirect associative catalyst scheme (AND-Gate). Step 1.) Staple strands are pre-hybridized to blocking strands to prevent initial hybridization to docking strand. Clamps placed on ends serve increase duplex stability and limit competitive hybridization (refer to **Figure 3.6**). **Step 2A.)** MicroRNA inputs strand displace onto the blocking strands. The staples are released through dissociation of clamps by mechanism of toehold exchange (refer to **Figure 3.7**). **Step 3A.)** Unblocked staples hybridize onto docking strand to activate split catalyst. Both miRNA inputs must be present for activation to occur. Refer to **Figure 3.6** for alternative reaction pathway.

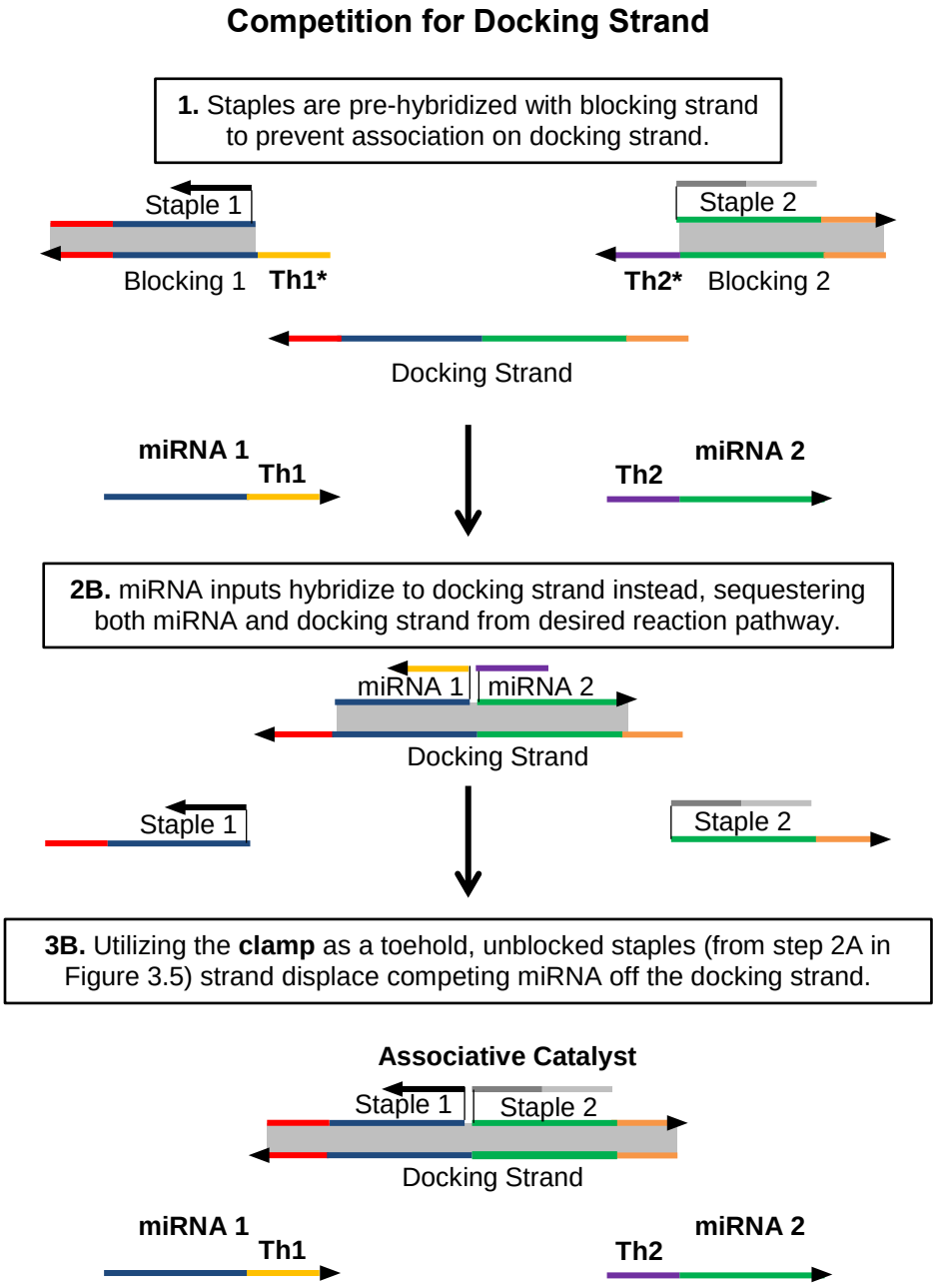

Figure 3.6 Reaction pathway of competitive hybridization to docking strand. Step 2B.) miRNA inputs hybridize to docking strand instead, sequestering both miRNA and docking strand from desired reaction pathway. **Step 3B.)** Utilizing the clamp as a toehold, unblocked staples (from step 2A in **Figure 3.5**) strand displace competing miRNA off the docking strand.

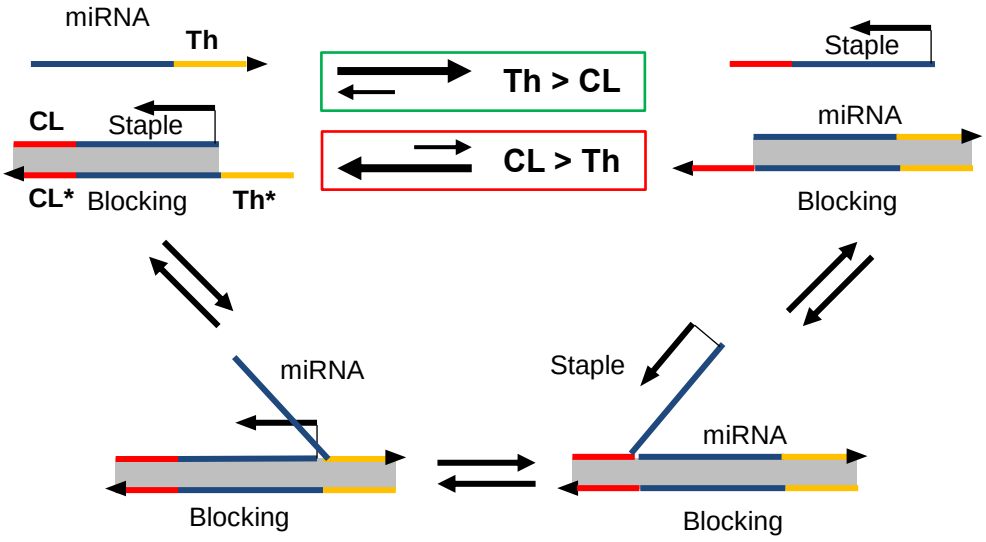

Figure 3.7 Mechanism of toehold exchange. Toehold exchange is a reversible process since both the invading miRNA and the incumbent staple have toehold domains, indicated as Th (yellow) and CL (red). Both can initiate toehold binding to the blocking strand and displace the other. Equilibrium will favor the strand with the stronger toehold. Therefore, in order to drive the reaction in the direction of staple release, the miRNA toehold needs to be stronger than that of the staple clamp.

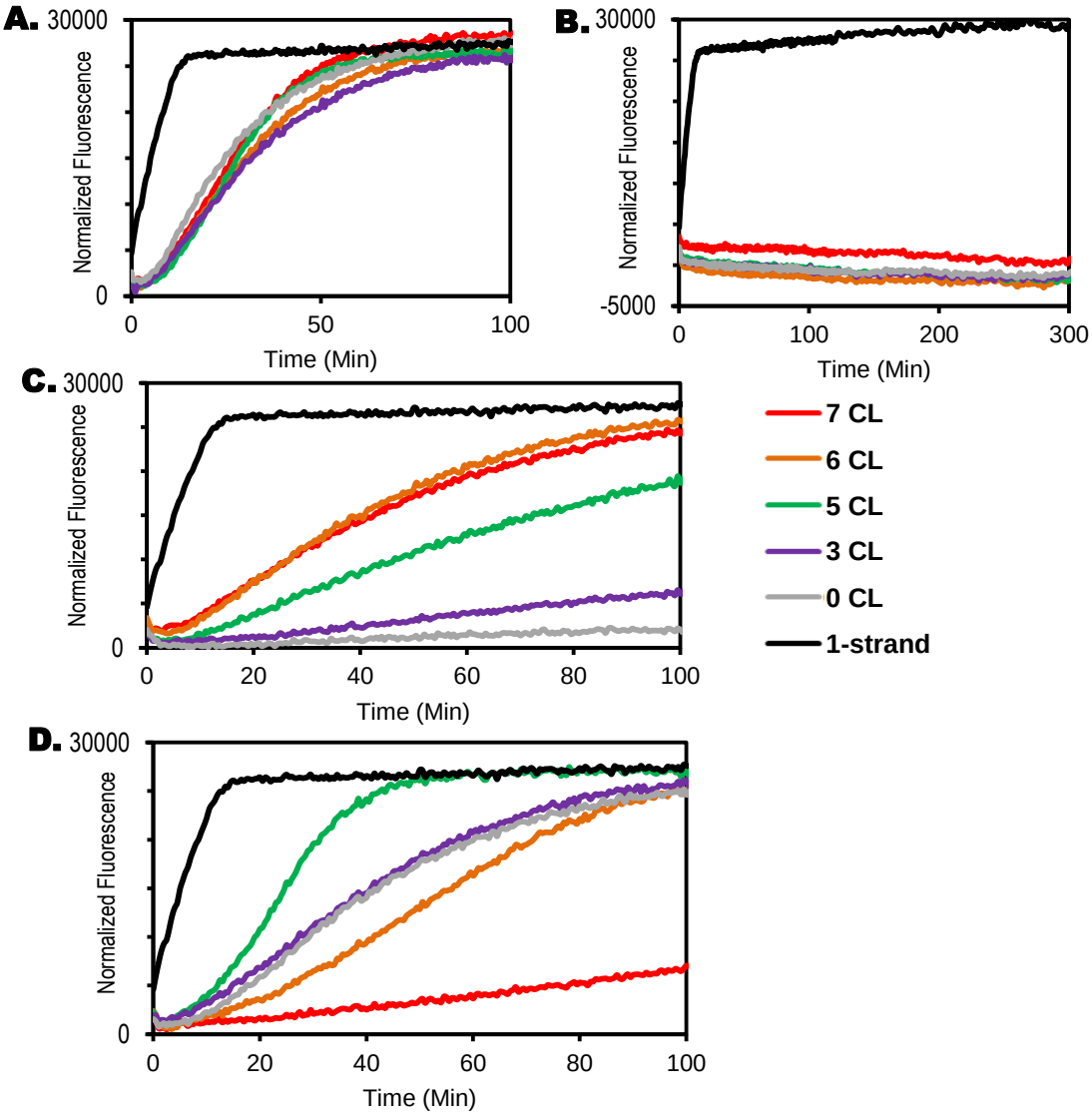

Figure 3.8 Effect of varying GC clamp length on indirect associative catalyst AND-gate. A.) Docking strand with staples. All clamp lengths exhibit similar kinetics. **B.)** Docking strand with blocked staples. In the absence of miRNA inputs, no leakage is observed in any of the samples. **C.)** Docking strand with staples and miRNA. Competitive hybridization between the staples and miRNA for the docking strand is minimized with longer clamp lengths. Without a clamp (grey), miRNA competition for the docking strand eliminated CHA. **D.)** Docking strand with blocked staples and miRNA inputs. A 5nt clamp provides the most optimal AND-gate kinetics. Clamps >5 suffer from slower dissociation of the staples from the blocking strands. Clamps <5 are slowed by greater competitive hybridization of the miRNA.

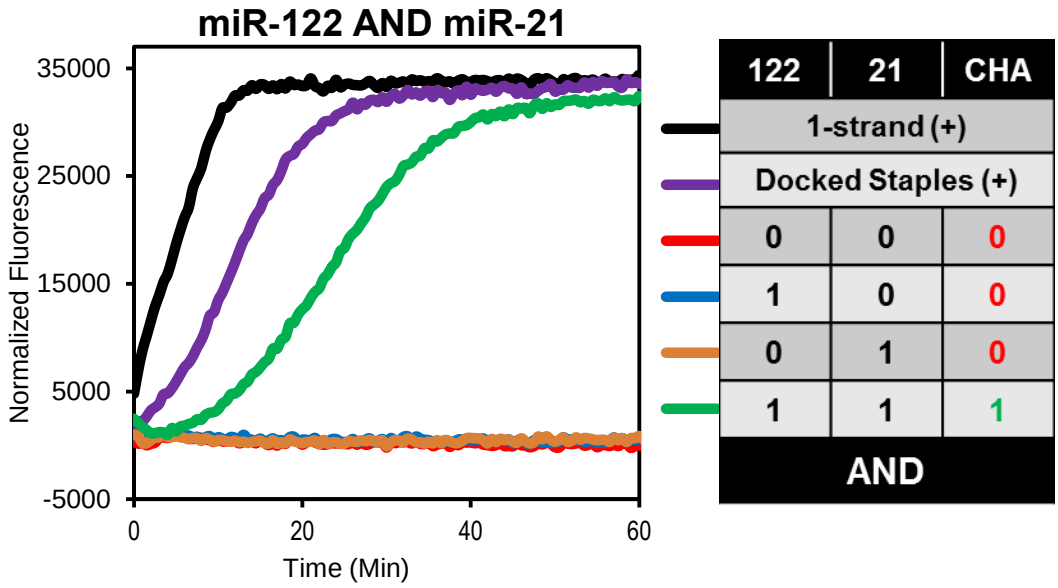

Figure 3.9 Response of AND-gated CHA. Behavior of two-input CHA circuit for the AND-gated logic table. Only in the presence of miR-21 AND miR-122 is CHA observed (green).

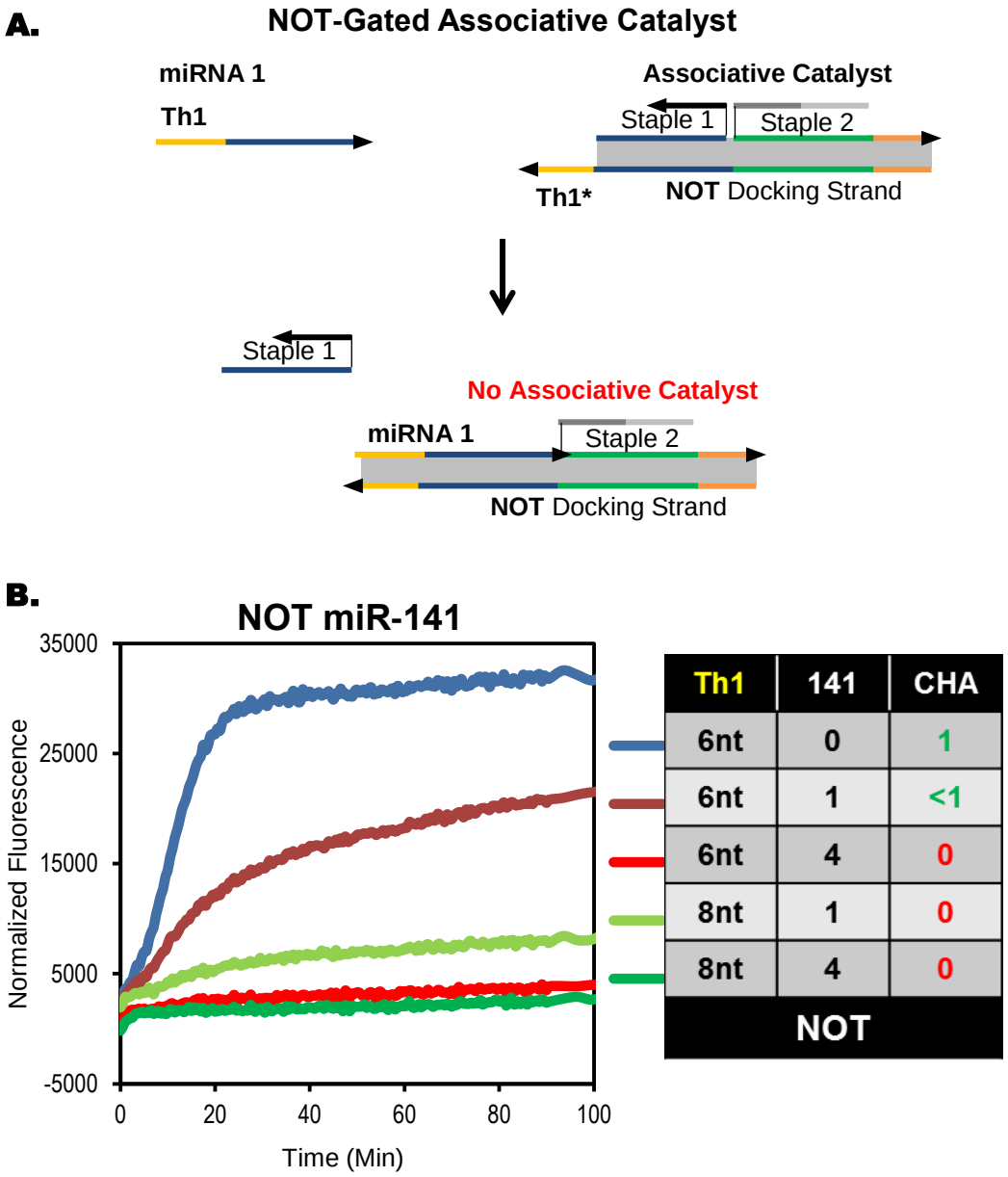

Figure 3.10 Tuning NOT-gated CHA. A.) Schematic of NOT-gated associative catalyst. The presence of a NOT-gated miRNA causes displacement of Staple 1 from the docking strand, thus inactivating the associative catalyst. **B.)** Fluorescence of the NOT-gated CHA circuit. Toehold length of 6nt and 8nt is compared. With a longer toehold, a lower threshold concentration of the NOT-gated miRNA (miR-141) is tolerated.

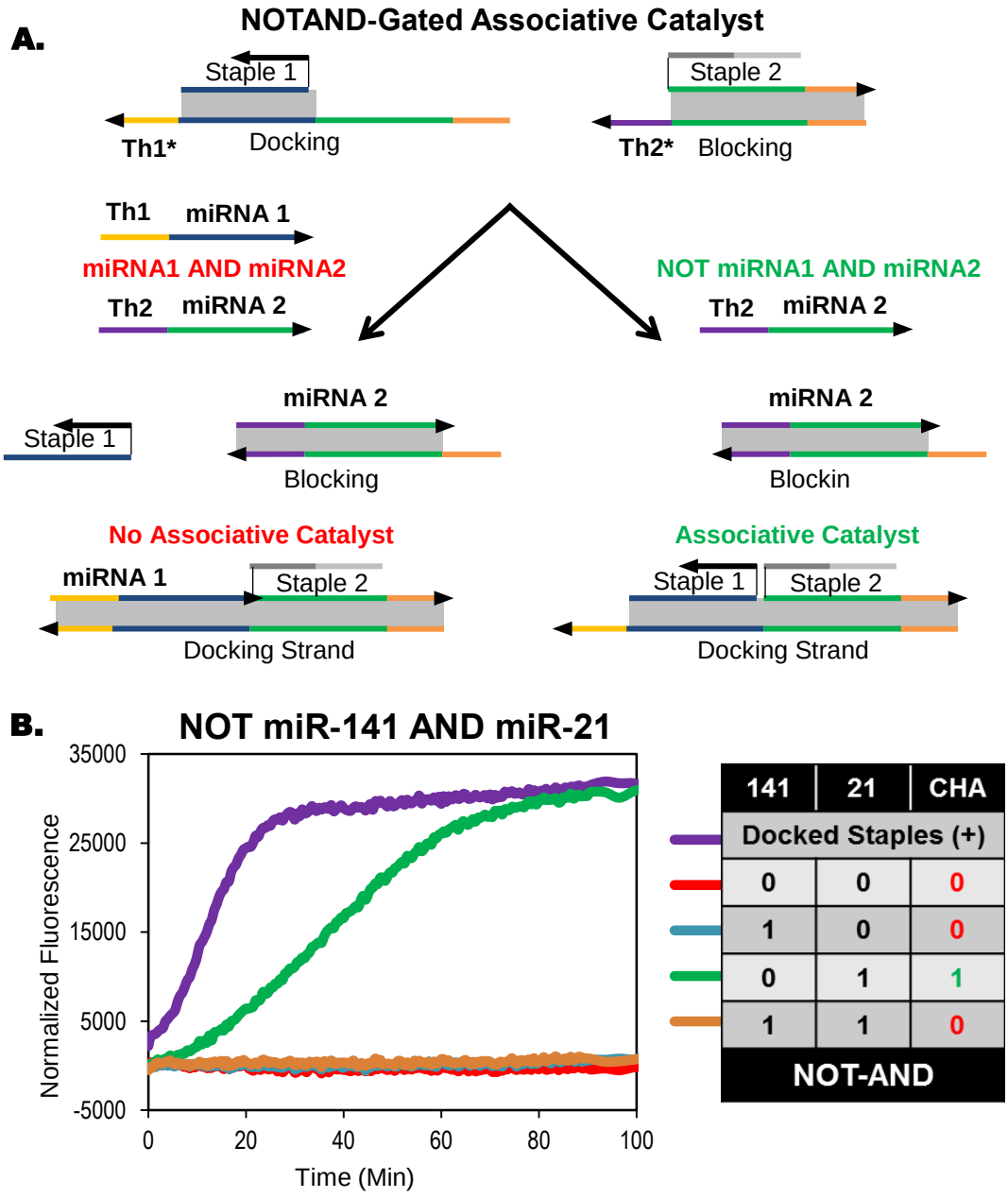

Figure 3.11 NOTAND-gated CHA. A.) Schematic of NOTAND-gated associative catalyst. The presence of miRNA 1 displaces staple 1 from the docking strand. miRNA 2 triggers the release of the staple 2 from the blocking strand. Only the right combination yields the associative catalyst. **B.)** Full logic table of the NOTAND-gated CHA circuit. Only the absence of miR-141 and presence of mir-21 is CHA observed (green).

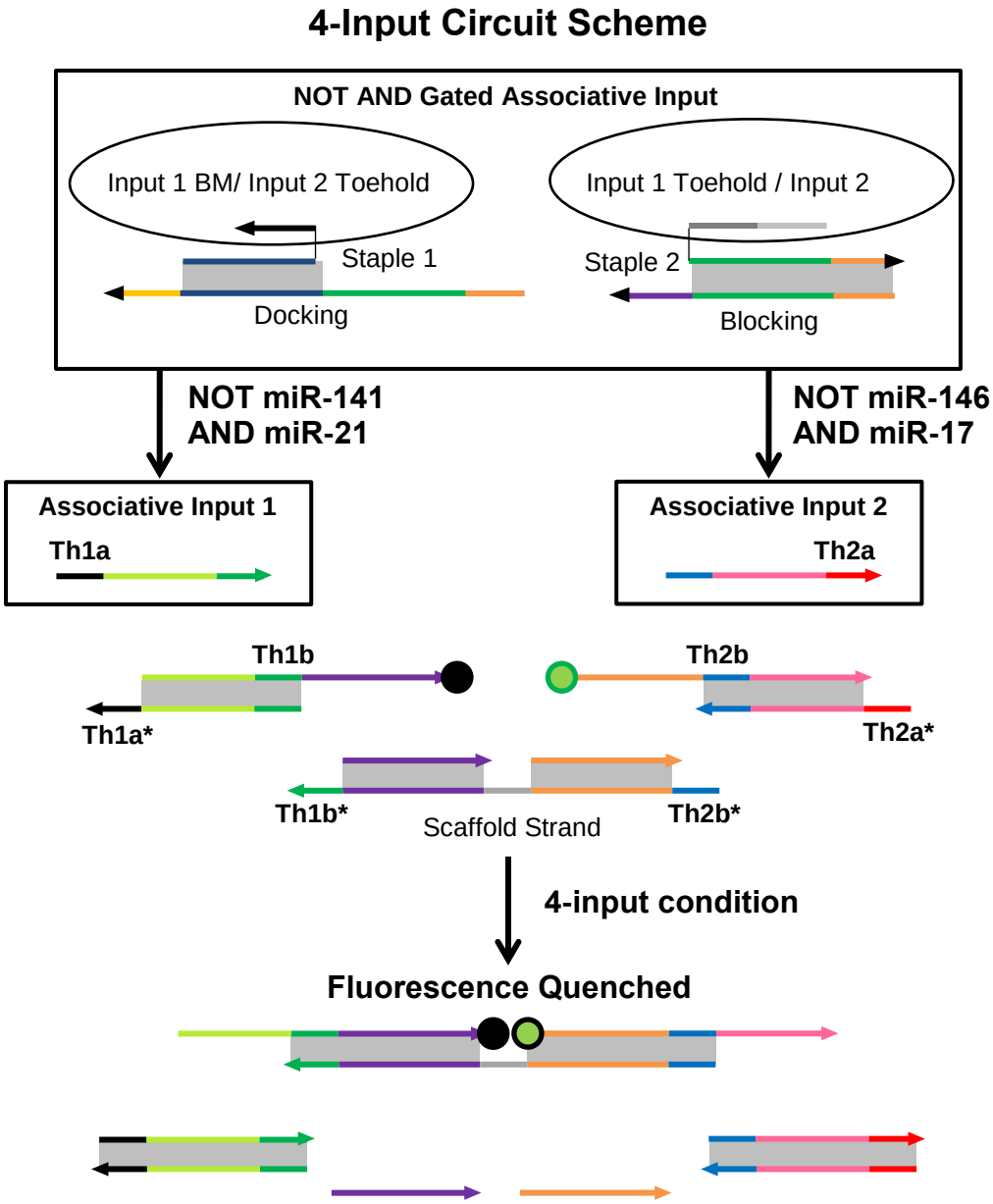

Figure 3.12 Schematic of four-input circuit reaction pathway. The two-input circuit from Chapter 2 (**Figure 2.3A**) is converted to a four-input circuit by turning the original inputs into split associative inputs. Each associative input is controlled by a NOTAND-gate. Associative input 1 and associative input 2 trigger strand displacement of a quencher-labeled strand and fluorophore-labeled strand onto the scaffold strand, respectively. Only the correct combination of all four miRNAs will result in the quenching of fluorescence.

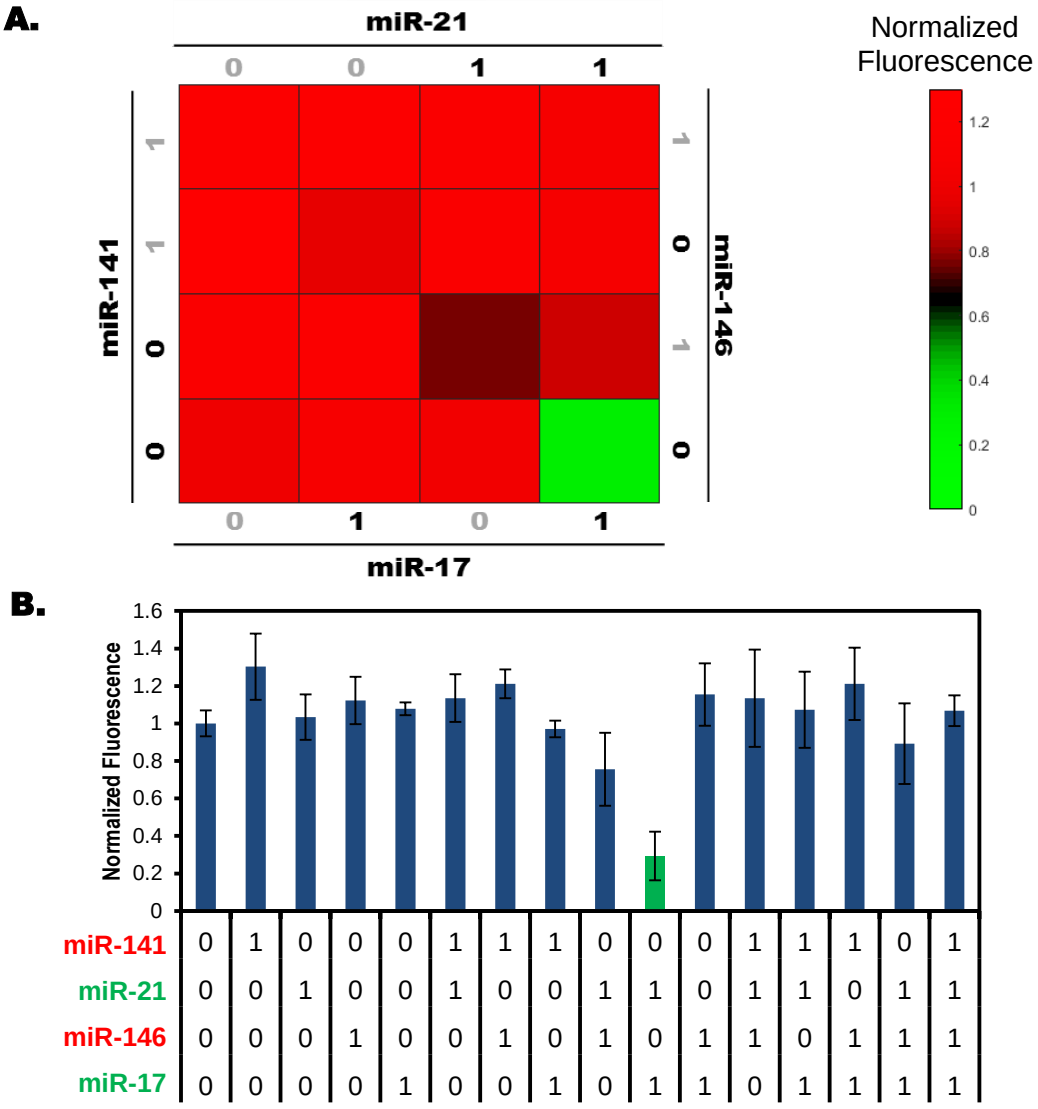

Figure 3.13 Four-input miRNA classifier. A.) Heat map representation of normalized fluorescence for full logic table 6 h after reaction. Fluorescence was normalized between the no input condition (set at 1) and the fluorophore and quencher prehybridized to the scaffold strand (set at 0). The Boolean value of each miRNA input for each square on the heat map is identified on the four sides. Only the correct four-input condition results in a signal quenched (green). Color axis was set to span the entire range of data values. However, as a diagnostic this threshold should be determined by internal standard controls of known input concentrations. **B.)** Bar graph representation of normalized fluorescence for full-logic table. Error bars depict standard deviation for three biological replicates.

REFERENCES

1. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature chemistry* **3**, 103-113 (2011).

2. Chen, Y., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nature nanotechnology* **10**, 748-760 (2015).

3. Jung, C. & Ellington, A. D. Diagnostic applications of nucleic acid circuits. *Accounts of chemical research* **47**, 1825-1835 (2014).

4. Carthew, R. W. Gene regulation by microRNAs. *Current Opinion in Genetics & Development* **16**, 203-208 (2006).

5. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120**, 15-20 (2005).

6. Landgraf, P. *et al*. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129**, 1401-1414 (2007).

7. Rie, D. d. *et al*. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnology* **35**, 872-878 (2017).

8. Li, Y. & Kowdley, K. V. MicroRNAs in Common Human Diseases. *Genomics, Proteomics & Bioinformatics* **10**, 246-253 (2012).

9. Paul, P. *et al*. Interplay between miRNAs and human diseases. *Journal of Cellular Physiology* **233**, 2007-2018 (2018).

10. Tong , A. W. & Nemunaitis, J. Modulation of miRNA activity in human cancer: a new paradigm for cancer gene therapy? *Cancer Gene Therapy* **15**, 341-355 (2008).

11. Cullen, B. R. Viruses and microRNAs: RISCy interactions with serious consequences. *Genes & Development* **25**, 1881-1894 (2011).

12. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-Free Nucleic Acid Logic Circuits. *Science* **314**, 1585-1588 (2006).

13. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333**, 1307-1311 (2011).

14. Mohammadi, P., Beerenwinkel, N. & Benenson, Y. Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy. *Cell systems* **4**, 207 (2017).

15. Zadeh, J. N. *et al*. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32**, 170-173 (2011).

16. Wolfe, B. R., Porubsky, N. J., Zadeh, J. N., Dirks, R. M. & Pierce, N. A. Constrained Multistate Sequence Design for Nucleic Acid Reaction Pathway Engineering. *Journal of the American Chemical Society* **139**, 3134 (2017).

17. Hemphill, J. & Dieters, A. DNA Computation in Mammalian Cells: MicroRNA Logic Operations. *Journal of American Chemical Society* **135**, 10512–10518 (2013).

18. Li, B., Ellington, A. D. & Chen, X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic acids research* **39**, e110 (2011).

19. Ang, Y. S., Tong, R. & Yung, L. L. Engineering a robust DNA split proximity circuit with minimized circuit leakage. *Nucleic acids research* **44**, e121 (2016).

20. Shekhawat, S. S. & Ghosh, I. Split-protein systems: beyond binary protein–protein interactions. *Current Opinion in Chemical Biology* **15**, 789-797 (2011).

21. Li, F., Lin, Y.& Le, C. Binding-Induced Formation of DNA Three-Way Junctions and Its Application to Protein Detection and DNA Strand Displacement. *Analytical Chemistry* **85**, 10835-10841 (2013).

22. Green, A. A. *et al*. Complex cellular logic computation using ribocomputing devices. *Nature* **548**, 117-121 (2017).

23. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. (2009).

24. Qian, L. & Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **332**, 1196-1201 (2011).

25. Srinivas, N. *et al*. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res* **41**, 10641-10658 (2013).

26. Šulc, P., Ouldridge, T., Romano, F., Doye, J. K. & Louis, A. Modelling Toehold-Mediated RNA Strand Displacement. *Biophysical Journal* **108**, 1238-1247 (2015).

27. Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807-10815 (1995).

Chapter 4

RNA STRAND DISPLACEMENT FOR DYNAMIC CAS6-GUIDED PROTEIN ASSEMBLY IN LIVE MAMMALIAN CELLS

Abstract

Toehold-mediated strand displacement has been predominately executed in a cell-free context, which has thus far limited its scope primarily to sensing and diagnostic purposes. Moving forward, the most powerful execution of this technology is to interrogate endogenous environments inside living cells, especially when coupled with our previous work of using strand displacement to actuate protein function. Towards executing an intracellular demonstration, we implemented a genetically-encoded device whose protein and RNA components self-assembled inside HeLa cells. Our strategy uniquely utilizes Cas6 endoribonucleases for their high binding affinity to efficiently dock protein components on a RNA scaffold, as well as for their cleavage properties to generate discrete RNA strands that can be dynamically controlled through strand displacement. We fused orthogonal Cas6 with split nanoluciferase output domains, and found Cas6-guided assembly to depend on RNA scaffold hybridization. Furthermore, a trigger strand was successfully designed to disassemble the scaffold through strand displacement. The opposite OFF to ON functionality was also achieved using a CHA circuit. Our strategy encompasses a unique interplay between protein and RNA components, where proteins not only serve as output modules, but are also used to generate the RNA circuit itself. Overall, this is a promising step in realizing nucleic acid nanotechnology for future intracellular applications.

4.1 Introduction

In the past decade, there have been a myriad of dynamic nucleic acid based devices developed using toehold-mediated strand displacement technology[1, 2]. Although wide-ranging in scope and strategy, these studies have been predominately performed in a cell-free context. The most practical application of strand displacement technology has thus far been limited to sensing and diagnostic purposes, where biological targets of interest can be pre-extracted from cells[3, 4]. However, the ultimate and most powerful execution of this technology is inside living cells. The ability to harness strand displacement to interrogate endogenous cellular environments and actuate a desired response would have tremendous implications for therapeutic applications. In Chapter 2, we demonstrated how linking proteins to DNA powerfully expanded the breadth of useful output functions a strand displacement based device could actuate. This chapter aims to take the concept an important step further by executing a dynamic RNA-protein device inside live mammalian cells.

Through a gene delivery approach, the RNA and protein components were expressed inside HeLa cells. A novel strategy of utilizing Cas6 endoribonucleases was executed to assemble the RNA-protein device (**Figure 4.1**). Not only was strong Cas6 binding used to guide efficient protein attachment to RNA, but its cleavage functionality was used to generate the individual discrete RNA components for running strand displacement in live cells.

There are two general strategies to achieve strand displacement intracellularly. The first is to deliver the nucleic acid components into cells directly; the second is to genetically encode and express the nucleic acid components from within the cell (**Table 4.1**). Since the central theme of the thesis is focused on harnessing nucleic acid

nanotechnology for dynamic protein assembly, we must consider delivery of not only a nucleic acid component, but a protein component as well.

A major advantage of the direct delivery strategy is that nucleic acids can be chemically modified with synthetic features (reporter dyes, backbone modifications, peptide/protein attachment, etc.) prior to delivery. If the device must be synthetically processed, such as the split yCD-DNA conjugates of Chapter 2, direct delivery is necessary. Delivery of short nucleic acid cargo is already a well-established field, primarily due the promise of RNA interference technologies[5, 6]. For the delivery of strand displacement based nanodevices, one can imagine that many of the strategies used for RNAi delivery can be hijacked or repurposed successfully. However, the challenge lies in how to ensure ubiquitous delivery of multiple unique components. Often there are multi-strand complexes and/or structural features whose integrity must be maintained. Previous studies have successfully used commercially available lipoplex-based transfection agents to deliver and execute simple strand displacement circuits inside mammalian cells[7-10]. Excitingly, the delivered devices were able to interact and utilize endogenous RNA as inputs. Other delivery techniques and vehicles that have been explored for nucleic acid circuits include streptolysin-O[10], DNA origami[1, 11], and gold nanoparticles[12].

The biggest obstacles limiting intracellular efficiencies are uneven subcellular distribution and nuclease degradation. Lipid-based transfection reagents often result in significant amounts of endosomal entrapment[10, 13]. Microinjection and streptolysin-O offer more ubiquitous distributions in the cytoplasm, but suffer from low throughput and low delivery efficiencies, respectively. Unmodified short nucleic acids are extremely vulnerable to degradation both in serum and inside the cell (half-life on the

order of minutes)[14]. Backbone modifications, in particular, inter-nucleotide phosphorothioate linkages and 2' O-methyl ribose modifications[10, 15], have been shown to greatly extend cellular half-life. However, these synthetic chemical modifications are costly and only serve to delay eventual degradation.

To address and overcome stability concerns, an alternative approach that is gaining momentum is to genetically encode the nucleic acid device. Packaging the entire device into a vector to deliver as a gene therapy offers long-term embedded expression. Once the genetic vector is delivered, it guarantees that all encoded components will be received by the target cell. Furthermore, because components are expressed by the cell itself, they will be distributed more ubiquitously. However, this does not negate the entrapment challenges gene delivery of the vector itself may pose.

Gene therapy also offers a more generalizable strategy, particularly in the context of protein-nucleic acid hybrid systems. For direct delivery, each device may need different packaging conditions and modifications[10], particularly if there is protein cargo attached. The delivery of protein-nucleic acid conjugates is in itself an underdeveloped field, the obstacles of which have been highlighted in Chapter 2. On the other hand, if packaged into a genetic vector, the delivery is unaffected by the specific protein and RNA components encoded.

A transcription based system means that the device must be RNA based and contain no synthetic modifiers. Thus far, DNA has been the preferred medium of choice for studying strand displacement, and moreover synthetic dyes are typically used to characterize and track behavior. While RNA behaves by the same principles of hybridization and strand displacement, adapting DNA circuits to RNA is not trivial. Since RNA-RNA interactions are more stable than DNA-DNA interactions,

significant energetic differences preclude direct sequence conversion from DNA to RNA[16]. Promisingly, the Ellington group has demonstrated that matching structural free energies can be employed to convert DNA circuits to RNA circuits[17]. By doing so, they were able to execute RNA-based catalytic hairpin assembly (CHA) with the same kinetic behavior as the original DNA version in a cell-free environment. Furthermore, in place of synthetic reporter dyes, they used a split fluorogenic RNA aptamer (Spinach) to track hairpin assembly. RNA aptamers such as Spinach bind to small molecule fluorophores resulting in fluorescence and serve as a promising alternative to synthetic reporters unavailable to transcription based systems[18, 19]. Other cell-free studies demonstrated that co-transcriptional folding is a viable mechanism for self-assembling complex RNA origami[20, 21].

Previously, genetically encoded RNA devices have been successfully executed in live bacterial cells. A group was able to express the RNA-based CHA described earlier in *E. coli* cells[22]. Upgrading to a more sensitive split Broccoli aptamer, they were able to track increased hairpin assembly inside cells in the presence of the catalyst strand. Furthermore, they designed molecular beacon-gated catalyst strands that were inactive until opened by the presence of endogenous RNA inputs. Another notable series of work involved the *de novo* design of toehold switches for riboregulation[23, 24]. Toehold-mediated strand displacement was used to open hairpin locked ribosome binding sites to allow for translation of a reporter protein. The toehold switch principle was expanded to have logic-gated behavior, and ultimately an impressive 12-input ribocomputing circuit was executed inside *E. coli*[24].

Particularly of interest here is a study in which a RNA scaffold with protein binding aptamers was genetically expressed and self-assembled in live bacterial

cells[25]. The RNA scaffold was designed to take on various discrete, one-dimensional, and two-dimensional structures, and the aptamer binding sites allowed for MS2 and PP7 fusion proteins to be docked. A split GFP and hydrogen-producing pathway enzymes were both shown to assemble on the RNA scaffold for increased fluorescence and increased hydrogen production, respectively. This framework of using protein-RNA binding interactions is adopted in this chapter to replace synthetic HaloTag-DNA conjugation. The stronger the protein-RNA interaction, the closer one can mimic a covalent attachment for greatest protein assembly efficiency. Previous studies have favored aptamer sequences such as MS2 and PP7, which offer low nanomolar dissociation constants[26]. We sought to improve upon this by using Type I Cas6 family proteins, which bind to their RNA substrate with picomolar affinity[27].

Cas6 family proteins belong to Type I CRISPR/Cas prokaryotic immune systems[28, 29]. The more famous and more widely utilized CRISPR/Cas9 is a Type II system, in which a single large protein (Cas9) is responsible for degrading foreign nucleic acids. However, in Type I CRISPR/Cas, multiple Cas proteins complex together to accomplish the same task[29]. Within this system, the Cas6 endoribonuclease is particularly suited for protein scaffolding. Cas6 proteins bind sequence-specific short RNA loops (20-30nt) with extremely high affinity[27, 30-32]. Furthermore, various monomeric Cas6 proteins have been isolated from different species, each recognizing unique RNA motifs[27]. Their small size (~20kDa) also makes them ideal fusion partners. By fusing orthogonal Cas6 to proteins of interests, Cas6 can be used to guide protein assembly on a RNA scaffold.

The scaffolding benefits of high affinity Cas6 are obvious, but a more intriguing feature is the single-turnover cleavage activity Cas6 proteins exhibit. Cas6

cleave at the 3' end of their RNA loop motif and afterward remain bound to the RNA substrate[30, 32]. Previous studies have successfully utilized Cas6 proteins for RNA cleavage processing in live mammalian cells[33-35]. For a static RNA scaffold, this cleavage functionality would not stand out or may even be undesired, but for a dynamic scaffold this cleavage is a convenient way to generate discrete RNA strands to serve as components which can undergo strand displacement for dynamic behavior.

Inspired by previous works to genetically encode RNA scaffolds and circuits in live bacterial cells, this chapter aims to execute a dynamic protein-RNA device in mammalian cells. In fact, toehold-mediated strand displacement induced protein assembly has never been shown in any type of live cells prior to this study. To accomplish this task, we introduced the novel strategy of using Cas6 not only for binding and protein assembly on the RNA scaffold, but also for cleaving RNA to form the discrete RNA components for dynamic control. First, a framework for Cas6-guided protein assembly on a RNA scaffold in HeLa cells was established. Next, strand displacement mediated disassembly of the scaffold was explored with various displacement trigger strand designs. Finally, a CHA circuit was adopted to show OFF to ON functionality for Cas6-guided protein assembly.

4.2 Materials and Methods

4.2.1 Molecular cloning and vector construction

All genetic manipulation and plasmid maintenance was performed in *E. coli* NEB5α. The pCDNA3.1(+) plasmid backbone was used for all vectors. Inserts were ligated through unique sticky ends into restriction digested pCDNA3.1(+) backbone using T4 ligase. All oligos and primers used can be found in **Table 4.2**.

The human codon optimized *P. aeruginosa* Csy4 gene was amplified through PCR off the PGK1p-Csy4-pA plasmid provided by the Tim Lu lab (Addgene). The *E.coli* Cse3 gene was ordered as a human codon optimized gBlock (Integrated DNA Technologies). LgBiT was amplified by PCR off pBiT1.1-C [TK-LgBiT] (Promega). SmBiT and T2A sequences were constructed through overlapping oligos. Annealed oligos were phosphorylated using T4 PNK. SmBiT and LgBiT inserts included a (GGSGGGS) linker on the N-terminal end and a C-terminal His6 tag. A four-piece ligation was first used to construct pCDNA3.1-T2A-Csy4-LgBiT. Next, a three-piece ligation was used to build pCDNA3.1-Cse3-SmBiT-T2A-Csy4-LgBiT, also referred to as Cas6-NanoBiT. For gRNA constructs, the overlapping oligo method was used exclusively to build inserts.

4.2.2 Mammalian cell lines and cell culture

HeLa cells (ATCC) were grown in MEM media (Corning) containing 10% FBS (Sigma), 100 × U ml–1 penicillin, and 100 × U ml–1 streptomycin (GE Healthcare). For luminescence transfection experiments, cells were seeded in 24-well plates at 50,000 cells per well in 0.5mL of MEM. For Western blot analysis, cells were seeded at 300,000 cells per well in a 6-well plate.

4.2.3 Plasmid DNA transfection

All plasmid DNA used for transfections were prepared using the ZymoPURE midiprep kit (Zymo Research). Transfection using Lipofectamine 3000 (Life Technologies) was carried out 20-24 h after seeding. Following the manufacturer's protocol, 500ng total DNA was transfected per well with 1µl of P3000 reagent and 1µl of Lipofectamine 3000 reagent. Transfections were allowed to proceed for 24 h prior

to measurement. For co-transfection of multiple plasmids, the various plasmids were mixed at desired ratios and normalized to 500ng total DNA before complexing with transfection agent. Plasmid mass ratios used in co-transfection studies can be found in **Table 4.3**. Samples harvested for western blotting were transfected with 2.5µg of plasmid DNA per well using 5µl of P3000 Reagent and 5µl of Lipofectamine 3000 Reagent.

4.2.4 Western blotting of Cas6 fusion protein expression

Cells were harvested by incubating on ice for 30 min with 250µl of lysis buffer (150nM NaCl, 50mM Tris, 1% Triton-X 100, pH 8.0), containing protease inhibitor cocktail 1 and PMSF. A cell scraper was used to manually scrape cells from the plate. Harvested cell lysate was centrifuged at 16,100g for 20 min to separate the soluble and insoluble fraction. Soluble fraction was quantified through Bradford assay and 20µg of total protein was loaded per lane. A 12% SDS-PAGE gel was run before protein transfer to nitrocellulose membrane. Mouse anti-his primary antibody and anti-mouse HRP conjugated secondary antibody were used. Chemiluminescence of HRP was used to image blot.

4.2.5 Nanoluciferase luminescence measurements

Nanoluciferase luminescence was measured 24 h post transfection. Each well was gently washed with 0.5mL of warm PBS, before adding 100µl PBS in each well. Next, an equal volume (100µl) of freshly reconstituted Nano-Glo Luciferase Assay Reagent (Promega) was added. Reconstituted Nano-Glo Luciferase Assay Reagent was prepared by diluting one volume of substrate in 50 volumes of room temperature assay buffer as recommended by the manufacturer. Next, the plate was incubated in

the dark with gentle shaking for 30 minutes. The sample liquid was then transferred from the clear 24-well plate to a black 96-well plate for luminescence reading under that plate reader. Because of the low expression of the split nanoluciferase measurements were taken at a high sensitivity setting (Sen: 255) on a Synergy H4 (BioTek) plate reader.

4.3 Results and Discussion

4.3.1 Cas6-guided RNA scaffold components and design

To create a genetically encoded dynamic protein-RNA scaffold, a two-plasmid system was devised in which one plasmid encoded for the proteins components while the other contained the RNA components (**Figure 4.1**). This division allowed for the relative expression of protein to RNA to be easily tuned by varying the ratio of each plasmid used for co-transfection. The ratio of protein to RNA scaffold is crucial for maximizing assembly efficiency.

Utilizing Cas6 family proteins as a means to link proteins to RNA, two orthogonal Cas6 proteins were chosen to be fused to two output proteins of interest. Well-characterized Cas6 proteins, Csy4 from *P. aeruginosa* and Cse3 from *E. coli*, were chosen for this study[30, 32]. Each Cas6 protein recognizes, binds, and cleaves its own unique RNA substrate (**Table 4.2**). The split nanoluciferase protein pair, known as NanoBiT, was chosen as the output because of its usefulness as a highly sensitive reporter[36]. When the split fragments are brought into close proximity, reconstitution allows them to regain activity and convert the substrate, furimazine, into blue light. NanoBiT has been engineered for high signal to background ratio, and this interaction is reversible as well. The split nanoluciferase fragments, referred to as LgBiT and

SmBiT, were fused to Csy4 and Cse3, respectively. Cas6 proteins conferred the ability to bind and cleave the RNA scaffold, while the split nanoluciferase offered a sensitive method to detect protein assembly and disassembly.

The initial RNA-based scaffold is comprised of two RNA strands each containing an orthogonal Cas6 binding loop on its 3' end (**Figure 4.1**). Hereafter, RNA molecules which possess these Cas6 binding motifs will be referred to as guide RNAs (gRNAs). This is not to be confused with guide RNAs used in CRISPR/Cas9 gene editing systems. Because the Cas6 also have endoribonuclease activity, the components were expressed in tandem as a single transcript. Upon cleavage, the scaffold/circuit will be discretized into individual RNA components. This tandem gRNA configuration is preferred because the intracellular environment is extremely crowded. Diffusional coefficients of nucleic acids have been reported to be 5-100 times slower in eukaryotic cells compared to cell-free solution conditions[37, 38]. Initial tethering of the RNA components minimizes the diffusional hurdle faced within the crowded cell. The 5' ends of the gRNAs contain extension sequences responsible for RNA scaffold/circuit functionality. For example, in the initial design 5' extension sequences were comprised of a simple hybridization domain that allowed the Cse3 gRNA and Csy4 gRNA to form a duplex. Lastly, a 3' BGH terminator was included to promote polyadenylation and prevent degradation of the RNA prior to Cas6 processing.

4.3.2 Expression of Cas6-NanoBiT fusion proteins

A vector was constructed that encoded for the Cas6 fusion proteins, Cse3-SmBiT and Csy4-LgBiT, separated by a self-cleaving T2A peptide sequence[39] for bi-cistronic expression driven by a constitutive CMV promoter (**Figure 4.2A**). This

protein pair will be hereafter referred to as Cas6-NanoBiT. HeLa cells were transfected with the Cas6-NanoBiT plasmid, and Western blotting of C-terminal His-tags located on each protein revealed two distinct bands as compared to lysate of HeLa cells transfected with the empty vector (**Figure 4.2B**). These bands correspond to the full-length sizes of Cys4-LgBiT (41.1 kDa) and Cse3-SmBiT (30.6 kDa). Furthermore, no truncation or uncleaved T2A product bands were observed. The Csy4-LgBiT band was of lower intensity than the Cse3-SmBiT band, which was expected since Csy4-LgBiT is the second gene on the T2A transcript[39]. Because LgBiT is known to confer background luminescence, we designed the order such that LgBiT would be limiting over SmBiT.

4.3.3 Optimization of protein expression level

Because one component is protein-based and the other is RNA-based, balance between the translation and transcription rate of the two was crucial. Too much protein and most would not be associated with the RNA scaffold; too little protein and the RNA scaffold would not be saturated. The ratio of protein plasmid to RNA plasmid used in co-transfection was varied to find this balance. Since one transcript will produce the translation of multiple proteins, it is intuitive that less protein plasmid was needed over the RNA plasmid. A 1:39 mass ratio of Cas6-NanoBiT plasmid to gRNA plasmid was found be optimal (**Figure 4.3B, Figure 4.4B**), and this ratio was used moving forward. Although lowering the Cas6-NanoBiT plasmid in half for a 1:78 ratio offered higher fold scaffolding enhancements, greater variability was observed. Luminescence measurements were noisier at such low expression, and transfection efficiency varied greatly experiment to experiment with such low dose of plasmid.

4.3.4 Effects of RNA hybridization length

In optimizing the RNA hybridization domain the most obvious considerations were hybridization efficiency and scaffold length. If the domain was too short, the RNA strands would not readily hybridize. Keeping this in mind, all hybridization domains tested had a melting temperature above 37°C. On the other hand, since both Cas6 proteins are located on the 3' end of the RNA, a long hybridization domain would spatially limit efficient protein assembly. Additionally, a more nuanced factor is the rotational plane the two RNA ends lie in. Taking these three factors into consideration, several gRNA constructs (**Figure 4.3B**) with varying hybridization domain lengths (13, 18, and 26 base pairs) were co-transfected with Cas6-NanoBiT to see which offered the highest amount of protein assembly. A hybridization length of 18bp was found to offer a 7.3-fold increase in luminescence over the empty vector negative control. Meanwhile the 13bp and 26bp lengths had lower fold increases of 3.8-fold and 4.2-fold, respectively. Given that approximately 11 base pairs make up a 360° turn of RNA[40], the 13bp and 26bp hybridization lengths offered similar planes of proteins assembly. On the other hand, the 18bp configuration would be rotated ~180° (**Figure 4.3C**). The changes observed are likely because the 18bp scaffold provides a more optimal rotational plane for split nanoluciferase reconstitution. This is in line with previous findings that even small changes in rotation can significantly change protein assembly on a RNA scaffold[41].

The order of the tandem gRNA for the 18bp scaffold was also switched by placing the Csy4 gRNA first rather than second (**Figure 4.3C**). Despite a slightly lower 5.8-fold enhancement, the Csy4 gRNA was kept first moving forward, since the cleavage of the first gRNA is most crucial for discretization. Because Csy4 is better characterized, it serves as a more viable target for future protein engineering[42.]

4.3.5 RNA hybridization induced protein assembly

To confirm that the luminescence increases observed were due to RNA hybridization, the hybridization domains of Cse3 gRNA were scrambled (**Figure 4.4A**). Both the 18bp and 26bp scrambled counterparts exhibited lower luminescence levels (**Figure 4.4B**). However, significant fold increases (~2.5 fold) over the empty vector control were observed in both, which we attribute to uncleaved transcripts. Prior to cleavage the split nanoluciferase is located on the same RNA molecule, and this conformation seemingly still allows for split nanoluciferase reconstitution (**Figure 4.4C**). Only upon cleavage do the scrambled gRNAs become untethered. The assembly observed in the scrambled constructs implies that Cas6 cleavage is a rate limiting step that could hinder proper protein-RNA device function. In fact, hybridization of the non-scrambled gRNAs through co-transcriptional folding likely occurs, independent of whether Cas6 cleavage has occurred. Csy4 has a reported cleavage rate (k_{obs}) of 3.8 min^{-1}. [31] While this is similar to self-cleaving ribozymes commonly used in synthetic biology, other endoribonucleases exhibit cleavage rates several orders of magnitude faster[43].

To improve the fold change between assembled and scrambled "disassembled" states moving forward, one would ideally increase the cleavage rate of the Cas6. Directed evolution or rational mutagenesis techniques could be applied to find mutant versions with faster cleavage kinetics. If protein engineering of the Cas6 does not sufficiently improve cleavage rate, alternative solutions are also available. The most obvious is to express the various Cas6 gRNAs on separate transcripts instead of in tandem. However, whether this is done by using multiple plasmids or multiple expression cassettes within a plasmid, there are concerns of transcriptional imbalance of the different RNA components. Separate transcripts also bring intracellular

diffusional hurdles back into play. A preferable compromise is to insert spacer regions in between tandem RNA components, so that even prior to cleavage the output proteins fused to Cas6 are too far apart to interact. This approach is likely to be highly successful when the proteins of interest have little or no background affinity.

4.3.6 Cas6 binding induced protein assembly

Next, gRNA mutants were constructed in which the Cas6 RNA loop substrates were replaced with their reverse complement sequence (**Figure 4.5A**). These non-binding gRNA mutants offered no luminescence increase compared to the empty vector (**Figure 4.5B**), strongly suggesting that Cas6 binding is responsible for complexing the proteins on the RNA scaffold. Future qPCR or Western blotting experiments are necessary to confirm that these changes are not due to expression differences of Cas6-NanoBiT upon co-transfection with various gRNA. Immunoprecipitation of the Cas6-RNA complex will also provide additional molecular proof.

4.3.7 Disassembly through RNA strand displacement

The initial demonstration of RNA scaffold hybridization was a static assembly. To explore a dynamic RNA scaffold, a toehold-mediated strand displacement event was used to trigger disassembly. An additional RNA trigger strand was constitutively expressed after the Cse3 gRNA. The RNA trigger binds to the toehold of the Cse3 gRNA, initiating branch migration for the formation of a trigger and Cse3 gRNA duplex, ultimately disassembling Csy4 from Cse3 (**Figure 4.6**).

4.3.8 Effect of toehold length on strand displacement induced protein disassembly

A basic principle in strand displacement is that longer toeholds promote faster strand displacement kinetics. In a study involving RNA toehold switches they found that for unwinding a 18bp stem, a ≤6nt toehold was ineffective[23]. However, starting at 8nt they began to observe effective toehold switch activation, with the increasing effects up to 12nt. Since the hybridization domain of our design is 18bp as well, a 13nt toehold trigger was initially tested. The trigger strand was found to cause minimal disassembly (**Figure 4.7B**). When luminescence was normalized between the 18bp assembled and 18bp scrambled gRNA controls, the addition of the 13nt toehold trigger only caused a 20% drop in assembly. When the toehold length was increased to 24nt, a significant increase in disassembly was observed (60%). Scrambled versions of the 13nt and 24nt toehold trigger strands yielded no drop in luminescence compared to the assembled control.

4.3.9 Competitive hybridization of trigger strand

Next, a trigger in which only the 24nt toehold region was scrambled was tested to probe if the luminescence decrease was due to a toehold-mediated strand displacement process or from competitive hybridization between Csy4 gRNA and trigger strand for Cse3 gRNA. With the scrambled toehold, a 30% decrease in assembly was observed (**Figure 4.7B**), indicating that both strand displacement and competitive hybridization events have approximately equal contributions in the 60% decrease observed. This is unsurprising since all three RNA components were placed in tandem, making both Csy4 gRNA and trigger strand hybridization to Cse3 possible. Given the rate limiting cleavage of the Cas6 gRNA, hybridization is likely to occur predominately in cis (**Figure 4.8A**). Competitive hybridization is further supported by

the fact that when the RNA sequence (Cas6 gRNAs with scrambled toehold trigger) was modeled in m-fold[44], both configurations were predicted with very similar structural free energies (**Figure 4.8B**). One caveat to these predicted structures is that the algorithm does not account for possible effects Cas6 binding may have on RNA folding.

Competitive hybridization could be removed by expressing the trigger strand on a separate transcript, since the cis-hybridization of Csy4 gRNA to Cse3 gRNA would be more favorable than the trans-hybridization of the trigger strand to Cse3 gRNA. Temporal control over trigger strand expression can be incorporated to further dictate a strand displacement mechanism over competitive hybridization.

4.3.10 Surrounding trigger strand with ribozymes

While increasing the toehold length resulted in a more effective trigger strand, there was still a discrepancy between these findings and the toehold lengths reported by previous studies of our own and others[23]. Since the toehold is located on the 3' end of the trigger, it lays adjacent to the BGH terminator, which contains bulky secondary structures to promote transcriptional termination[45]. We suspected that steric hindrance of the 3' end caused the apparent toehold length to be shortened, resulting in the empirical discrepancies with previous studies.

To explore this hypothesis, self-cleaving ribozymes were used to create clean ends on the trigger strand. The well-characterized hammerhead[46] (HH) and hepatitis delta virus[47] (HDV) ribozymes were incorporated before and after the trigger strand, respectively (**Figure 4.9A**). The addition of a 5' HH ribozyme resulted in a 30% decrease in luminescence, which is smaller than the original trigger with no ribozymes (**Figure 4.9B**). The HH ribozyme is a structured 49nt sequence, and its insertion

possibly reduced the competitive hybridization contribution of the trigger strand. It acts as both a physical spacer and a temporal delay so that co-transcriptional hybridization of Cse3 gRNA to Csy4 gRNA is more favorable than to the trigger strand. In fact, when a scrambled toehold version of the 5'HH trigger strand was modeled in m-fold, all the predicted structures had Cse3 gRNA hybridized to Csy4 gRNA (data not shown). Given this model, it seems reasonable that the 5'HH trigger resulted in a 30% decrease; the same decrease attributed to strand displacement in the ribozyme-less trigger.

To create a trigger unhindered by the 3' BGH terminator, an additional HDV ribozyme was inserted. This dual ribozyme trigger resulted in a 90% decrease in protein assembly (**Figure 4.9B**), a significant improvement over the trigger with no ribozymes or only a 5'HH. This supports the hypothesis that previous triggers suffered from steric hindrance of the 3' toehold region. Furthermore, since the 5'HH is still present, we attribute the 90% disassembly primarily to a strand displacement mechanism. Future evaluation of scrambled toehold versions of the ribozyme triggers is necessary to confirm our model that the 5'HH prevents competitive hybridization and the 3'HDV removes steric hindrance of the toehold region.

4.3.11 OFF to ON circuit using catalytic hairpin assembly

Our initial design demonstrated a dynamic RNA-protein scaffold that went from an assembled to disassembled state (ON to OFF). Next, we wanted to accomplish the opposite functionality (OFF to ON). The ability to dynamically shuttle in both directions would establish the foundation for building higher order logic circuits with this technology. Again, we looked towards utilizing catalytic hairpin assembly (CHA). In previous chapters, the use of CHA was touted for its signal amplification properties.

While amplification is still a valuable attribute, the main attraction here is that the hairpins offer a method for kinetically trapping the RNA scaffold in an initial OFF state without the need for interstrand hybridization.

We adopted the RNA-based CHA circuit previously executed inside live *E. coli* cells[22]. The split Broccoli was removed and Cas6 gRNA motifs were added to the 3' end of each hairpin (**Figure 4.10A**). Again, the gRNAs were expressed in tandem (**Figure 4.11A**); co-transcriptional folding allows for proper hairpin formation without separating the hairpins onto different transcripts. The catalyst strand was expressed from a separate plasmid instead of in tandem for easy modulation of hairpin to catalyst ratio. Poly-A stretches were placed on the ends of the catalyst strand to avoid steric hindrance from surrounding RNA stretches. Two different catalyst designs were tested, a linear catalyst and a molecular beacon-gated catalyst (**Figure 4.10B**). The molecular beacon-gated catalyst is inactivated by a stem loop structure that blocks off the catalyst's toehold. However, if the RNA target is present, it will hybridize to the complementary loop region, opening up the beacon. When the molecular beacon is opened the toehold will no longer be sequestered and the catalyst strand is activated. This design offers a simple strategy for decoupling the CHA sequences from the biologically relevant target. The loop region can be easily switched out for any biological target of interest without affecting the CHA sequences. A miR-21 gated version (MB21), that had been previously been tested *in vitro* successfully, was chosen for this study[22].

A three plasmid co-transfection, with equal amounts of the hairpin and catalyst plasmids, was carried out in HeLa cells. The linear catalyst and MB21-gated catalyst had modest luminescence increases of 35% and 19%, respectively, over the scrambled

catalyst negative control. These enhancements were less impressive than the linear hybridization designs, but this is unsurprising because by switching to a three plasmid system, it was necessary to dilute the amount of Cas6 gRNA in half. Another possible limiting factor is hairpin assembly in the absence of the catalyst. Since H1 and H2 are in a tandem configuration, any improper formation of the hairpins or breathing in the stem region could lead to H1 and H2 hybridization. In the future, a control in which one of the CHA hairpins is replaced with an orthogonal non-interacting hairpin sequence of same size and structure is necessary to determine how much CHA leakage is present and if hairpin formation occurs rapidly enough through co-transcriptional folding.

For the MB21-gated catalyst, further characterization with proper controls will be necessary to fully evaluate its behavior. There are multiple possibilities as to why the MB21 catalyst has less assembly than the linear catalyst. If there is limiting amounts of miR-21 in the cell, not all of the MB21 catalysts will be activated. Ill-formed MB21 or hairpin breathing (spontaneous opening at the ends of the stem) can also result in activated catalyst. This would be concerning because in that case the CHA observed would not be mir-21 gated, but rather a form of leak. In order to rule out these undesired mechanisms, a molecular beacon in which the loop region is scrambled is necessary. Specificity can also be further evaluated by testing the circuit in a different cell line, such as HEK293, where miR-21 expression is low, or by knocking down expression of miR-21, either by expressing its antagomir or using a small-molecule inhibitor[7, 22].

Although the results were modest, the ability to execute Cas6-guided CHA inside HeLa cells confirms the feasibility of hairpin-based RNA computation in

mammalian cells. Furthermore, the demonstration of OFF to ON behavior is crucial towards the goal of building therapeutic devices (i.e. a default setting of OFF rather than ON is more desirable when discriminating between healthy and sick cells). Ultimately, more optimization and characterization of Cas6-guided CHA will be necessary to achieve assembly yields approaching those observed in cell-free settings.

4.4 Conclusion

This chapter establishes a novel framework for using Cas6 endoribonucleases as a means for dynamic protein assembly through RNA computation in live mammalian cells. The picomolar affinity of Cas6 facilitates highly efficient docking of Cas6-NanoBiT fusion proteins onto Cas6 gRNA, and 5' extension sequences on Cas6 gRNA can be designed to run RNA circuitry. Both the Cas6 protein and the RNA circuit components were genetically encoded and transfected into HeLa cells for evaluation of the system.

Various RNA scaffolds were shown to have 4 to10-fold increases in luminescence, but scrambled hybridization controls also had significant protein assembly (2.5-fold). The slow cleavage rate of the Cas6 was found to be a major bottleneck, and moving forward Cas6 engineered to cleave faster will likely dramatically improve the fold changes observed between assembled and disassembled states. Helical rotation of the RNA scaffold was also found to play an important role in split nanoluciferase reconstitution. Next, toehold-mediated strand displacement was used to trigger disassembly of the RNA scaffold. Varying toehold length and incorporating ribozymes revealed that in order to have effective strand displacement RNA components must be carefully designed. A competitive hybridization mechanism was found to have a major contribution. However, incorporation of a HH ribozyme

spacer seemed to provide both a spatial and temporal block for this mechanism and guarantee a predominantly strand displacement process. We also found that flanking RNA structures can sterically hinder strand displacement. When a HDV ribozyme was incorporated to remove the BGH terminator from the toehold end of the trigger, 90% disassembly of the scaffold was achieved. Finally, a Cas6-guided CHA circuit was executed, and while hairpin assembly was modest, it confirms the feasibility of running complex hairpin based RNA circuitry with our Cas6-guided strategy.

There are two avenues moving forward we believe to be most exciting to explore. The first is the interface of RNA circuitry with endogenous signals. The molecular beacon strategy for the miR-21 gated catalyst strand is a promising avenue for creating adapter RNA components that modularly translate endogenous biological cues into standard device inputs. The second avenue is exchanging the split nanoluciferase with other output protein systems. While split nanoluciferase served as a highly useful reporter for initial characterization purposes, other proteins may prove a much more interesting and valuable application of our Cas6-RNA guided technology. FRET fluorescent proteins could be used to allow for further spatial and temporal characterization the system. But more importantly, cell killing outputs, such as split yCD, should be tried in the future to fully realize this strategy's potential for creating therapeutic computing devices.

TABLES

Table 4.1 Comparison of direct delivery and genetically encoded strategies.

	Direct Delivery	Encode Genetically
Pros	• Can be modified with synthetic features. • Interstrand complexes and structures can be precisely prepared. • Covalent attachment of protein cargo.	• Long-term embedded control. • Ubiquitous distribution in cytoplasm once expressed. • All device components (including proteins) can be packaged into single genetic vector.
Cons	• Short half-life inside cells. Expensive modifications to increase stability. • Uneven subcellular distribution. Compartmental entrapment. • Each component must reach cell. Limits complexity of device. • Delivery conditions must be optimized for each device, particularly for a protein-nucleic acid hybrid system.	• No synthetic modifications. Must adopt biological replacements. • Must be RNA based. RNA less characterized than DNA. • Interstrand complexes cannot be pre-formed. Must self-assemble. Co-transcriptional folding of hairpin structures to kinetically trap and inactive RNA components.

Table 4.2 Sequences of primers and gRNA inserts. Csy4 and Cse3 binding motifs are highlighted in blue and orange, respectively. HH and HDV ribozyme sequences are depicted in green and purple, respectively.

Primer/Oligo Name	**Sequence (5' to 3')**
Cse3-FOR	GATGAT **GCTAGC** ATG TAC CTG AGC AGG ATC ACA CTG
Cse3-REV	GATGAT **AAGCTT** CAC CTC GGC GCC GGG CTT AAT C
SmBiT-F1	**AGCTT** GGGAGTTCCGGTGGTGGCGGGAGCGTGACCGGCTACCG
SmBiT-F2	GCTGTTCGAGGAGATTCTGGGTGGCGGGAGCcaattgCATCA CCACCACCATCAC **GGTAC**
SmBiT-R1	**C** GTGATGGTGGTGGTGATGcaattgGCTCCCGCCA
SmBiT-R2	CCCAGAATCTCCTCGAACAGCCGGTAGCCGGTCACGCTCC CGCCACCACCGGAACTCCC **A**
T2A-F1	**GATCC** gagggcaggggctccctgctgacctgcggcgacgtggaggagaaccccggcccc **C**
T2A-R1	**TCGAG** ggggccggggttctcctccacgtcgccgcaggtcagcagggagcccctgccctc **G**
Csy4-FOR	GATGAT **CTCGAG** GAC CAC TAT CTG GAC ATC AGA CTG
Csy4-REV	GATGAT **ggatcc** attttcacc GAA CCA GGG CAC GAA GCC
LgBiT-FOR	GATGAT **ggatcc** GGGAGTTCCGGTGGTGGCGGGAGC GTC TTC ACA CTC GAA GAT TTC G
LgBiT-REV	GATGAT **GGGCCC** TTA GTGATGGTGGTGGTGATG GCTCCCGCCACC ACT GTT GAT GGT TAC
gRNA Construct	**Sequence (5' to 3')**
13bp (Cse3,Csy4)	GCCAGTCCTAATCGACACTGGTACATGAGTTCCCCGCGCC AGCGGGGATGTACCAGTGTCCGTTCACTGCCGTATAGGCA GCCTGCAGG
18bp (Cse3,Csy4)	GCCAGTCCTAATCGAACGGACACTGGTACATGAGTTCCCC GCGCCAGCGGGGATGTACCAGTGTCCGTTCACTGCCGTAT AGGCAGCCTGCAGG
26bp (Cse3,Csy4)	GCCAGTCCTAATCGAACGGACACTGGTACATCGCACTACG AGTTCCCCGCGCCAGCGGGGGTAGTGCGATGTACCAGTGT CCGTTCACTGCCGTATAGGCAGCCTGCAGG
Scram 26bp (Cse3,Csy4)	GCCAGTCCTAATCGAACGGACACTGGTACATCGCACTACG AGTTCCCCGCGCCAGCGGGGAGTCGAATTCTGCGGCAGTG TCGTTCACTGCCGTATAGGCAGCCTGCAGG

18bp (Csy4, Cse3)	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGCCA GTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGC CAGCGGGGATAAACCG
Scram 18bp (Csy4, Cse3)	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGCCA GTCCTAATCGGATCAGACATATCAGGCGAGTTCCCCGCGC CAGCGGGGATAAACCG
13_18_trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGCCA GTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGC CAGCGGGGATAAACCGCATCagagacgAagctT ATGTACCAGTGTCCGTTCGATTAGGACTGGCTTTTGgatc
13_18_scram _trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGCCA GTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGC CAGCGGGGATAAACCGCATCagagacgAagctT GCAGCTACTCTGATTGCTTCGATAAGGTGGCTTTTGgatc
24_18	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATA ACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGT TCCCCGCGCCAGCGGGGATAAACCGCATCagagacgAagctT
24_18 Scram	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATA ACCATGGCCAGTCCTAATCGGATCAAGCATATCAGGCGAGT TCCCCGCGCCAGCGGGGATAAACCGCATC
24_18 (Rev Comp)	ATGTACCAGTGTCCgttcGCTGCCTATACGGCAGTtaagaaacat cGAATAACCATGGCCAGTCCTAATCGAACGGACACTGGTAC ATCGGTTTATCCCCGCTGGCGCGGGGAACTCCATCagagacg AagctT
24_18 Scram (Rev Comp)	ATGTACCAGTGTCCgttcGCTGCCTATACGGCAGTtaagaaacat cGAATAACCATGGCCAGTCCTAATCGGATCAAGCATATCAG GCCGGTTTATCCCCGCTGGCGCGGGGAACTCCATC
24_18_trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATA ACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGT TCCCCGCGCCAGCGGGGATAAACCGCATCagagacgAagctTA TGTACCAGTGTCCGTTCGATTAGGACTGGCCATGGTTATTC
24_18_scram _trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATA ACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGT TCCCCGCGCCAGCGGGGATAAACCGCATCagagacgAagctTG TTGCTTGCGTGTGAGATACCATTACTGTCTGAGATTCCACG
24_18_scram _toehold_trig ger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATA ACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGT TCCCCGCGCCAGCGGGGATAAACCGCATCagagacgAagctTA TGTACCAGTGTCCGTTCGCGGTATCGCCTATTGTTGATAGA

24_18_HH_trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATAACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGCCAGCGGGGATAAACCGCATCagagacgCGACTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCtagtcgAagctTATGTACCAGTGTCCGTTCGATTAGGACTGGCCATGGTTATTC
24_18_HH_scram_trigger	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATAACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGCCAGCGGGGATAAACCGCATCagagacgCGACTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCtagtcgAagctTGTTGCTTGCGTGTGAGATACCATTACTGTCTGAGATTCCACG
24_18_HH_trigger_HDV	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATAACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGCCAGCGGGGATAAACCGCATCagagacgCGACTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCtagtcgAagctTATGTACCAGTGTCCGTTCGATTAGGACTGGCCATGGTTATTCGCggccGCtttGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC
24_18_HH_scram_trigger_HDV	ATGTACCAGTGTCCgttcactgccgtataggcagctaagaaacatcGAATAACCATGGCCAGTCCTAATCGAACGGACACTGGTACATGAGTTCCCCGCGCCAGCGGGGATAAACCGCATCagagacgCGACTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCtagtcgAagctTGTTGCTTGCGTGTGAGATACCATTACTGTCTGAGATTCCACGGCggccGCtttGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC
HP1_HP2	aaaaGAGCGATACTGTTGGTAGATGTCGCCATGTCTTGCGACATCTACCAACAGgttcactgccgtataggcagctaagaaaGgatcCACTAGTCCAGTGTGGTGGaattCaaaaAGATGTCGCAAGACATGGCGACATCTACCAACAGCCATGTCTTGgagttccccgcgccagcgggataaaccgC
catalyst	aaaaaaGACATCTACCAACAGTATCGCTCaaaaaa
MB21_catalyst	aaaaaaGACATCTACCAACAGTATCGCTCTCAACATCAGTCTGATAAGCTAAGCGATAaaaaaa
scram_catalyst	aaaaaaTCATGACGCAGACCCATACTTACaaaaaa

Table 4.3 Co-transfection plasmid mass ratios. The distribution of various plasmids used for transfection studies. Total DNA per well is constant at 500ng per well.

System	Plasmid	Amount (ng/well)
Hybridization Length (1:39)	Cas6-NanoBiT	12.5
	gRNA	487.5
Hybridization Length (1:78)	Cas6-NanoBiT	6.25
	gRNA	493.75
Scrambled hybridization/ Rev. Comp. gRNAs/ Trigger Disassembly	Cas6-NanoBiT	12.5
	gRNA	487.5
Catalytic Hairpin Assembly (CHA)	Cas6-NanoBiT	12.5
	H1-H2 gRNA	243.75
	Catalyst/MB21 Cat/ Scram Cat	243.75

FIGURES

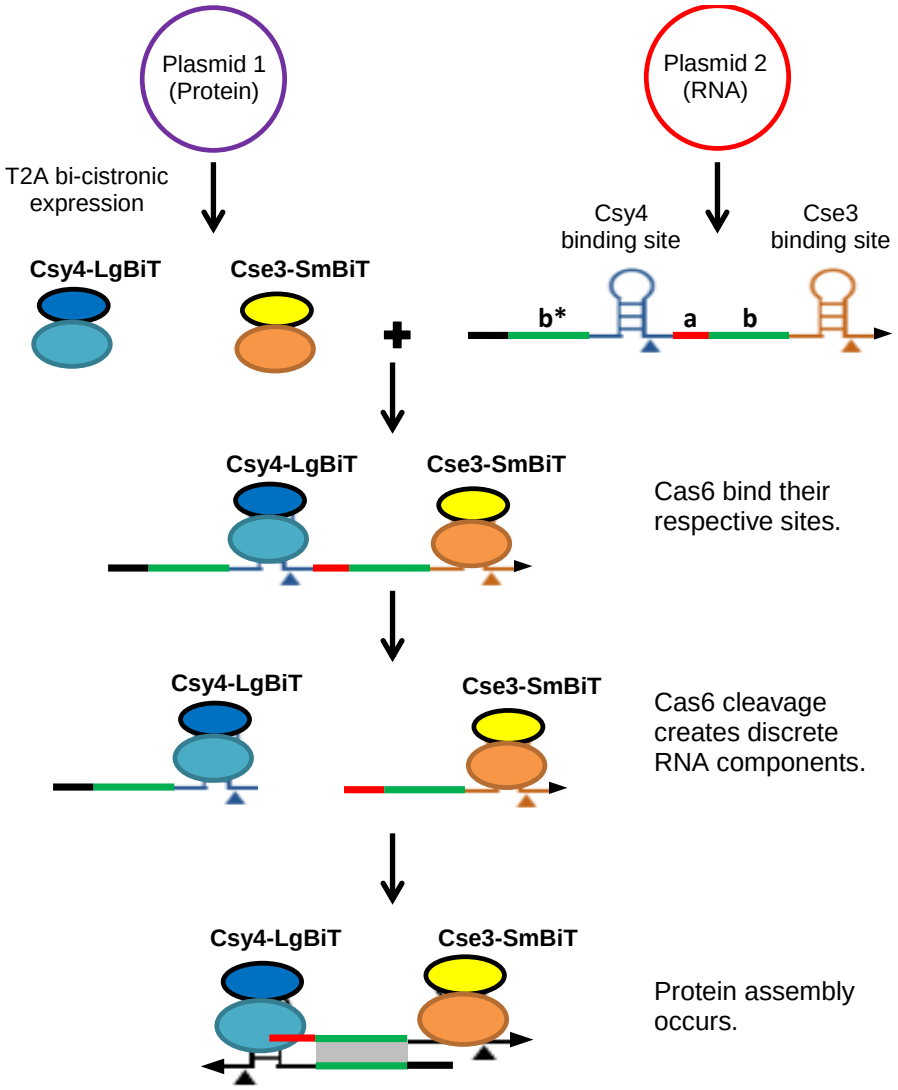

Figure 4.1 Genetically encoded RNA scaffold for Cas6-guided protein assembly. A two plasmid system is utilized, in which the first expresses the Cas6 fusion proteins while the second expresses the Cas6 gRNA scaffold. A single RNA transcript contains all Cas6 gRNA components. Upon binding and cleavage (colored triangles) of the orthogonal Cas6 (Csy4 and Cse3) at their respective RNA sites, the RNA components are discretized. Hybridization of the gRNAs allows for protein assembly and reconstitution of nanoluciferase activity.

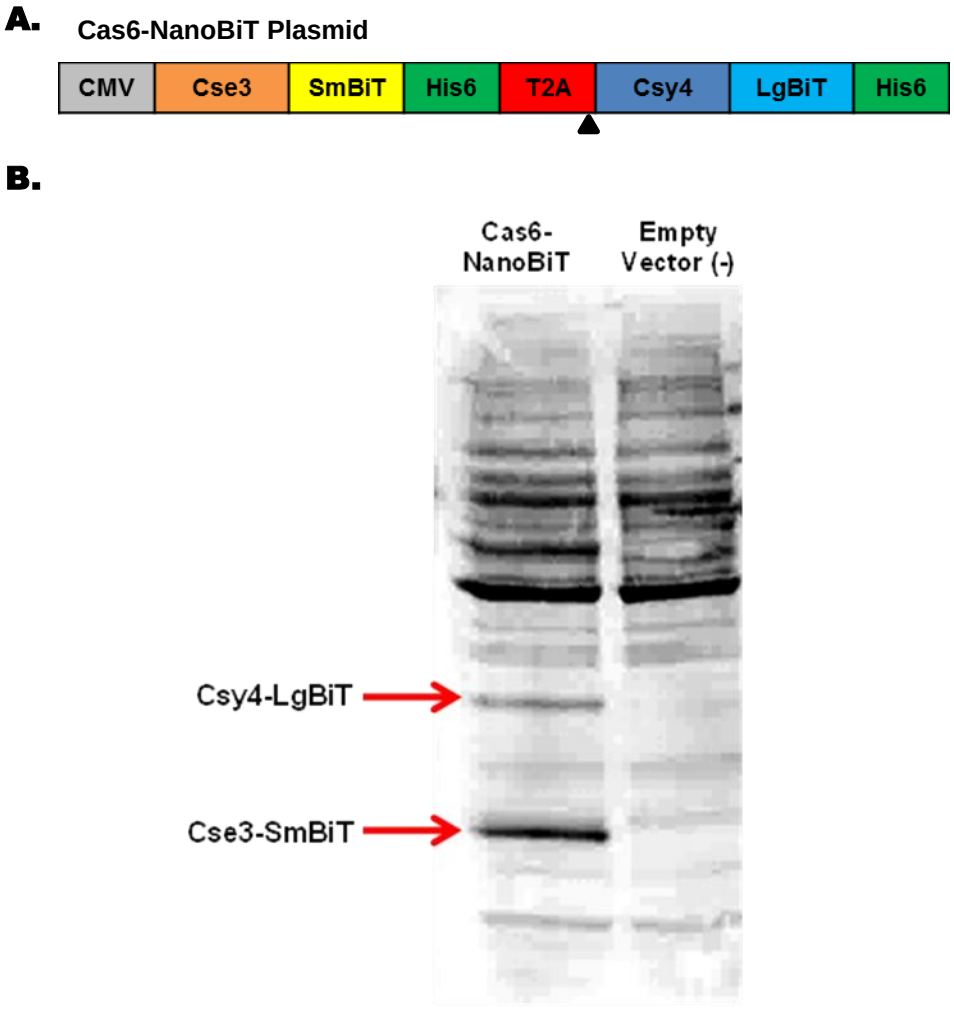

Figure 4.2 Expression of Cas6-NanoBiT fusion proteins in HeLa cells. A.) Bi-cistronic expression of Cas6-NanoBiT. The black triangle represents where T2A cleavage occurs. **B.)** Chemiluminescent anti-his Western blot of HeLa cell lysate transfected with the Cas6-NanoBiT plasmid. Compared to the empty vector, two distinct bands appear corresponding to the full-length sizes of Csy4-LgBiT-his6 (41.1kDa) and Cse3-SmBiT-his6 (30.6kDa). No uncleaved T2A or truncation product bands are present. There is a larger band that appears to be unique to the Cas6-NanoBiT sample, but this is attributed to a blot artifact at the same location in the negative control. Furthermore, its size is too large (>75kDa band on ladder) to be the uncleaved product.

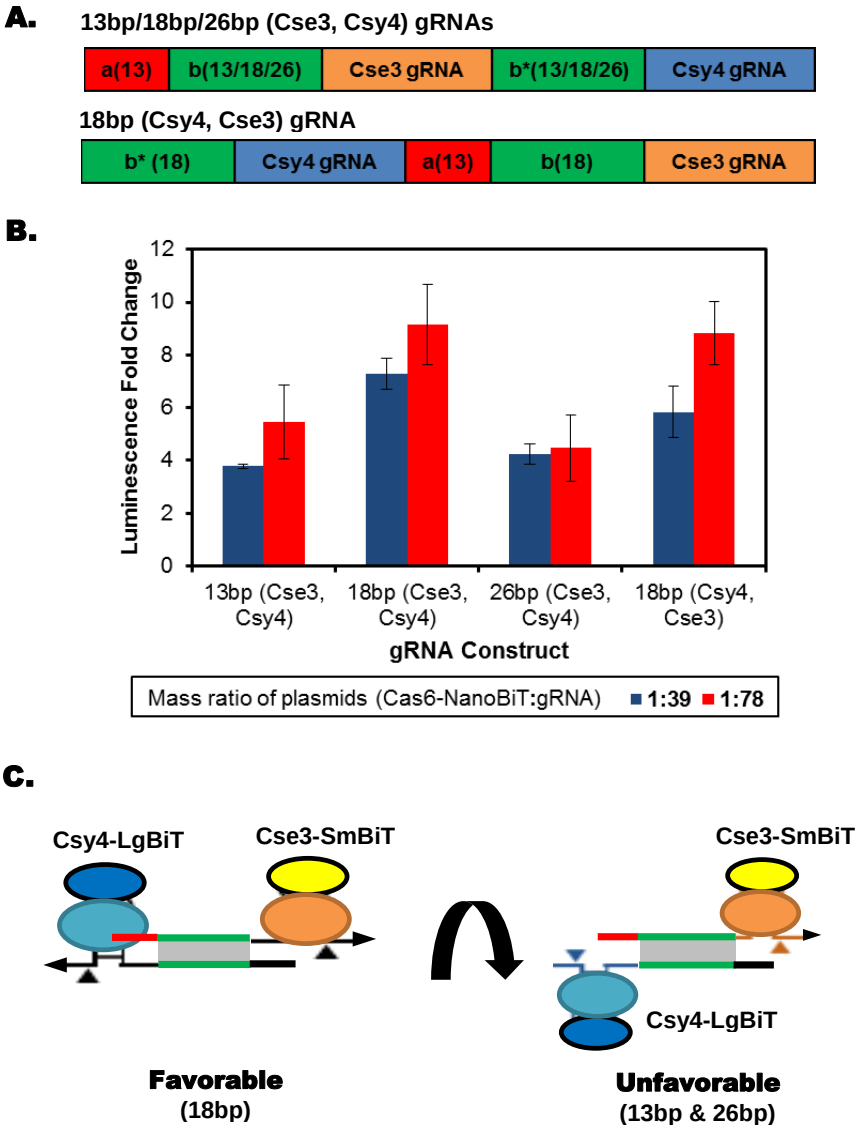

Figure 4.3 Varying ratio of protein to gRNA and varying hybridization domain length. A.) Schematic of features on various gRNA constructs. Lengths (nt) of toehold (red) and hybridization domains (green) are indicated in parentheses. **B.)** Luminescence measured in HeLa cells transfected with a 1:39 (blue) or 1:78 (red) ratio of Cas6-NanoBiT to gRNA. Constructs of 13bp, 18bp, and 26bp hybridization domain lengths were compared. Luminescence fold change is defined as the luminescence measured in the gRNA construct over that of the empty vector. Error bars indicate standard deviation from three biological replicates. **C.)** Schematic of helical rotation and its effect on split nanoluciferase reconstitution. The 13bp and 26bp constructs lie in an unfavorable orientation as compared to the 18bp construct.

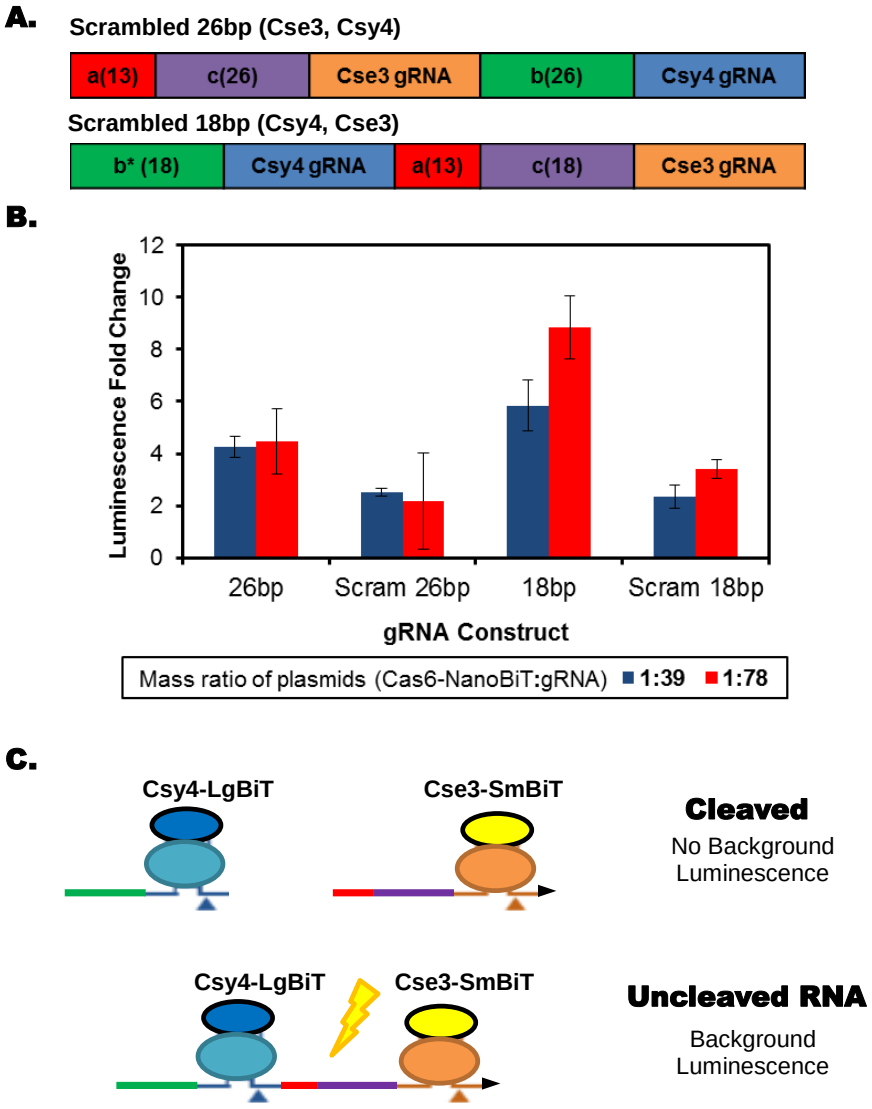

Figure 4.4 Luminescence enhancement in scrambled hybridization domain gRNA constructs. A.) Schematic of features on gRNA constructs. The scrambled hybridization region is shown in purple. **B.)** Luminescence fold change of various gRNA constructs over the empty vector control measured from HeLa cell transfection. Scrambled hybridization domain constructs have lower fold changes than their unscrambled counterparts, but they are still elevated over the empty vector control. Error bars represent standard deviation from three biological replicates. **C.)** Model of increased luminescence in scrambled gRNA constructs from uncleaved transcripts.

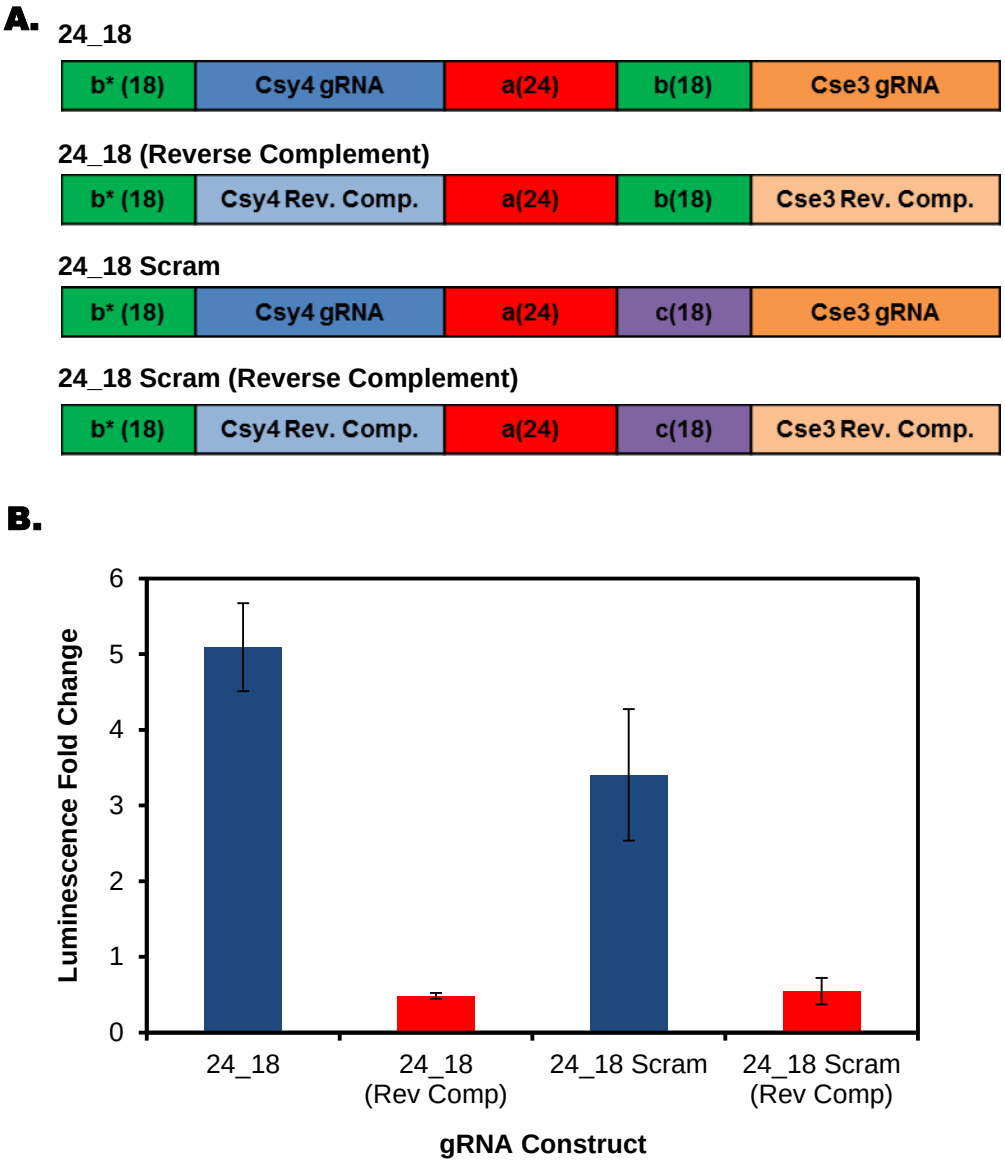

Figure 4.5 Cas6 binding induced protein assembly. A.) Schematic of features on gRNA constructs. The 24_18 architecture refers to a 24nt toehold and an 18nt hybridization region. Reverse complement of Cas6 binding motifs allow similar loop structure to form, but Cas6 no longer bind these sequences. **B.)** Luminescence fold changes measured in transfected HeLa cells. gRNA constructs with correct Cas6 binding motifs (blue) have increased luminescence over the empty vector control. Whereas, in gRNA constructs with reverse complement Cas6 motifs no increase is observed. This indicates that split nanoluciferase assembly is induced through Cas6 binding to RNA. Error bars depict standard deviation from three biological replicates.

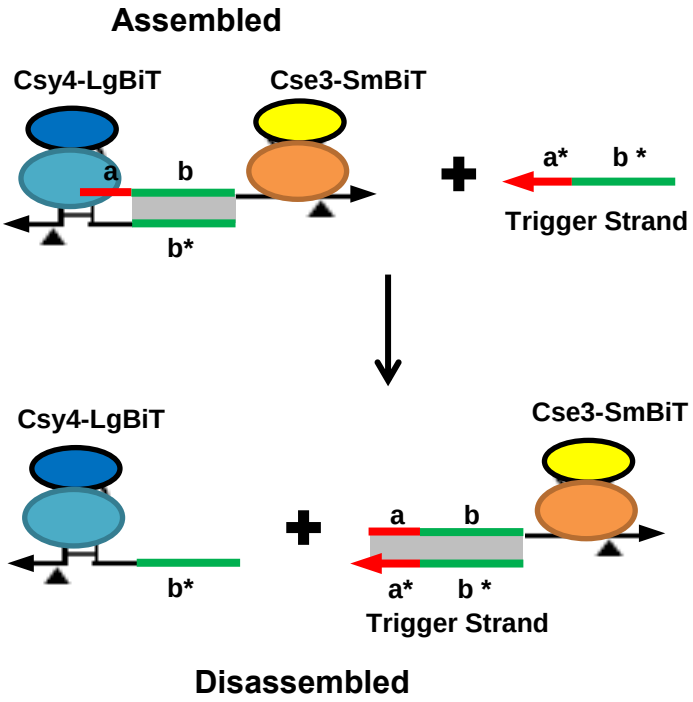

Figure 4.6 Toehold-mediated strand displacement induced disassembly. The RNA trigger strand binds to the toehold (red) on the Cse3 gRNA inducing branch migration and displacement of the Csy4 gRNA. This results in protein disassembly and loss of nanoluciferase activity.

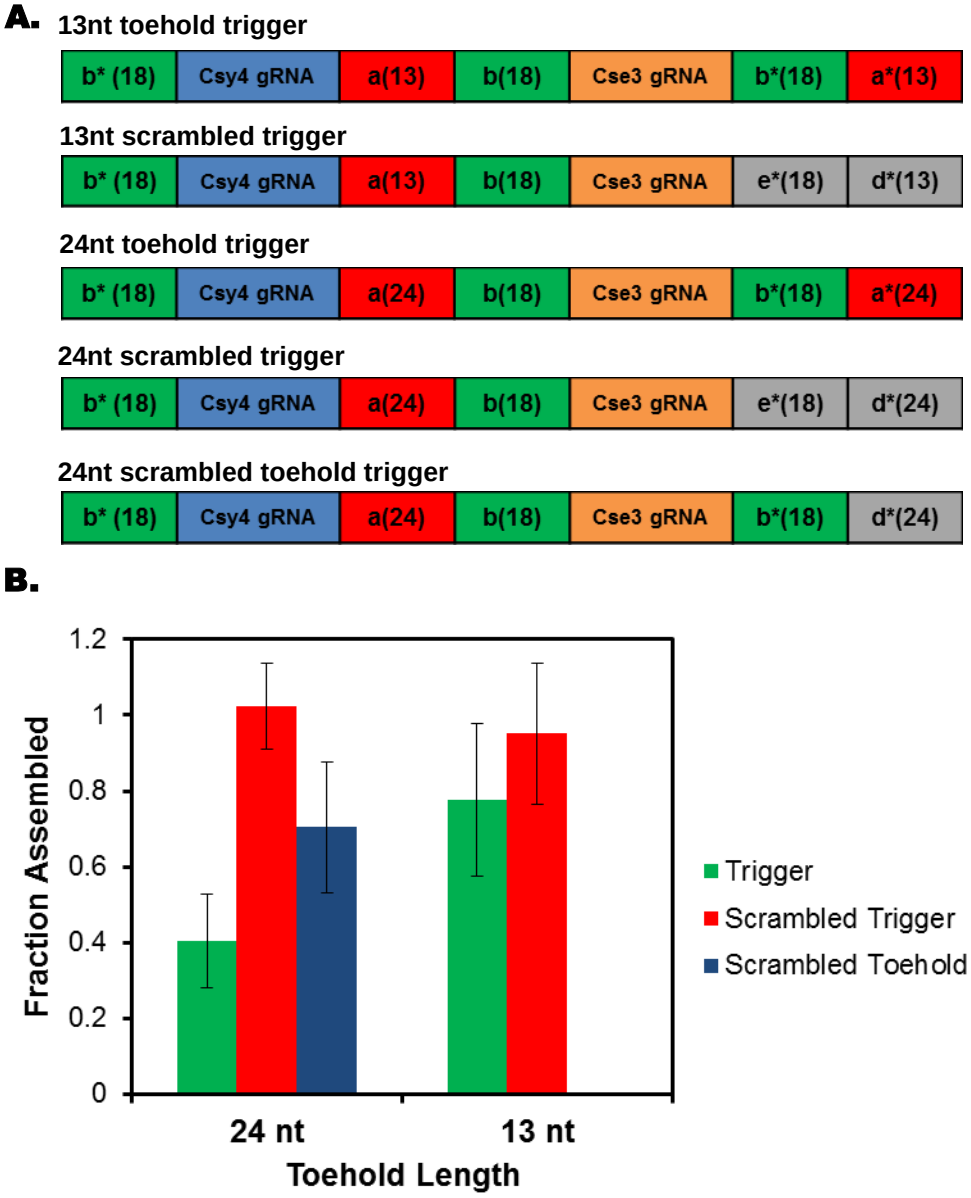

Figure 4.7 Varying toehold length on trigger strand. A.) Schematic of the various RNA scaffolds with constitutive triggers tested. Numbers in parenthesis indicate the length of each domain. In all designs, the hybridization domain has a length of 18nt. **B.)** The luminescence measured in HeLa cells transfected with various trigger constructs. Fraction assembled is defined as the luminescence normalized between the assembled construct with no trigger (set at 1) and the scrambled hybridization domain construct (set at 0) for each toehold length. Error bars represent standard deviation from three biological replicates.

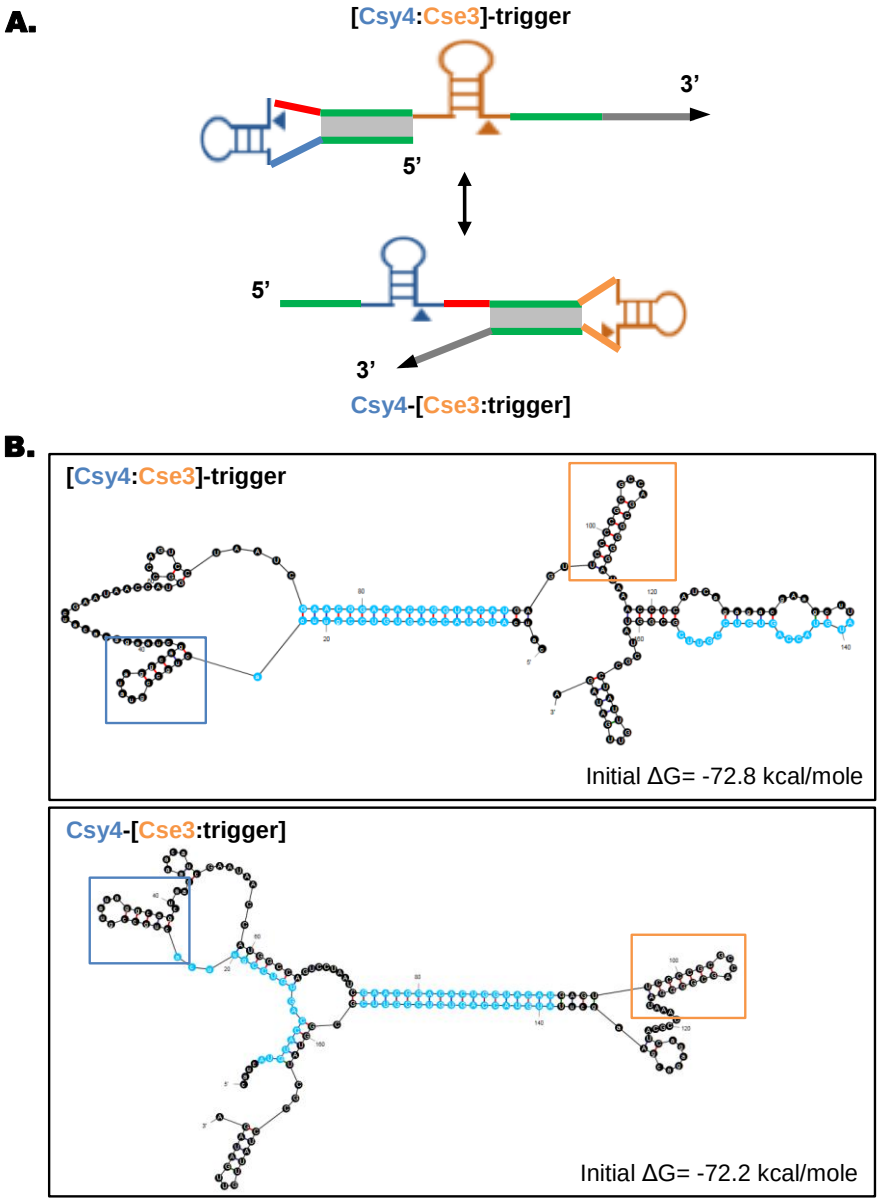

Figure 4.8 Competitive hybridization of trigger strand. A.) Model of competitive hybridization between Csy4 gRNA and the trigger strand for Cse3 gRNA. This is a strand displacement independent process. The two competing configurations are depicted with the duplex indicated by brackets. **B.)** Predicted RNA structures of the scrambled toehold trigger construct. Both duplex conformations were predicted by m-fold with similar structural free energies (Initial ΔG). The hybridization domains are highlighted in cyan and Csy4 and Cse3 loop motifs are boxed in blue and orange, respectively. Images were generated on the m-fold web server.

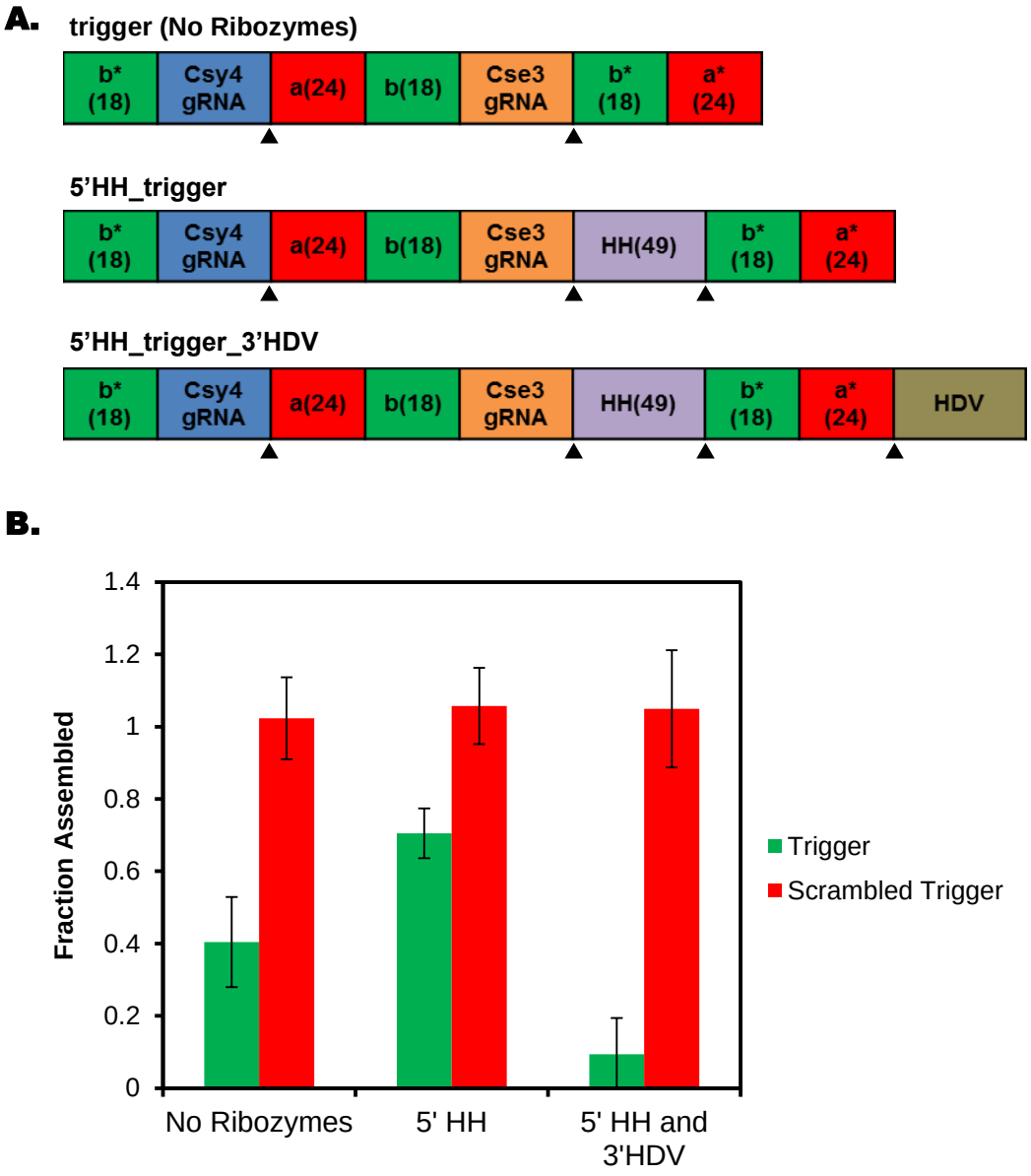

Figure 4.9 Addition of ribozymes to trigger strand. A.) Schematic of features on various trigger designs. Black triangles indicate RNA cleavage locations. All designs have the same architecture of a 24nt toehold and 18nt hybridization domain. Numbers within parenthesis describe the length of the domain. **B.)** The luminescence measured in HeLa cells transfected with various ribozyme designs. Fraction assembled is defined as the luminescence normalized between the assembled construct with no trigger (set at 1) and the scrambled hybridization domain construct (set at 0). Error bars represent standard deviation from three biological replicates.

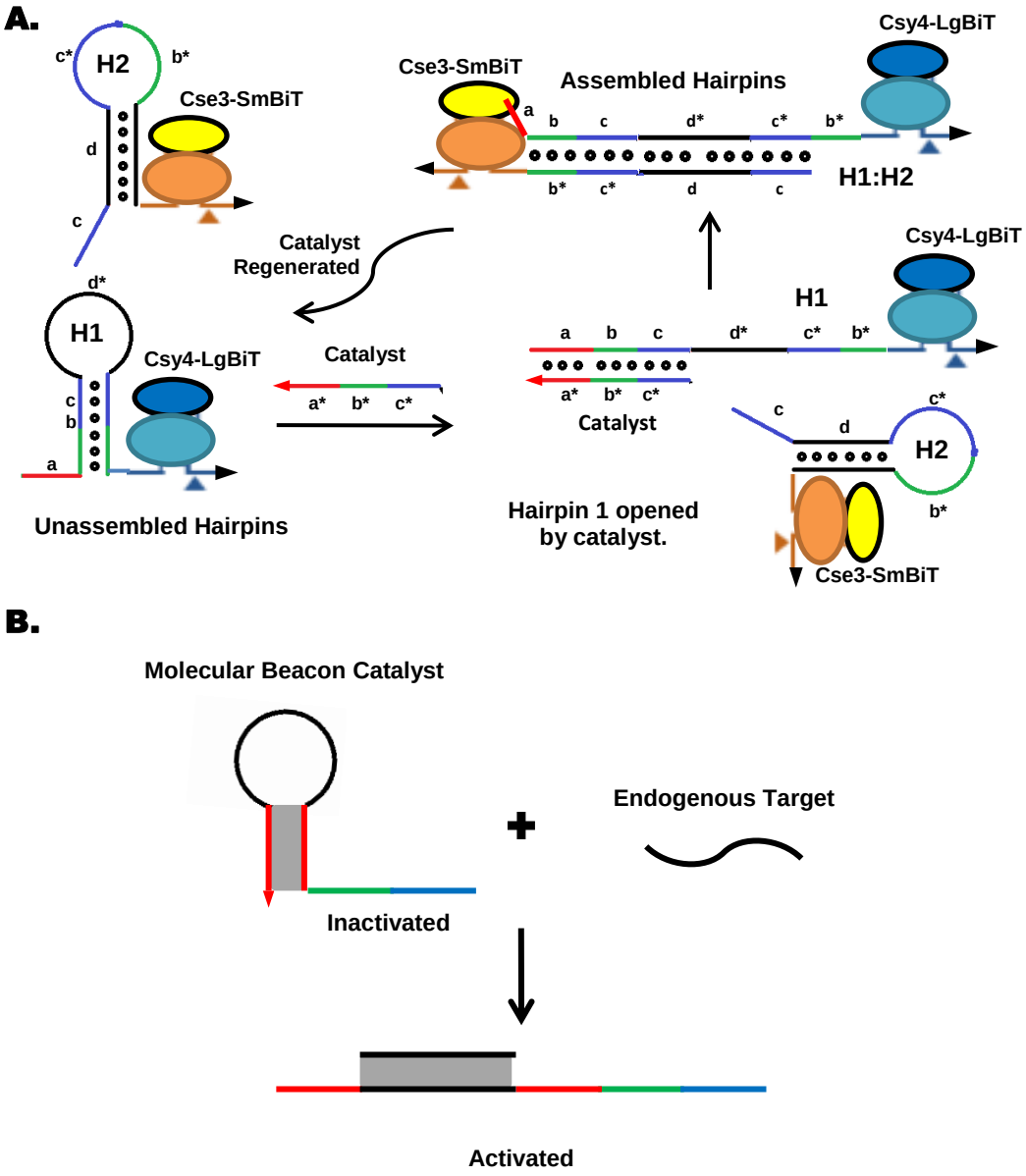

Figure 4.10 Mechanism of Cas6-guided protein assembly using CHA. A.) Schematic of Cas6-guided CHA circuit. Csy4 and Cse3 gRNA motifs are added to the 3' end of the CHA hairpins. In the absence of the catalyst strand the CHA hairpins are kinetically trapped and cannot interact. When the catalyst strand is present it triggers hairpin assembly and docks Csy4-LgBiT and Cse3-SmBiT on the same RNA scaffold. **B.)** Schematic of molecular beacon gated catalyst. The hairpin inactivates the toehold region (shown in red). When an endogenous target binds to the complementary loop region (shown in black) of the molecular beacon it opens up the stem, exposing the toehold and activating the catalyst strand.

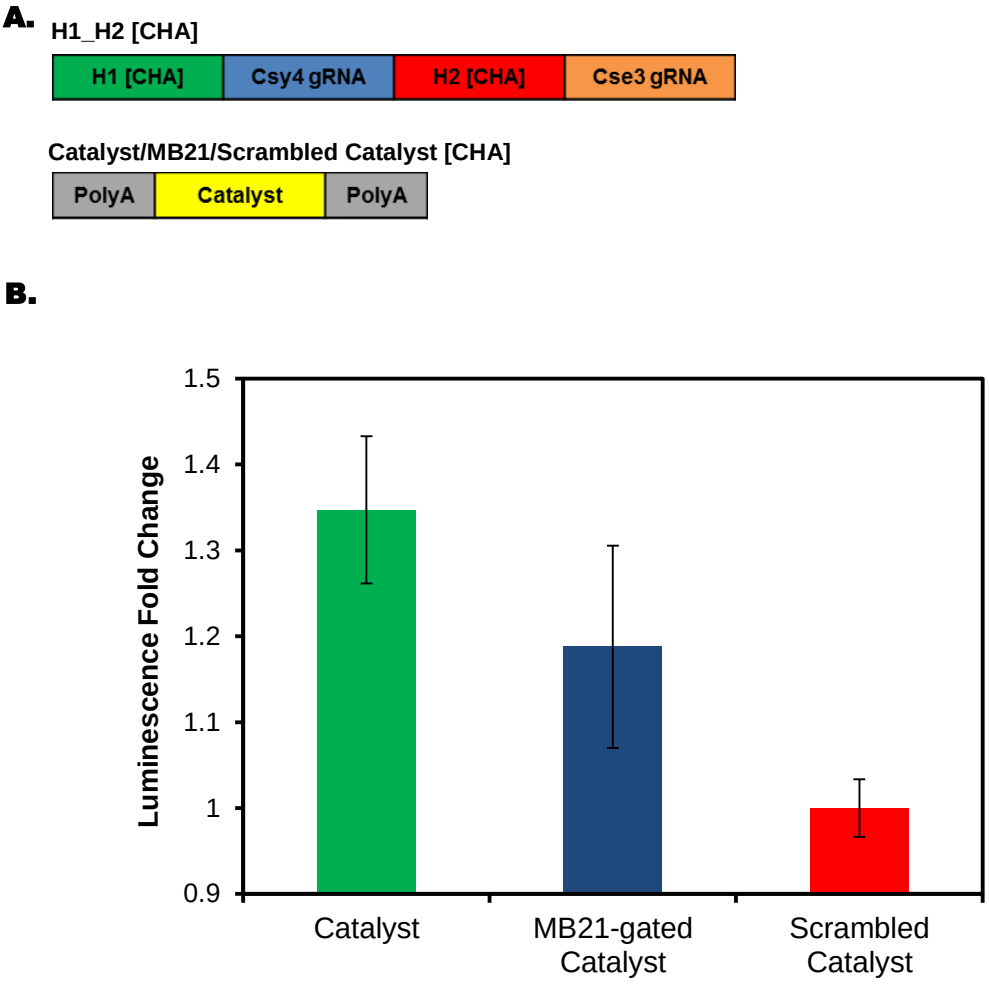

Figure 4.11 Cas6-guided protein assembly using CHA. A.) Schematic of features on CHA circuit plasmids. The CHA hairpins are located in tandem on a single transcript while the catalyst strand in expressed from a second plasmid. **B.)** Luminescence observed when various catalyst constructs were co-transfected with Cas6-NanoBiT and H1-H2 [CHA] plasmid. Fold change was normalized to the scrambled catalyst (set at 1). Error bars represent standard deviation from three biological replicates.

REFERENCES

1. Chen, Y., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nature nanotechnology* **10**, 748-760 (2015).

2. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature chemistry* **3**, 103-113 (2011).

3. Jung, C. & Ellington, A. D. Diagnostic applications of nucleic acid circuits. *Accounts of chemical research* **47**, 1825-1835 (2014).

4. Lopez, R., Wang, R. & Seelig, G. A molecular multi-gene classifier for disease diagnostics. *Nature chemistry* **10**, 746-754 (2018).

5. Kanasty, R., Dorkin, J. R., Vegas, A., & Anderson D.. Delivery materials for siRNA therapeutics. *Nature Materials* **12**, 967-977 (2013).

6. Tatiparti, K., Sau, S., Kashaw, S. K. & Iyer, A. K. siRNA Delivery Strategies: A Comprehensive Review of Recent Developments. *Nanomaterials* **7**, 77 (2017).

7. Hemphill, J. & Dieters, A. DNA Computation in Mammalian Cells: MicroRNA Logic Operations. *Journal of American Chemical Society* **135**, 10512–10518 (2013).

8. Wu, C. *et al*. A Nonenzymatic Hairpin DNA Cascade Reaction Provides High Signal Gain of mRNA Imaging inside Live Cells. *Journal of the American Chemical Society* **137**, 4900-4903 (2015).

9. Cheglakov, Z., Cronin, T. M., He, C. & Weizmann, Y. Live Cell MicroRNA Imaging Using Cascade Hybridization Reaction. *Journal of the American Chemical Society* **137**, 6116–6119 (2015).

10. Groves, B. *et al*. Computing in mammalian cells with nucleic acid strand exchange. *Nature nanotechnology* **11**, 287-294 (2015).

11. Shawn M. Douglas, Ido Bachelet & George M. Church. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831-834 (2012).

12. Rosi, N. L. *et al*. Oligonucleotide-Modified Gold Nanoparticles for Intracellular Gene Regulation. *Science* **312**, 1027-1030 (2006).

13. Sahay, G. *et al*. Efficiency of siRNA delivery by lipid nanoparticles is limited by endocytic recycling. *Nature biotechnology* **31**, 653-658 (2013).

14. Fisher, T. L., Terhorst, T., Cao, X. & Wagner, R. W. Intracellular disposition and metabolism of fluorescently-labeled unmodified and modified oligonucleotides microinjected into mammalian cells. *Nucleic acids research* **21**, 3857-3865 (1993).

15. Watts, J. K., Deleavey, G. F. & Damha, M. J. Chemically modified siRNA: tools and applications. *Drug Discovery Today* **13**, 842-855 (2008).

16. Lesnik, E. A. & Freier, S. M. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry* **34**, 10807-10815 (1995).

17. Bhadra, S. & Ellington, A. D. Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. *Nucleic acids research* **42**, e58 (2014).

18. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642-646 (2011).

19. Ouellet, J. RNA Fluorescence with Light-Up Aptamers. *Frontiers in chemistry* **4**, 29 (2016).

20. Jaeger, L. *et al*. In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nature Nanotechnology* **5**, 676-682 (2010).

21. Geary, C., Rothemund, P. W. K. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804 (2014).

22. Karunanayake Mudiyanselage, A. P. K. K. *et al*. Genetically Encoded Catalytic Hairpin Assembly for Sensitive RNA Imaging in Live Cells. *J. Am. Chem. Soc.* **140**, 8739-8745 (2018).

23. Green, A., Silver, P., Collins, J. & Yin, P. Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell* **159**, 925-939 (2014).

24. Green, A. A. *et al*. Complex cellular logic computation using ribocomputing devices. *Nature* **548**, 117-121 (2017).

25. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **333**, 470-474 (2011).

26. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W. J. Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res* **44**, 9050-9070 (2016).

27. Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences* **40**, 58-66 (2015).

28. Horvath, P. & Barrangou, R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **327**, 167-170 (2010).

29. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Brouns, S. J. J. & Oost, V. D., John. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology* **13**, 722-736 (2015).

30. Haurwitz, R. E., Sternberg, S. H. & Doudna, J. A. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO Journal* **31**, 2824-2832 (2012).

31. Sternberg, S. H., Haurwitz, R. E. & Doudna, J. A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA (New York, N.Y.)* **18**, 661-672 (2012).

32. Brouns, S. J. J. *et al*. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**, 960-964 (2008).

33. Guo, R. *et al*. Inhibition of HIV-1 Viral Infection by an Engineered CRISPR Csy4 RNA Endoribonuclease. *PLoS One* **10**, e0141335 (2015).

34. Nissim, L., Perli, S., Fridkin, A., Perez-Pinera, P. & Lu, T. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Molecular Cell* **54**, 698-710 (2014).

35. Borchardt, E. K. *et al*. Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4. *RNA (New York, N.Y.)* **21**, 1921-1930 (2015).

36. Dixon, A. S. *et al*. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS chemical biology* **11**, 400-408 (2016).

37. Schoen, I., Krammer, H. & Braun, D. Hybridization kinetics is different inside cells. *Proceedings of the National Academy of Sciences* **106**, 21649 (2009).

38. Lukacs, G. L. *et al*. Size-dependent DNA Mobility in Cytoplasm and Nucleus. *Journal of Biological Chemistry* **275**, 1625-1629 (2000).

39. Liu, Z. *et al*. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific reports* **7**, 2193-9 (2017).

40. Taylor, P., Rixon, F. & Desselberger, U. Rise per base pair in helices of double-stranded rotavirus RNA determined by electron microscopy. *Virus Research* **2**, 175-182 (1985).

41. Sachdeva, G., Garg, A., Godding, D., Way, J. C. & Silver, P. A. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic acids research* **42**, 9493-9503 (2014).

42. Davis, K. M. Directed Evolution and Engineering of CRISPR-Associated Nucleases. (2016).

43. Emilsson, G. M., Nakamura, S., Roth, A. & Breaker, R. R. Ribozyme speed limits. *RNA (New York, N.Y.)* **9**, 907-918 (2003).

44. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415 (2003).

45. Goodwin, E. C. & Rottman, F. M. The 3'-flanking sequence of the bovine growth hormone gene contains novel elements required for efficient and accurate polyadenylation. *Journal of Biological Chemistry* **267**, 16330-16334 (1992).

46. Flaherty, K. M., Pley, H. W. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74 (1994).

47. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-574 (1998).

Chapter 5

DISSERTATION CONCLUSIONS AND FUTURE WORK

This dissertation builds towards creating a cancer computing device based on the combined strength of nucleic acids and proteins. In such a device, nucleic acids provide input sensing and decision making, which is then transduced to protein actuation. To achieve this, a platform technology was developed to harness toehold-mediated strand displacement for programmable dynamic protein assembly. In each chapter we have focused on important aspects for realizing this platform as a novel class of smart therapeutics.

We first established the technology by synthesizing protein-DNA conjugates to be tested within strand displacement circuits. DNA strand displacement was used to control the spatial proximity and the corresponding FRET response of CFP and YFP. The attachment of proteins to DNA was found to have no negative effects on strand displacement function or kinetic behavior, highlighted by successful execution of multi-input, reversible, and amplification (CHA) architectures. Next, to prove the technology's powerful utility, DNA strand displacement was used to dynamically drive the reconstitution of split yCD activity in the presence of cancer-specific miRNA. Significant prodrug activation capable of cell killing was only observed in the proper "diseased" state. A miRNA-gated CHA circuit was also demonstrated for amplified yCD activity.

In designing the miRNA-gated CHA, we realized that a widespread hurdle exists for adapting *de novo* strand displacement architectures for biologically relevant

contexts. Many designs, especially those involving hairpins, suffer compromises in function and efficiency when integrated with target sequence constraints. To address this, we developed a novel strategy called split associative strand displacement, the benefits of which were demonstrated to be two fold. First of all, by splitting the catalyst strand into its toehold and branch migration regions, strand displacement was conditional on hybridization to a docking strand. Toehold-mediated strand displacement was applied to control this association. Secondly, the docking regions for assembly are sequence decoupled from the split catalyst. Without changing any CHA sequences, miRNAs were used to run CHA with AND, NOT, and NOTAND behavior. This was also generalized successfully to a four-input miRNA classifier design. As opposed to traditional integration, we established a novel approach that allows biological inputs to be elegantly interfaced with *de novo* circuit components through a simple set of design principles. In the study, fluorophore and quencher dyes were used for characterization, but this strategy can be readily incorporated into nucleic acid-protein devices in the future.

Thus far, we had addressed how to design and build nucleic acid-protein hybrids to exhibit features such as, dynamic protein control through strand displacement, therapeutic output functionality, and effective coordination between biological inputs and *de novo* circuitry. In the last part of the dissertation, we sought to implement a nucleic acid-protein device in live mammalian cells. The protein-DNA conjugates that were successfully tested *in vitro* are certainly worth future intracellular delivery studies to further explore if endogenous miRNA can induce effective prodrug activation for cell death.

However, the alternative approach of delivering the device through gene therapy also holds great potential. To investigate this, we implemented a genetically-encoded device, in which expressed protein and RNA components self-assembled to take on dynamic RNA strand displacement induced protein function. Efficient self-assembly between protein and RNA was achieved by harnessing a special family of endoribonucleases called Cas6. The picomolar binding affinity they exhibit towards their RNA substrates was exploited to achieve efficient and specific attachment between protein and RNA. Furthermore, Cas6 cleavage was used to discretize an initial RNA transcript into individual RNA strands. By using orthogonal Cas6 as fusion partners, proteins could be attached to specific Cas6 gRNA which contained 5' extension sequences for running RNA circuitry. When the system was expressed in HeLa cells, we showed that RNA hybridization could be used for Cas6-guided protein assembly. Furthermore, by expressing a trigger strand the proteins could be disassembled. A CHA circuit was also implemented with modest results to demonstrate OFF to ON functionality. This study is the first execution of a genetically-encoded strand displacement circuit for protein assembly inside live mammalian cells.

Future investigation into engineering Cas6 for faster cleavage rates will improve strand discretization and allow this strategy to operate at higher efficiencies. The success of simple RNA strand displacement reactions also warrants the exploration of more complex reaction cascades, such as the associative strand displacement scheme we developed. Finally, the split nanoluciferase reporter can be exchanged for other functional modules to carry out therapeutic action and/or to interact with native cellular pathways.

Ultimately, our technology shows the powerful utility of combing nucleic acid and proteins into hybrid devices, especially when toehold-mediated strand displacement is used to generate computing power. Moving forward, the goal is to further integrate the designs and strategies developed in this dissertation to create and improve upon our first-generation devices. Beyond disease therapeutics, this technology has widespread applicability and can be expanded to generate synthetic programmable protein switches for any biological system of interest.

Appendix A

REPRINT PERMISSIONS

SPRINGER NATURE

Title: Dynamic protein assembly by programmable DNA strand displacement
Author: Rebecca P. Chen et al
Publication: Nature Chemistry
Publisher: Springer Nature
Date: Mar 12, 2018
Copyright © 2018, Springer Nature

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials.

Already **a RightsLink user** or want to learn more?

Author Request

If you are the author of this content (or his/her designated agent) please read the following. If you are not the author of this content, please click the Back button and select no to the question "Are you the Author of this Springer Nature content?".
Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:
To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).
The author and any academic institution, where they work, at the time may reproduce the contribution for the purpose of course teaching.
To reuse figures or tables created by the Author and contained in the Contribution in oral presentations and other works created by them.
To post a copy of the contribution as accepted for publication after peer review (in locked Word processing file, of a PDF version thereof) on the Author's own web site, or the Author's institutional repository, or the Author's funding body's archive, six months after publication of the printed or online edition of the Journal, provided that they also link to the contribution on the publisher's website.
Authors wishing to use the published version of their article for promotional use or on a web site must request in the normal way.

If you require further assistance please read Springer Nature's online author reuse guidelines.

For full paper portion: Authors of original research papers published by Springer Nature are encouraged to submit the author's version of the accepted, peer-reviewed manuscript to their relevant funding body's archive, for release six months after publication. In addition, authors are encouraged to archive their version of the manuscript in their institution's repositories (as well as their personal Web sites), also six months after original publication.

v1.0

BACK CLOSE WINDOW