7-1-2014

Engineering Catalytic Molecular Logic Devices for Biodetection

Carl Brown III

Carl W. Brown, III
Candidate

Biomedical Sciences
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Steven W. Graves, Ph.D , Chairperson

Darko Stefanovic, Ph.D.

Larry A. Sklar, Ph.D.

Dave Whitten, Ph.D.

Bridget Wilson, Ph.D.

ENGINEERING CATALYTIC MOLECULAR LOGIC DEVICES FOR BIODETECTION

BY

CARL W. BROWN, III

B.S. Biomedical Engineering, University of Rochester, 2008

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biomedical Sciences

The University of New Mexico
Albuquerque, New Mexico

July 2014

DEDICATION

I would like to dedicate this dissertation to my wife, Karen Brown. Her love and support for me over the years has been one of the few constants in an ever-changing graduate experience for me. She was there through exciting discoveries and disappointing letdowns, and waited patiently as I repeatedly underestimated how long it would take me to finish my experiments. Her support for me never waivered, even as late nights grew later and deadlines drew closer. She served as a constant reminder of keeping things in perspective, and that sometimes, the right answer really is “potato”. Nothing more need be said.

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor and mentor Dr. Steve Graves for all of his advice, input, and guidance over the years. He gave me a tremendous amount of intellectual freedom on each of the many projects I have worked on, displaying a trust in me even when experimental results might have suggested a different course. Being allowed to take risks helped me gain a knowledge and understanding of my work that could not have been achieved otherwise and a wonderful satisfaction of the work we accomplished together. I would also like to thank my co-advisor Dr. Darko Stefanovic, who broadened my perspectives on the more theoretical and abstract ideas and has been instrumental in my professional development. I am grateful and honored to have worked for both mentors, who have made my graduate career a positive and rewarding experience. I would like to thank my committee members, Dr. Larry Sklar, Dr. Dave Whitten, and Dr. Bridget Wilson, for their guidance over the years.

I would like to thank Dr. Matthew Lakin, who worked on these projects side-by-side with me for years. Assuming the role of post-doc, mentor, co-worker, and friend, he has influenced nearly every aspect of my graduate development and I am a better researcher for it. I would also like to thank Dr. Menake Piyasena, whose guidance and support was invaluable to my early development.

I have also had the privilege to mentor many fantastic researchers over the years, which has made me a better scientist, mentor, and person. I would especially like to thank the two undergraduates, Hannah West and Eli Horwitz, who each worked with me for a considerable amount of time. I appreciate their dedication, hard work, and great conversation over that time. They were mature beyond their years. I cannot thank them enough.

Finally, I would like to thank my friends and family who have supported me throughout the years. I would like to thank my parents, for instilling in me the confidence and the competitive spirit to never settle for anything less than what I could accomplish. I would also like to thank my in-laws, Keith and Lynne Miller, for their unconditional support and encouragement, and my grandmother Lenora Brown, who is the epitome of a lifelong learner. Meliora.

ENGINEERING CATALYTIC MOLECULAR LOGIC DEVICES FOR BIODETECTION

BY

Carl W. Brown, III

B.S. Biomedical Engineering, University of Rochester, 2008
Ph.D. Biomedical Sciences, University of New Mexico, 2014

ABSTRACT

This dissertation describes the development of DNA computing techniques and molecular logic devices specifically engineered for direct translation to biological sample detection. As disease states originate at the molecular level, it is critical to design diagnostic and therapeutic devices that are capable of molecular-scale sensing and decision-making in the cellular environment. The predictable nature of DNA hybridization and secondary structure formation enables programmable interactions, providing a stable, cost-effective, and biocompatible mechanism for making decisions on the molecular scale. The incorporation of DNAzymes, DNA strands that can perform a variety of chemical reactions, adds innate catalysis and a rich biochemical diversity to DNA logic. By regulating DNAzyme activity via hybridization-based approaches, we have developed a new mechanism for implementing DNA logic, referred to as DNAzyme displacement. This mechanism was used for the construction of DNA logic gates, extended logic cascades, and sensitive biosensors, each capable of operating in non-pristine conditions and under minimal purification and setup restrictions. Logic cascades were constructed through the development of a signal propagation molecule known as

a structured chimeric substrate (SCS), which was able to pass a signal between any DNAzyme pair, resulting in the longest synthetic DNA cascade to date. A multi-step DNAzyme displacement reaction was developed for the construction of modular biosensor gates, capable of rapidly multiplexing samples with a limit of detection of 7.4 pM. Other innovative experimental characterization included high-throughput screening efforts of a DNAzyme and alternative methods of compartmentalization including surface-based and lipid-conjugated DNA and protein reactions. This work shows the potential of using DNA to implement molecular logic for the development of intelligent biosensors.

Table of Contents

List of Figures

List of Tables

Chapter 1.
Introduction

1.1 Biodetection, Theranostics, and the Biological Computer

The ideal of personalized medicine and point-of-care detection has generated considerable excitement in the biomedical community. There is a considerable allure to an individual-specific approach to understanding the interaction and progression of disease and infection to provide the most advanced and relevant treatment possible. This is a noble goal indeed, but if we are to provide truly personalized medicine – to detect the onset of disease, the source of infection, or the mutation of an oncogenic protein – we must set our sights on the molecular underpinnings of these events and seek to identify and treat disease states in the complex biochemical interactions of the cellular milieu. On the front end of this problem lies the issue of biodetection. Significant time and effort has been devoted to the symptomatic detection of diseases with considerable success, but for effective treatment many diseases require identification before symptoms appear. Biodetection at the molecular level is a tantalizing prospect, but a much more intractable problem; issues such as sequence mutations, low target concentrations, specificity, and sensitivity each present their own unique challenges to the development of biosensors. On the back end lies the prospect for treatment and the development of targeted therapeutics to be released under the proper conditions. Unification of both detection and treatment aspects under a single device may then culminate in an autonomous theranostic device capable of operating at the cellular or molecular level, a true biological computer. The first step to fulfilling this vision of personalized point-of-care medicine requires the

development of molecular scale detection capabilities in order to directly interface with the molecules indicative of disease states.

To improve biodetection capabilities to this degree, we require a technological advance akin to the molecular valves of Isaac Asimov's Multivac in *The Last Question* – a mechanism for making sophisticated decisions using the information available to us at the level of individual molecules. Cells, experts in this field for four billion years, rely on complex networks of proteins; our knowledge currently falls short of using *de novo* protein design for synthetic decision networks. However, the advent of DNA nanotechnology, self-assembly, and computation through the predictable formation of nucleic acid interactions has provided us with the principles and building blocks needed to realize the vision of "molecular valves" using natural biochemistry.

Already an impressive amount of work has demonstrated the ability of DNA to make decisions, spawning the field of DNA computation, but the application of these computational principles for biodetection remains an untapped potential. There have been many different mechanisms developed to use DNA to make decisions; here we focus on the use of innately catalytic DNA strands called DNAzymes. My thesis is that precise thermodynamic and kinetic design of nucleic acid structures will enable the development of scalable DNA computational circuits robust enough for use in biodetection assays. This has applications for basic scientific advances in DNA nanostructure design, dynamic DNA interactions, and intermolecular interaction regulation, and lays the groundwork for continued work towards a new generation of biological-inspired

and targeted theranostic devices, yet another step toward a true biological computer[1].

1.2 DNA computation and Molecular Logic

Even with the initial discovery of DNA, Watson and Crick immediately recognized the structure being conducive to the transmission of information[2]. Since then, the thermodynamics of DNA hybridization has been studied extensively[3-6], and the simplicity of Watson-Crick base pairing (A binds to T, G binds to C) has led to the development of modeling software[7-9], which can accurately predict DNA sequence interactions and secondary structure formation. As science has yet to unlock the complex thermodynamics that underlie protein folding, the predictability of DNA hybridization and secondary structure has generated considerable excitement for the prospects of engineering specific and directed molecular interactions, giving rise to the fields of DNA nanotechnology[10,11] and, later, DNA computation.

The field of DNA computation began 20 years ago with the seminal paper by Leonard Adleman, in which he demonstrated that DNA hybridization could be programmed to devise solutions to computationally intractable problems, in this case an instance of the directed Hamiltonian path problem[12]. Since this initial demonstration of computation, DNA nanotechnology has been applied to many other dynamic nanoscale processes such as self-assembly[13-15], actuation[16-19], molecular walkers[17,20], and motors[18]. Although there have been other implementations of molecular computing such as using naturally occurring proteins[21-24], the programmability of base pair interactions and the predictability

of secondary structure through established hybridization models[6] makes DNA an ideal substrate for the development of *de novo* molecular logic gates, circuits, and computers.

To date, there have been several different implementations using DNA to create logic gates and circuits, including tile self-assembly[25,26 27,28], hairpin assembly[29-31], ribozymes[32,33], DNAzymes[34-37], and strand displacement[38-40]. The computation of tile self-assembly implements an algorithm for the construction of highly ordered DNA nanostructures, which can encode results of an arbitrary computation, emulating the capabilities of silicon-based digital computers[41,42]. DNAzyme-based and DNA strand displacement-based systems both compute using dynamic interactions between separate nucleic acid complexes to propagate signals. Unfortunately, there has been very little integration between logic gates of different biochemical mechanisms[43]. Circuits tend to be homogenous, consisting only of the base technology (*e.g.* strand displacement, loop-regulated E6 DNAzymes, etc.), limiting the potential responses that could be achieved by the integration of different types of DNA logic gates. This scientific and technological gap must be bridged for the full biodetection potential of molecular computing algorithms to be realized.

1.3 DNAzymes and functional nucleic acids for protein-free biosensing

The use of functional nucleic acids as an enzyme-free biosensor platform holds considerable promise for the development of low-cost biodetection diagnostics and the subsequent evolution of such devices for the implementation of biomedical theranostics[44]. DNA itself is a natural information carrier through

sequence-specific hybridization; its biocompatibility, stability, and decreasing synthesis costs make it an ideal candidate for use in biosensors. Although the use of hybridization-based probes has seen considerable interest and success [45,46], the use of functional nucleic acids richly expands the potential utility for DNA-based probes and the diversity of their targets.

Functional nucleic acids (FNA), which include ribozymes, DNAzymes (also known as deoxyribozymes or DNA enzymes in the literature), aptamers, and aptazymes, contain specific sequences that confer additional functionality[44]. In the case of ribozymes and DNAzymes, these sequences act as a catalyst for a specific biochemical reaction. Aptamers, on the other hand, are sequences that possess an intrinsic affinity for a selected target, which are typically small molecules like ATP[47] but can be larger targets as well. Aptazymes are hybrid sequences of aptamers and DNAzymes in which the binding of the aptamer target leads to the modulation of DNAzyme activity[48-50]. Assays utilizing FNAs have been developed for a diverse set of input modalities, including heavy metal sensing[51-53], small molecule sensing[47-49], nucleic acid-based pathogen detection[54,55], cellular binding and analysis[1,56,57], and molecular logic and computation[34-37,58-60]. FNAs have also demonstrated a wide variety of output modalities[30], such as optical[61], surface plasmon resonance[62], fluorescence[36], electrochemical[63-66], *in vivo* mRNA knockdown[67-75], and gold nanoparticle aggregation[51,52,76-81]. The outputs can also include the release of therapeutic compounds[1,56,82], such as the release of a caged anti-tumorgenic compound through aptameric conformation change[83-85] or DNAzyme to cleave a target

mRNA[71] has already been demonstrated. The diversity of input and output alternatives holds promise for the development of a wide variety of biosensors tailored to specific conditions or targets[86], which share common chemical and computational principles.

One of the biggest challenges in biosensor development is achieving the sensitivity necessary to detect the low target concentrations of pathogenic sequences typically available in prepared solutions. One class of FNAs that appear ideally suited to address this challenge is DNAzymes, which are single DNA oligomers capable of catalyzing chemical reactions[68,72,87] previously thought to be catalyzed only by proteins. Their existence was hypothesized after the discovery of naturally occurring ribozymes, although no naturally occurring DNAzymes have been discovered to date. All known DNAzyme catalytic motifs have been artificially selected for using the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) process[58,88]. Since the discovery of the first DNAzyme[89], many more catalytic sequences have been discovered, capable of catalyzing a broad set of biochemical reactions[58] including RNA hydrolysis[89,90], RNA ligation[91,92], DNA ligation[93], DNA phosphorylation[94], HRP-like activity[95,96], and DNA hydrolysis[97,98]. Due to the stability, ease of use, and low cost of synthesis, DNAzymes have generated a significant amount of interest in the scientific community.

The family of RNA-cleaving DNAzymes have some of the fastest catalytic rates of all DNAzymes, although this is still orders of magnitude slower than their protein counterparts. For example, the 8-17 DNAzyme achieves a 10^8 rate

enhancement over spontaneous RNA hydrolysis, while RNase A is capable of a 10^{14} rate enhancement. This fact has made them the best characterized and most widely used. Many different sequences have been discovered that can perform this reaction, generally named for the round of selection in which they were discovered. Of these, the most prominent DNAzymes are E6[89], 8-17[90,99], and 10-23[100], most frequently used due to their small size and relatively high catalytic rate. The structure of these DNAzymes consist of a fixed catalytic core flanked by two substrate binding arms. Mutational analysis has been conducted for the 8-17 DNAzyme core[101], which showed the highly conserved nature in general as well as the necessity of certain bases within the core, which likely play a major role in catalysis of the RNA cleavage reaction. Although RNA folding software[9,102] predicts a small stem forming through hybridization within the catalytic core, it is likely that this structure is more complex. The RNA-cleaving DNAzymes are highly dependent on divalent metal cation cofactors, which likely assist directly in the RNA hydrolysis by associating with the catalytic core[103,104]. The thermodynamic tables used as the basis for thermodynamic structure prediction software such as NUPACK are not parameterized for these salt conditions, which makes prediction of the structure of the catalytic core challenging and research on the interaction and conformation of this secondary structure is ongoing[105,106]. Due to this divalent metal cofactor dependence, many of these DNAzyme sequences can be used as natural heavy metal sensors[107,108]. Although originally selected for in the presence of lead, subsequent selection experiments have found sequences highly active in other metals such as

magnesium, manganese, and zinc, and moderately active with many other divalent metal cations[109].

While substrate specificity in protein enzymes is achieved through tertiary conformational interactions, which are difficult to predict[110], DNAzymes rely on secondary conformation and hybridization via the substrate binding arms. It is typically assumed that the substrate binding arms are completely unconstrained, although it is possible that single bases near the catalytic core may still be necessary for function[101]. These interactions are easily modeled using nucleic acid folding software such as NUPACK[9], which uses nearest neighbor stacking interactions[6] to predict folding and the design of secondary structures[102]. This property makes these DNAzymes very attractive for scientific and engineering purposes, as it allows for the sequence specific programming of DNAzyme/substrate interactions. This interaction is highly specific and requires the hybridization of both substrate binding arms; a single arm binding is insufficient to obtain substrate cleavage. The cleavage reaction can be monitored using a variety of methods including PAGE analysis, radiolabeling, and FRET-based fluorescence measurements (**Figure 1.1**).

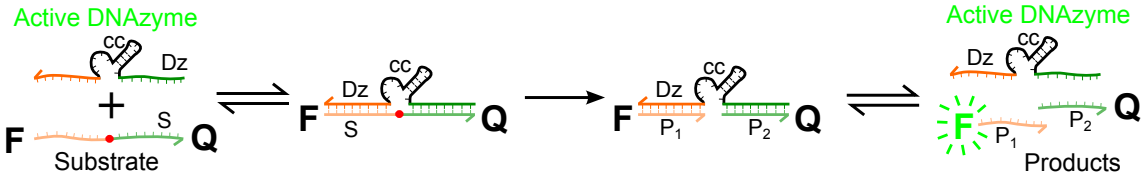

Figure 1.1 – The mechanism of DNAzyme-catalyzed RNA cleavage. Here, the 8-17 DNAzyme binds to its corresponding substrate through hybridization with the substrate binding arms. The catalytic core (cc) is black, while the individual substrate binding arms are colored orange and green. The RNA base is hydrolyzed, which results in the formation of two shorter product strands, which freely dissociate from the DNAzyme. Fluorescence increase results from a loss of FRET.

We focus our efforts on the 8-17 DNAzyme[99], originally discovered in 1997 by Santoro and Joyce using Mg^{2+} as the metal ion cofactor[90]. A Zn^{2+} optimized version was later discovered in 2000 by Li *et al*[109], and this sequence serves as the basis for nearly all of the RNA-catalysis experiments conducted herein. This DNAzyme is highly active around a pH of 7 and activity has been characterized in presence of a wide variety of divalent metal cations. It has a catalytic core of 15 nucleotides (nt), one of the smallest DNAzymes cores discovered to date, making it easy to manipulate. It also has one of the highest catalytic rates of all RNA-cleaving DNAzymes, at ~1/min under typical high salt buffer conditions. Given the rate of 10^{-7}/min for spontaneous RNA hydrolysis, this corresponds to roughly a 10^{8} rate enhancement of RNA hydrolysis[111]. Conversely, the rate enhancement attained by the near-perfect protein catalyst RNAse A is roughly a 10^{12} rate enhancement. The limitation of the DNAzyme on rate enhancement compared to RNAse A does not appear to be a hard biochemical limit[112]; indeed, efforts to improve the cleavage rate by engaging additional biochemical

mechanisms would be highly desirable. The biochemical mechanism of this DNAzyme is discussed in detail in **Chapter 7**.

1.4 DNAzymes for molecular computation

The first experimental demonstration of using DNAzymes for molecular logic gate construction was in 2002 by Stojanovic and Stefanovic in the *Journal of the American Chemical Society*[36], an extension of their earlier work combining DNAzymes with molecular beacons[113]. Here, they use both the 8-17 and E6 DNAzymes for the development of Boolean logic gates, specifically YES, NOT, AND, and XOR. In the case of YES and AND gates, inhibition was achieved through molecular beacon-type stem loops, which bind back on one or both substrate binding arms to create an intramolecular interaction, thereby blocking any interaction with the substrate. The addition of an input oligomer sequence would bind to these loops, creating a more favorable duplex between the input and loop than the competing hybridization between the loop and DNAzyme. This opens up the loop and allows the substrate to productively bind the DNAzyme, resulting in an output signal via loss of FRET after substrate cleavage. Although the cleavage rate of E6 is notably slower than that of 8-17 (0.04/min vs. 1/min), the catalytic core of the E6 DNAzyme is larger and more flexible, which allowed an additional stem loop to be added to the core of the E6 DNAzyme without significantly affecting DNAzyme catalysis. This properly enabled the construction of a NOT gate, in which the binding of an input oligomer would hybridize to the loop and deform the catalytic core to deactivate the DNAzyme.

Because the substrate specificity of DNAzymes is achieved through base-pair hybridization, multiple such gates could be run in parallel in the same solution. This was initially implemented in a large-scale automaton[114], capable of playing tic-tac-toe, and subsequently improved upon in a second iteration by removing many of the restrictions of the original design[115]. This was a critical step in the use of DNAzymes to construct automated decision-making, and demonstrated the potential complexity of information process by biomolecular circuits[116]. However, the loop-based approach to DNAzyme inhibition generally uses high (μM) target input concentrations, orders of magnitude above the sensitivity necessary for biosensor detection. These interactions are typically concentration driven; although the input binding may be more thermodynamically favorable, the intramolecular binding of the loop to the DNAzyme is kinetically favorable. This necessitates the high input concentrations in order to bias the reaction to activate.

Since this demonstration, several other groups have demonstrated alternative molecular logic gate implementations using DNAzymes, including DNA ligation[37], DNA-cleavage[34], and split DNAzymes[117]. Of particular interest is the split DNAzymes, which have been developed into large libraries of computing subunits[35,118], even demonstrating operation in a cellular environment[119]. This mechanism works by separating the catalytic core of the DNAzyme (either E6 or 10-23) into two subunits. The structure of the DNAzyme is modified slightly in this case; in addition to the catalytic core there are two input recognition arms opposite the substrate arms. With the addition of a target input, these arms are

brought together to reform the catalytic core, which can now cleave the chimeric substrate. Although this mechanism is highly specific, it also slows down the catalysis rate of the DNAzyme, likely due to inefficiencies in the reforming of the catalytic core.

The use of DNAzymes for biomolecular circuit development holds several key advantages over strand displacement. As DNAzymes can catalyze RNA hydrolysis in multiple turnover reactions, they provide each logic gate with an innate signal amplification capability without requiring additional strands in solution. Signal amplification is especially important for biodetection, where target strands may be present in low concentrations, which can be anywhere from nanomolar to femtomolar. The minimization of strands in solution is an important factor for the development of complex circuits capable of operating in a random DNA background, as well as for reducing device costs. Furthermore, the incorporation of DNAzymes into biodetection circuits provides access to the wide array of DNAzyme-catalyzed reactions, which could lead to the development of hybrid circuits capable of regulating many different interactions via covalent chemical modifications.

However, the use of RNA-cleaving DNAzymes poses a significant barrier to the development of serial logic gates, which is necessary for the implementation of more complex logic functions. To connect DNAzyme-based logic gates, there must be an intrinsic mechanism for passing a signal from one gate to another – a process referred to as cascading. Success in this area has been limited to split DNAzymes; the first instance sequestered downstream

activator sequence in a multi-strand complex,[35,118] while another developed two-layer cascades where the downstream DNAzyme generated a colorimetric readout.[96] However, the first case uses of a multi-strand complex as the mediator, which increases the number of strands and the complexity of circuit preparation, while the second case uses a downstream DNAzyme that prevents further signal propagation within the molecular circuit, limiting the potential size and complexity of the circuit. Additionally, the structure of the core and the interactions with the divalent metal ion cofactors make modeling problematic. Although the use of DNAzymes presents many exciting opportunities for the implementation of DNA computation, these technological hurdles must be addressed for this to remain a viable technology.

1.5 DNA strand displacement

An alternative method for DNA computing is called DNA strand displacement (or toehold-mediated strand displacement, TMSD)[120]. This method relies solely on the thermodynamic favorability of base pair hybridization between multiple DNA strands[121,122]. The mechanism is outlined in **Figure 1.2**. Here, a preformed DNA duplex, consisting of strands B and C are present in solution. Although most of the two strands are complementary, there is a small, single-strand region left unhybridized on strand C, referred to as a toehold. The addition of a third strand (strand A), fully complementary to strand C, binds to the toehold in a reversible process. Because strand A and strand B are both complementary to strand C, they are able to compete for the same binding domain. However, as strand A is also complementary to strand C in the toehold region, it is able to form a longer,

more thermodynamically favorable complex. After strand A binds to the strand C toehold, it initiates a process known as branch migration. After strand A fully competes off strand B, strand B cannot rebind to the complex as all available regions are now hybridized, and is therefore released into solution. In this way, the addition of strand A can be considered the input, while the release of strand B can be considered the output. This reaction is thermodynamically favorable, typically through an enthalpy gain through the formation of additional base pairs[123] although entropic gain through the release of strands into solution can be achieved as well[121].

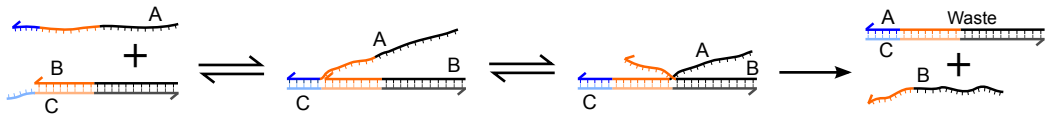

Figure 1.2 – The mechanism of toehold-mediated strand displacement.

Due to the predictability between hybridization reactions and the specificity with which these reactions proceed, TMSD has become the dominant method for implementing programmable DNA nanotechnology, including DNA computing. This mechanism was initially used in the construction of DNA tweezers by Yurke *et al.*[16], which showed how a nanomechanical process could be cycled multiple times by the addition of fuel oligonucleotides. TMSD has also been used extensively in DNA computation, for the construction of Boolean logic gates[38], as well as more complex circuit behavior such as thresholding and amplification[30,31,39,121,124-127]. Signal amplification has also been demonstrated

using a related DNA hybridization technology, catalytic hairpin assembly[29-31]. These principles have led to circuits of impressive performance, such as the calculation of square roots[39,124], most notably through the use of the standardized "seesaw" gate design developed by Qian and Winfree[40].

Strand displacement has several advantages over DNAzymes. The use of thermodynamic principles ensures that reactions can be highly specific and sensitive without relying on concentration effects to drive the reaction. As the reactions rely solely on hybridization, strand interactions are relatively easy to predict, and are amenable to rapid scaling into larger circuits because a displaced strand can serve directly as the input of another logic gate. Prediction of these interactions has been aided by the development of structure prediction and strand displacement modeling software, such as NUPACK[4,5,9] and Visual DSD[7,128,129]. Using these programs, unwanted interactions can be identified and addressed. Despite these significant advances, the use of strand displacement alone in biosensors poses challenges. Although signal amplification has been demonstrated using strand displacement, it requires an explicit design with additional strands and reactions[38,120,121,126 30,31]. This is undesirable for both *in vitro* and *in vivo* biosensors, as it can increase biosensor cost, circuit complexity, and provide more failure modes for the reaction. The most successful biosensors, such as molecular beacons[130], are single step reactions, illustrating the importance of minimizing complexity to achieve optimal performance. In addition, as each of the gates relies only on hybridization this can cause problems with larger scale circuits, where each additional strand and complex increases the

probability of cross-reactive sequences. Finally, performance of these circuits has been significantly improved through the use of PAGE purification of DNA complexes[39], but this is an expensive and time-consuming process. This method would likely be cost-prohibitive in the development of biosensors and not easily amenable to large-scale preparation of biodetection devices. It would be highly desirable to simplify the circuit composition while retaining the signal amplification properties, all while using low-cost unpurified DNA strands.

1.6 Molecular logic for in vitro *biosensing*

Although a significant effort has been devoted to the development of DNA logic gates and many have extolled their potential for *in vitro* biosensing applications, the demonstrations of this key concept have been surprisingly limited. Some forms of DNA computation have been shown to incorporate biological proteins for the development of *in vitro* and *in vivo* synthetic transcriptional circuits[56,131-134]. While these circuits display a wide range of interesting computational behaviors, the use of proteins is less advantageous in the development of biosensors. As previously mentioned, the complexity of protein folding has prevented the development of *de novo* protein-based biosensors, limiting the use of proteins to the availability of characterized interactions. Although proteins have a wide diversity of structure and can catalyze many chemical reactions, they are also much more expensive than nucleic acid probes, requiring a considerable time and cost investment for the expression, purification, and storage. They typically operate at physiological temperatures or require the use of temperature cycles and are typically are stored at cold temperatures, and use at room temperature

often leads to protein degradation. This requires additional equipment infrastructure, which may not be available outside of a laboratory setting. DNA, on the other hand, is highly stable, can easily be lyophilized for long-term storage, and assays using DNA, even DNA enzymes, can be operated at isothermal conditions.

There have been several groups that have utilized DNAzymes for *in vitro* nucleic acid detection. Gerasimova *et al* developed a two layer DNAzyme cascade for the visualization of bacterial RNA. Here, the presence of an RNA target brought together both detection arms of a split 10-23 DNAzyme[117]. This now functional DNAzyme can cleave an inhibited peroxidase-like (G-quadruplex) DNAzyme, activating the sequence, which can now produce a colored substrate. Other groups have also used the DNAzymes as visualization probes of DNA samples[54,135-137], however, these examples do not employ molecular logic. Although there is significant promise in the use of DNA or DNAzyme biosensors, several basic parameters must be optimized before *in vitro* and *in* vivo detection of real world disease or pathogen signatures. These parameters include, but are not limited to, detection sensitivity, availability of target sequences, selection of additional aptamer sequences for small molecules, multi-input integration, gate and network signal reproducibility, and the compartmentalization of signal transduction. These wide ranging needs necessitate the continued development of a variety of DNA computational implementations.

1.7 Multi-input integration and multiplex analysis

Logic-based integration of multiple inputs is critical for DNA computation to truly distinguish itself from the myriad of other biosensors for the construction of intelligent *in vitro* biosensors. Utilization of multiple input signatures can provide a more comprehensive understanding of the biochemical environment, beyond the presence or absence of a given target. Many complex target applications could take advantage of this property, including *in vitro* applications such as multi-strain typing and differentiation as well as *in vivo* applications such as cancer or autism. Many conditions are defined by the many targets, due to the up- or down-regulation of many different RNA transcripts or the presence of multiple single nucleotide polymorphisms (SNPs)[138,139]. While there are many alternative biodetection methods that offer competitive or even superior sensitivity for individual targets, including PCR and ELISAs, the use of DNA in molecular computing devices places it in a unique position for the development of sophisticated decision networks integrating multiple input signatures.

Although the aforementioned *in vivo* applications require significant technological advances for detection and computation to be truly feasible, many such *in vitro* sensing applications are possible with current technology. Particularly of interest are multi-strain pathogens, such as the single-stranded RNA virus Dengue or Shiga-toxin bearing *E. coli* (STEC) bacteria. The presence of four serotypes of Dengue has led to difficulties with the development of vaccines covering all of the serotypes simultaneously[140], and the predominance of infection in tropical and sub-tropical third world nations necessitates the

development of comprehensive, low-cost diagnostics assays[141-143]. Serotyping is critical as the onset of severe disease states such as dengue hemorrhagic fever and dengue shock syndrome is generally induced through the infection of multiple serotypes[144]. Although O157 is the most prominent and well-characterized STEC serotype[145,146], over 100 different serotypes have been implicated in enterohemorrhagic illness[147-149]. In addition, positive STEC identification is also dependent on various targets, which differ across serotypes[150-154]. Concurrent sensing of all serotypes, including multiple diagnostic targets, would be a significant benefit for agriculture and food processing companies, where bacterial culturing can take upwards of two weeks.

Although DNA computation appears well suited to the integration of multiple inputs, there are significant technological hurdles that must be addressed for this to be a viable technology. As previously mentioned, the use of RNA-cleaving DNAzymes presents a significant challenge to the development of large, multi-input decision networks due to the limitations in gate interactions. The use of parallel gates limits the logical computation to the number of inputs a single gate can accept, necessitating the construction of molecular signal intermediaries that can pass a signal between gates to scale up the complexity of the circuit. Work in this area remains limited[35,96]. Although complex strand displacement circuits have been demonstrated, indicating the potential of solution-based multi-input detection approach, their use against physiologically relevant targets has not been demonstrated.

1.8 Scaling up circuit complexity through compartmentalization

One of the biggest challenges facing DNA computing is the ability to segregate gates and sequences to prevent spurious interactions. This has hindered the scaling of size and complexity of solution-based cascades, stunting their development for multi-input integration and intelligent biosensing applications. In DNAzyme cascading, compartmentalization has been achieved through the construction of multi-strand intermediary complexes[35], and in TMSD through the blocking of toehold sequences, referred to as toehold inactivation[120]. However, both of these technologies are typically run in a well-mixed solution, which means that molecules are free to interact at any time. This puts a significant pressure on the circuit design to sequester hybridization interactions until a specified temporal activation is desired. Additionally, each component must be designed separately, typically through the use of different, non-interacting sequences to ensure the activation of a specific pathway via directed hybridization. This also makes the scaling of circuits difficult; silicon-based circuits can contain millions of identical transistors due to their physical segregation on a microchip. Biology employs a similar solution, achieving physical separation of components into organelles through the use of biological membranes. It would be highly desirable to apply the concept of physical compartmentalization to DNA computing as well, which may enable for the reuse of circuit components.

This concept has been explored by several groups. In the initial demonstration of DNA logic component compartmentalization, DNA oligomers may be physically separated on the surface of particles using a non-covalent

biotin-streptavidin conjugation. In the first demonstration of this approach, Frezza *et al.* constructed two layer TMSD cascades attached to sepharose beads, transmitting a DNA signal from one bead population to another via the addition of an input strand[155]. While this demonstrated the feasibility of this approach, the authors noted a decrease in expected input-to-output stoichiometry as a result of an increased reaction time. Additionally, the authors noted that supernatant filtration was required at each step of the cascade to prevent non-specific interaction. In related work, Stojanovic and colleagues demonstrated a two layer E6 DNAzyme cascade attached to microspheres, in which the addition of an input DNA oligomer deprotects a chimeric substrate (RNA/DNA hybrid oligomer), allowing it to be cleaved from the microsphere surface[156]. This cleavage product then can bind to the stem loop of a second E6 DNAzyme on another bead population, opening the stem loop and activating the DNAzyme. This DNAzyme can now cleave its corresponding substrate, and the signal exchange between beads can be measured via fluorescence. Although this was an innovative method to implement DNAzyme cascades, this required the use of high concentrations of input, DNAzyme, and substrate. As the beads are immobilized on the surface, this approach ensures that DNAzymes can only cleave substrates within their physical proximity, thus a high concentration of DNAzymes relative to substrates was necessary to ensure this condition. This resulted in the loss of signal amplification ability, negating one of the strongest advantages of using DNAzymes.

Nevertheless, this implementation enabled cascades to be performed using these DNAzymes while simultaneously simplifying gate structure for such interactions. These results argue strongly for the potential of physical compartmentalization for DNAzyme-based cascades and circuits. To overcome the biggest limitation of this system, the signal amplification issue, the system would require dynamic interaction of separate components on the surface. This would enable DNAzyme strands to seek out and bind with multiple substrates on the surface, while still being constrained to the surface. One possible solution to this is the use of supported lipid bilayers.

Originally developed by Bayerl and Bloom in 1990[157], this technology uses silica microspheres as a template to adsorb a lipid bilayer to the surface. This produces a single bilayer mimicking the structure typically found in biology, as opposed to alternative technologies that use multi-lamellar vesicles. Importantly, the lipids also retain their natural fluidity[158]. These bilayers have been used for many applications, including the study of transmembrane and membrane-associated proteins and dynamics, with potential applications in biosensing as well[159,160]. This system appears amenable to bead-based DNAzyme circuits, in which the lipids provide the lateral fluidity necessary to encourage productive interactions between multiple substrate molecules via lateral diffusion of DNA circuits components attached to functionalized lipid head groups.

There are other notable efforts to explore the interface between DNA and lipid membranes, including both DNA origami[161] as well as DNA logic on cell surfaces. Several groups have constructed transmembrane pores through DNA

origami[162,163], most notably the creation of a large synthetic lipid membrane channel detailed in a 2012 *Science* report from the Simmel group[164]. For computation and diagnostic purposes, Rudchenko and colleagues constructed a strand displacement cascade conjugated to antibodies for cell typing[82]. As this field continues to expand and mature, a more detailed understanding of the DNA-lipid bilayer interface will be critical for the development of compartmentalized DNA computers. The continued development of bead-based computation, combined with multiplexed flow cytometry analysis, may also provide an additional method for scaling up input detection. Multi-input analysis of microspheres has been well characterized for flow cytometry, where multiplex beads sets are commercially available and are able to distinguish between many different bead populations[165]. The use of physical separation of computational and/or detection elements may improve the scaling of target sequences.

1.9 Present Studies

In the work described here, we have created a hybrid logic gate mechanism, using toehold-mediated strand displacement to regulate the activity of an RNA-cleaving DNAzyme. This approach combines the advantages of strand displacement, particularly the highly specific, thermodynamically driven, programmable reactions, with the innate catalysis of the 8-17 DNAzyme. We then utilized this regulation mechanism to create robust multi-layer DNAzyme cascades through the rational design of a structured chimeric substrate intermediary, in which we demonstrate the longest DNAzyme cascade to date. We also developed a new DNAzyme-based gate structure specifically designed

as an efficient biosensor, capable of integrating directly with our DNAzyme cascades. Finally, we explore basic principles for the future development of compartmentalized DNA circuit components using both bead-based and supported lipid bilayer-based platforms.

1.10 References

(1) Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E. *Nature* **2004**, *429*, 423.

(2) Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737.

(3) Bois, J. S.; Venkataraman, S.; Choi, H. M.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A. *Nucleic Acids Res* **2005**, *33*, 4090.

(4) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev* **2007**, *49*, 65.

(5) Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2003**, *24*, 1664.

(6) SantaLucia, J., Jr. *Proc Natl Acad Sci U S A* **1998**, *95*, 1460.

(7) Lakin, M. R.; Youssef, S.; Polo, F.; Emmott, S.; Phillips, A. *Bioinformatics* **2011**, *27*, 3211.

(8) Zuker, M. *Nucleic Acids Research* **2003**, *31*, 3406.

(9) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 170.

(10) Seeman, N. C. *Nature* **2003**, *421*, 427.

(11) Seeman, N. C. *Journal of theoretical biology* **1982**, *99*, 237.

(12) Adleman, L. M. *Science* **1994**, *266*, 1021.

(13) Yan, H.; LaBean, T. H.; Feng, L.; Reif, J. H. *Proc Natl Acad Sci U S A* **2003**, *100*, 8103.

(14) Woo, S.; Rothemund, P. W. *Nat Chem* **2011**, *3*, 620.

(15) Zhang, F.; Nangreave, J.; Liu, Y.; Yan, H. *Nano Lett* **2012**, *12*, 3290.

(16) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605.

(17) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. *Angew Chem Int Ed Engl* **2004**, *43*, 4906.

(18) Bath, J.; Green, S. J.; Allen, K. E.; Turberfield, A. J. *Small* **2009**, *5*, 1513.

(19) Franco, E.; Friedrichs, E.; Kim, J.; Jungmann, R.; Murray, R.; Winfree, E.; Simmel, F. C. *Proc Natl Acad Sci U S A* **2011**, *108*, E784.

(20) Pei, R.; Taylor, S. K.; Stefanovic, D.; Rudchenko, S.; Mitchell, T. E.; Stojanovic, M. N. *J Am Chem Soc* **2006**, *128*, 12693.

(21) Katz, E.; Privman, V. *Chem Soc Rev* **2010**, *39*, 1835.

(22) Katz, E.; Wang, J.; Privman, M.; Halamek, J. *Anal Chem* **2012**, *84*, 5463.

(23) Wang, J.; Katz, E. *Anal Bioanal Chem* **2010**, *398*, 1591.

(24) Zhou, J.; Arugula, M. A.; Halamek, J.; Pita, M.; Katz, E. *J Phys Chem B* **2009**, *113*, 16065.

(25) Lin, C.; Liu, Y.; Rinker, S.; Yan, H. *ChemPhysChem* **2006**, *7*, 1641.

(26) Rothemund, P. W.; Papadakis, N.; Winfree, E. *PLoS biology* **2004**, *2*, e424.

(27) Winfree, E. *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2003 NAE Symposium on Frontiers of Engineering* **2004**, 105.

(28) Winfree, E.; Eng, T.; Rozenberg, G. In *DNA computing*; Springer Berlin Heidelberg: 2001, p 63.
(29) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.
(30) Li, B.; Ellington, A. D.; Chen, X. *Nucleic Acids Res* **2011**, *39*, e110.
(31) Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2013**, *110*, 5386.
(32) Penchovsky, R. *ACS Synth Biol* **2012**, *1*, 471.
(33) Penchovsky, R.; Breaker, R. R. *Nat Biotechnol* **2005**, *23*, 1424.
(34) Chen, X.; Wang, Y. F.; Liu, Q.; Zhang, Z. Z.; Fan, C. H.; He, L. *Angew Chem Int Ed Engl* **2006**, *45*, 1759.
(35) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.
(36) Stojanovic, M. N.; Mitchell, T. E.; Stefanovic, D. *J Am Chem Soc* **2002**, *124*, 3555.
(37) Stojanovic, M. N.; Semova, S.; Kolpashchikov, D.; Macdonald, J.; Morgan, C.; Stefanovic, D. *J Am Chem Soc* **2005**, *127*, 6914.
(38) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585.
(39) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.
(40) Qian, L.; Winfree, E. *J R Soc Interface* **2011**, *8*, 1281.
(41) Winfree, E., California Institute of Technology, 1998.
(42) Winfree, E. **1998**.
(43) Zhang, D. Y.; Hariadi, R. F.; Choi, H. M.; Winfree, E. *Nat Commun* **2013**, *4*, 1965.
(44) Li, Y.; Lu, Y. *Functional Nucleic Acids for Analytical Applications*; Springer, 2009.
(45) Gerasimova, Y. V.; Kolpashchikov, D. M. *Biosens Bioelectron* **2013**, *41*, 386.
(46) Kolpashchikov, D. M.; Gerasimova, Y. V.; Khan, M. S. *ChemBioChem* **2011**, *12*, 2564.
(47) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656.
(48) Achenbach, J. C.; Nutiu, R.; Li, Y. *Analytica Chimica Acta* **2005**, *534*, 41.
(49) Famulok, M. *Current opinion in molecular therapeutics* **2005**, *7*, 137.
(50) Tang, J.; Breaker, R. R. *Chemistry & biology* **1997**, *4*, 453.
(51) Liu, J.; Lu, Y. *J Fluoresc* **2004**, *14*, 343.
(52) Liu, J.; Lu, Y. *J Am Chem Soc* **2004**, *126*, 12298.
(53) Aragay, G.; Pons, J.; Merkoçi, A. *Chemical reviews* **2011**, *111*, 3433.
(54) Ali, M. M.; Aguirre, S. D.; Lazim, H.; Li, Y. *Angewandte Chemie International Edition* **2011**, *50*, 3751.
(55) Darius, A. K. L.; Ling, N. J.; Mahesh, U. *Molecular BioSystems* **2010**, *6*, 792.

(56) Xie, Z.; Wroblewska, L.; Prochazka, L.; Weiss, R.; Benenson, Y. *Science* **2011**, *333*, 1307.
(57) Zhu, X.; Cao, Y.; Liang, Z.; Li, G. *Protein & cell* **2010**, *1*, 842.
(58) Silverman, S. K. *Acc Chem Res* **2009**, *42*, 1521.
(59) Liu, J.; Cao, Z.; Lu, Y. *Chemical reviews* **2009**, *109*, 1948.
(60) Zhu, J.; Li, T.; Zhang, L.; Dong, S.; Wang, E. *Biomaterials* **2011**, *32*, 7318.
(61) Wang, G.; Wang, Y.; Chen, L.; Choo, J. *Biosensors and Bioelectronics* **2010**, *25*, 1859.
(62) Zagorovsky, K.; Chan, W. C. *Angewandte Chemie International Edition* **2013**, *52*, 3168.
(63) Shen, L.; Chen, Z.; Li, Y.; He, S.; Xie, S.; Xu, X.; Liang, Z.; Meng, X.; Li, Q.; Zhu, Z. *Analytical chemistry* **2008**, *80*, 6323.
(64) Baker, B. R.; Lai, R. Y.; Wood, M. S.; Doctor, E. H.; Heeger, A. J.; Plaxco, K. W. *Journal of the American Chemical Society* **2006**, *128*, 3138.
(65) Xiao, Y.; Piorek, B. D.; Plaxco, K. W.; Heeger, A. J. *Journal of the American Chemical Society* **2005**, *127*, 17990.
(66) Xiao, Y.; Rowe, A. A.; Plaxco, K. W. *Journal of the American Chemical Society* **2007**, *129*, 262.
(67) Abdelgany, A.; Wood, M.; Beeson, D. *J Gene Med* **2007**, *9*, 727.
(68) Baum, D. A.; Silverman, S. K. *Cell Mol Life Sci* **2008**, *65*, 2156.
(69) Dass, C. R. *Trends Pharmacol Sci* **2004**, *25*, 395.
(70) Fokina, A. A.; Meschaninova, M. I.; Durfort, T.; Venyaminova, A. G.; Francois, J. C. *Biochemistry* **2012**, *51*, 2181.
(71) Mitchell, A.; Dass, C. R.; Sun, L. Q.; Khachigian, L. M. *Nucleic Acids Res* **2004**, *32*, 3065.
(72) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.
(73) Chan, C.; Khachigian, L. *Internal medicine journal* **2009**, *39*, 249.
(74) Dass, C. R.; Choong, P. F.; Khachigian, L. M. *Mol Cancer Ther* **2008**, *7*, 243.
(75) Peracchi, A. *Rev Med Virol* **2004**, *14*, 47.
(76) Zhao, W.; Ali, M. M.; Aguirre, S. D.; Brook, M. A.; Li, Y. *Anal Chem* **2008**, *80*, 8431.
(77) Huttanus, H. M.; Graugnard, E.; Yurke, B.; Knowlton, W. B.; Kuang, W.; Hughes, W. L.; Lee, J. *Biosens Bioelectron* **2013**, *50*, 382.
(78) Liu, J.; Lu, Y. *Anal Chem* **2004**, *76*, 1627.
(79) Reynolds, R. A., 3rd; Mirkin, C. A.; Letsinger, R. L. *Pure and Applied Chemistry* **2000**, *72*, 229.
(80) Wu, P.; Hwang, K.; Lan, T.; Lu, Y. *J Am Chem Soc* **2013**, *135*, 5254.
(81) Zhao, W.; Lam, J. C.; Chiuman, W.; Brook, M. A.; Li, Y. *Small* **2008**, *4*, 810.
(82) Rudchenko, M.; Taylor, S.; Pallavi, P.; Dechkovskaia, A.; Khan, S.; Butler, V. P., Jr.; Rudchenko, S.; Stojanovic, M. N. *Nat Nanotechnol* **2013**, *8*, 580.

(83) Bagalkot, V.; Farokhzad, O. C.; Langer, R.; Jon, S. *Angewandte Chemie International Edition* **2006**, *45*, 8149.
(84) Taghdisi, S. M.; Lavaee, P.; Ramezani, M.; Abnous, K. *European Journal of Pharmaceutics and Biopharmaceutics* **2011**, *77*, 200.
(85) Zhang, L.; Radovic-Moreno, A. F.; Alexis, F.; Gu, F. X.; Basto, P. A.; Bagalkot, V.; Jon, S.; Langer, R. S.; Farokhzad, O. C. *ChemMedChem* **2007**, *2*, 1268.
(86) Famulok, M.; Hartig, J. S.; Mayer, G. *Chemical Reviews* **2007**, *107*, 3715.
(87) Li, Y.; Breaker, R. R. *Curr Opin Struct Biol* **1999**, *9*, 315.
(88) Willner, I.; Shlyahovsky, B.; Zayats, M.; Willner, B. *Chem Soc Rev* **2008**, *37*, 1153.
(89) Breaker, R. R.; Joyce, G. F. *Chem Biol* **1995**, *2*, 655.
(90) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.
(91) Flynn-Charlebois, A.; Prior, T. K.; Hoadley, K. A.; Silverman, S. K. *J Am Chem Soc* **2003**, *125*, 5346.
(92) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. *J Am Chem Soc* **2003**, *125*, 2444.
(93) Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.
(94) Li, Y.; Breaker, R. R. *Proc Natl Acad Sci U S A* **1999**, *96*, 2746.
(95) Eckhoff, G.; Codrea, V.; Ellington, A. D.; Chen, X. *J Syst Chem* **2010**, *1*, 13.
(96) Gerasimova, Y. V.; Cornett, E. M.; Edwards, E.; Su, X.; Rohde, K. H.; Kolpashchikov, D. M. *ChemBioChem* **2013**, *14*, 2087.
(97) Chandra, M.; Sachdeva, A.; Silverman, S. K. *Nat Chem Biol* **2009**, *5*, 718.
(98) Gu, H.; Furukawa, K.; Weinberg, Z.; Berenson, D. F.; Breaker, R. R. *J Am Chem Soc* **2013**, *135*, 9121.
(99) Schlosser, K.; Li, Y. *Chembiochem* **2010**, *11*, 866.
(100) Santoro, S. W.; Joyce, G. F. *Biochemistry* **1998**, *37*, 13330.
(101) Peracchi, A.; Bonaccio, M.; Clerici, M. *J Mol Biol* **2005**, *352*, 783.
(102) Zadeh, J. N.; Wolfe, B. R.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 439.
(103) Faulhammer, D.; Famulok, M. *J Mol Biol* **1997**, *269*, 188.
(104) Okumoto, Y.; Sugimoto, N. *J Inorg Biochem* **2000**, *82*, 189.
(105) Kim, H.-K.; Rasnik, I.; Liu, J.; Ha, T.; Lu, Y. *Nat Chem Biol* **2007**, *3*, 763.
(106) Kim, H. K.; Li, J.; Nagraj, N.; Lu, Y. *Chem Eur J* **2008**, *14*, 8696.
(107) Lan, T.; Lu, Y. In *Interplay between Metal Ions and Nucleic Acids*; Springer: 2012, p 217.
(108) Xiang, Y.; Lu, Y. *Inorganic chemistry* **2013**.
(109) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.

(110) Martin, C. H.; Nielsen, D. R.; Solomon, K. V.; Prather, K. L. *Chem Biol* **2009**, *16*, 277.
(111) Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. *RNA* **2003**, *9*, 907.
(112) Breaker, R. R.; Emilsson, G. M.; Lazarev, D.; Nakamura, S.; Puskarz, I. J.; Roth, A.; Sudarsan, N. *RNA* **2003**, *9*, 949.
(113) Stojanovic, M. N.; de Prada, P.; Landry, D. W. *ChemBioChem* **2001**, *2*, 411.
(114) Stojanovic, M. N.; Stefanovic, D. *Nat Biotechnol* **2003**, *21*, 1069.
(115) Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N. *Nano Lett* **2006**, *6*, 2598.
(116) Pei, R.; Matamoros, E.; Liu, M.; Stefanovic, D.; Stojanovic, M. N. *Nat Nanotechnol* **2010**, *5*, 773.
(117) Kolpashchikov, D. M. *Chem Rev* **2010**, *110*, 4709.
(118) Elbaz, J.; Moshe, M.; Shlyahovsky, B.; Willner, I. *Chemistry* **2009**, *15*, 3411.
(119) Kahan-Hanum, M.; Douek, Y.; Adar, R.; Shapiro, E. *Sci Rep* **2013**, *3*.
(120) Zhang, D. Y.; Seelig, G. *Nat Chem* **2011**, *3*, 103.
(121) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121.
(122) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2009**, *131*, 17303.
(123) Srinivas, N.; Ouldridge, T. E.; Šulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P.; Winfree, E. *Nucleic acids research* **2013**, *41*, 10641.
(124) Qian, L.; Winfree, E.; Bruck, J. *Nature* **2011**, *475*, 368.
(125) Zhang, D. Y.; Winfree, E. *Nucleic Acids Res* **2010**, *38*, 4182.
(126) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2008**, *130*, 13921.
(127) Li, B.; Jiang, Y.; Chen, X.; Ellington, A. D. *J Am Chem Soc* **2012**, *134*, 13918.
(128) Lakin, M. R.; Parker, D.; Cardelli, L.; Kwiatkowska, M.; Phillips, A. *J R Soc Interface* **2012**, *9*, 1470.
(129) Lakin, M. R.; Youssef, S.; Cardelli, L.; Phillips, A. *J R Soc Interface* **2012**, *9*, 470.
(130) Tyagi, S.; Kramer, F. R. *Nature biotechnology* **1996**, *14*, 303.
(131) Bonnet, J.; Yin, P.; Ortiz, M. E.; Subsoontorn, P.; Endy, D. *Science* **2013**.
(132) Elowitz, M. B.; Leibler, S. *Nature* **2000**, *403*, 335.
(133) Kim, J.; White, K. S.; Winfree, E. *Mol Syst Biol* **2006**, *2*, 68.
(134) Kim, J.; Winfree, E. *Mol Syst Biol* **2011**, *7*, 465.
(135) Neo, J. L.; Aw, K. D.; Uttamchandani, M. *Analyst* **2011**, *136*, 1569.
(136) Neo, J. L.; Kamaladasan, K.; Uttamchandani, M. *Curr Pharm Des* **2012**, *18*, 2048.
(137) Li, Y. *Future microbiology* **2011**, *6*, 973.
(138) Lazarus, R.; Vercelli, D.; Palmer, L. J.; Klimecki, W. J.; Silverman, E. K.; Richter, B.; Riva, A.; Ramoni, M.; Martinez, F. D.; Weiss, S. T. *Immunological reviews* **2002**, *190*, 9.

(139) Martin, E. R.; Lai, E. H.; Gilbert, J. R.; Rogala, A. R.; Afshari, A.; Riley, J.; Finch, K.; Stevens, J.; Livak, K.; Slotterbeck, B. D. *The American Journal of Human Genetics* **2000**, *67*, 383.

(140) Cardosa, M. *British medical bulletin* **1998**, *54*, 395.

(141) Bhatt, S.; Gething, P. W.; Brady, O. J.; Messina, J. P.; Farlow, A. W.; Moyes, C. L.; Drake, J. M.; Brownstein, J. S.; Hoen, A. G.; Sankoh, O.; Myers, M. F.; George, D. B.; Jaenisch, T.; Wint, G. R.; Simmons, C. P.; Scott, T. W.; Farrar, J. J.; Hay, S. I. *Nature* **2013**, *496*, 504.

(142) Bhattacharya, M. K.; Maitra, S.; Ganguly, A.; Bhattacharya, A.; Sinha, A. *Int J Biomed Sci* **2013**, *9*, 61.

(143) Guzman, M. G.; Halstead, S. B.; Artsob, H.; Buchy, P.; Farrar, J.; Gubler, D. J.; Hunsperger, E.; Kroeger, A.; Margolis, H. S.; Martinez, E.; Nathan, M. B.; Pelegrino, J. L.; Simmons, C.; Yoksan, S.; Peeling, R. W. *Nat Rev Microbiol* **2010**, *8*, S7.

(144) Kliks, S. C.; Nisalak, A.; Brandt, W. E.; Wahl, L.; Burke, D. S. *Antibody-dependent enhancement of dengue virus growth in human monocytes as a risk factor for dengue hemorrhagic fever*, DTIC Document, 1989.

(145) Burland, V.; Shao, Y.; Perna, N. T.; Plunkett, G.; Sofia, H. J.; Blattner, F. R. *Nucleic Acids Res* **1998**, *26*, 4196.

(146) Ge, B.; Larkin, C.; Ahn, S.; Jolley, M.; Nasir, M.; Meng, J.; Hall, R. H. *Mol Cell Probes* **2002**, *16*, 85.

(147) Brooks, J. T.; Sowers, E. G.; Wells, J. G.; Greene, K. D.; Griffin, P. M.; Hoekstra, R. M.; Strockbine, N. A. *J Infect Dis* **2005**, *192*, 1422.

(148) Johnson, K. E.; Thorpe, C. M.; Sears, C. L. *Clin Infect Dis* **2006**, *43*, 1587.

(149) Kappeli, U.; Hachler, H.; Giezendanner, N.; Beutin, L.; Stephan, R. *Emerg Infect Dis* **2011**, *17*, 180.

(150) Auvray, F.; Lecureuil, C.; Tache, J.; Leclerc, V.; Deperrois, V.; Lombard, B. *Lett Appl Microbiol* **2007**, *45*, 646.

(151) Newton, H. J.; Sloan, J.; Bulach, D. M.; Seemann, T.; Allison, C. C.; Tauschek, M.; Robins-Browne, R. M.; Paton, J. C.; Whittam, T. S.; Paton, A. W.; Hartland, E. L. *Emerg Infect Dis* **2009**, *15*, 372.

(152) Schmidt, H.; Beutin, L.; Karch, H. *Infect Immun* **1995**, *63*, 1055.

(153) Schmidt, H.; Kernbach, C.; Karch, H. *Microbiology* **1996**, *142 (Pt 4)*, 907.

(154) Serra-Moreno, R.; Jofre, J.; Muniesa, M. *J Bacteriol* **2007**, *189*, 6645.

(155) Frezza, B. M.; Cockroft, S. L.; Ghadiri, M. R. *J Am Chem Soc* **2007**, *129*, 14875.

(156) Yashin, R.; Rudchenko, S.; Stojanovic, M. N. *J Am Chem Soc* **2007**, *129*, 15581.

(157) Bayerl, T. M.; Bloom, M. *Biophysical journal* **1990**, *58*, 357.

(158) Davis, R. W.; Flores, A.; Barrick, T. A.; Cox, J. M.; Brozik, S. M.; Lopez, G. P.; Brozik, J. A. *Langmuir* **2007**, *23*, 3864.

(159) Buranda, T.; Huang, J.; Ramarao, G.; Ista, L. K.; Larson, R. S.; Ward, T. L.; Sklar, L. A.; Lopez, G. P. *Langmuir* **2003**, *19*, 1654.

(160) Chemburu, S.; Fenton, K.; Lopez, G. P.; Zeineldin, R. *Molecules* **2010**, *15*, 1932.
(161) Perrault, S. D.; Shih, W. M. *ACS nano* **2014**.
(162) Burns, J. R.; Gopfrich, K.; Wood, J. W.; Thacker, V. V.; Stulz, E.; Keyser, U. F.; Howorka, S. *Angew Chem Int Ed Engl* **2013**.
(163) Burns, J. R.; Stulz, E.; Howorka, S. *Nano Lett* **2013**, *13*, 2351.
(164) Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H.; Simmel, F. C. *Science* **2012**, *338*, 932.
(165) Khan, S. S.; Smith, M. S.; Reda, D.; Suffredini, A. F.; McCoy, J. P. *Cytometry Part B: Clinical Cytometry* **2004**, *61*, 35.

Chapter 2.
Goals and Overview of this Work

The goal of this work is to develop and characterize DNA computational principles using the class of functional nucleic acids known as DNAzymes, engineered specifically for the purpose of biodetection and biosensor development. These implementations would have advantages over existing DNA circuits as they benefit from a ground-up, rationally designed approach that contains advantages and considerations specific to use in real-world biodetection devices. The focus on practical applications ensures that designs have been tested using minimal purification standards, ensuring reduced cost and time of preparation. The principles explored here also lay the groundwork for expanded development of DNA computation for biosensing, and opens the potential for new DNA computational frameworks.

The studies in **Chapter 3**, published as a communication in *ChemBioChem*, describe the mechanism of *DNAzyme displacement*, which combines the strengths of DNA strand displacement with inclusion of DNAzymes for the construction of DNA logic gates. Strand displacement ensures highly specific and sensitive strand hybridization, relieving the previous issues of high concentration and competitive hybridization that hindered earlier DNAzyme regulation mechanisms. The combination of DNAzymes and strand displacement regulation ensures that the addition of a single input is able to generate multiple output signals through multiple turnover reactions while minimizing the circuit architecture necessary to execute this response. This work also introduced the concept of arbitrary sequence detection through the rational introduction of

mismatches between the DNAzyme and inhibitor. This has general applications for strand displacement circuits, whereby the release of a sequence is triggered by the addition of two unrelated sequences. The functionality of these gates was demonstrated through the construction of a circuit to evaluate a three-input logic function. The format of this chapter is as follows: abstract, body of manuscript (individual sections are not delineated in communications), materials and methods, acknowledgements, and references. Additional controls, along with the Supporting Information of the manuscript, are included in Appendix 1. Additional chapters are formatted in the same manner to ensure consistency throughout this dissertation.

The work in **Chapter 4**, accepted as a communication for *Angewandte Chemie International Edition*, introduces the design and function of a structured chimeric substrate (SCS), a single-stranded molecule that can transmit a DNA signal between two DNAzyme displacement gates. This enabled the straightforward scaling of a DNAzyme cascade, for the construction of a five-layer cascade, the longest such cascade to date. This has important implications for DNA computing and protein signaling cascades, as communication between DNAzymes offers a model framework for the design and construction of synthetic enzyme networks. This work also demonstrated that the SCS molecule could communicate between logic gates other than 8-17 DNAzyme displacement gates. This indicates the feasibility of developing hybrid DNA computational circuits, including DNAzymes containing different core sequences. This work also demonstrated the benefit of strand displacement regulation by showing signal

amplification for each successive layer of a four layer cascade. The SCS signal transmission was also highly resistant to a random DNA background, which holds promise for biosensor performance of these cascades in non-pristine conditions. Finally, a bioassay circuit was developed to demonstrate the potential application for pathogen typing, in which a patterned circuit was implemented against each of the four Dengue virus serotypes. Here, three synthetic target sequences taken directly from the Dengue genome were used to test for the presence of a given serotype, with a positive signal attained only in the presence of all three targets.

Chapter 5 details the rational design process for the design of the SCS structure. This effort confirmed two key hypotheses for the SCS and DNAzyme cascades. First, we hypothesized that by focusing on the thermodynamic balance of the structure between the pre- and post- cleaved states, we could execute a function specific to a given structure while not specific to the individual sequences of the structure. In short, the design of a single SCS structure could be altered to match any two gate sequences. The five layer cascades proved this hypothesis correct, as each layer used the same SCS structure, each with a different sequence corresponding to the connecting gates. Second, we hypothesized that through an iterative process of structural design, we can achieve the desired kinetic rates in the absence of appropriate modeling software. This is an important principle for dynamic DNA nanotechnology, as most software determines DNA hybridization and structure at thermodynamic equilibrium, we are often unable to reliably predict dynamic structural

interactions. However, with an understanding of the underlying processes that govern structural formation and hybridization interactions, the desired results can still be attained.

After the success of DNAzyme displacement and subsequent cascading, we redesigned the gate structure to decouple the detection and reporter domains, referred to as modular gates. Implementation and characterization of these gates are detailed in **Chapter 6**. Each gate can now detect an arbitrary input sequence and serve as an individual biosensor. By disconnecting the detection and reporter domains, we are free to independently vary either domain. By fixing the reporter domain while varying the detection domain, we are able to create a massively multiplexable OR gate, in which the presence of any target sequence will result in a positive signal. The reporter domain can also vary and differing substrates can be used to read out results in multiple fluorescent channels. This new modular gate structure is highly sensitive, achieving an LOD of 7.4 pM. Gates were suitable for detection of ssDNA, dsDNA, RNA, and small molecules and initial efforts to cascade modular gates showed promise, as did their performance in a random DNA background. Structural optimizations reduced spurious activation and additional improvements are detailed. The new gate design offers unique advantages for biosensor implementation.

As each DNAzyme is selected under specific circumstances, we hypothesized that the presence of a scaffold library consisting of millions of small molecule compounds may elucidate new DNAzyme activity. RNA-cleaving DNAzymes are orders of magnitude slower than their protein counterparts. This

is likely due to the underutilization of all available biochemical mechanisms for the hydrolysis of the RNA base. In **Chapter 7**, we discuss the results of a high-throughput small molecule screen of the 8-17 DNAzyme. Although we did not find any compounds that elicited an increase in the catalytic rate of the DNAzyme, we did find several inhibitors of unknown mechanism. This work may lead to a deeper understanding of the RNA world hypothesis and the discovery of new biochemical mechanisms as well as improved practical applications that may benefit from enhanced DNAzyme catalysis.

The work in **Chapter 8** describes initial efforts in DNAzyme and DNA-based circuit compartmentalization. We explore various attachment strategies to microspheres of both DNA and proteins, including direct bead attachment using biotin-streptavidin, 6xHis tags/Ni-NTA headgroup labeled lipids for non-covalent protein attachment and thiolated DNA/maleimide headgroup labeled lipids to form thioether bonds for covalent attachment to supported lipid bilayers. We demonstrate the ability to perform strand displacement reactions off microsphere surfaces and DNAzyme cleavage off supported lipid bilayers. Future work will expand these efforts to standardize circuit components and improve the potential of DNA computation through the use of physical separation of components.

Taken as a whole, this work represents a concerted effort to advance the field of DNA computation for the practical use in biodetection through the understanding and direction of both the underlying basic science and engineering principles of DNAzyme activity and regulation. These advances should be broadly applicable to DNAzymes of similar size and structure, with future

directions and other targets for biodetection optimization detailed in **Chapter 9**. Construction of complex circuit designs and behaviors are ongoing, as are efforts to use existing designs and cascades for the detection of real-world pathogenic targets under physiological conditions. Hence this dissertation provides a comprehensive report of the initial efforts and results for the design, characterization, and demonstration of DNAzyme-based logic circuits for biodetection applications.

Chapter 3.
Catalytic Molecular Logic Devices By DNAzyme Displacement

Carl W. Brown, III[1], Matthew R. Lakin[3], Darko Stefanovic*[1,3], and Steven W. Graves*[1,2]

[1] Center for Biomedical Engineering, University of New Mexico, Albuquerque, NM 87131, USA

[2] Department of Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, NM 87131, USA

[3] Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA

(Published in *ChemBioChem* 2014 May; **15** (7): 950-954)

3.1 Abstract

Chemical reactions catalysed by DNAzymes offer a route to programmable modification of biomolecules for therapeutic purposes. To this end, we have developed a new type of catalytic DNA-based logic gates in which DNAzyme catalysis is controlled via toehold-mediated strand displacement reactions. We refer to these as DNAzyme displacement gates. The use of toeholds to guide input binding provides a favourable pathway for input recognition, and the innate catalytic activity of DNAzymes allows amplification of nanomolar input concentrations. We demonstrate detection of arbitrary input sequences by rational introduction of mismatched bases into inhibitor strands. Furthermore, we illustrate the applicability of DNAzyme displacement to compute logic functions involving multiple logic gates. This work will enable sophisticated logical control of a range of biochemical modifications, with applications in pathogen detection and autonomous theranostics.

3.2 Body of Manuscript

Biomolecular computing devices show promise for the integrated detection, analysis, and processing of signals from the chemical environment,[1-3] with applications in directed nanoscale assembly[4-6] and actuation,[7-10] and autonomous nucleic acid biosensors and theranostics.[11-13] A variety of DNA-based computational logic gates have been developed and characterized, including strand displacement,[14-16] hairpin assembly,[17-19] ribozymes,[20,21] and DNAzymes[22-25]. DNA strand displacement is a particularly robust, well-known method with many advantages for programming reaction pathways in

biomolecular computing devices.[26] Single-stranded domains called toeholds serve as nucleation sites for invading strand hybridization[27,28] and allow subsequent branch migration reactions to be programmed based on sequence specificity. The kinetics of toehold-mediated strand displacement and strand exchange reactions, and their dependence on toehold lengths, have been previously studied.[29] With a judicious choice of toehold lengths, the reaction can be biased in a particular direction based on thermodynamic considerations[14,30] without relying on concentration-driven effects. While circuits based solely on hybridization and strand displacement are known to exhibit signal amplification capabilities,[15,16,18,19,30-33] these typically require additional strands or structures, thereby increasing circuit complexity. Furthermore, the ability of pure DNA strand displacement devices to interface with other chemistries is limited because strand displacement gates communicate exclusively by means of strand release.

The ability to control other chemical reactions, such as RNA cleavage and phosphorylation, using DNA strand displacement would be of great interest for biomedical applications. To address this need, in this paper we use strand displacement reactions to control the activity of inherently catalytic DNAzymes. This approach could be applied to exercise control over a range of chemical reactions catalyzed by DNAzymes, such as cleavage of RNA[34,35] and DNA,[22,36-38] phosphorylation,[39,40] ligation,[41] synthesis of branched nucleic acids[42] and peroxidase-like oxidation reactions.[43,44] This has potential applications in synthetic logic circuits for conditional gene silencing, and for support of failing cellular metabolisms. Our DNAzyme-based approach is much simpler than

tackling similar problems via *de novo* protein engineering[45], although previous work has demonstrated computational capabilities using existing protein enzymes.[2,46-48] As shown in **Figure 3.1**, we inactivate a DNAzyme (Dz) through hybridization with a partially complementary inhibitor strand (Inh), which binds to at least one of the substrate recognition arms (the s_1 domain) and also deforms the catalytic core of the DNAzyme (the cc domain). This prevents the DNAzyme from binding to, and cleaving, its complementary FRET substrate, sequestering the DNAzyme in a catalytically inactive complex (Dz-Inh). Activating inputs are single-stranded oligomers (Act) that bind to a complementary toehold (t*) on the inhibitor strand and initiate branch migration across the s_1 and cc domains. This results in displacement of a catalytically active DNAzyme strand (Dz) from the complex, which can cleave the complementary substrate (Sub) to produce two shorter cleavage products ($Prod_1$ and $Prod_2$). We call this process *DNAzyme displacement*. Here we focus on the 8-17 DNAzyme,[35,49] due to its small size and high catalytic efficiency (1.35 min^{-1}) in the presence of Zn^{2+} metal ion cofactors.[50,51] However, our approach is broadly applicable to structurally similar DNAzymes, such as the RNA-cleaving 10-23 motif,[35,52] DNA-cleaving DNAzymes,[22,36-38] ligases,[25,41] and peroxidases.[43,44]

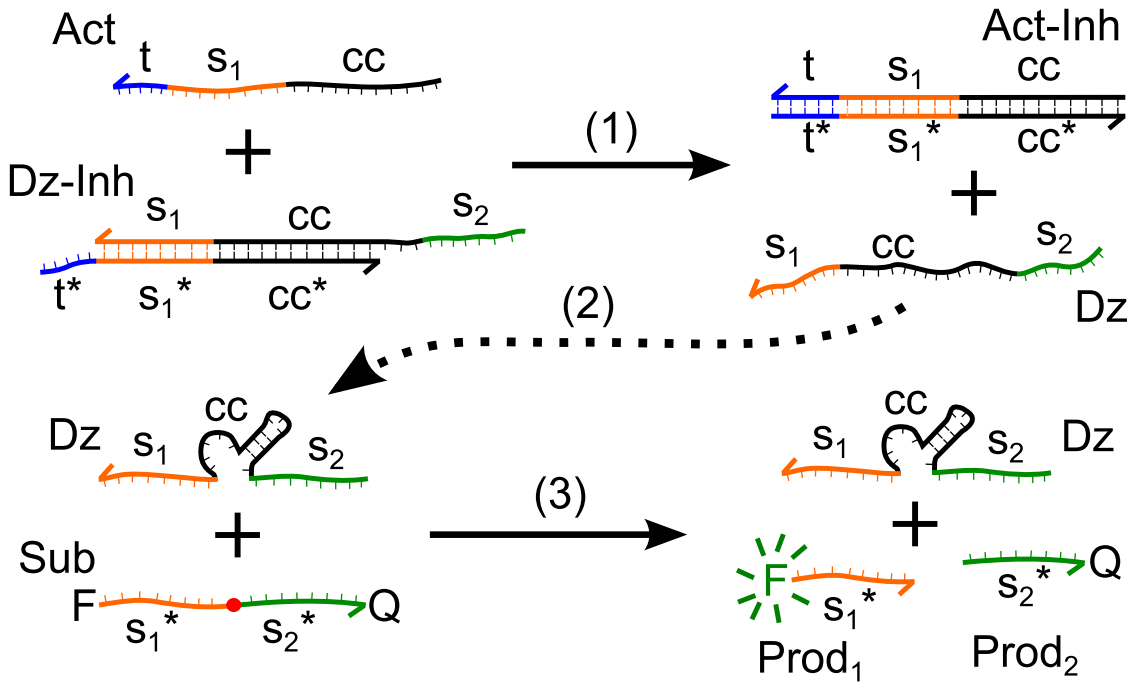

Figure 3.1 DNAzyme displacement reaction mechanism. DNAzyme displacement reaction mechanism. The catalytic activity of the DNAzyme strand (Dz) is inhibited by hybridization to a partially complementary inhibitor strand (Inh) with a short overhanging toehold (t*), resulting in an inactive DNAzyme complex (Dz-Inh). The DNAzyme is activated by a toehold-mediated strand displacement reaction: an activator strand (Act) binds to the complex (Dz-Inh) via the t toehold (step 1). The input initiates a branch-migration reaction across the s_1 and cc domains that eventually displaces the DNAzyme strand (Dz), leaving an inert waste complex (Act-Inh). The DNAzyme strand then folds into a catalytically active conformation (step 2) and proceeds to bind to substrate molecules (Sub) and cleave them, producing shorter cleavage products (step 3). The cleavage reaction causes separation of the fluorophore-quencher pair attached to the cleavage products ($Prod_1$ and $Prod_2$), which causes an increase in bulk fluorescence due to loss of FRET.

We have implemented a set of DNAzyme displacement logic gates complete for Boolean logic. This expands on previous work in the Ellington group[43] by coupling input logic to the gates, and by displacing DNAzymes which catalyze nucleic acid reactions that could engage in further logic functions, rather than producing a colorimetric readout. Here, "true" (logic 1) is encoded by the presence of a particular input species, and "false" (logic 0) is encoded by the absence of that species. **Figure 3.2A** shows the action of a sensor (YES) gate that accepts an input complementary to the inhibitor strand. When the active DNAzyme is displaced from the complex, it cleaves a chimeric RNA/DNA substrate molecule labeled with a FRET pair, producing a fluorescent readout. **Figure 3.2B** demonstrates a NOT logic gate, where the input is an inhibitor strand that binds to the DNAzyme and prevents it from cleaving the FRET substrate. In **Figure 3.2C** we implement AND logic by extending the inhibitor to be complementary to the entire DNAzyme strand, with a toehold at each end (t_1 and t_2). The active DNAzyme can only be displaced by *both* input strands simultaneously, in cooperative strand displacement reactions across the s_1 and s_2 substrate binding domains and both halves of the catalytic core (cc_1 and cc_2).[53] The AND gate may respond more slowly because two input strands must simultaneously bind and completely displace the Dz strand; additionally we observed direct activator-substrate interactions, likely due to sequence overlap (**Figure A1.1**). Importantly, all gates and inputs were ordered and used without additional purification. This is especially important for cost control and ease of use in real-world bioassays. We found that adding the inhibitor in 10-20% excess

relative to the DNAzyme produced optimal performance, as this provides a sufficient margin of error to accommodate concentration variations between stocks without the excess inhibitor slowing the response to low input concentrations (**Figure A1.2**). DNAzyme displacement gates function well at near-equimolar concentrations of the gate, inhibitor, and input species because they rely on thermodynamically driven hybridization to drive the reactions. This allows us to use multiple-turnover output reactions to signal the presence of low concentrations of inputs relative to the gate concentration (**Figure A1.2**), which is a key consideration in the development of practical bioassays.

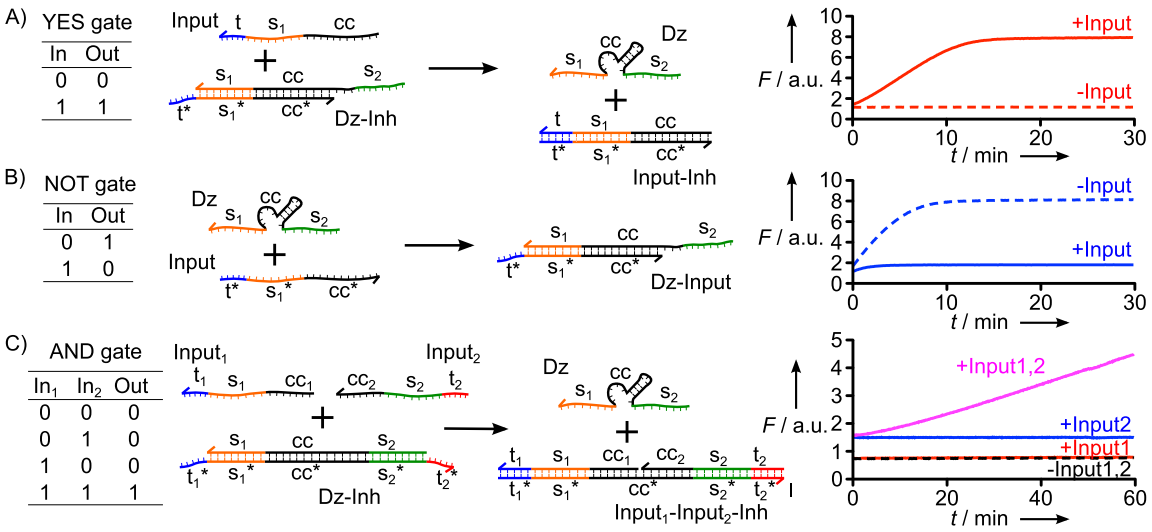

Figure 3.2 Logic gates implemented using DNAzyme displacement reactions. A) The YES gate detects its input via the reaction mechanism shown in Figure 1B. B) The NOT gate accepts an inhibitor strand as input, which deactivates a previously active DNAzyme. C) The AND gate has a full-length inhibitor with toeholds (t_3 and t_4) on each end, and both input strands must be present to release the active DNAzyme via a cooperative strand displacement reaction.

In the experiments discussed above, the input strands had certain sequence restrictions. To be able to displace the DNAzyme strand from the inhibitor complex, the input strands shared some sequence with the fixed catalytic core of the DNAzyme. For practical applications that involve detecting the presence of arbitrary sequences, such as pathogen markers,[11,13] it is necessary to avoid sequence restrictions. We addressed this by developing a DNAzyme displacement AND gate that can detect arbitrary input sequences, as shown in **Figure 3.3A**. (Similar techniques could be used to develop a YES gate.) The input strands bind cooperatively to the AND gate as before, but they do not contain the conserved catalytic core sequence. Hence, they cannot completely displace the DNAzyme from the complex, and the 15 nt catalytic core of the 8-17 DNAzyme is too long for rapid spontaneous dissociation. Therefore we rationally introduced mismatched bases in the cc* domain of the inhibitor, producing a mismatched cc^*_{MM} domain. This destabilized the binding between the inhibitor and the DNAzyme core[54] and promoted unbinding of the active DNAzyme while discouraging rebinding. We also extended the input toeholds to 10 nt to provide an additional bias towards activation. **Figure 3.3B** shows the operation of DNAzyme displacement AND gates based on this principle, with 0, 1, 2, or 3 mismatched bases in the core inhibitor. Since the AND gate inhibitor binds the entire DNAzyme, mismatched bases in the core do not significantly destabilize the complex in the absence of inputs and thus we see no signal under these conditions for each inhibitor variant. As expected, with a fully complementary inhibitor there is very little activation of the DNAzyme in the presence of both

inputs. Increasing the number of mismatches increases the activation rate, with 3 mismatched bases giving the fastest response. **Figure 3.3C** demonstrates that very little activation is seen with just one input present, hence we obtain AND logic. **Appendix 1** contains a more detailed analysis of the mismatched AND gate design.

Figure 3.4 illustrates the application of DNAzyme displacement gates to evaluate a three-input logic function. This circuit uses the AND gate with 3 mismatches from **Figure 3.3** in conjunction with the NOT gate from **Figure 3.2B**, and an implicit logical OR function is achieved because gates cleave substrates that are monitored in the same fluorescent channel. This demonstrates the potential of our approach for precise control of output chemistries by the calculation of non-trivial logic functions.

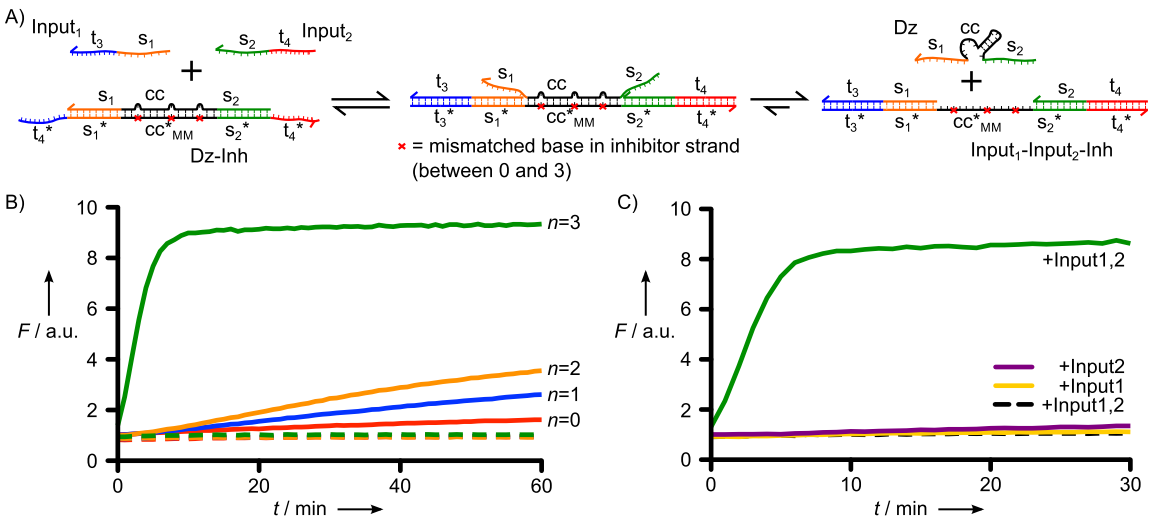

Figure 3.3 Detection of arbitrary input sequences using mismatched inhibitors. A) Mechanism for an AND gate that detects two arbitrary input sequences. Since these inputs no longer displace the catalytic core, mismatched bases are added to the part of the inhibitor strand that binds to the core, to encourage unbinding of the DNAzyme strand in the presence of both inputs. B) Kinetic traces for the AND gate design using mismatched inhibitors, for different numbers n=0,1,2,3 of mismatched bases in the cc^*_{MM} domain. For clarity, we only plot the responses with both inputs present (solid lines) and with neither input present (broken lines with corresponding colors). C) Complete characterization of the AND gate with 3 mismatched bases in the cc^*_{MM} domain. We still see very strong inhibition of the AND gate in the presence of a single input, even when 3 mismatches are present in the inhibitor

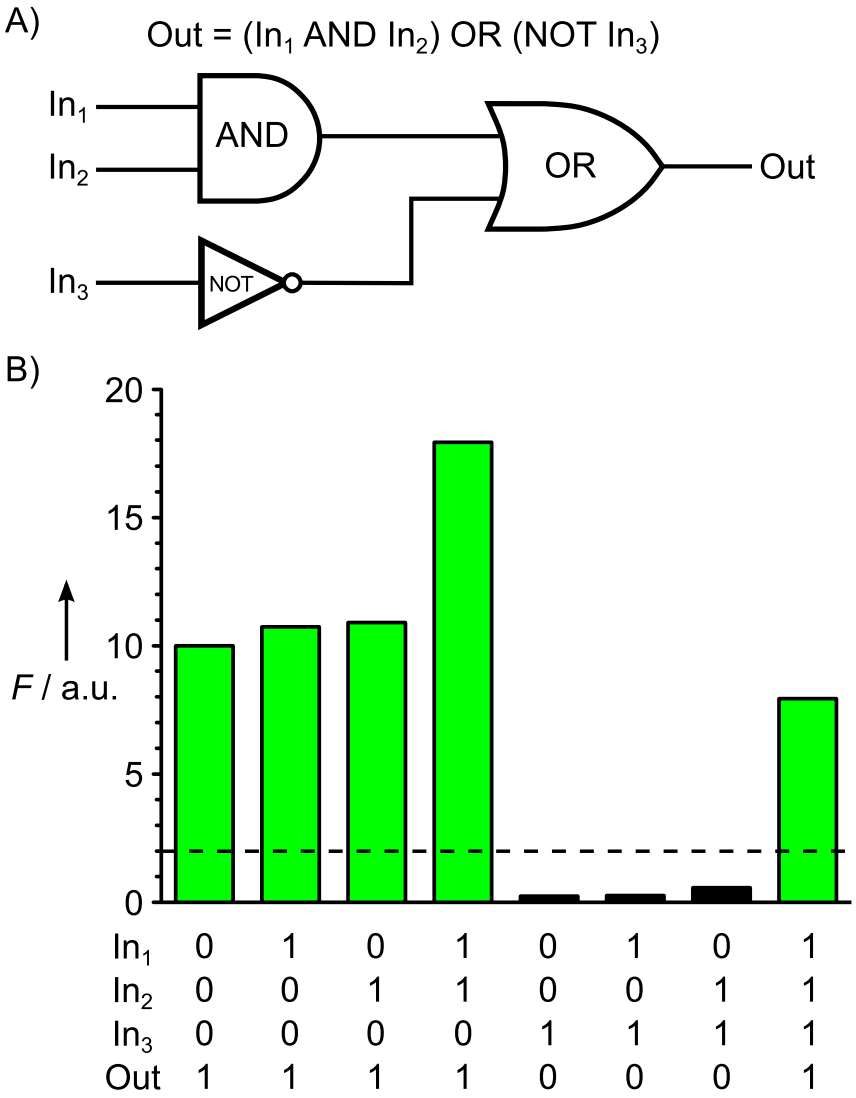

Figure 3.4 Demonstration of a logic computation using DNAzyme displacement gates. A) Diagram of example logic circuit. B) Experimental validation of the corresponding DNAzyme displacement circuit. Responses below the broken red line are interpreted as Out=0 and those above are interpreted as Out=1. The circuit responds correctly for each input combination, with a high signal-to-noise ratio. The fluorescence value in the fourth column is higher because both the AND gate and the NOT gate are active simultaneously in this case, resulting in both substrate populations being cleaved to produce a higher overall fluorescence level.

In summary, we have demonstrated that toehold-mediated strand displacement reactions are a robust, programmable means of regulating the catalytic activity of DNAzymes. Previous work on DNAzyme-based logic has used a number of control mechanisms, such as intramolecular loop-based inhibition[24,55] and self-assembly of multi-component DNAzymes in the presence of inputs.[56-58] Loop-inhibited DNAzymes have been used in DNAzyme logic circuits[23,59-61] and decision-making automata.[55,62,63] While the scale and performance of such circuits has been impressive, they typically require high (μm) input concentrations so that input binding overcomes the intramolecular loop re-inhibition reaction, which makes them less suited to biosensor applications. Applications of multi-component DNAzymes to amplified detection of low concentrations of DNA typically introduce autocatalysis to compensate for this, for example using ligase DNAzymes[64-66] or assembly of larger structures.[67-69] In our approach, controlling toehold lengths allows reaction pathways to be directed based primarily on thermodynamic considerations rather than concentration gradients, which also allows detection and amplification of nanomolar concentrations of inputs using DNAzyme and inhibitor concentrations also in the nanomolar range. Rational introduction of mismatched bases allows rapid detection of arbitrary input sequences, and the inherent catalytic activity of DNAzymes allows amplifying sequence detectors to be constructed with minimal design effort and without extensive purification of strands or complexes. This work will enable DNA strand displacement circuits to be used to control a range of DNAzyme-catalyzed chemical reactions.

The mechanism and potential utility of DNAzymes for *in vivo* theranostics remains an open area of research, due to their dependence on millimolar concentrations of divalent metal cations.[70-74] In particular, the 8-17 DNAzyme performs optimally in the presence of Zn^{2+},[51] which may not be available in the intracellular environment. However, the performance of the 10-23 DNAzyme has been investigated in physiological conditions[75-77] and, due to its size and structural similarity to the 8-17 DNAzyme, the 10-23 DNAzyme should also be amenable to regulation by DNAzyme displacement[35,78]. *In vitro* evolution experiments under stringent physiological conditions or using directed targets may also yield additional potential candidates.[79] If these potential issues are resolved, these techniques may find applications in synthetic cellular regulatory networks and isothermal, nucleic acid-based, amplified detection of viruses and other pathogens.

3.3 Materials and Methods

3.3.1 Oligonucleotide sequences and sequence design.

Conserved sequences for the catalytic core of the 8-17 DNAzyme were obtained from the literature[51]. Sequences for the remaining domains were analyzed using the NUPACK web server[80,81] and manually optimized to limit the formation of unwanted secondary structure. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). DNAzymes, inhibitors, and input strands were ordered purified with standard desalting. DNA/RNA chimeric FRET reporter substrates were ordered purified using RNase-free HPLC. All sequences are listed in **Appendix 1**, along with their respective concentrations in each

experiment. Oligonucleotides were resuspended in RNase-free H_2O (Sigma-Aldrich) in accordance with the manufacturer-provided specifications at a stock concentration (50 μM). Working stocks were made by adding the resuspended oligonucleotide solution (50 μL) into buffer (950 μL). All reactions were run in a buffer of NaCl (1M), HEPES (50 mM), and $ZnCl_2$ (1mM), at pH 7.0.

3.3.2 Logic Gate Preparation.

DNAzyme-inhibitor complexes were prepared by annealing the DNAzyme and inhibitor strands at 95 °C for 3 minutes on a heat block and cooled to room temperature over a minimum of 90 minutes to anneal.

3.3.3 Logic Gate Characterization Assays.

Characterization of logic gate behavior was monitored as a time-based kinetic loss of FRET assay using a chimeric DNA substrate with an RNA base at the cleavage site. Dequenching of a 5' FAM group by the 3' TAMRA group indicated cleavage. Reagents were added in the order of logic gate, input, and subsequent addition of substrate to initiate the reaction. Characterization of individual logic gates (**Figure 3.2**) was performed on a PTI (Birmingham, NJ) Quantamaster-40 fluorimeter at an excitation wavelength of 492 nm and an emission wavelength of 518 nm. Characterization of concentration profiles (**Figure A1.2**), the AND gate for detection of arbitrary sequences (**Figure 3.3**), and the logic circuit demonstration (**Figure 3.4**) were taken on a Spectramax M2e plate reader (Molecular Devices, Sunnyvale, CA). In the experiment demonstrating input signal amplification (**Figure A1.2**), each kinetic trace was baseline-subtracted from the first point of that trace.

3.3.4 Logic Circuit Demonstration.

The circuit was set up in a manner similar to the logic gate characterization experiments. Gates were added first, followed by input, in the concentrations denoted in **Table A1.5**. To assess the final state of the circuit, inputs were allowed to react with gate complexes in the absence of reporter for 15 minutes. Upon addition of substrate, an endpoint fluorescent value was taken after 15 minutes. The t=0 fluorescence value for the case where IN_1=0, IN_2=0 and IN_3=1 was used as a baseline (since in this case we would expect minimal DNAzyme activity) and this value was subtracted from all of the endpoint fluorescence values.

3.4 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 1027877 and 1028238, and by INCBN IGERT DGE-0549500. The authors thank Milan N. Stojanovic for helpful discussions, and Hannah E. West and Eli K. Horwitz for experimental assistance.

3.5 References

(1) *Biomolecular Information Processing - From Logic Systems to Smart Sensors and Actuators*; Katz, E., Ed.; Wiley-VCH: Weinheim, Germany, 2012.

(2) Katz, E.; Privman, V. *Chem Soc Rev* **2010**, *39*, 1835.

(3) Willner, I.; Shlyahovsky, B.; Zayats, M.; Willner, B. *Chem Soc Rev* **2008**, *37*, 1153.

(4) Yan, H.; LaBean, T. H.; Feng, L.; Reif, J. H. *Proc Natl Acad Sci U S A* **2003**, *100*, 8103.

(5) Woo, S.; Rothemund, P. W. *Nat Chem* **2011**, *3*, 620.

(6) Zhang, F.; Nangreave, J.; Liu, Y.; Yan, H. *Nano Lett* **2012**, *12*, 3290.

(7) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605.

(8) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. *Angew Chem Int Ed Engl* **2004**, *43*, 4906.

(9) Bath, J.; Green, S. J.; Allen, K. E.; Turberfield, A. J. *Small* **2009**, *5*, 1513.

(10) Franco, E.; Friedrichs, E.; Kim, J.; Jungmann, R.; Murray, R.; Winfree, E.; Simmel, F. C. *Proc Natl Acad Sci U S A* **2011**, *108*, E784.

(11) Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E. *Nature* **2004**, *429*, 423.

(12) Rudchenko, M.; Taylor, S.; Pallavi, P.; Dechkovskaia, A.; Khan, S.; Butler, V. P., Jr.; Rudchenko, S.; Stojanovic, M. N. *Nat Nanotechnol* **2013**, *8*, 580.

(13) Xie, Z.; Wroblewska, L.; Prochazka, L.; Weiss, R.; Benenson, Y. *Science* **2011**, *333*, 1307.

(14) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585.

(15) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.

(16) Qian, L.; Winfree, E.; Bruck, J. *Nature* **2011**, *475*, 368.

(17) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.

(18) Li, B.; Ellington, A. D.; Chen, X. *Nucleic Acids Res* **2011**, *39*, e110.

(19) Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2013**, *110*, 5386.

(20) Penchovsky, R. *ACS Synth Biol* **2012**, *1*, 471.

(21) Penchovsky, R.; Breaker, R. R. *Nat Biotechnol* **2005**, *23*, 1424.

(22) Chen, X.; Wang, Y. F.; Liu, Q.; Zhang, Z. Z.; Fan, C. H.; He, L. *Angew Chem Int Ed Engl* **2006**, *45*, 1759.

(23) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.

(24) Stojanovic, M. N.; Mitchell, T. E.; Stefanovic, D. *J Am Chem Soc* **2002**, *124*, 3555.

(25) Stojanovic, M. N.; Semova, S.; Kolpashchikov, D.; Macdonald, J.; Morgan, C.; Stefanovic, D. *J Am Chem Soc* **2005**, *127*, 6914.

(26) Zhang, D. Y.; Seelig, G. *Nat Chem* **2011**, *3*, 103.
(27) Xing, Y.; Yang, Z.; Liu, D. *Angew Chem Int Ed Engl* **2011**, *50*, 11934.
(28) Yurke, B.; Mills Jr, A. P. *Genet Prog Evol Mach* **2003**, *4*, 111.
(29) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2009**, *131*, 17303.
(30) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121.
(31) Zhang, D. Y.; Winfree, E. *Nucleic Acids Res* **2010**, *38*, 4182.
(32) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2008**, *130*, 13921.
(33) Li, B.; Jiang, Y.; Chen, X.; Ellington, A. D. *J Am Chem Soc* **2012**, *134*, 13918.
(34) Parker, D. J.; Xiao, Y.; Aguilar, J. M.; Silverman, S. K. *J Am Chem Soc* **2013**, *135*, 8472.
(35) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci USA* **1997**, *94*, 4262.
(36) Chandra, M.; Sachdeva, A.; Silverman, S. K. *Nat Chem Biol* **2009**, *5*, 718.
(37) Gu, H.; Furukawa, K.; Weinberg, Z.; Berenson, D. F.; Breaker, R. R. *J Am Chem Soc* **2013**, *135*, 9121.
(38) Xiao, Y.; Wehrmann, R. J.; Ibrahim, N. A.; Silverman, S. K. *Nucleic Acids Res* **2012**, *40*, 1778.
(39) Li, Y.; Breaker, R. R. *Proc Natl Acad Sci U S A* **1999**, *96*, 2746.
(40) Walsh, S. M.; Sachdeva, A.; Silverman, S. K. *J Am Chem Soc* **2013**, *135*, 14928.
(41) Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.
(42) Lee, C. S.; Mui, T. P.; Silverman, S. K. *Nucleic Acids Res* **2011**, *39*, 269.
(43) Eckhoff, G.; Codrea, V.; Ellington, A. D.; Chen, X. *J Syst Chem* **2010**, *1*, 13.
(44) Gerasimova, Y. V.; Cornett, E. M.; Edwards, E.; Su, X.; Rohde, K. H.; Kolpashchikov, D. M. *ChemBioChem* **2013**, *14*, 2087.
(45) Martin, C. H.; Nielsen, D. R.; Solomon, K. V.; Prather, K. L. *Chem Biol* **2009**, *16*, 277.
(46) Katz, E.; Wang, J.; Privman, M.; Halamek, J. *Anal Chem* **2012**, *84*, 5463.
(47) Wang, J.; Katz, E. *Anal Bioanal Chem* **2010**, *398*, 1591.
(48) Zhou, J.; Arugula, M. A.; Halamek, J.; Pita, M.; Katz, E. *J Phys Chem B* **2009**, *113*, 16065.
(49) Peracchi, A.; Bonaccio, M.; Clerici, M. *J Mol Biol* **2005**, *352*, 783.
(50) Bonaccio, M.; Credali, A.; Peracchi, A. *Nucleic Acids Res* **2004**, *32*, 916.
(51) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.
(52) Joyce, G. F. *Methods Enzymol* **2001**, *341*, 503.
(53) Zhang, D. Y. *J Am Chem Soc* **2011**, *133*, 1077.

(54) Jiang, Y. S.; Bhadra, S.; Li, B.; Ellington, A. D. *Angew Chem Int Ed* **2014**, *53*, 1845.
(55) Stojanovic, M. N.; Stefanovic, D. *Nat Biotechnol* **2003**, *21*, 1069.
(56) Elbaz, J.; Moshe, M.; Shlyahovsky, B.; Willner, I. *Chemistry* **2009**, *15*, 3411.
(57) Kolpashchikov, D. M. *Chem Rev* **2010**, *110*, 4709.
(58) Mokany, E.; Bone, S. M.; Young, P. E.; Doan, T. B.; Todd, A. V. *J Am Chem Soc* **2010**, *132*, 1051.
(59) Elbaz, J.; Wang, F.; Remacle, F.; Willner, I. *Nano Lett* **2012**, *12*, 6049.
(60) Orbach, R.; Remacle, F.; Levine, R. D.; Willner, I. *Proc Natl Acad Sci USA* **2012**, *109*, 21228.
(61) Lederman, H.; Macdonald, J.; Stefanovic, D.; Stojanovic, M. N. *Biochemistry* **2006**, *45*, 1194.
(62) Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N. *Nano Lett* **2006**, *6*, 2598.
(63) Pei, R.; Matamoros, E.; Liu, M.; Stefanovic, D.; Stojanovic, M. N. *Nat Nanotechnol* **2010**, *5*, 773.
(64) Lu, C. H.; Wang, F.; Willner, I. *J Am Chem Soc* **2012**, *134*, 10651.
(65) Wang, F.; Elbaz, J.; Teller, C.; Willner, I. *Angew Chem Int Ed Engl* **2011**, *50*, 295.
(66) Wang, F.; Elbaz, J.; Willner, I. *J Am Chem Soc* **2012**, *134*, 5504.
(67) Gerasimova, Y. V.; Kolpashchikov, D. M. *Angew Chem Int Ed Engl* **2013**, *52*, 10586.
(68) Kolpashchikov, D. M.; Gerasimova, Y. V.; Khan, M. S. *ChemBioChem* **2011**, *12*, 2564.
(69) Wang, F.; Elbaz, J.; Orbach, R.; Magen, N.; Willner, I. *J Am Chem Soc* **2011**, *133*, 17149.
(70) Young, D. D.; Lively, M. O.; Deiters, A. *J Am Chem Soc* **2010**, *132*, 6183.
(71) Fokina, A. A.; Meschaninova, M. I.; Durfort, T.; Venyaminova, A. G.; Francois, J. C. *Biochemistry* **2012**, *51*, 2181.
(72) Peracchi, A. *Rev Med Virol* **2004**, *14*, 47.
(73) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.
(74) Baum, D. A.; Silverman, S. K. *Cell Mol Life Sci* **2008**, *65*, 2156.
(75) Dass, C. R. *Trends Pharmacol Sci* **2004**, *25*, 395.
(76) Dass, C. R.; Choong, P. F.; Khachigian, L. M. *Mol Cancer Ther* **2008**, *7*, 243.
(77) Dass, C. R.; Saravolac, E. G.; Li, Y.; Sun, L. Q. *Antisense Nucleic Acid Drug Dev* **2002**, *12*, 289.
(78) Santoro, S. W.; Joyce, G. F. *Biochemistry* **1998**, *37*, 13330.
(79) Sriram, B.; Banerjea, A. C. *Biochem J* **2000**, *352 Pt 3*, 667.
(80) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 170.

(81) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev* **2007**, *49*, 65.

Chapter 4.
Signal Propagation in Multi-Layer DNAzyme Cascades using Structured Chimeric Substrates

Carl W. Brown, III[1], Matthew R. Lakin[3], Eli K. Horwitz[1], M. Leigh Fanning[3], Hannah E. West[1], Darko Stefanovic*[1,3], and Steven W. Graves*[1,2]

[1] Center for Biomedical Engineering, University of New Mexico, Albuquerque, NM 87131, USA

[2] Department of Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, NM 87131, USA

[3] Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA

(Accepted to *Angewandte Chemie International Edition*)

4.1 Abstract

Signal propagation through enzyme cascades is a critical component of information processing in cellular systems. Although such systems have potential as biomolecular computing tools, rational design of synthetic protein networks remains infeasible. DNA strands with catalytic activity (DNAzymes) are an attractive alternative, enabling rational cascade design through predictable base-pair hybridization principles. Multi-layered DNAzyme signaling and logic cascades are now reported. Signaling between DNAzymes was achieved using a structured chimeric substrate (SCS) that releases a downstream activator after cleavage by an upstream DNAzyme. The SCS can be activated by various upstream DNAzymes, can be coupled to DNA strand-displacement devices, and is highly resistant to interference from background DNA. This work enables the rational design of synthetic DNAzyme regulatory networks, with potential applications in biomolecular computing, biodetection, and autonomous theranostics.

4.2 Body of Manuscript

Cells use enzymatic signaling pathways for a number of critical functions, including detection of environmental stimuli, signal amplification, and regulated information propagation through the intracellular environment. Cells typically implement these functions using proteins, but the complexity of protein folding makes the rational design of protein-based signaling cascades infeasible[1]. Although prior work on biocomputing devices using naturally occurring proteins shows promise[2-6], this approach is limited by the possible protein-protein

interactions. DNA, on the other hand, is an ideal alternative engineering material for *de novo* design of synthetic enzymatic cascades, thanks to predictable Watson-Crick base pairing and secondary structure formation. Synthetic analogs of some basic cellular processes have been implemented in DNA, including computation,[7-9] self-assembly,[10,11] locomotion,[12-17] small molecule sensing,[18-22] and catalysis[23-25]. Here we focus on DNAzymes[26] (also known as deoxyribozymes), which are single-stranded DNA molecules that can catalyze many of the same reactions as protein enzymes[27-30] and have been used for computation in parallel gate arrays.[9,31,32] We report a DNAzyme cascade system that uses structured, single-stranded substrates to sequester activating sequences and to propagate an activating signal to a downstream DNAzyme when cleaved by an upstream DNAzyme. We have developed multi-layer signaling cascades and logic circuits, in which a conformational change in a molecule propagates information downstream, mimicking biological systems that rely on modifications such as phosphorylation of downstream enzymes to propagate information.

We based our designs on the most widely used family of DNAzymes: RNA-cleaving DNAzymes. With appropriate metal cation cofactors, these DNAzymes cleave RNA or chimeric DNA/RNA substrates in a multiple-turnover reaction, providing built-in signal amplification capabilities. For a given catalytic motif, DNAzyme-substrate pairs can be designed by simply choosing appropriate complementary sequences for the substrate and the substrate-binding arms of the DNAzyme. This is considerably simpler than designing enzyme-substrate

pairs *de novo* by protein engineering. We used the 8-17 RNA-cleaving DNAzyme due to its compact size and efficient catalytic rate.[33-35]

This work is built on previous work on ribozyme circuits[36,37] and on DNAzyme signaling cascades that either sequestered the downstream effector sequence in a partially complementary complex,[38,39] or built two-layer cascades where the downstream DNAzyme generates a colorimetric readout.[40] In the first case, the use of a multi-strand complex as the mediator increases the number of strands and the complexity of circuit preparation. In the second case, the downstream DNAzyme cannot propagate the signal further within the molecular circuit. Although signal amplification has also been demonstrated using DNA strand displacement[25,41-43] and catalytic hairpin assembly,[44,45] these circuits must be specifically designed to obtain catalysis, for example, using seesaw gates.[7,8,46] We use DNAzyme displacement reactions,[47] which combine the advantages of strand displacement to program reaction pathways with the inherent catalytic ability of DNAzymes. This reduces the number of DNA strands needed to achieve signal amplification.

In cellular enzymatic signaling cascades, an activation signal is typically passed from one enzyme to another through chemical modifications. In this work, information propagation between enzymatic units through the covalent modification of a structured chimeric substrate (SCS). The SCS uses a metastable dual stem-loop design[48,49] (**Figure 4.1a**) and comprises several domains that make up interchangeable input and output modules. The use of a modular intermediary simplifies the design process by removing the need for

direct enzyme-enzyme interactions, which are often found with protein-based cascades, such as phosphorylation in the MAPK pathway.[50] The inner 7bp stem and 8bp loop constitute the output module, whose secondary structure weakly sequesters a downstream activator. The outer 7bp stem and 6bp loop stabilize the structure and protect the activator toehold in the outer loop, thus preventing unwanted interactions with the downstream DNAzyme before cleavage. The outer stem and loop also constitute the input module, with a substrate-binding and -cleavage domain for an upstream DNAzyme. We minimized the size of the outer loop to better protect the toehold, which led to a 5bp overlap between the upstream DNAzyme binding arm and the downstream inhibitor toehold sequences. As shown in **Figure 4.1b**, an upstream DNAzyme interacts with the SCS when one of the 8bp substrate binding arms hybridizes with the 4bp outermost toehold and opens the outer stem via a toehold-mediated strand displacement reaction.[42] The second arm binds the outer loop, linearizing the substrate domain and correctly positioning the SCS cleavage site opposite the catalytic core of the DNAzyme. The subsequent cleavage reaction causes the outer stem to dissociate as waste, freeing the protected toehold in the outer loop of the SCS, which can now hybridize to its complement more effectively. The relatively weak secondary structure in the activator released by SCS cleavage allows it to interconvert between hairpin and linear structural forms. Thus, downstream interactions are not impeded by secondary structure in the activator.

This mechanism is particularly suited for use with our previously reported DNAzyme displacement logic gates,[47] in which DNAzyme catalysis is controlled

using toehold-mediated DNA strand displacement reactions.[42] The activator released by SCS cleavage binds to the toehold of the downstream DNAzyme-inhibitor complex and undergoes branch migration to displace a catalytically active DNAzyme strand, producing an inert waste complex (**Figure 4.1c**). The displaced DNAzyme refolds into a catalytically active conformation and can then cleave its own substrate. Thus, activation of one DNAzyme species causes the activation of a second DNAzyme species, implementing signal propagation.

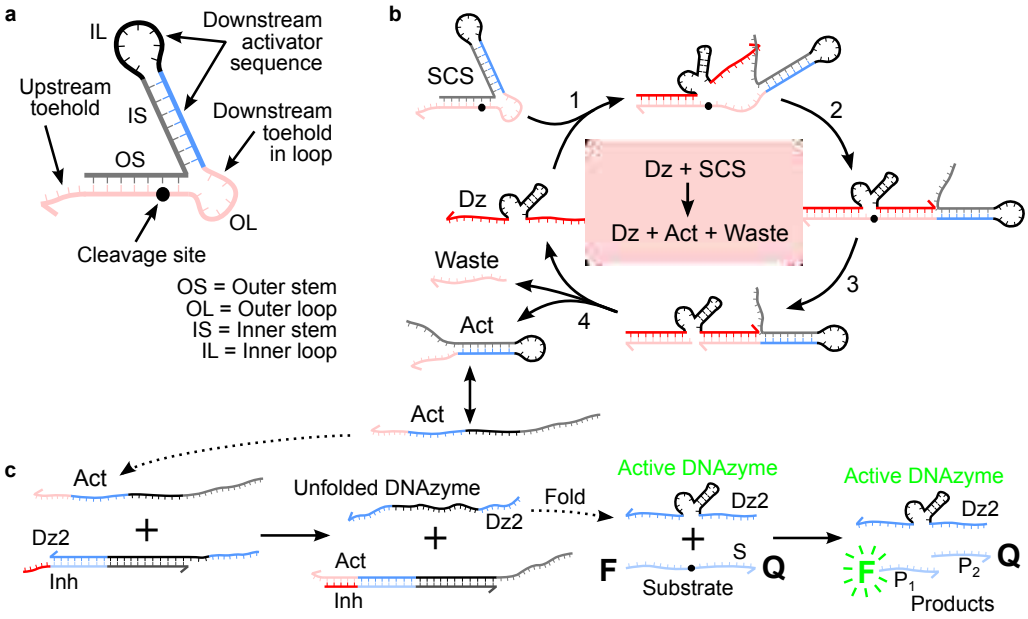

Figure 4.1 - SCS design and mechanisms for SCS cleavage and DNAzyme displacement. a) Design of an SCS to enable signaling between DNAzymes. The SCS consists of an outer stem and loop, which make up the upstream DNAzyme binding domain (red), and an inner stem and loop, which sequester a downstream activator sequence (blue and black). The cleavage site is located towards the inner end of the outer stem. The grey cage sequence is chosen to fold into the desired structure, producing a topological constraint on the downstream reaction kinetics that is undone when the SCS is cleaved by the upstream DNAzyme. b) Mechanism of cleavage of the SCS by an upstream DNAzyme (Dz). The upstream DNAzyme binds to the outer stem and loop by toehold-mediated strand displacement. The cleavage reaction produces a waste strand and an activator strand (Act). In the activator structure, the outer loop has been released from the topological constraint previously imposed by the outer stem, making the downstream toehold in the outer loop available to bind with a downstream circuit element. c) DNAzyme displacement reaction mechanism. The catalytic activity of the downstream DNAzyme strand (Dz2) is inhibited by hybridization to a partially complementary inhibitor strand (Inh) with a short overhanging toehold. Activation occurs by a toehold-mediated strand displacement reaction: An input strand (Act) binds to the complex (Dz2-Inh) through the toehold. The input initiates a branch-migration reaction that eventually displaces a catalytically active downstream DNAzyme strand (Dz2), leaving an inert waste complex (Act-Inh). The DNAzyme strand then folds into a catalytically active conformation and proceeds to bind to substrate molecules (S) and cleave them, producing shorter cleavage products (P_1 and P_2). The cleavage reaction causes separation of the fluorophore-quencher pair attached to the two ends of the substrate, which isobserved as an increase in bulk fluorescence due to a loss of fluorescence resonance energy transfer (FRET).

For correct behavior in synthetic multi-enzyme systems, each enzyme must interact with its intended substrate with high specificity. In protein-based enzymatic cascades, specificity is derived from complex interactions between the secondary and tertiary conformations of both enzyme and substrate, rendering rational design of such interactions infeasible. DNAzyme-based cascades achieve specificity through sequence-specific hybridization to substrates. We can modify the SCS input and output modules to enable signaling between DNAzymes with different substrate binding arms while keeping the SCS structure intact. As the SCS does not need to be redesigned for each subsequent layer, this enables rapid construction of two-, three-, four-, and five-layer linear DNAzyme signaling cascades, each initiated by the addition of active top-layer DNAzymes (**Figure 4.2**). Each cascade uses the same reporter layer (layer 1), with layer *n*+1 added upstream of layer *n* to extend the cascade. This naming system reflects the sequence commonality between each *n* layer, irrespective of cascade length. Our five-layer cascade is the longest DNAzyme signaling cascade implemented to date. The development of extended catalytic signaling cascades with a high signal-to-noise ratio is challenging because unwanted signal generated in the absence of input (leakage) is also amplified by downstream circuit elements. Kinetic traces of multi-layer cascades (**Figure 4.2b**) show that the time taken for cascade execution increase with the number of layers (Supplementary Discussion in **Sections A2.2.11-13**). Lower DNAzyme concentrations reduce leakage at the expense of activation speed by relying on multiple-turnover cleavage for signal amplification (**Figure A2.1**). In particular,

using lower concentrations in the upstream layers of the cascade with increasing concentrations in each downstream layer can reduce leakage without affecting the maximum output level or a significant sacrifice in speed (**Figure 4.2c**). Additional controls using uncleavable SCS molecules demonstrate that cleavage is necessary for signal propagation (**Figure 4.2**). Therefore, we have demonstrated that chemical modification of a structured substrate by a DNAzyme can be used to propagate information in a signaling cascade.

As DNA interactions are sequence-specific, the SCS can interact with any upstream or downstream circuit components with the correct sequence. We implemented signaling cascades between a variety of DNA logic components in both the upstream and downstream positions, including various DNAzyme logic gates and a strand displacement reporter gate (**Figure A2.3**). This demonstrates the flexibility of DNAzyme-based interactions via the SCS, which enables development of hybrid DNA circuits comprising components from multiple architectures, which is currently a significant challenge.

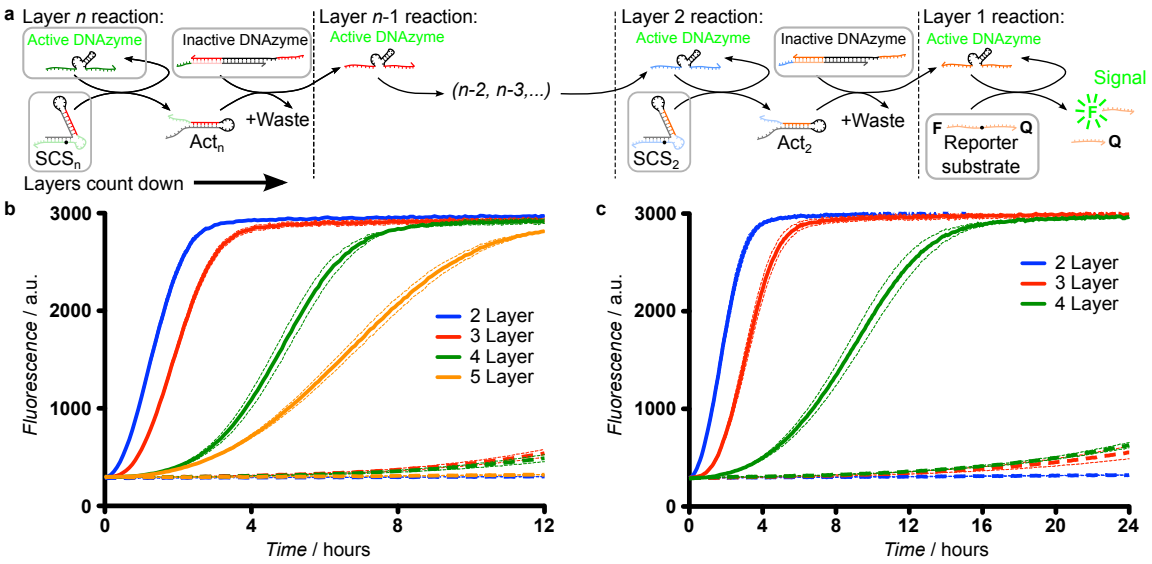

Figure 4.2 - Demonstration of DNAzyme signaling cascades. a) Multi-layer DNAzyme signaling cascades using DNAzyme displacement reactions. Initial species for each layer of the cascade are highlighted in grey boxes. In each layer, an active DNAzyme cleaves the corresponding SCS, producing an activator that releases the downstream DNAzyme from its catalytically inactive enzyme-inhibitor complex via a DNAzyme displacement reaction, thereby propagating the activating signal to the next layer of the cascade. b) The mean fluorescence signal (solid lines) from two-layer (blue), three-layer (red), four-layer (green) and five-layer (orange) linear DNAzyme signaling cascades with equimolar (100 nM) DNAzyme concentrations in each layer. The dashed line represents the same reaction without the top-layer active DNAzyme, which measures the non-specific activation (leakage) of the downstream circuit. The dotted lines represent the 95% confidence interval from three replicate experiments. c) Kinetic traces for multi-layer linear DNAzyme signaling cascades with increasing DNAzyme concentrations in each layer (25 nM in fourth layer, 50 nM in third, 75 nM in second, and 100 nM in first) to demonstrate signal amplification. In both plots, dotted lines represent the 95% confidence interval from three replicate experiments.

Multi-layer synthetic DNAzyme logic cascades offer a route to increasing the sophistication of biomolecular logic circuits, with the long-term aim of enabling robust, isothermal detection of disease states via sequence-specific nucleic acid detection[40,51-55] or aptamer-based detection of small molecules.[18-20] Incorporating logic into enzymatic cascades enables the integration of multiple input signals, which can reduce false positives in bioassays and enable detection of disease states where a single target is insufficient for an accurate diagnosis. To illustrate the potential of DNAzyme logic cascades for detecting multiple pathogenic targets in extracted DNA, we implemented multi-layer circuits for typing representative pathogen signatures from all four dengue virus serotypes (DEN1-4). Dengue is a major global health concern,[56] and accurate serotyping isimportant because sequential infection with different serotypes is a risk factor for dengue hemorrhagic fever and dengue shock syndrome, both of which can be fatal.[57]

We exploited the modularity of the SCS to design a two-layer, three-input AND circuit template, in which two DNA oligomers derived from conserved sequences within the ssRNA dengue genomes and a serotype-specific DNA oligomer must be present to produce a fluorescent output. As shown in **Figure 4.3a**, each layer of the circuit is an AND gate activated by two inputs in a cooperative displacement reaction.[58] The use of mismatches in the inhibitor is required for rapid release of the DNAzyme, because the catalytic core is not displaced by either input.[47] One of the inputs of the downstream gate is released upon cleavage of the SCS. We replicated this template, modifying the highlighted

parts of the upstream AND gate and the SCS, to produce four circuits, each sensitive to a different serotype-specific target sequence (**Figure A2.4**). We observed strong positive responses from all four circuits in the presence of all three signatures, at least 2.5 times the maximum response seen in the absence of one or more signatures (**Figure 4.3b**). Higher leakage is seen in the presence of the downstream DengueB input, suggesting that there is some interaction between the SCS and the downstream AND gate prior to SCS cleavage. Misfolding of SCS or enzyme strands reduced system performance (**Figure A2.5**), showing that optimization of the predicted secondary structure is important for efficient circuit operation.

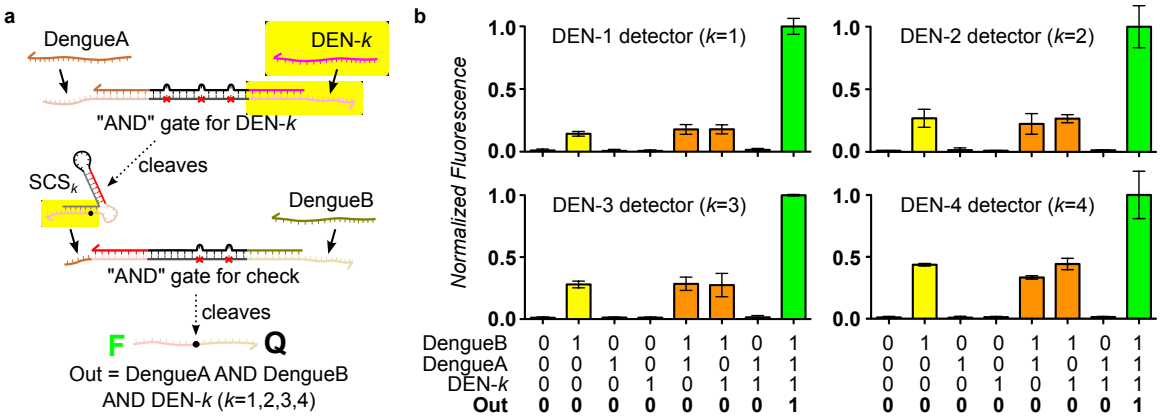

Figure 4.3 – Exemplary application of the SCS in a multi-layer diagnostic logic circuit. a) Design of multi-layer diagnostic logic circuits for detection of sequences from the genomes of all four dengue serotypes. The circuit template for serotype DEN-*k* (*k*=1,2,3,4) requires the presence of two conserved sequences from the dengue viral genome (DengueA and DengueB) and one sequence specific to the serotype of interest (DEN-*k*). This is implemented by DNAzyme displacement "AND" gates with mismatched inhibitors,[47] which are connected by a SCS molecule. When both upstream inputs are present, the active upstream DNAzyme cleaves the SCS, producing an activator that serves as one input to the downstream gate. If the second input to the downstream gate is also present, the downstream DNAzyme will be activated, which is observed by a loss of FRET following substrate cleavage. The upstream AND gate uses the three-mismatch design, characterized in our previous work,[47] whereas the downstream gate uses an asymmetric pattern of mismatches because the activator produced by the SCS only partially displaces the catalytic core. We derived detection circuits for all four dengue serotypes (DEN1-4) by modifying only the domains highlighted in yellow. b) Demonstrations of serotyping circuits for DEN1-4, which show correct operation of all four instantiations of the three-input AND circuit template. Each serotyping circuit was characterized using all eight combinations of the two conserved sequences and the correct serotype-specific sequence. Variations in the normalized fluorescence levels (i.e., different levels of activation and leakage) may be attributed to variations in the stability of the corresponding SCS_k structure in each case. Error bars represent the 95% confidence interval from three replicate experiments.

To test our approach in a minimal biological background, we implemented a two-layer DNAzyme cascade using the SCS with increasing amounts of random background DNA (**Figure A2.6**). This models a common detection scenario in which all the nucleic acids have been extracted from a sample for analysis. We showed that the SCS design is sufficient for this minimal assay detection environment. Furthermore, all experiments described herein were performed using minimal oligonucleotide purification techniques, which is essential for the development and use of low-cost bioassays. Therefore, we have demonstrated two key properties for a practical bioassay: robust operation in background and straightforward preparation.

In summary, we have developed a method to design extended DNAzyme signaling cascades that exhibit many of the functionalities of cellular cascades: integration of multiple input signals, signal amplification, transduction, and propagation. The combination of DNAzymes, strand displacement and rationally designed, structured chimeric substrates enabled us to implement synthetic signaling cascades compatible with a variety of DNA logic gates, including the longest DNAzyme signaling cascade demonstrated to date. These DNAzyme cascades hold promise for practical applications such as pathogen detection. We illustrated this by demonstrating that our circuits resist background interference and can implement multi-input, multi-layer detection of multiple pathogen signatures.

Future work will explore the operation of DNAzyme cascades in physiologically relevant conditions[59] such as cell lysate[60] or serum, which may be

challenging due to the presence of nucleases that may degrade circuit components, or because of insufficient concentrations of the metal ion cofactors required for efficient DNAzyme catalysis.[61,62] Furthermore, the modular design of the SCS should allow the implementation of increasingly complex synthetic DNAzyme signaling networks, incorporating network motifs such as feedforward and feedback cycles.[63] These circuits could exhibit non-trivial dynamic behaviors to enable more sophisticated decision-making for diagnostic and therapeutic applications, possibly connected to alternative readout technologies such as gold nanoparticles[64-67] or paperfluidic devices.[51,68]

4.3 Materials and Methods

4.3.1 Description of Materials

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Oligonucleotide sequences are listed in **Tables A2.1-A2.8**. DNAzymes and inhibitors were purchased with standard desalting whenever possible, with the exception of oligonucleotides that exceeded 60 base pairs in length (which were PAGE purified by the manufacturer, in accordance with the manufacturer's recommended procedures). All DNA/RNA chimeric substrate molecules (SCS molecules and fluorescent reporter substrates) were purified by RNase-free HPLC by the manufacturer. The fluorescent reporter substrates were labeled with a 5' FAM quenched by a 3' TAMRA fluorophore. Oligonucleotides were resuspended in RNase-free H_2O (Sigma-Aldrich) in accordance with the manufacturer-provided specifications at a stock concentration of 50 μM. Working

stocks were made by adding 50 µL of the resuspended oligonucleotide solution into 950 µL buffer.

4.3.2 Preparation of DNAzyme-inhibitor complexes and SCS molecules

DNAzyme strands and inhibitor strands were pre-complexed by heating the DNAzyme and inhibitor strands together at 95 °C for 3 minutes on a heat block, and subsequently annealing by cooling to room temperature over a minimum of 90 minutes. In many cases, an excess of inhibitor relative to DNAzyme was used, to ensure complete inhibition of the DNAzymes – in these cases, the resulting solution of DNAzyme-inhibitor complexes and excess free inhibitor strands was used without further purification. Single-stranded SCS molecules (and loop-inhibited DNAzymes) were prepared using the same heating and annealing protocol.

4.3.3 Assay conditions and instrumentation

All assays were performed at room temperature (23 °C) in a buffer of 1M NaCl, 50 mM HEPES, 1 mM $ZnCl_2$, pH 7.0. Fluorescence was read either on a Quantamaster 40 fluorimeter (PTI, Binghamton, NJ) in a 300 µL reaction volume or Spectramax M2e fluorescent plate reader (Molecular Devices, Sunnyvale, CA) in a 200 µL reaction volume. In all cases, fluorescein emission was monitored at 492 nm excitation and 518 nm emission wavelengths. Full details of assay conditions for individual experiments are listed in **Appendix 2**.

4.4 References

(1) Martin, C. H.; Nielsen, D. R.; Solomon, K. V.; Prather, K. L. *Chem Biol* **2009**, *16*, 277.

(2) *Biomolecular Information Processing - From Logic Systems to Smart Sensors and Actuators*; Katz, E., Ed.; Wiley-VCH: Weinheim, Germany, 2012.

(3) Katz, E.; Privman, V. *Chem Soc Rev* **2010**, *39*, 1835.

(4) Katz, E.; Wang, J.; Privman, M.; Halamek, J. *Anal Chem* **2012**, *84*, 5463.

(5) Wang, J.; Katz, E. *Anal Bioanal Chem* **2010**, *398*, 1591.

(6) Zhou, J.; Arugula, M. A.; Halamek, J.; Pita, M.; Katz, E. *J Phys Chem B* **2009**, *113*, 16065.

(7) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.

(8) Qian, L.; Winfree, E.; Bruck, J. *Nature* **2011**, *475*, 368.

(9) Stojanovic, M. N.; Stefanovic, D. *Nat Biotechnol* **2003**, *21*, 1069.

(10) Woo, S.; Rothemund, P. W. *Nat Chem* **2011**, *3*, 620.

(11) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.

(12) Bath, J.; Green, S. J.; Allen, K. E.; Turberfield, A. J. *Small* **2009**, *5*, 1513.

(13) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. *Angew Chem Int Ed Engl* **2004**, *43*, 4906.

(14) Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605.

(15) Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. *Nature* **2010**, *465*, 206.

(16) Pei, R.; Taylor, S. K.; Stefanovic, D.; Rudchenko, S.; Mitchell, T. E.; Stojanovic, M. N. *J Am Chem Soc* **2006**, *128*, 12693.

(17) Gu, H.; Chao, J.; Xiao, S. J.; Seeman, N. C. *Nature* **2010**, *465*, 202.

(18) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656.

(19) Stojanovic, M. N.; Kolpashchikov, D. M. *J Am Chem Soc* **2004**, *126*, 9266.

(20) Teller, C.; Shimron, S.; Willner, I. *Anal Chem* **2009**, *81*, 9114.

(21) Achenbach, J. C.; Nutiu, R.; Li, Y. *Analytica Chimica Acta* **2005**, *534*, 41.

(22) Li, Y.; Lu, Y. *Functional Nucleic Acids for Analytical Applications*; Springer, 2009.

(23) Seelig, G.; Yurke, B.; Winfree, E. *J Am Chem Soc* **2006**, *128*, 12211.

(24) Stojanovic, M. N.; Mitchell, T. E.; Stefanovic, D. *J Am Chem Soc* **2002**, *124*, 3555.

(25) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121.

(26) Li, Y. F.; Breaker, R. R. *Curr Opin Struct Biol* **1999**, *9*, 315.
(27) Breaker, R. R.; Joyce, G. F. *Chem Biol* **1995**, *2*, 655.
(28) Chandra, M.; Sachdeva, A.; Silverman, S. K. *Nat Chem Biol* **2009**, *5*, 718.
(29) Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.
(30) Li, Y.; Breaker, R. R. *Proc Natl Acad Sci U S A* **1999**, *96*, 2746.
(31) Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N. *Nano Lett* **2006**, *6*, 2598.
(32) Pei, R.; Matamoros, E.; Liu, M.; Stefanovic, D.; Stojanovic, M. N. *Nat Nanotechnol* **2010**, *5*, 773.
(33) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.
(34) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.
(35) Schlosser, K.; Li, Y. *Chembiochem* **2010**, *11*, 866.
(36) Penchovsky, R. *ACS Synth Biol* **2012**, *1*, 471.
(37) Penchovsky, R.; Breaker, R. R. *Nat Biotechnol* **2005**, *23*, 1424.
(38) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.
(39) Elbaz, J.; Moshe, M.; Shlyahovsky, B.; Willner, I. *Chemistry* **2009**, *15*, 3411.
(40) Gerasimova, Y. V.; Cornett, E. M.; Edwards, E.; Su, X.; Rohde, K. H.; Kolpashchikov, D. M. *ChemBioChem* **2013**, *14*, 2087.
(41) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585.
(42) Zhang, D. Y.; Seelig, G. *Nat Chem* **2011**, *3*, 103.
(43) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2008**, *130*, 13921.
(44) Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2013**, *110*, 5386.
(45) Li, B.; Ellington, A. D.; Chen, X. *Nucleic Acids Res* **2011**, *39*, e110.
(46) Qian, L.; Winfree, E. *J R Soc Interface* **2011**, *8*, 1281.
(47) Brown, C. W., III; Lakin, M. R.; Stefanovic, D.; Graves, S. W. *ChemBioChem* **2014**, *15*, 950.
(48) Bois, J. S.; Venkataraman, S.; Choi, H. M.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A. *Nucleic Acids Res* **2005**, *33*, 4090.
(49) Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P.; Blakey, M. I.; Simmel, F. C. *Phys Rev Lett* **2003**, *90, 118102*.
(50) Seger, R.; Krebs, E. G. *FASEB J* **1995**, *9*, 726.
(51) Allen, P. B.; Arshad, S. A.; Li, B.; Chen, X.; Ellington, A. D. *Lab Chip* **2012**, *12*, 2951.
(52) Lu, C. H.; Wang, F.; Willner, I. *J Am Chem Soc* **2012**, *134*, 10651.
(53) Wang, F.; Elbaz, J.; Teller, C.; Willner, I. *Angew Chem Int Ed Engl* **2011**, *50*, 295.
(54) Wang, F.; Elbaz, J.; Willner, I. *J Am Chem Soc* **2012**, *134*, 5504.
(55) Weizmann, Y.; Patolsky, F.; Willner, I. *Analyst* **2001**, *126*, 1502.

(56) Simmons, C. P.; Farrar, J. J.; Nguyen, V. V. C.; Wills, B. *New Engl J Med* **2012**, *366*, 1423.
(57) Sangkawibha, N.; Rojanasuphot, S.; Ahandrik, S.; Viriyapongse, S.; Jatanasen, S.; Salitul, V.; Phanthumachinda, B.; Halstead, S. B. *Am J Epidemiol* **1984**, *120*, 653.
(58) Zhang, D. Y. *J Am Chem Soc* **2011**, *133*, 1077.
(59) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.
(60) Kahan-Hanum, M.; Douek, Y.; Adar, R.; Shapiro, E. *Sci Rep* **2013**, *3, 1535*.
(61) Fokina, A. A.; Meschaninova, M. I.; Durfort, T.; Venyaminova, A. G.; Francois, J. C. *Biochemistry* **2012**, *51*, 2181.
(62) Young, D. D.; Lively, M. O.; Deiters, A. *J Am Chem Soc* **2010**, *132*, 6183.
(63) Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; Chapman & Hall / CRC, 2007.
(64) Zhao, W.; Lam, J. C.; Chiuman, W.; Brook, M. A.; Li, Y. *Small* **2008**, *4*, 810.
(65) Liu, J.; Lu, Y. *J Fluoresc* **2004**, *14*, 343.
(66) Liu, J.; Lu, Y. *J Am Chem Soc* **2004**, *126*, 12298.
(67) Liu, J.; Lu, Y. *Anal Chem* **2004**, *76*, 1627.
(68) Zhao, W.; Ali, M. M.; Aguirre, S. D.; Brook, M. A.; Li, Y. *Anal Chem* **2008**, *80*, 8431.

Chapter 5.
Design Principles of DNAzyme Cascading

5.1 Introduction

The innate ability of DNAzymes to catalyze a variety of reactions[1-4] makes them an attractive candidate to make molecular-scale decisions about the biochemical environment[5]; as such, use of DNAzymes for molecular logic has been well represented in the scientific literature[6-16]. Despite a few notable exceptions[17,18], the RNA-cleaving family of DNAzymes remains the most widely used and best characterized, due to their small size, rapid turnover rate, and potential therapeutic interest[16,19]. While these DNAzymes make excellent individual logic gates, their catalytic mechanism hinders the development of more complex circuits that require the interaction of multiple logical elements.

For multiple RNA-cleaving DNAzyme logic gates to interact, the cleavage of one DNAzyme (referred to as the upstream position) must result in the modulation of activity of a second DNAzyme (downstream position). However, this is a difficult engineering challenge, as the cleavage of a DNAzyme substrate results in the formation of two shorter product strands. Thus, the longer, pre-cleaved substrate must be prevented from interacting with the downstream DNAzyme, while the shorter, cleaved product must rapidly initiate this reaction. While several groups have sought to address this problem, they require high input concentrations, additional strands (which increase circuit complexity), or are unable to further propagate a logic signal, each of which are undesirable for use in biodetection assays[13,20,21].

The approach to implementing a DNAzyme cascade was detailed in **Chapter 4**. Briefly, we used the 8-17 DNAzyme[22-24], regulated by DNAzyme displacement[25], as the logical element. To connect two of these gates, we rationally designed a structured chimeric substrate (SCS) to act as a signal intermediary between the two DNAzymes. This chapter focuses on the engineering challenges and rational design approach employed to construct an efficient, modular SCS design for DNAzyme displacement gates.

5.2 Results

5.2.1 Definition of Initial Objectives and Constraints

The first step in DNAzyme communication was to outline the desired objectives, identify associated constraints imposed by these objectives, and propose structural features that could theoretically satisfy these constraints. The overarching objective was to create a mechanism by which DNAzyme-based logic gates could be easily scaled into complex decision networks. The performance of these networks must also be suitable for biodetection; this required circuits to operate at equimolar component concentrations and to be amenable to low input and gate concentrations (nM, pM) with little to no loss in circuit performance. This prevented a reliance on concentration effects to force binding interactions, instead forcing the design of the system to handle these interactions. The design of a system capable of executing these two main objectives required a careful balance of interacting components; therefore we decided to create a separate molecule capable of transmitting information between one DNAzyme and another. While other systems were amenable to

direct modification of the DNAzyme gate structure, this would make the task of scaling up much more difficult[6,7,9]. However, by designing each interaction separately, the identification of a mechanism that achieved productive interaction should lend itself to rapid scaling, so long as each interaction satisfied the same basic structural and mechanistic constraints. Thus, we removed the emphasis on specific sequences and thermodynamic stability, increasing the potential interactions that could be successfully employed by this regime.

After we designed the DNAzyme displacement method of regulating DNAzyme activity to be compatible with these larger objectives (detailed in **Chapter 3**), we identified the specific objectives necessary for the communication of two DNAzyme displacement gates. As the inhibitor of the downstream DNAzyme gate is removed by a complementary activator sequence through strand displacement, we decided to sequester the activator sequence using the secondary structure of the intermediary. This intermediary molecule also had to contain an RNA base and a complementary sequence to the substrate binding arms of an upstream DNAzyme. Binding and cleavage by the upstream DNAzyme to our intermediary molecule then had to result in a structural change that released the activator sequence, making it available to now bind the downstream gate. Hence, we named this intermediary a structured chimeric substrate (SCS), reflecting the nature and function of this molecule.

The first two objectives for the SCS design were mechanistic: an upstream DNAzyme must cleave the SCS and this cleavage product must activate a downstream DNAzyme. The third objective was to optimize the kinetic rates

corresponding to these mechanistic objectives with regards the interaction with the downstream gate: the pre-cleaved SCS should bind at a low kinetic rate (leakage), while the post-cleaved SCS product should bind at a high kinetic rate (activation). These rates are determined by the relative thermodynamic stability of the hybridization interactions. Before SCS cleavage, the retention of the secondary structure of the SCS via intramolecular interactions should be thermodynamically favorable. After cleavage, the interaction of the SCS product and the downstream inhibitor should be thermodynamically favorable. Therefore, the SCS structure was designed to balance the thermodynamic stability of the pre-cleaved state to minimize leakage and the post-cleavage state to maximize activation.

However, these rates are not mechanistically symmetrical (**Figure 5.1**). The rate of activation is complex, multi-step process, initiated via a binding step between the upstream DNAzyme and the SCS. After the DNAzyme is stably bound, it must then hydrolyze the RNA base. This is the rate of cleavage by the DNAzyme, which is affected by many factors, such as the type of DNAzyme used, buffer conditions, and orientation of the DNAzyme-substrate complex, as detailed in **Chapter 7**. The DNAzyme must then dissociate from the cleaved products, a rate dependent on the length of the DNAzyme binding arms. Rebinding of these sequences after dissociation is highly unlikely, due to the short product hybridization lengths and relatively low concentration in solution. Finally, the activator is now available to hybridize to the toehold of the downstream inhibitor and undergo strand displacement. After this DNAzyme is

released, there is a subsequent binding, cleavage, and product dissociation step of the FRET substrate. The rate of substrate cleavage of the activation pathway reflects the combined rate from all of these steps and can be treated as a single rate of activation.

Opposite the rate of activation is the rate of gate leakage, defined only by the relative thermodynamic stability of the SCS secondary structure and its ability to bind to the downstream inhibitor toehold. If fluctuations or imperfections in the SCS structure were to expose the toehold, the activator sequence would be able to displace the downstream inhibitor (Dz/INH), leading to the productive release of the downstream DNAzyme (Dz). Although the individual rates of each of these steps correspond to the same steps in activation, the lack of SCS cleavage means the entire sequence remains intact during this process. As the additional sequence and structure likely ensures a different rate constant than with the cleavage product, this interaction may not occur in exactly the same manner. Although binding to the toehold remains the most likely mechanism for inhibitor displacement, invasion through the core sequence from DNA breathing may also occur. The rational design process to obtain the structure that best satisfies these constraints is detailed below. This chapter characterizes major design iterations; detailed modifications of each version can be found in **Appendix 3**.

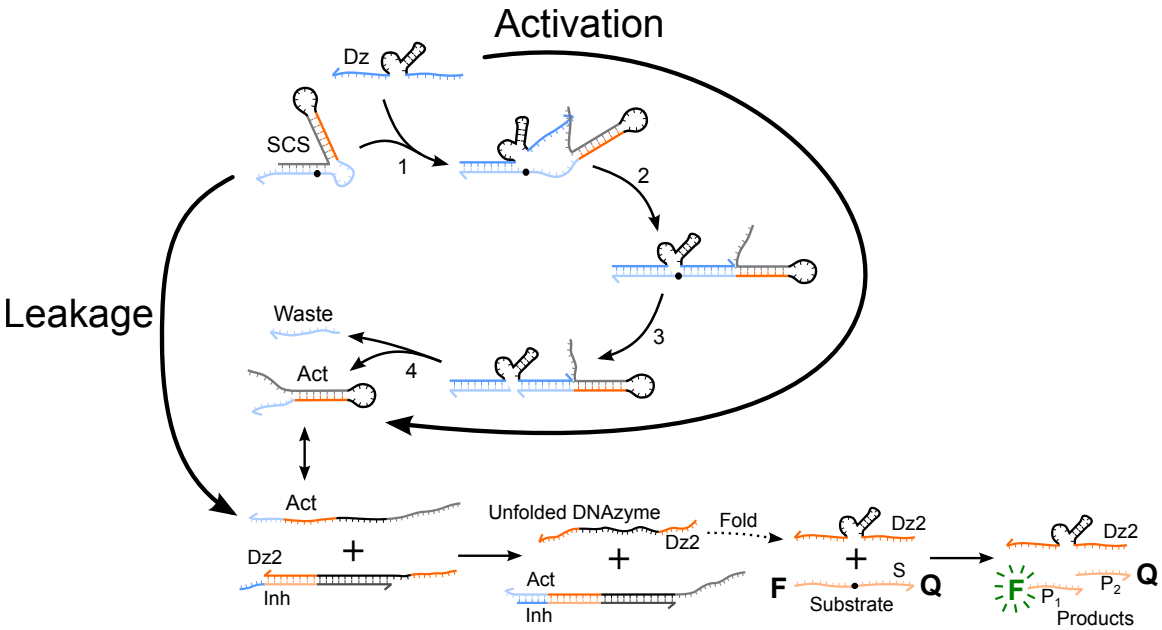

Figure 5.1 – A schematic mechanism depicting the physical processes associated with both activation (right side) and leakage (left side) rates derived from the SCS structure. Although the SCS structure is drawn as the dual stem loop in this figure (**Design 8**), reflecting the most successful structure, the actual structure varies according to each specific design. The activation rate is defined as the kinetic rate of signal activation, beginning with the addition of the upstream DNAzyme (red). The leakage rate is defined as the kinetic rate of spurious activation of the downstream gate by the SCS in the absence of the upstream DNAzyme. In both cases, the activator is sequestered in the SCS and can hybridize to the downstream inhibitor in the inactivate DNAzyme complex (Inh) which releases the downstream DNAzyme. This DNAzyme can now hybridize and cleave the reporter substrate, leading to a loss of FRET and an increase in fluorescence. The rational design process is employed in this chapter to optimize these rates, to achieve maximum activation while minimizing leakage.

5.2.2 Rational design of SCS structure for DNAzyme displacement gates

The rational design process of the SCS structure was to determine the most efficient way to sequester the activator sequence, while also enabling cleavage to release the activator. Initially, the activator sequence was 24 bases long, made up of a 5bp toehold, an 8bp substrate binding arm, and an 11 bp core sequence. The first two major design iterations used this activator sequence. Subsequent optimization of DNAzyme displacement gates reduced the necessary length of the activator to 20 bases, removing 4 bases from the core displacement sequence. As the DNAzyme displacement gates are regulated by toehold-mediated strand displacement, the availability of the activator toehold was identified as the most likely method for designing the SCS structure that satisfied all prior objectives.

The first attempt to design an SCS for communication between two DNAzyme gates used a stem loop structure, with a 26 bp loop and a 13 bp stem. Using the 24bp activator, the toehold, along with a significant portion of the activator, was sequestered in the stem. The remainder of the activator continued into the loop. The loop also contained the RNA cleavage site and substrate binding arms of the upstream DNAzyme (**Design 1, Figure 5.2A**). The postulated mechanism for this SCS structure was for the upstream DNAzyme to bind to the loop and cleave the RNA base in the loop. This would split the stem loop into two strands, which could then dissociate and diffuse away from each other. Once the strands unbound, the toehold of the activator was free to

hybridize to the downstream Dz-Inh complex and displace the inhibitor, releasing an active DNAzyme.

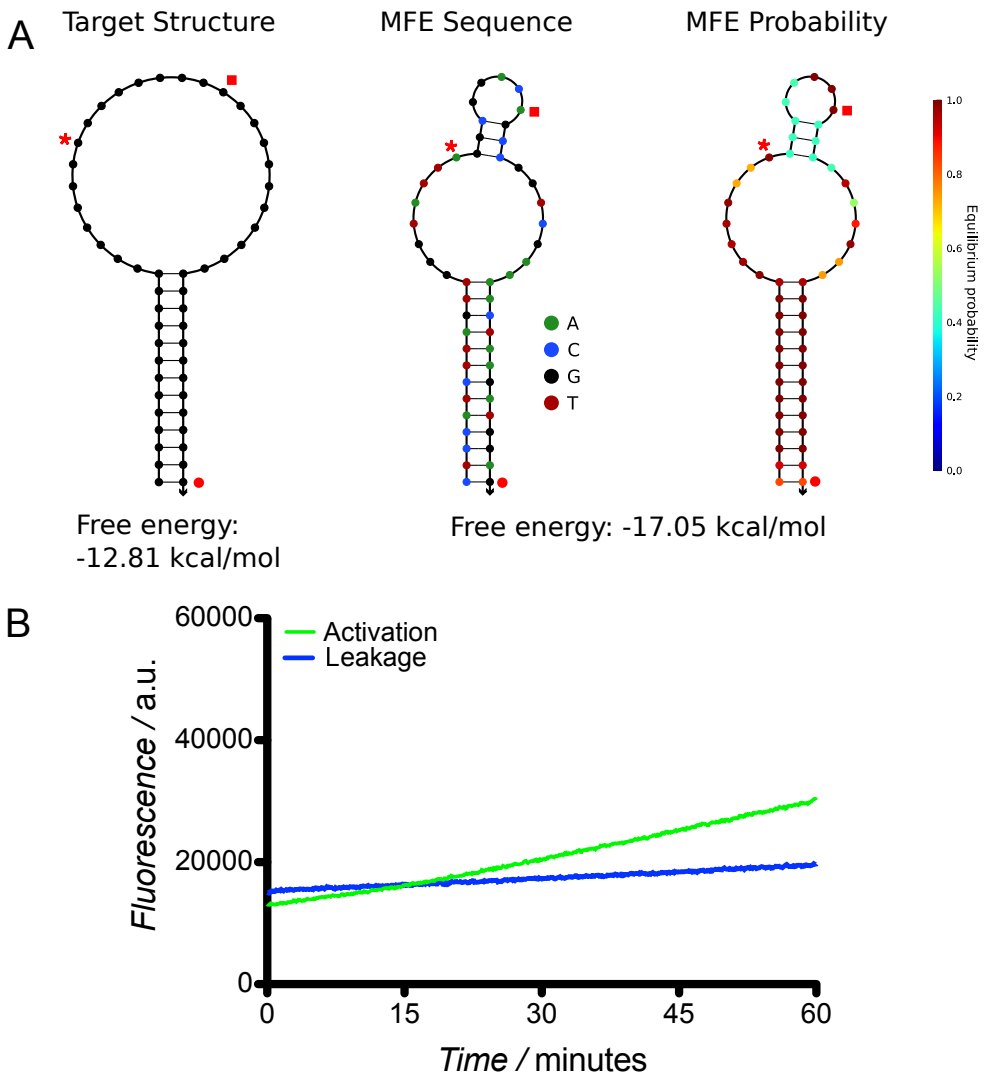

Figure 5.2 – Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a stem loop design, in which the upstream DNAzyme bound to the loop of the SCS. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 1** over 60 min. Although activation was achieved, the rate is very slow, likely due to the inefficiency of stem loop cleavage and the stability of the long stem.

Although this design did result in successful cascading (**Figure 5.2B**), the rate of activation was far too slow for practical applications and modifications yielded no appreciable improvement on cascade response (**Appendix A3.2**). We surmised this was due to slow rate of product dissociation due to the high stability of the long SCS stem length. Although the SCS structure was likely being cleaved, evidenced by the positive results, the protected toehold was not available to bind to the downstream inhibitor, as the complementary stem sequence remained bound even after cleavage. The next major design tested this by shortening the stem loop to 5 base pairs. The activator was tested in a reverse orientation in SCS structure; the enzyme binding arm and core sequence of the activator was left single-stranded, extending from the 5' side of the SCS while the toehold remained bound in the stem (**Design 2, Figure 5.3A**). Since the complementary sequence on the inhibitor is normally complexed with the downstream DNAzyme, we hypothesized that having the activator single stranded for these domains would not result in much activation, as the toehold binding is still required to initiate the reaction. However, this did not significantly improve the cascade signal (**Figure 5.3B**).

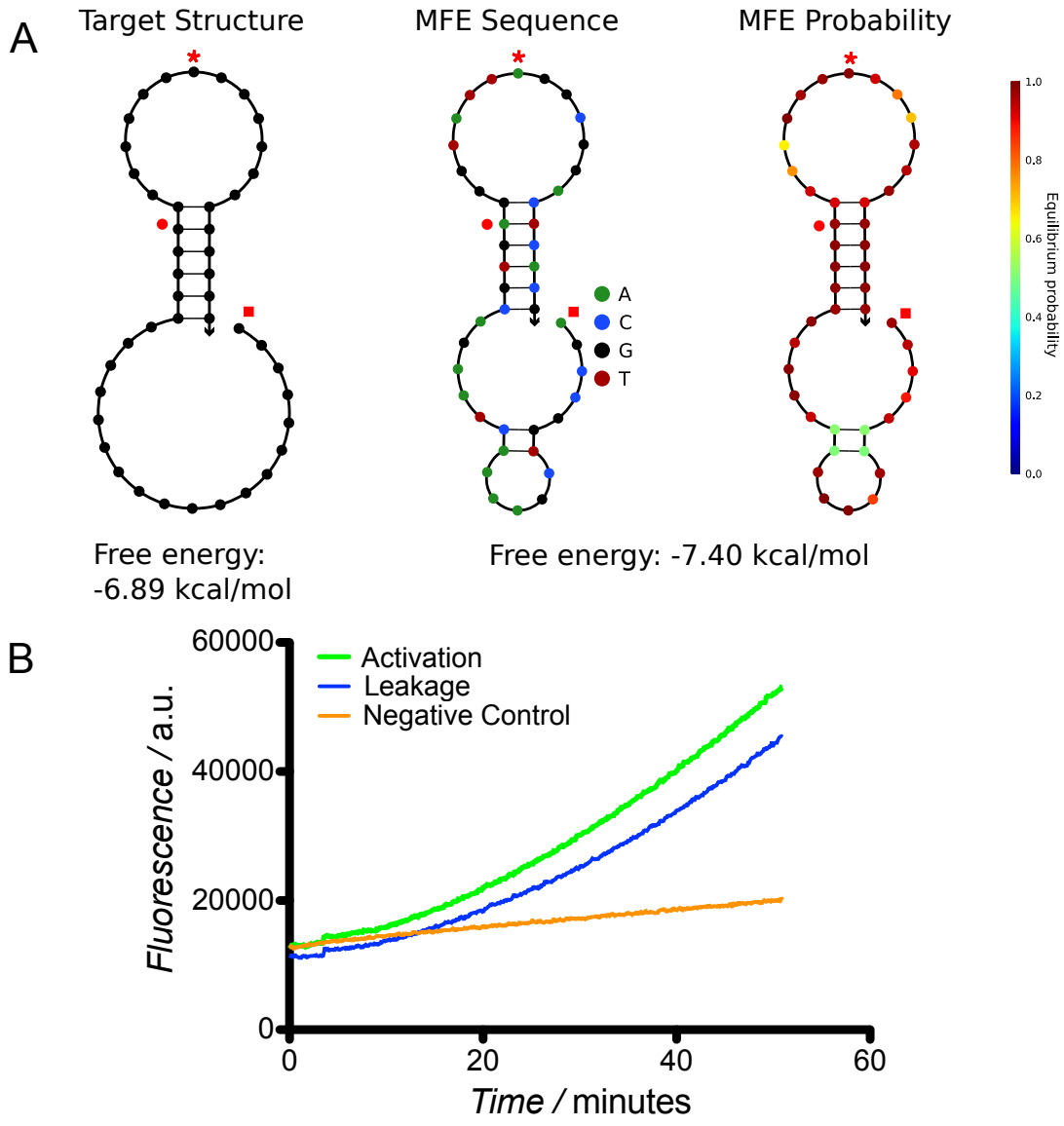

Figure 5.3 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a stem loop design, in which the upstream DNAzyme bound to the loop of the SCS. The main difference between this and the previous design is the orientation of the activation sequence, which is now at the 5' end of the SCS. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 2** over 50 min. Although activation was achieved, the rate is very slow, likely due to the inefficiency of stem loop cleavage. The shorter stem decreased SCS stability, resulting in the increase in gate leakage over time.

Due to the relatively poor performance of both the long and short stems of Designs 1 & 2, we refocused our design efforts on the location of the cleavage site. The two original designs placed the cleavage site in the middle of the loop, which required the DNAzyme to bind to a structured substrate, as opposed to the typical unstructured FRET substrate. As discussed in depth later in **Chapter 7**, the efficacy of the DNAzyme-catalyzed RNA hydrolysis reaction is highly dependent on the DNAzyme holding the RNA base in a specific conformation to rapidly facilitate the base catalysis. We hypothesized that the torsional strain of the loop likely alters this natural conformation and interaction of the RNA base. Binding of the DNAzyme directly to the loop also likely altered the DNAzyme's ability to properly orient the RNA base. Thus, we redesigned the SCS structure to utilize the principle of strand displacement to release the activator (**Design 3, Figure 5.4A**). Here, the 3' binding arm of the upstream DNAzyme hybridizes to a 5' toehold extending from the SCS stem. The substrate binding arm now acts as an invasion strand, displacing the stem loop. The loop then opens and the second binding arm is able to bind to its complementary sequence in the loop, creating a linear substrate properly oriented for RNA cleavage. The cleavage site, now much closer to the stem, can now be efficiently cleaved.

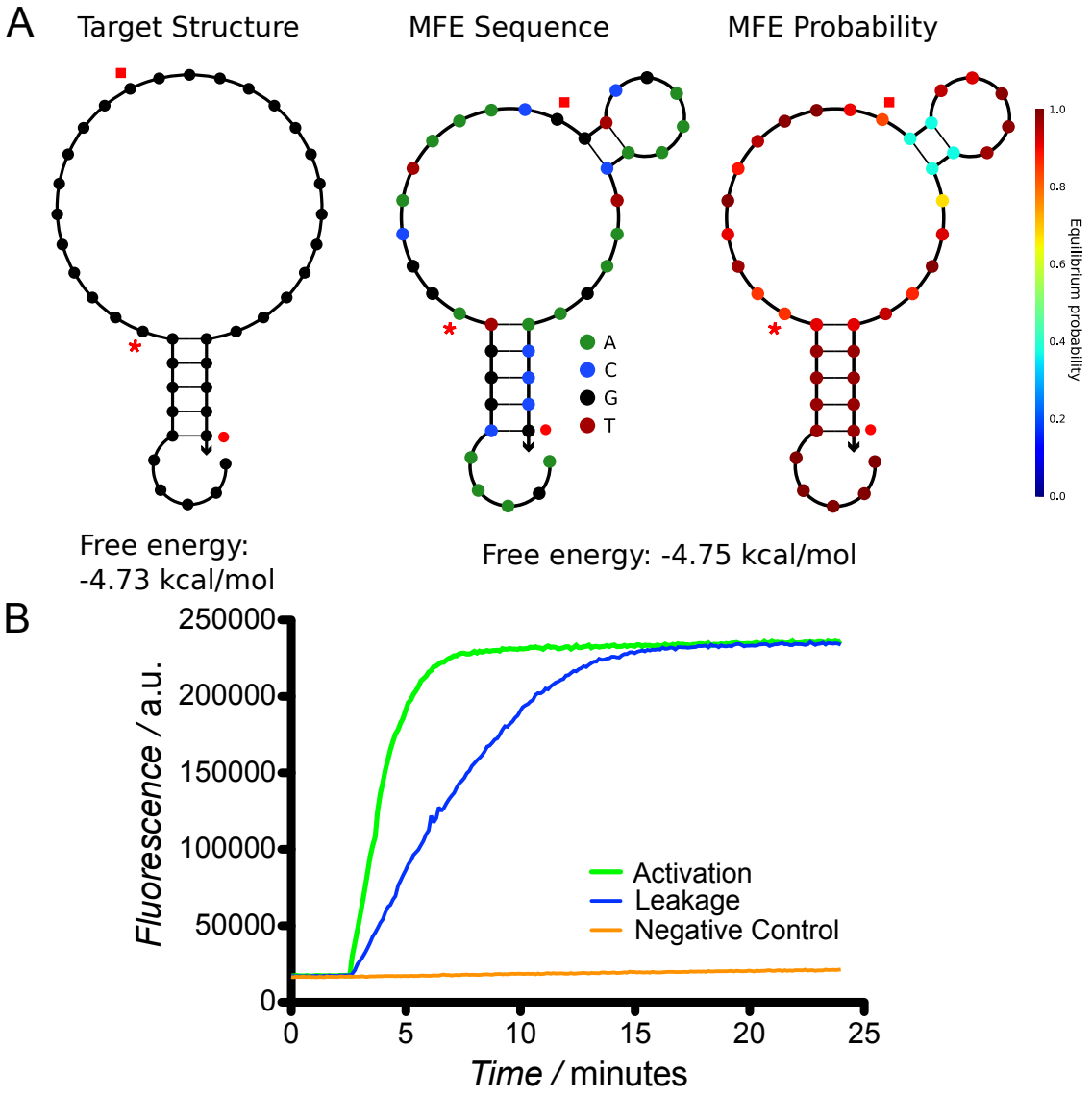

Figure 5.4 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a stem loop design, in which the upstream DNAzyme displaced the stem through hybridization and the activator is located on the 3' side of the SCS. The main difference between this and the previous design is the use of strand displacement for binding of the upstream DNAzyme to its complementary substrate sequence. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 3** over 25 min. Although rapid activation was achieved, the rate of leakage also increased dramatically over previous versions, indicating the protection of the toehold was insufficient. This is likely due to the short stem and large loop.

This design, however, now introduces some sequence constraints into the system, through overlap between the upstream and downstream gates. The stem sequence now performs two functions: it protects the downstream toehold and serves as the binding arm displacement sequence for the upstream enzyme. Thus, these two domains must contain the same sequence, which restricts the design of this SCS structure. Overall, this was viewed as a small but necessary restriction. As the downstream toehold and substrate binding arms are normally free to vary, this was determined to be an acceptable constraint as such sequences pose little restriction on the system as a whole.

This design also resulted in successful cascading. However, the response of this system was completely different from the original designs (**Figure 5.4B**). The system activated extremely quickly compared to **Design 1**, the leakage of the system was also extremely high. Clearly, this new structure provided a significant rate enhancement of downstream gate activation, demonstrating the activator could indeed be quickly released. The high rate of leakage indicated this SCS structure design was insufficient to stably sequester the activator from the downstream toehold. We surmised this was influenced both by the length of the stem and the size of the loop. A 5bp stem is likely affected by DNA breathing and a large stem (25 bp for **Design 3**) may make it difficult for SCS to maintain its structure through intramolecular binding by increasing the spacing between hybridization sequences thus reducing the probability of the ends binding and interacting. Taken together, it is unsurprising the activator was easily deprotected even in the absence of upstream cleavage, as these results are in line with the

findings of Bonnet *et al*.[26] In their paper, they demonstrated that the rate of stem closure is highly dependent on loop size, in which bigger loops close more slowly than smaller loops. This has important implications, as large stem loops that spontaneously dissociate are likely to remain in an open conformation for a significantly longer time, increasing the probability of serendipitous hybridization to the downstream toehold.

Increasing the stem size to 7bp appeared to marginally slow leakage while maintaining rapid activation (**Design 4**, **Figure 5.5**), but the use of these longer stems from the previous 5bp length was incompatible with the upstream enzyme strand displacement mechanism necessary to achieve a linear substrate. The effects of DNAzyme binding arm length on enzyme activity has been well characterized,[27] and the optimal length for rapid product dissociation was determined to be 8bp, which imposed an additional constraint on stem loop design. Had the DNAzyme arms been allowed to extend further, we could have potentially extended the stem to stabilize the structure, which would have been easily displaced by the extended binding arms. However, an 8bp arm limit ensures that binding, cleavage, and product dissociation can occur at optimal rates, which is essential to achieve multiple turnover. The ability to obtain multiple turnover is critical, as it is the main advantage of using DNAzymes for such reactions.

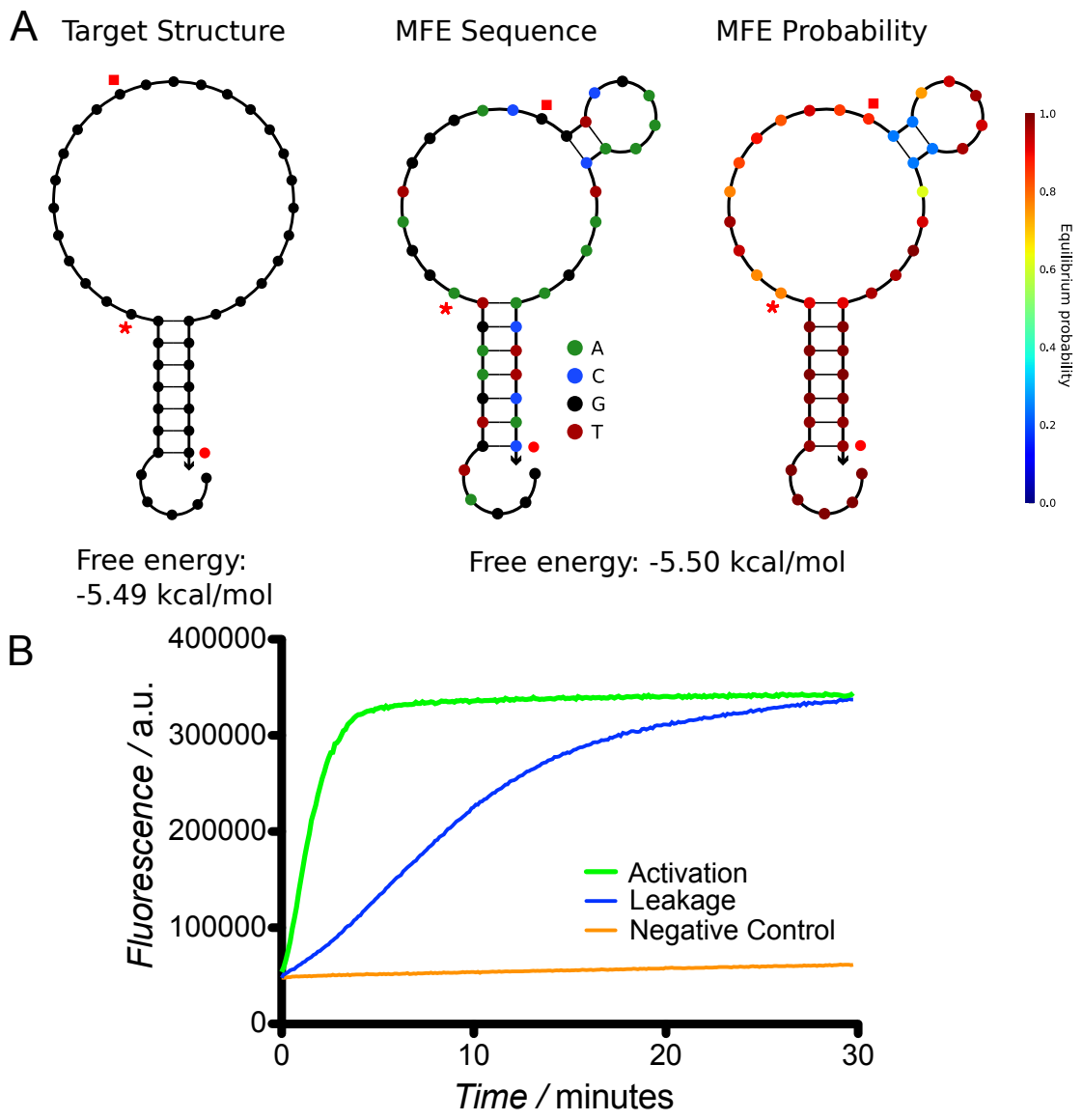

Figure 5.5 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a stem loop design, in which the upstream DNAzyme displaced the stem through hybridization and the activator is located on the 3' side of the SCS. The main difference between this and the previous design is the elongation of the stem and subsequent extension of the upstream DNAzyme binding arms. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 4** over 30 min. Although rapid activation was achieved, the rate of leakage was still very high, indicating the protection of the toehold remained insufficient. This is likely due to the short stem and large loop.

As the first two designs excelled at activator sequestration and the second two designs excelled at activator release via cleavage, it was evident that a productive design was indeed achievable, but a matter of determining the right structure to balance each of the complex rates making up the reaction. Therefore, we decided to test a hybrid structure for **Design 5** (**Figure 5.6A**), which combined the strand displacement and linear substrate alignment mechanism of **Design 3** with the reversed activator orientation of **Design 2**. An earlier variant of this structure SCS-D5v2 was unsuccessful (**Appendix 3.1.5**), as a single 5bp stem was deemed insufficient to prevent leakage. This stem length is likely insufficient to properly retain the SCS structure for extended periods of time. With such little overall structure, any isoforms would likely result in incomplete sequestration of the toehold, and therefore would result in activation of the downstream gate.

Positioning of the activator as a single-stranded overhang enabled the requisite size of the loop to be significantly reduced. This increases the probability for stem rehybridization after spontaneous dissociation due to DNA breathing and other thermodynamic effects. As short stems were insufficient to properly sequester the activator, **Design 5** increased the overall stability of the structure by using additional hybridization to reduce loop size and increase free energy. The cleavage site was left unhybridized, creating a 2bp bubble, resulting in a dual stem and loop structure. This ensured the retention of 5bp stems, beneficial for rapid activation, while augmenting the structure with a second short stem to increase overall structure rigidity. The separation into two stems would

likely keep the structure intact through avidity interactions, as the degradation of the structure would only occur after two separate stem dissociation events, the first initiated at the toehold and the second initiated in the inner loop. After cleavage, only the first stem dissociates, which releases the toehold domain, while the second one can refold on itself. As this stem does not participate in the downstream interactions, this is a desirable result. However, the DNA breathing of each stem may also result in a faster displacement, as well as the increase of the loop. In this design, the 5' arm of upstream DNAzyme binds to the 3' toehold of the SCS, initiating displacement of the outer stem. The 3' arm binds to the inner loop, and displaces the inner stem (**Figure 5.6B**). Although significantly more stable than previously designs, this design only moderately improved upon gate leakage.

To further increase the stability of the structure to suppress gate leakage, we removed the 2bp bubble in **Design 6** so the cleavage site was also hybridized. The loop size was decreased to a minimal 4bp, creating a single long stem with a very short loop (**Figure 5.7A**). After binding to the toehold and partially displacing the stem, the second arm is able to bind to the loop of the SCS and displace the rest of the stem, with the cleavage site in the middle of the stem. Although this design performed exceedingly well (**Figure 5.7B**), it was limited in that it required the use of 10bp upstream DNAzyme binding arms. This design would present difficulties for scaling up circuit complexity, and 10bp binding arms also slow the rate of product release and therefore enzyme turnover.

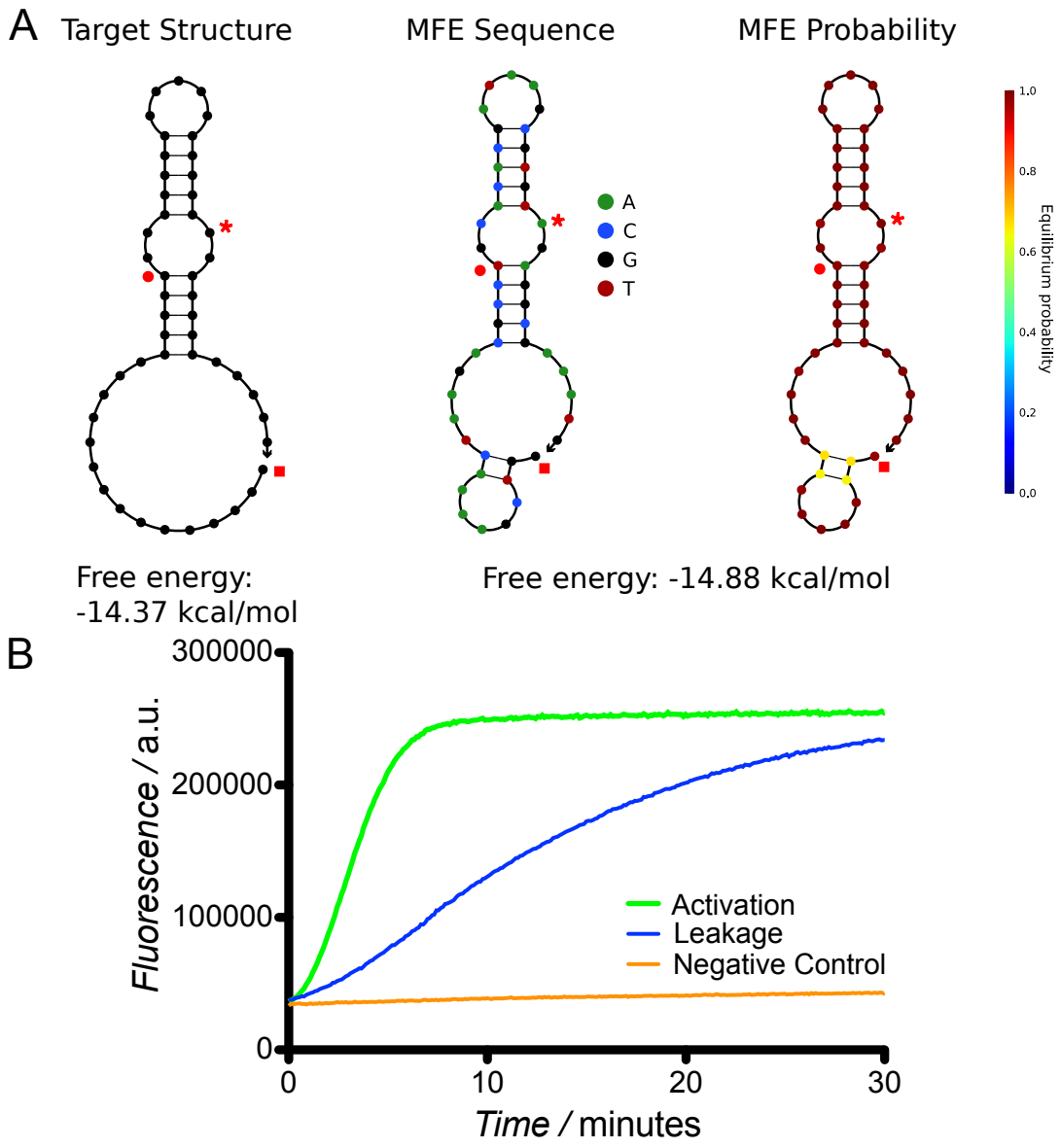

Figure 5.6 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a dual stem loop design, in which the upstream DNAzyme displaced the stem through hybridization and the activator is located on the 5' side of the SCS. The main difference between this and the previous design is the protection of the inner loop with a second stem. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 5** over 30 min. This design slightly decreased both the activation and leakage rates from the two previous designs, but there was still insufficient distinction between the two traces.

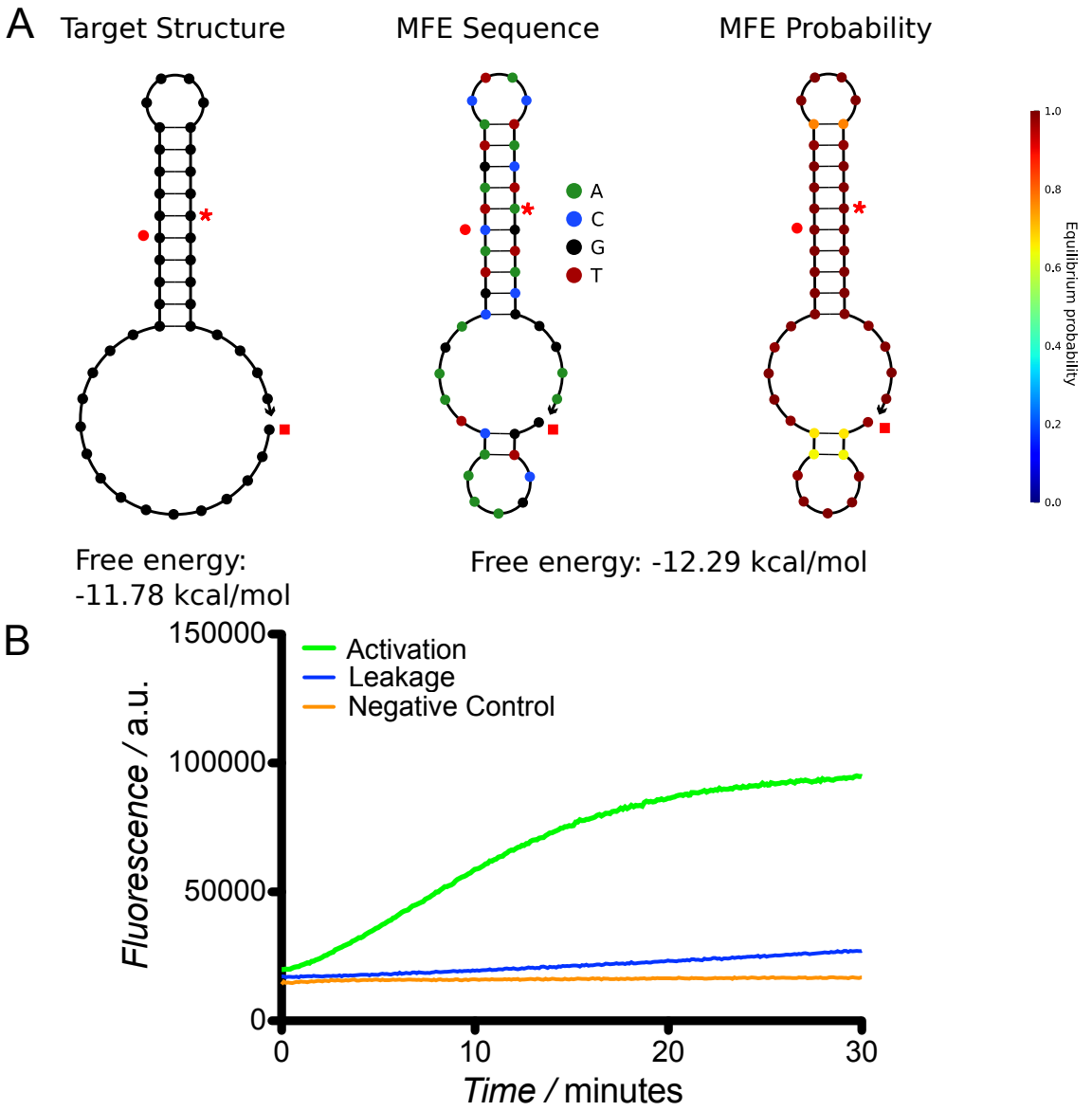

Figure 5.7 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a single stem loop design, in which the upstream DNAzyme displaced the stem through hybridization and the activator is located on the 5' side of the SCS. The main difference between this and the previous design is the removal of the 2bp bubble at the cleavage site, instead ensuring that sequence was also hybridized. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 6** over 30 min using DNA only strands. This design was among the best designs in terms of kinetic rate differences between activation and leakage, although certain design constraints made scaling up somewhat problematic.

Design 7 moved the activator sequence back to the 3' side of the SCS, ensuring the activator sequence was now bound back into a loop. In this design, each of the stems were 5bp, which formed two loops – an inner loop which contained the activator sequence, and an outer loop, which separated the two stems (**Figure 5.8A**). The cleavage site was placed in the middle of the outer stem. Here, the upstream substrate binding arm hybridizes to the toehold on the 3' side of the stem loop and initiates strand displacement of the outer stem. The other substrate arm binds the outer loop and displaces through the inner stem. Cleavage renders the outer stem as a waste product, while the inner stem containing the toehold remains intact. This design relies on the relative instability of the inner stem and loop, so that after the cleavage and dissociation of the outer stem, the inner stem will still activate the downstream gate, despite the toehold theoretically being protected in the stem loop.

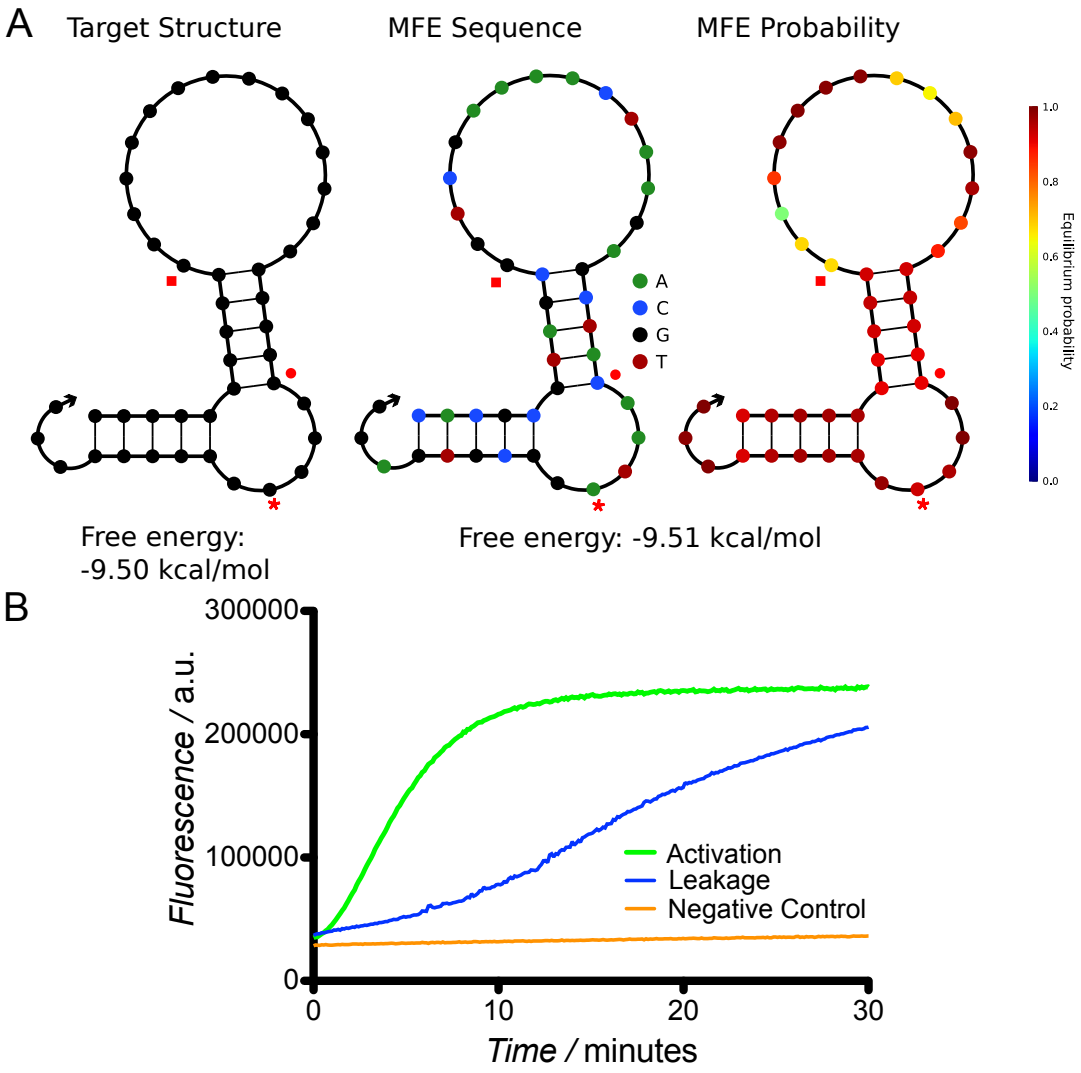

Figure 5.8 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a dual stem loop design, in which the upstream DNAzyme displaced the stem through hybridization and the activator is located on the 3' side of the SCS. The main difference between this and the previous design was the orientation of the activator, which reinstated the dual stem and loop design. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 7** over 30 min. This design had a good activation rate and a much decreased leakage rate compared to **Designs 3-5**, but further optimization was still required.

Once again, although this design yielded moderate improvement of leakage (**Figure 5.8B**), further optimization was required. We tested several variants of this design, including alterations to the inner and outer stem length, outer loop, upstream toehold, and upstream DNAzyme substrate binding arms. Details of these potential variants and their accompanying results can be found in **Appendix 3.3.1,** listed as variant designs of **Design 8**. This design (**Figure 5.9A)** was a further derivative of **Design 7**, optimizing the structure to maximize many of the kinetic rates listed in **Figure 5.1**. One of the main iterations, **Design 8v1**, was tested as in its DNA form as well as its RNA form (**Figure 5.9B**), a process discussed in detail in **Section 5.2.3**. Although the rate of activation was somewhat slow after 30 min, this design retained good signal-to-background over 120 min (**Figure 5.9C**). We examined the effect of substrate binding arm length; shorter arms resulted in a significantly increased activation rate, consistent with the findings of other groups on the optimal binding arm length for maximal product release rate. Therefore, we targeted the final optimization to ensure each DNAzyme used 8bp binding arms.

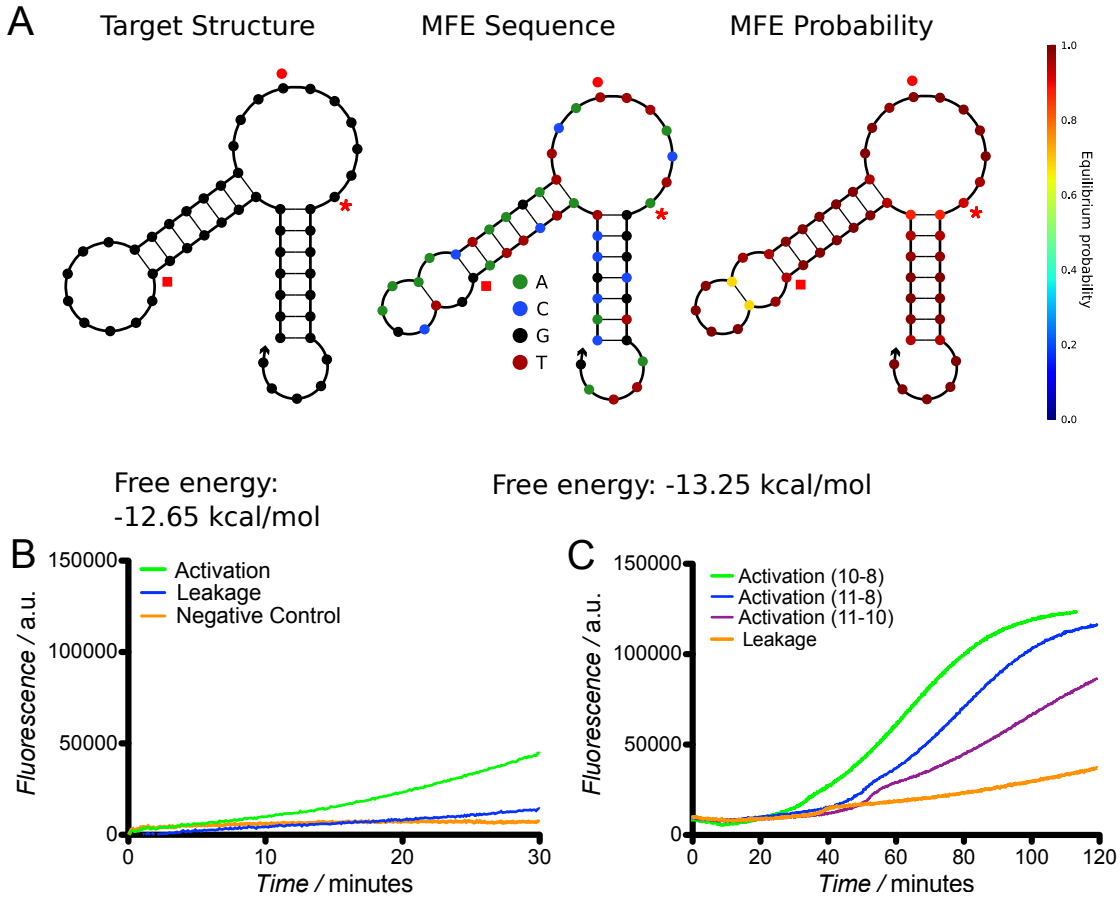

Figure 5.9 - Representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a dual stem loop design, an optimized version of the previous design. The main difference between this and the previous design was the length of each stem, which also resulted in longer DNAzyme substrate binding arms. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 8v1** over 30 min using DNA only strands. This design significantly decreased the leakage rate at the expense of activation. (C) Optimization of the upstream DNAzyme substrate binding arms (RNA SCS), demonstrating the effect on activation rate, likely through the rate of product dissociation.

The final structure and sequence of **Design 8** is depicted in **Figure 5.10A**, and had several modifications from the other variants. Most notable was the placement of the cleavage site into the outer stem similar to **Design 6**. This had several notable effects. First, the position of the cleavage site fixed the DNAzyme binding arms at 8bp each, which is critical for rapid product dissociation, as previously discussed. Second, this enabled the size of the outer loop to be minimized, ensuring high structural stability to reduce leakage. Third, this also enabled the use of the longer 7bp stems, which were found in this design to be optimal to stabilize the structure to ensure low leakage. Lastly, the hybridization of the cleavage site in the outer stem also served to protect the RNA base from degradation, which could have also been a contributing factor to gate leakage. Using a DNA form of a pre- and post-cleaved SCS molecules offered a good approximation of the performance of the RNA form and upstream DNAzyme cleavage, as seen in **Figures 5.10B** and **5.10C**. Thus, the design of an SCS structure to facilitate DNAzyme communication through a two layer cascade was finally achieved, producing a robust activation response in the presence of an upstream DNAzyme and minimal gate activation (leakage) in the absence.

This design also benefitted from a 3bp extension of the inhibitor further into the core while retaining the length of the sequestered activator. This extension acts as a clamp, which had been shown to reduce gate leakage by preventing blunt end stacking of the activator and inhibitor in the core region.[28] By not fully displacing the inhibitor, we achieved a much improved circuit response.

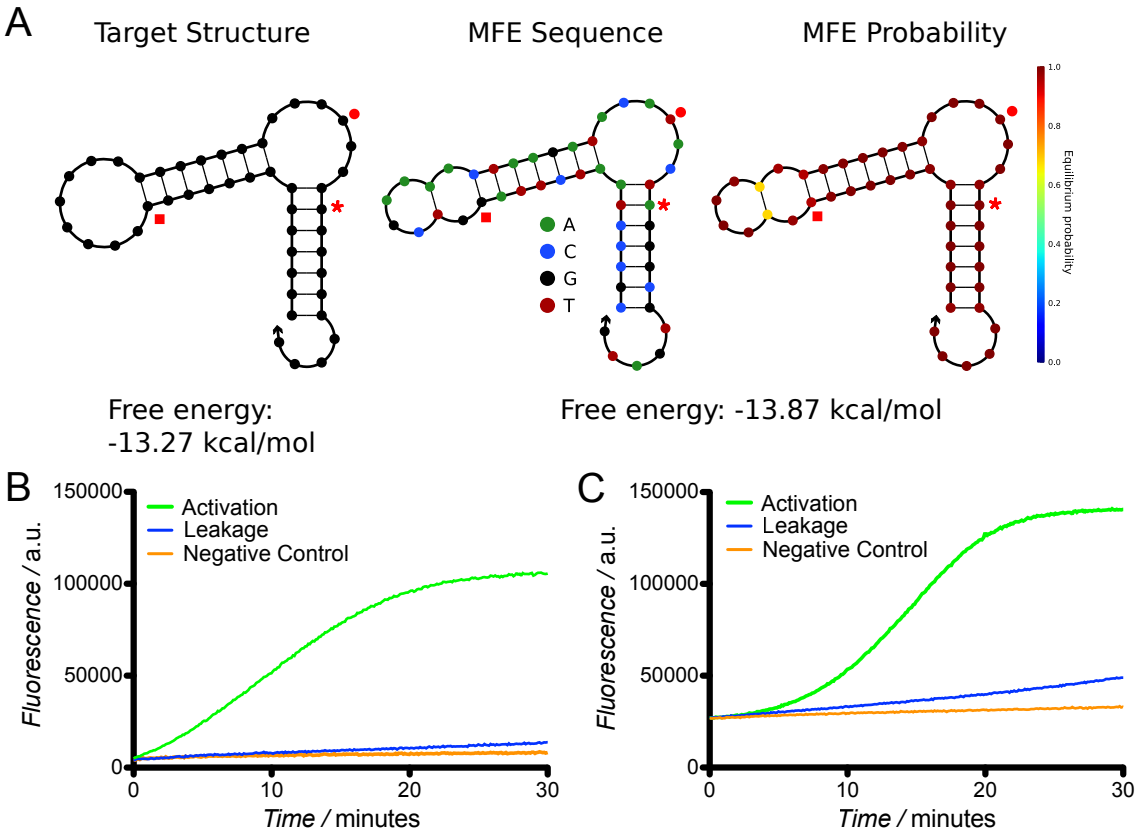

Figure 5.10 – Cartoon representation of the activation and leakage mechanisms can be found in **Figure 5.1**. (A) This SCS design was a dual stem loop design, an optimized version of the previous design, and the final design and sequence of the two layer DNAzyme cascade. The main difference between this and the previous design was the length of each stem, which was optimized by moving the RNA cleavage site into the outer stem. The target structure (left) was implemented with the given sequence (center) producing the MFE structure (right). The red circle denotes the beginning of the activator toehold; the end of the activator sequence is denoted by the red square. The red star marks the cleavage site. (B) Response of **Design 8** over 30 min using DNA only strands. This design showed a marked improvement in activation rate while retaining a low rate of leakage. (C) Implementation of the DNA analysis using the RNA form and upstream DNAzyme to active the SCS. Although the shapes of the curves are slightly different, the lag time of the RNA form likely corresponds to the rate of UE binding and cleavage, while the DNA form provides activator directly to the system. The similarity between the two graphs demonstrates the cost-effectiveness of this approach; approximation of RNA cleavage by a "pre-cleaved" DNA form provides a reliable qualitative assessment of the performance of the circuit.

5.2.3 NUPACK Coding and Structural Predications

The primary motivation for utilizing an SCS molecule for DNAzyme communication was the ability to design a single structure and replicate that structure between each two DNAzymes that must be connected in a signaling cascade. As evidenced by **Chapter 4**, this was a sound strategy, enabling the construction of a five layer cascade, with the opportunity to extend it further. However, after the minimal success of **Design 1** and the focus on structure rather than sequence, it became immediately apparent that there was a large solution set to each stem loop design. As upstream DNAzyme substrate binding arms were free to vary in sequence, there are many such base pair combinations that would allow us to achieve the desired structures. Each of these potential sequence solutions may influence the overall stability of the structure and fold into unproductive isoforms that could negatively affect the cascading reaction. Hence, we used the Nucleic Acids Package (NUPACK) folding program[29,30] to code for the desired structure, which would then provide sequences that would satisfy the constraints. Codes for each of the stem loop variants can be found in **Appendix 3.2**.

To program NUPACK to solve individual SCS structures, we input several sequences, including the downstream DNAzyme, the downstream inhibitor, the SCS stem loop, the SCS activator, and the upstream DNAzyme. The SCS stem loop is the desired SCS structure, whereas the SCS activator is the sequence available to the system after SCS cleavage and product dissociation. We found it was critical to specify both structures to ensure that both folded properly.

Because the SCS for Design 8 folded into a dual stem and loop, the cleavage product of this would have two single-stranded domains from both the 5' and the 3' side. Specifying the structure of the SCS activator structure post-cleavage ensured a constraint was placed on the sequence so that these overhangs would not hybridize to each other. As the 3' domain contains the sequence complementary to the downstream toehold, any hybridization to the 5' sequence would block this toehold from binding, drastically slowing the reaction. This can be observed in the structural optimization figure from **Appendix 2.5**.

Use of this activator sequence significantly sped up iteration time for each stem loop design while also significantly decreasing cost. While early designs were purchased directly as a DNA structure with RNA cleavage site, necessitating a surcharge for RNAse-free purification nearly three times the cost of the strand itself, later iterations and designs were tested using a DNA only form of both the SCS and ACT (SCS post-cleavage). In this way, we were able to assess the hypothetical minimum and maximum circuit response, where gate response in the presence of the SCS would be a measure of leakage, and the ACT would be the positive response, assuming all SCS molecules are cleaved. This was a reasonably good approximation of SCS behavior, and helped rule out nonviable structures before they were purchased with the RNA base. That it was both time and cost effective as well was highly beneficial, and this approach significantly sped up the testing time for each new stem loop.

After the initial coding of the NUPACK program, we used a custom Python script written by M. Leigh Fanning to analyze all of the potential candidate

sequences, comparing such parameters as structural metrics, activation, and leakage[31]. The calculation of activation and leakage is the relative percent of binding between the SCS or ACT to the downstream inhibitor, in the presence of the downstream DNAzyme. For a well-protected SCS structure, there will be minimal interaction with the inhibitor, as it is more favorable to bind to the downstream enzyme. For an optimal ACT structure, there will be significant interaction between the inhibitor, resulting in a high percentage of free downstream DNAzyme. These two parameters were viewed as relative percent of downstream DNAzyme released, with optimal SCS structures yielding nearly <1% DNAzyme release, while good ACT candidates yielded roughly 40-60% DNAzyme release and excellent ACT candidates yielded around 60-80%.

It is important to note here some of the potential shortcomings of this approach. First, NUPACK calculates all binding interactions at equilibrium conditions[32-34]. As upstream DNAzyme binding, cleavage, ACT folding and DNAzyme displacement are all dynamic processes, it was difficult at times to assess the relation between equilibrium predictions and *in vitro* circuit responses. Although the majority of experiments followed closely to predicted responses, deviations from expected results occurred frequently enough that to suggest there are additional complexities not accounted for by NUPACK. Second, while looking at relative binding rates between the SCS/ACT and downstream inhibitor was beneficial, they do not take into account the downstream DNAzyme and inhibitor existing in a pre-formed complex. Equilibrium predictions assume all strands are separate. This can introduce error in our predictions, as

complementary sections of the downstream DNAzyme and inhibitor are normally unavailable for binding, leaving just the toehold single-stranded. This fact may result is some of the aforementioned deviations between prediction and behavior. Third, despite proper NUPACK coding, some of the sequences generated still failed to fold into target structures. Therefore it was important to check each potential sequence set for structural integrity. Even in sequences that folded properly, certain sequences (particularly the presence of A-T pairs, especially at the ends of the stems) resulted in weakened structures or indications that isoforms could be present at significant percentages. As nature is not without a sense of irony, the most successful SCS structure for the original two-layer cascade was actually a misfolded structure from the original code, found in **Appendix 3.3.9**. Due to the excellent performance of this structure, all subsequent layers (**Appendix 3.3.10-12**) were modeled off this structure, thus requiring a slightly modified code from the original **Design 8** version.

Because of the high cost and time of synthesis of the normal FRET substrate, we fixed our output/reporter layer, adding additional layers upstream via SCS molecules. Because of this, our downstream DNAzyme sequence was always fixed while the upstream DNAzyme was free to vary. This imposes a constraint on the SCS sequence, and some DNAzyme sequences result in more viable SCS sequence candidates than others. While the optimized design of an entire cascade is highly desirable, this comes at a significant computational cost, as the number of potential interactions increases with each additional strand. For each additional layer, the upstream DNAzyme sequence from the previous layer

was used as the new downstream DNAzyme, allowing the next layer to be optimized to the DNAzyme ahead of it. While this clearly worked, given the performance of the five layer cascade, there was no optimization between DNAzymes, inhibitors, and SCS molecules of different layers. As the number of potential interactions is enormous, we focused on designing efficient interactions such that they would be the favored thermodynamic pathways. However, all inhibitors share a common DNAzyme core binding sequence. As each inhibitor is present in solution at a 25 nM excess, the addition of five enzyme ensures a significant (125 nM total) amount of free inhibitor. This may be the reason that the five layer cascade has very little leakage, even though the third and fourth layers have significantly more leakage. Further testing will be required to better understand this phenomenon, for instance, using of streptavidin-coated magnetic beads to remove excess inhibitor (described in **Chapter 6**).

5.3 Conclusions

The rational design process we followed in search of an SCS structure for the DNAzyme-based cascades sought to balance the complex kinetic rates making up the activation and leakage processes. Due to the minimal gate structure and toehold availability of the DNAzyme displacement gates, the desired kinetic rates were optimized through the thermodynamic favorability between the pre- and post- cleavage secondary structure of the SCS molecule. This approach is in contrast to the SCS design for the modular gate cascades, in which the structure is built into the gate itself, which relieved many of the design constraints on the SCS.

Overall, the design process was illuminating in several facets. First, this demonstrated the number of potential design variants available. Although many of them displayed only moderate success, the individuality of each design demonstrates that there is a highly diverse potential structure space that even small sequences can adopt, resulting in a wide variety of behaviors. Second, although several design had large structural differences, their performance was quite comparable, indicating that thermodynamic interactions can be achieved through many different pathways. This also shows the need for granularity in the design process: small alterations often resulted in large changes to the structural stability and therefore the kinetic rates of subsequent reactions. Finally, development of a five layer cascade demonstrated the success of the rational design approach for the dynamic modification of DNA nanostructures. By heavily focusing on the structure itself to execute the thermodynamic pathways, we were able to successively iterate the design to rapidly scale up the size and complexity of DNAzyme cascading interactions.

5.4 Future Directions

The success of the rational design for SCS structure, both with the DNAzyme displacement gates and the modular gates, holds significant promise for the future development and use of DNAzyme signaling cascades. The continued development of DNAzyme cascades may serve as a basis to construct synthetic enzymatic cascades as well as more complex computational architectures. This represents a significant development in the development of catalytic molecular logic devices, which can now implement serial interaction of logic gates rather

than solely parallel arrays in a reliable, predictable, and reproducible manner. This enables the integration of many input signatures into a single DNA circuit, anticipated to be one of the primary advantages over existing biomedical diagnostic devices.

The process of rational design is critical for the continued development of DNA nanostructures. Although modeling and predictive software can aid process[29,30,35-37], such software is currently incapable of predicting complex structure formation, dynamic multi-strand interaction, and deviations from traditional Michaelis-Menten enzyme kinetics using DNAzymes; typical interactions are generally modeled in a static, low salt, equilibrated biochemical environment[29,30,33,34]. Many DNA nanostructures contain dynamic interactions, such as self-assembly processes or molecular motors and walkers, and rational design may be useful when modeling software is insufficient for the prediction of such complex behaviors. Importantly, the rational design process for the SCS structure relied heavily on the modularity of the system and the matching of structural components to kinetic rates. In this way, we could optimize the structure for specific kinetic steps and identify rate-limiting steps solely with experimental characterization. Thus, this work shows that a robust design process capable of manipulating complex DNA nanostructures is feasible in the absence of appropriate *in silico* modeling software and holds promise for the integration of circuit response cascades and dynamic interactions into static DNA nanostructures such as DNA origami[38,39].

5.5 Materials and Methods

5.5.1 Materials

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Substrate molecules (DNA-RNA chimeras) were purified using RNAse-free HPLC. Sequences for all oligonucleotides used herein are presented in Table S1 in the Supporting Information. ATP was purchased from Sigma-Aldrich (St. Louis, MO). DNA strands purified using standard desalting were resuspended in RNAse-free H_2O (Sigma-Aldrich, St. Louis, MO) at a stock concentration of 50 μM. These original stocks were diluted to working stocks of 2.5 μM, by diluting 50 μL stock DNA into 950 μL assay buffer (**Section 5.5.3**). RNAse-free HPLC strands were resuspended directly at 2.5 μM in RNAse-free H_2O.

5.5.2 Gate preparation

Typically, 60 μL of DNAzyme and 75 μL inhibitor (25% excess inhibitor) of 2.5 μM working stock solutions were added together and heated together at 95 °C for 3 minutes on a heat block, and subsequently annealed by cooling to room temperature over a minimum of 90 minutes. All other strands that required an initially hybridized state, including all SCS and ACT molecules, were also annealed using the same protocol.

5.5.3 Assay conditions and instrumentation

All assays were performed at room temperature (23 °C) in a buffer of 1M NaCl, 50 mM HEPES, 1 mM $ZnCl_2$, pH 7.0. Order of strand addition was as follows: Substrate, Dz/INH, SCS or ACT, depending on the experiment. Upstream

DNAzyme (UE) was added last to SCS designs with an RNA cleavage site. Fluorescence was read on either a Spectramax M2e fluorescent plate reader (Molecular Devices, Sunnyvale, CA) in a 200 µL reaction volume or Quantamaster 40 fluorimeter (PTI, Binghamton, NJ) in a 300 µL reaction volume. Fluorescence was monitored at 492 nm excitation and 518 nm emission wavelengths.

5.6 References

(1) Baum, D. A.; Silverman, S. K. *Cell Mol Life Sci* **2008**, *65*, 2156.
(2) Silverman, S. K. *Acc Chem Res* **2009**, *42*, 1521.
(3) Silverman, S. K.; Baum, D. A. *Methods Enzymol* **2009**, *469*, 95.
(4) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.
(5) Adleman, L. M. *Science* **1994**, *266*, 1021.
(6) Lederman, H.; Macdonald, J.; Stefanovic, D.; Stojanovic, M. N. *Biochemistry* **2006**, *45*, 1194.
(7) Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N. *Nano Lett* **2006**, *6*, 2598.
(8) Pei, R.; Matamoros, E.; Liu, M.; Stefanovic, D.; Stojanovic, M. N. *Nat Nanotechnol* **2010**, *5*, 773.
(9) Stojanovic, M. N.; Mitchell, T. E.; Stefanovic, D. *J Am Chem Soc* **2002**, *124*, 3555.
(10) Stojanovic, M. N.; Nikic, D. B.; Stefanovic, D. *Journal of the Serbian Chemical Society* **2003**, *68*, 321.
(11) Stojanovic, M. N.; Stefanovic, D. *Nat Biotechnol* **2003**, *21*, 1069.
(12) Stojanovic, M. N.; Stefanovic, D. *Journal of the American Chemical Society* **2003**, *125*, 6673.
(13) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.
(14) Elbaz, J.; Moshe, M.; Shlyahovsky, B.; Willner, I. *Chemistry* **2009**, *15*, 3411.
(15) Elbaz, J.; Wang, F.; Remacle, F.; Willner, I. *Nano Lett* **2012**, *12*, 6049.
(16) Willner, I.; Shlyahovsky, B.; Zayats, M.; Willner, B. *Chem Soc Rev* **2008**, *37*, 1153.
(17) Stojanovic, M. N.; Semova, S.; Kolpashchikov, D.; Macdonald, J.; Morgan, C.; Stefanovic, D. *J Am Chem Soc* **2005**, *127*, 6914.
(18) Chen, X.; Wang, Y. F.; Liu, Q.; Zhang, Z. Z.; Fan, C. H.; He, L. *Angew Chem Int Ed Engl* **2006**, *45*, 1759.
(19) Dass, C. R.; Saravolac, E. G.; Li, Y.; Sun, L. Q. *Antisense Nucleic Acid Drug Dev* **2002**, *12*, 289.
(20) Eckhoff, G.; Codrea, V.; Ellington, A. D.; Chen, X. *J Syst Chem* **2010**, *1*, 13.
(21) Levy, M.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2003**, *100*, 6416.
(22) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.
(23) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.
(24) Schlosser, K.; Li, Y. *Chembiochem* **2010**, *11*, 866.
(25) Brown, C. W., III; Lakin, M. R.; Stefanovic, D.; Graves, S. W. *ChemBioChem* **2014**, *15*, 950.

(26) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc Natl Acad Sci U S A* **1998**, *95*, 8602.
(27) Cairns, M. J.; Hopkins, T. M.; Witherington, C.; Sun, L. Q. *Antisense Nucleic Acid Drug Dev* **2000**, *10*, 323.
(28) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.
(29) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 170.
(30) Zadeh, J. N.; Wolfe, B. R.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 439.
(31) Fanning, M. L.; Macdonald, J.; Stefanovic, D. In *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine*; ACM: 2011, p 404.
(32) Bois, J. S.; Venkataraman, S.; Choi, H. M.; Spakowitz, A. J.; Wang, Z. G.; Pierce, N. A. *Nucleic Acids Res* **2005**, *33*, 4090.
(33) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev* **2007**, *49*, 65.
(34) Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2003**, *24*, 1664.
(35) Lakin, M. R.; Parker, D.; Cardelli, L.; Kwiatkowska, M.; Phillips, A. *J R Soc Interface* **2012**, *9*, 1470.
(36) Lakin, M. R.; Youssef, S.; Cardelli, L.; Phillips, A. *J R Soc Interface* **2012**, *9*, 470.
(37) Lakin, M. R.; Youssef, S.; Polo, F.; Emmott, S.; Phillips, A. *Bioinformatics* **2011**, *27*, 3211.
(38) Woo, S.; Rothemund, P. W. *Nat Chem* **2011**, *3*, 620.
(39) Zhang, F.; Nangreave, J.; Liu, Y.; Yan, H. *Nano Lett* **2012**, *12*, 3290.

Chapter 6.
A Versatile, Modular DNA Biosensor

6.1 Introduction

Development of low cost bioassays is an important goal of research and medical diagnostics. The total cost of an assay comes from reagents, readout equipment, sample preparation, and research and development required to alter assay parameters for different targets. This makes construction of a universal biosensor a significant challenge as the number of bioassay targets can vary widely, as with the detection of multi-strain pathogen sequences or even by type (i.e. DNA, RNA, or small molecules). These factors necessitate the development of a modular, scalable biosensor that uses common protocols, reagents, and analytical output methods easily customized to detect different targets. Recent advances in DNA nanotechnology have enabled the construction of nucleic acid devices that use the predictable nature of Watson-Crick hybridization to perform complex decisions[1-4] and to detect target sequences with high specificity[5,6]. These properties, together with inherent biocompatibility and rapidly decreasing synthesis costs[7], make DNA an attractive engineering material for the development of low-cost bioassays.

Detection of multi-strain pathogens is an important problem in many areas of medical and agricultural science and represents a significant challenge for biosensor design and implementation. Many viral and bacterial pathogens occur in multiple strains, whose genetic differences can be exploited to design strain-specific tests. Depending on the application, it may be important for a test to identify a particular pathogen strain, or it may be sufficient to produce a generic

response to a range of pathogen signatures. It is also important that new sensing devices can be designed quickly and straightforwardly to detect newly discovered pathogen signatures. For example, Shiga-toxin bearing *E. coli* (STEC) is a major public health concern that has been associated with several outbreaks of gastrointestinal disease[8]. However, there are more than 100 STEC strains [9], and although O157 is the most prevalent, non-O157 STEC have also been implicated in human infections[9-11]. Simple, low-cost field tests that determine whether any of the STEC strains is present would allow contaminated meat or produce to be identified and removed from the food chain at an early stage to reduce the risk to public health are highly desirable. However, design and construction of such a biosensor is challenging. The simultaneous detection of large numbers of strains requires a highly modular design, capable of rapidly exchanging one strain detection design for another.

Here we present a versatile, modular DNA-based sensor design, expanding our previously characterized mechanism of *DNAzyme displacement*[12] **(Chapter 3)**. This combines toehold-mediated strand displacement with DNAzyme-catalyzed cleavage of substrate molecules. Toehold-mediated DNA strand displacement[13] (TMSD) is a specific and powerful DNA computing technique that has been used to implement digital logic circuits[2,14], neural networks[3], enzyme-free catalytic networks[15,16], hairpin assembly systems[17-20] and molecular walkers[21,22]. The kinetics and thermodynamics of strand displacement reactions can be controlled based on the lengths of the toeholds that nucleate the binding reactions[23]; this makes them ideal for programming detection

pathways for targets at low concentrations. DNAzymes[24] (also known as deoxyribozymes) are single-stranded DNA enzymes that can catalyze a variety of chemical reactions[25-32], in particular, cleavage of an RNA or chimeric DNA/RNA substrate molecule[33,34]. DNAzymes provide isothermal signal amplification via multiple turnover reactions. DNAzyme-based logic gates are DNAzymes whose catalytic activity is conditional on the presence (or absence) of one of more target species. Large-scale parallel DNAzyme gate arrays have been constructed for molecular decision-making[1,4,35-38]. Several alternative methods have been used to turn DNAzymes into logic gates, such as loop-based inhibition[1,4,35-38] and triggered assembly of multi-component DNAzymes[39-44]. However, these approaches require high concentration gradients to drive the forward inter-molecular reaction, to overcome the kinetically favorable intramolecular reactions in the case of loop-inhibition or the entropic penalty of assembling multiple components into an active DNAzyme. In the case of multi-component DNAzymes, this may be overcome at the cost of additional circuit complexity[45-48].

The continued use of DNAzyme displacement enables our circuits to have high input specificity, low input concentrations, and kinetically controllable reactions, while retaining a simplified gate design through the inherent catalytic ability of a DNAzyme. Although a new, more versatile and modular gate design is detailed in this chapter, it is also important to retain the principles discovered in the previous chapters. Thus, this chapter has two foci, the first being gate design and performance for use in bioassay implementation, and the second being gate

design for use in multi-layered cascades. The combination of these principles should lead to the development of more sophisticated biodetection algorithms, capable of integrating multiple inputs in large-scale decision networks.

6.2 Results

6.2.1 Modular gate design, structure, and behavior

Figure 6.1 depicts our modular sensor design, which consists of a catalytically inactive DNAzyme-inhibitor complex that can be divided into two orthogonal modules: a *detection module* (DM) and a *reporter module* (RM). When the correct target sequence is present, it binds to the DM toehold and branch migration occurs up to the beginning of the RM. This partially displaces the DNAzyme strand and opens up the RM toehold, which was previously sequestered in a 5-nucleotide (nt) bulge between the two modules. A fuel strand then binds to the free RM toehold and displaces the remainder of the DNAzyme strand. The free DNAzyme then folds into a catalytically active conformation and cleaves its complementary fluorogenic substrate, producing an increase in fluorescence due to loss of FRET. If substrate is present in excess, each DNAzyme catalyzes the cleavage of many substrate molecules in the multiple-turnover kinetic regime. This provides isothermal signal amplification in the readout module. Here, we base all designs on the 8-17 DNAzyme[34] because of its small size and high catalytic efficiency[49].

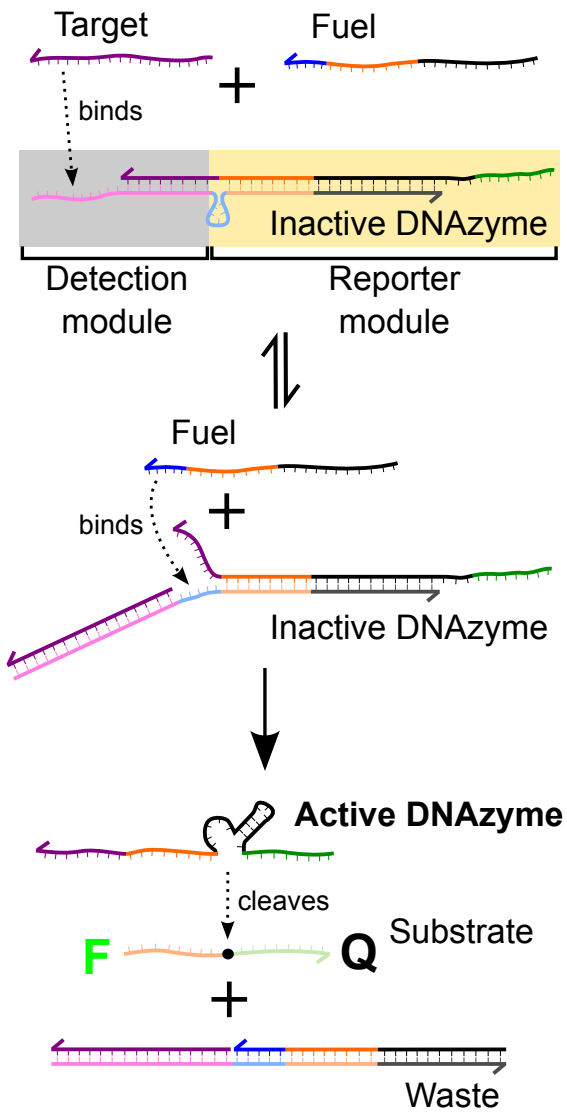

Figure 6.1 – Mechanism of the modular gate utilizing an advanced version of DNAzyme displacement. Here, an 8-17 DNAzyme is sequestered in an DNAzyme-inhibitor complex, rendering the DNAzyme inactive. This complex is sequentially separated into two orthogonal domains – a detection domain which binds the desired target sequence and a reporter domain which contains the DNAzyme and binds the fuel strand. The two domains are separated on the inhibitor by a toehold sequestered in a loop, which reduces the favorability of fuel strand binding while in complex form. Activation of the gate is a two-step process. First, the target strand binds to the external toehold of the inhibitor strand and removes it from the detection domain, which releases the looped toehold. This results in a linearization of the internal toehold, which now favorably binds the fuel strand present in solution. The fuel displaces the inhibitor from the reporter domain, which results in the release of an active DNAzyme, now able to cleave a FRET substrate.

6.2.2 Optimization of Gate Structure to Reduce Non-specific Activation

Sequestering the inner toehold in a loop reduces the thermodynamic favorability of binding between the fuel and inhibitor strands relative to the linear conformation. This constraint alone is insufficient to prevent non-specific activation of the gate; that is, the fuel is able to displace the RM allowing substrate cleavage, even in the absence of the target (**Figure 6.2A**), an effect that varies in a concentration dependent manner (**Figure A4.1**). While the rate of non-specific activation (leakage) was relatively low over short times, this signal is problematic for extended time courses. Our DNAzyme displacement AND gate contained mismatches (**Figure 6.3**) sufficient to destabilize the interaction between the DNAzyme and inhibitor; therefore, we systematically looked at the addition of mismatches to destabilize the binding of the fuel strand to the inhibitor in the absence of target, taking care to ensure that we retained activation in the presence of target. Single mismatches were introduced at the first, second, and third positions (P1, P2, P3) from the DM in the loop and compared to the response without any mismatches. Double mismatches were added in the first and second positions (P1, P2) and were tested using a 5bp loop as well as an 8bp loop (8L). (**Figure 6.2**). Experiments were run at typical gate concentrations (100 nM) over a short, 20 min time course. Compared to the control with no mismatch (**Figure 6.2A**) each addition of a mismatch showed significant improvement on leakage. The best options appeared to be a single mismatch in the second position (**Figure 6.2C**), and the double mismatches in the first and second positions, both in the five and eight base pair loops (**Figures 6.2E and**

6.2F, respectively). In each of these three cases, the leakage of the gate was the lowest over background.

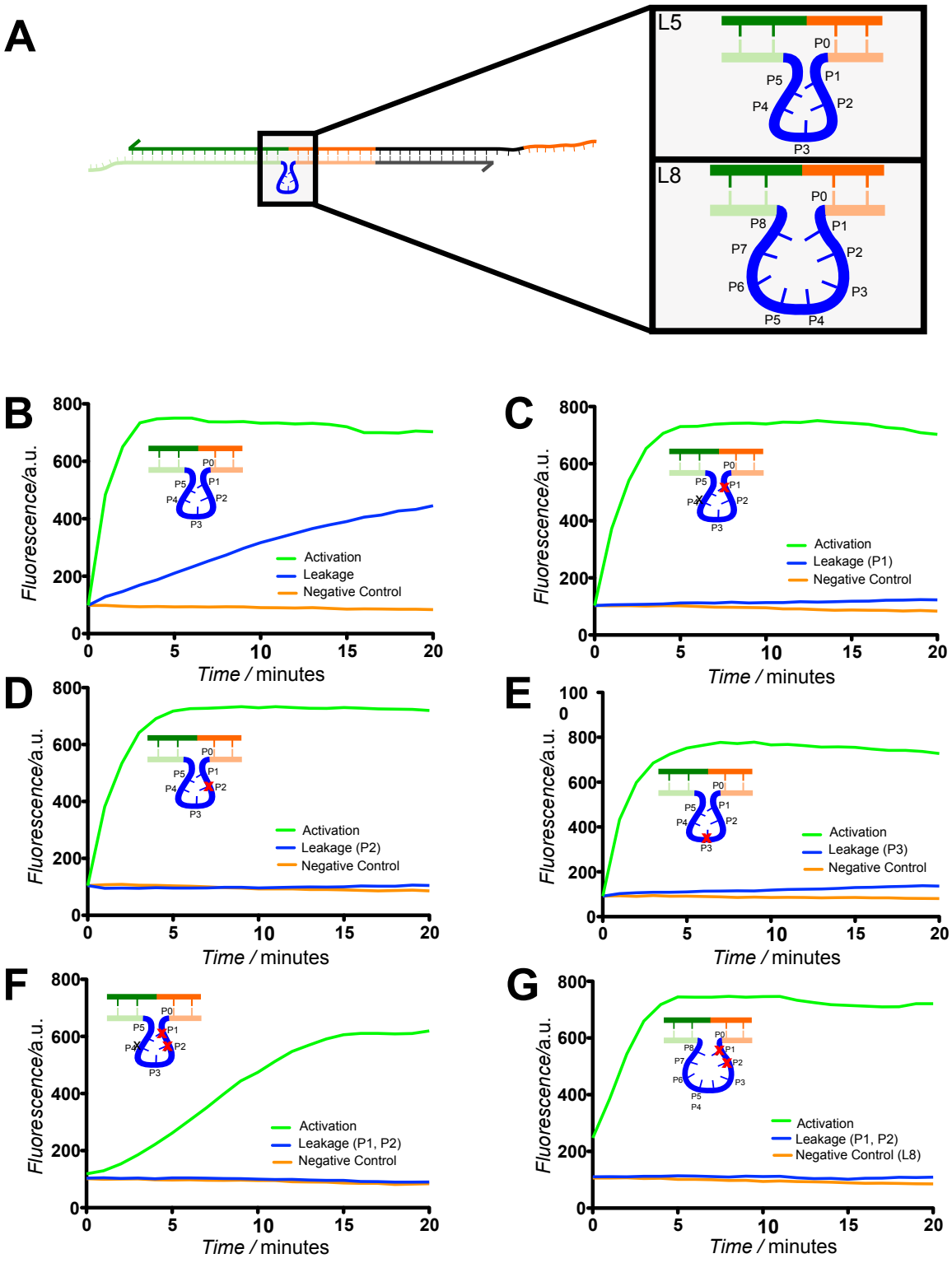

Figure 6.2 – The effect of mismatches on modular gate performance. (A) Rational mismatches were introduced to destabilize the binding of the free single-stranded fuel and the corresponding toehold on the inhibitor, sequestered in a loop. The position of the mismatch plays an important role in leakage and activation profiles. (B) Modular gate response with no mismatches results in high leakage. (C-E) The addition of a single mismatch significantly reduces this leakage. (F) The addition of two mismatches eliminates leakage, at the cost of activation rate. (G) Alternatively, activation rate can be increased through a larger loop, added more bases for the fuel to bind to the inhibitor. Insets for graphs depict the specific location for each mismatch.

However, experiments are often run on much longer time scales. Thus, it was critical to determine which mismatch profile best performed over extended time courses. To observe this effect, we ran the same gate concentration (100 nM) for 12 hours. (**Figure 6.3**). Over this time course, we found that the rate of leakage is quite significant over such extended time durations, the best performing gate being the double mismatch of the 5 bp loop. Although gate activation is very fast compared to the leakage, high leakage profiles indicate further optimization was necessary to ensure these gates were suitable for longer experiments. However, we previously found that reducing gate concentrations was sufficient for the reduction of gate leakage (**Figure A2.1)**. To see whether these gates followed a similar profile, we reduced the gate concentration to 40 nM and ran them for 12 hours. Here, we found that leakage was indeed reduced, without a significant reduction in activation. Again, the double mismatch of the 5 bp loop showed the lowest level of leakage over 12 hours, although the activation is noticeably slower than the other three gates. The reduction of concentration showed a marked reduction in leak over time, which is important for the determination of values like limit of detection (LOD), a critical parameter for real bioassay sample detection. A detailed discussion in **Section 6.2.4** will examine these parameters more closely.

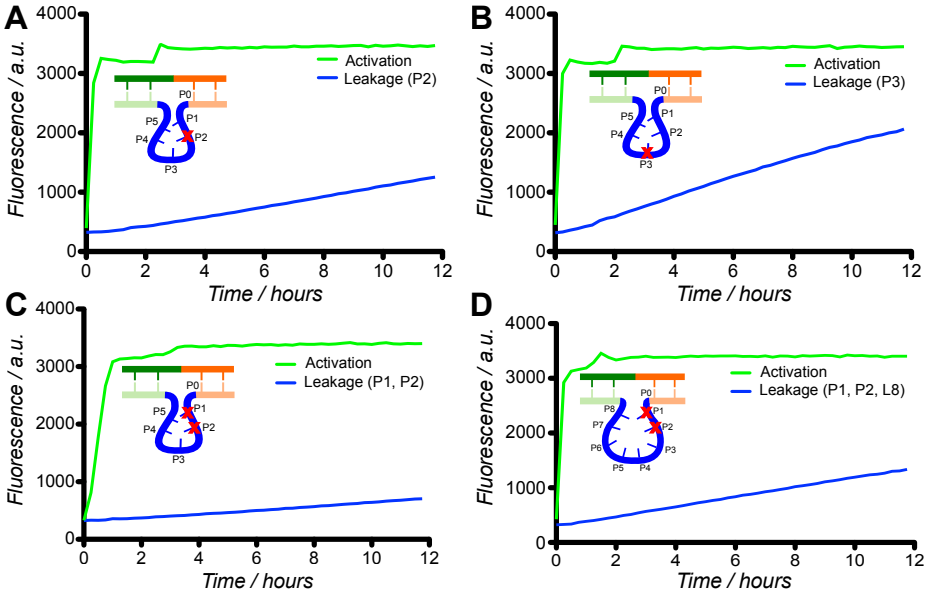

Figure 6.3 – Extended time profiles at 100 nM gate concentration, comparing the four best performing gates, labeled based on the position of their mismatch. (A) P2 gate (B) P3 gate (C) P1, P2 and (D) P1, P2, L8 gate.

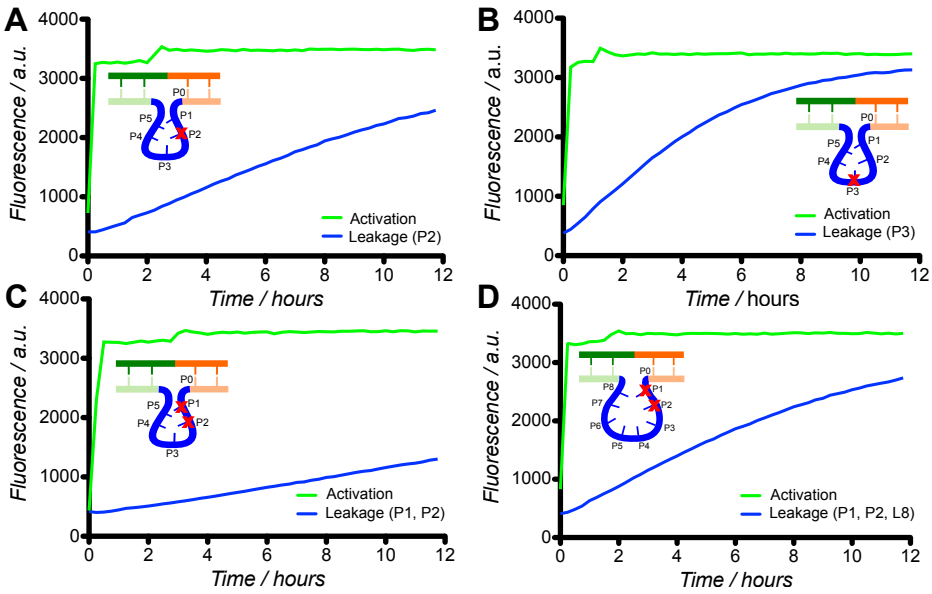

Figure 6.4 - Extended time profiles at 40 nM gate concentration, comparing the four best performing gates, labeled based on the position of their mismatch. (A) P2 gate (B) P3 gate (C) P1, P2 and (D) P1, P2, L8 gate.

As the addition of two mismatches at a reduced concentration still resulted in a robust activation signal, we explored the addition of an additional mismatch to each of the existing versions to see if even lower leakage over 12 hours could be obtained. All previous mismatched bases between the fuel and the inhibitor strand had been placed in the toehold region, which destabilized the binding interaction but likely had little effect on the displacement of the inhibitor strand. Here, the additional mismatch was placed at the last position in the double-stranded region of the DM (denoted P0), immediately adjacent to the first position of the loop. Thus, for the double mismatch versions, this third mismatch would result in three consecutive mismatches, two in the loop and one in the hybridized region of the inhibitor. With the addition of a mismatch as the P0 position, we looked at two versions of this. The first was a full mismatch, replacing an adenine with a cytosine, to mismatch with the paired thymine. We also explored a pseudo-mismatch, in which the adenine was replaced with a guanine, thus creating the opportunity for a wobble base pair with the thymine. As shown in **Figure 6.5**, the leakage was drastically reduced, even without a mismatch in the loop (**Figure 6.5A**). There were several notable observations with the addition of this new mismatch. The most striking feature is the reduction in activation. In **Figure 6.4**, all activation times using the DNA oligonucleotide (oligo) target input were on the order of minutes, with complete activation being achieved in less than one hour. As the P1 mismatch now likely slows the rate of strand displacement initiation, activation rates are now on the order of hours.

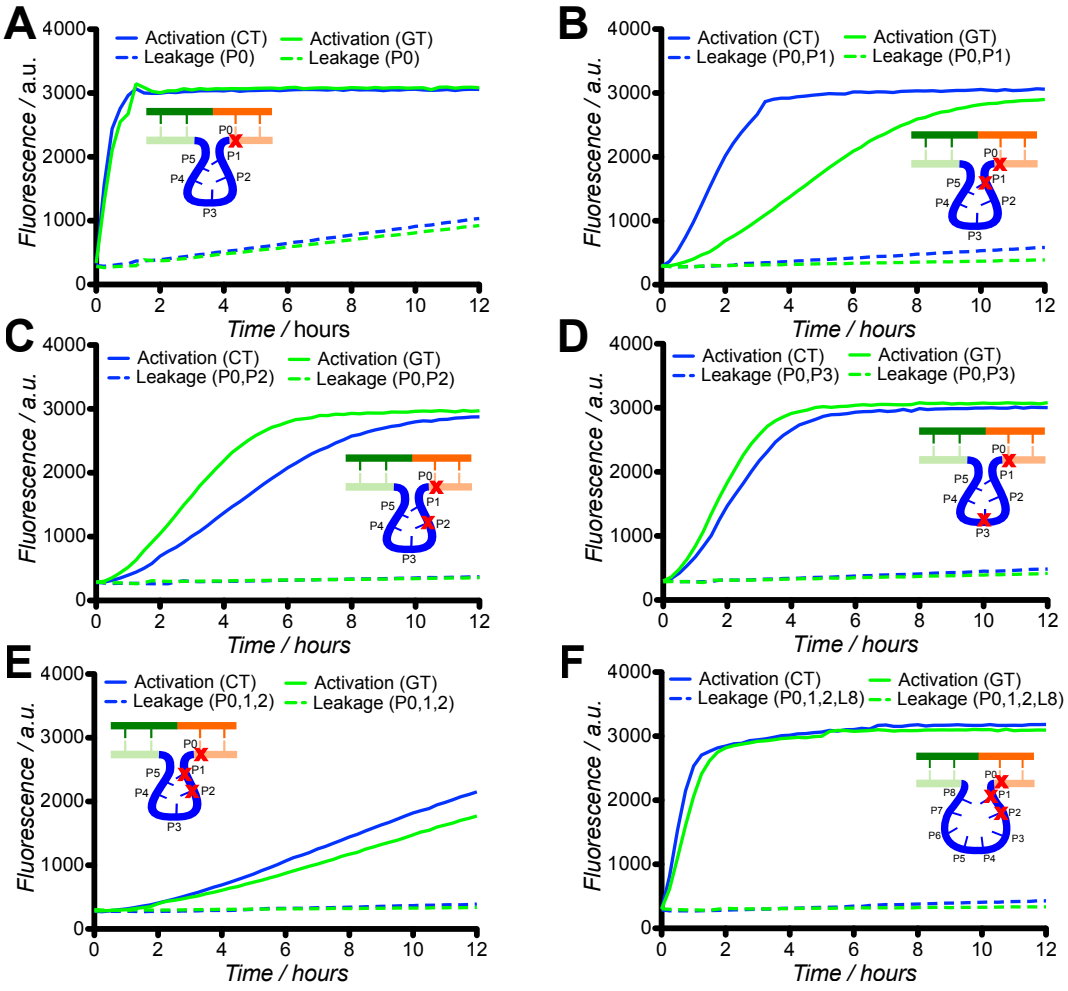

Figure 6.5 – Effect of an additional mismatch, at the P0 position. The mismatch was either a full mismatch (C-T) or a wobble mismatch (G-T). Each previously characterized gate was run with each new fuel containing either the CT or GT mismatch. (A) No mismatch gate (B) P1 gate (C) P2 gate (D) P3 gate (E) P1, P2 gate and (F) P1, P2, L8 gate.

Not surprisingly, the performance of the double mismatch (P1, P2) with the 5 bp loop suffered the most, reaching only half completion after 12 hours. The most promising gates appeared to be the single mismatch gate P2 and the double mismatch gate of P1, P2 with the 8 bp loop (8L). However, given the significant difference in activation rate and time to completion, the gate P1, P2, 8L was deemed the optimal choice for further experimentation.

6.2.3 Characterization of Input Adaptability (ssDNA, dsDNA, RNA, ATP)

The separation of target detection (DM) and reporter modules (RM) in our sensor design allows the sequences of the two modules to be varied independently of each other. This is important because the chimeric DNA/RNA oligo labeled with both a fluorophore and a quencher is typically the most expensive single element in our assays. Therefore, the ability to use the same substrate to report on the presence of different detection targets is of great practical utility. This property also makes these gates highly versatile, as the DM can be easily exchanged for alternative inputs. As the detection of nucleic acids relies only on hybridization principles, the structure of these gates enables straightforward detection of both ssDNA and RNA inputs. However, DNA is typically found in double-stranded form, making detection a challenge. Therefore, we designed gates to be used in the development of a protocol for the detection of dsDNA.

We developed four sensor gates to each detect independent subsequences from the pRSET emGFP plasmid (Life Technologies) by varying the DM, while keeping the RM fixed for each gate. We characterized the performance of each gate to the ssDNA oligomer targets corresponding to the

four target sequences on the plasmid (**Figure 6.6 A-D**). Individually, the sensors all performed similar to each other for the detection of ssDNA oligo targets. As these sensors detect single-stranded nucleic acid targets by Watson-Crick base pairing, they are also able to detect the corresponding RNA oligomer targets (**Figure 6.6 E,F**). This property expands the versatility of these modular gates, which could potentially be employed to detect mRNA or siRNA targets in addition to DNA targets.

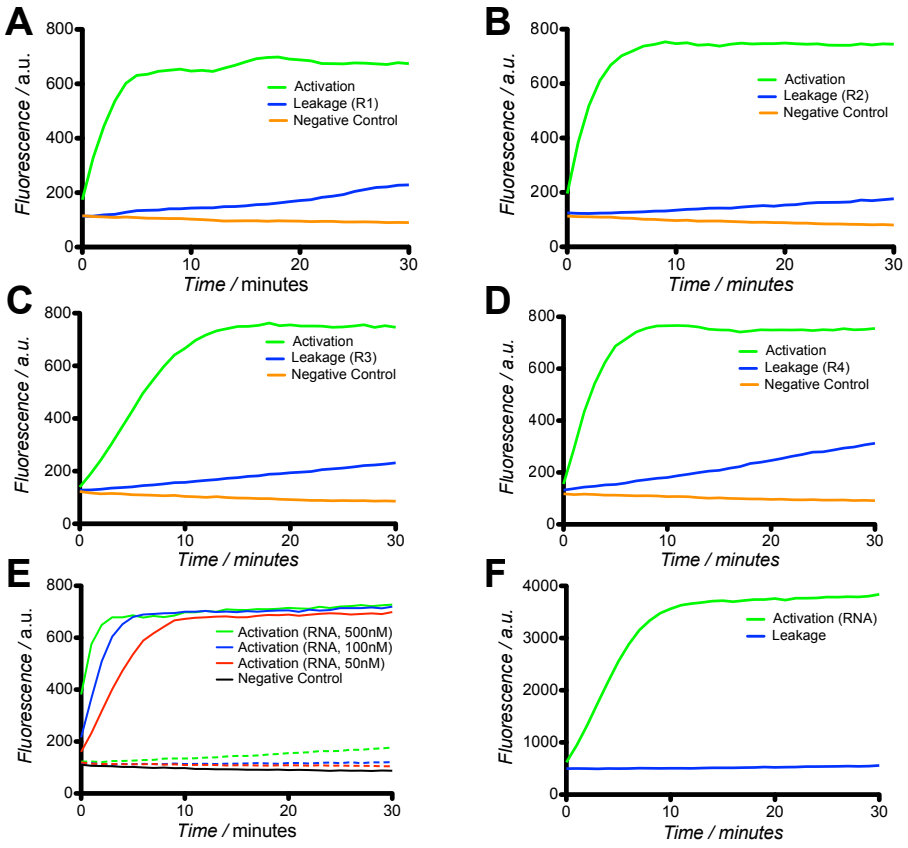

Figure 6.6 – Characterization of plasmid region gates. (A-D) Gates corresponding four different regions of the pRSET emGFP plasmid were designed and characterized. Each uses the P3 gate design. (E) R2 gate performance against an RNA-only target using various concentrations of fuel. (F) The R2 gate and RNA-only target at multiple turnover (250 nM substrate) conditions, with 100 nM fuel.

Although the detection of single-stranded DNA signatures is important, many of the desired targets are typically found in double stranded form. Detection of dsDNA requires additional processing to render targets single-stranded, so that the individual strands can activate our sensors via hybridization. After extracting and purifying the plasmid from SCS110 *E. coli* cells, we used an isothermal alkaline denaturation protocol to break apart the two strands. After gate structure optimization (**Section 6.2.2**), we successfully demonstrated the direct detection of denatured plasmid DNA with the R2 gate (**Figure 6.7**). Leakage remains at a low level and activation continues even after one week (**Figure 6.7D**).

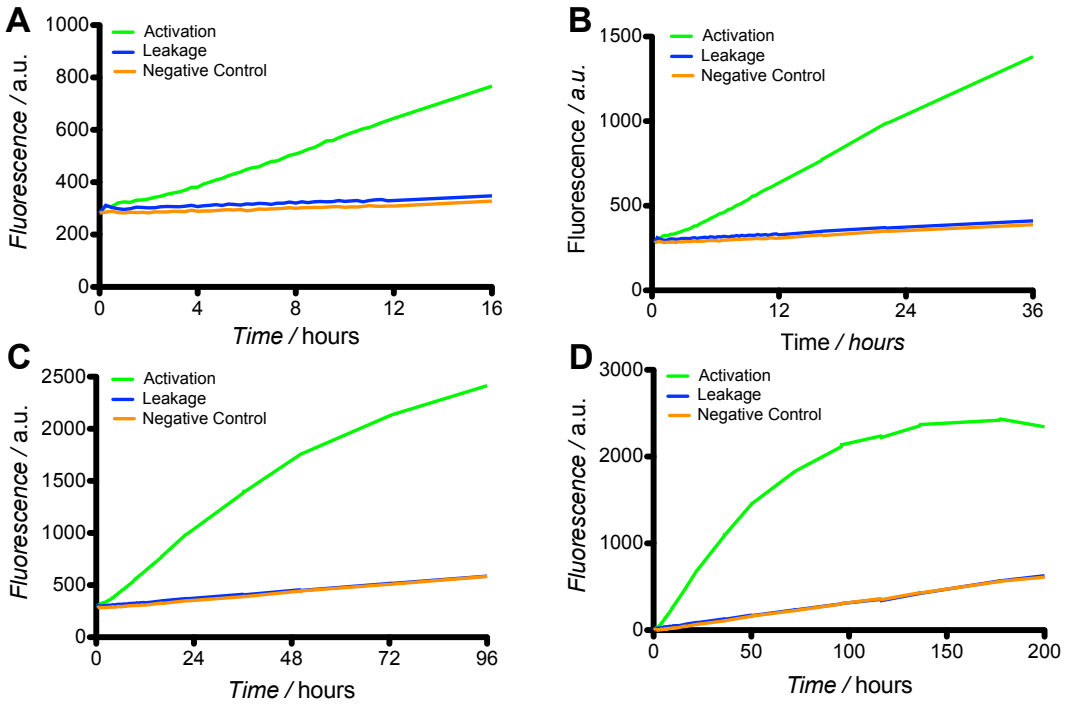

Figure 6.7 – Plasmid detection by the R2 (P1, P2, L8) gate. Gate, fuel, and plasmid concentrations were at 40 nM. (A) Fluorescence after 16 hours (B) Fluorescence after 36 hours (C) Fluorescence after 96 hours (D) Fluorescence after 204 hours.

Although the size of the plasmid ensures that the two strands will not rehybridize after neutralizing the pH, the target regions may still have local *cis* binding to form secondary structure, rendering potential target sites inaccessible. However, because our four sensors use a common readout module, we can perform multiplexed detection of the four target subsequences from the plasmid by using multiple sensors in parallel. This provides detection redundancy to overcome any secondary structure in the denatured plasmid strands, as any one of the four targets can activate the corresponding sensor. This could also provide additional signal amplification from low concentrations of plasmid, since each individual plasmid can activate four sensors, each capable of amplification via multiple turnover.

In addition to detecting other nucleic acids by direct hybridization, DNA strands can bind to small molecules, resulting in a conformational change[50]. We exploited this fact to further demonstrate the modularity of our sensor design by replacing the detection module with a partially blocked ATP aptamer sequence[51]. **Figure 6.8A** depicts a reaction scheme whereby binding of two ATP molecules to the aptamer displaces the blocking strand. This conformational change frees the secondary toehold and allows the fuel strand to bind as before. As shown in in **Figure 6.8B**, we observed significant signal over background in the presence of 1 mM ATP. In this experiment the leak rate was higher than in the earlier experiments, which was most likely because the blocking sequence for the aptamer detection module was shorter than the blocking sequence used in the earlier DNA detection modules: 8 nucleotides as opposed to 15. This meant that

the DNAzyme-inhibitor complex in the ATP-sensing gate was less stable than in the DNA-sensing gates, which would make it easier for the fuel strand to invade the toehold in the absence of the target signal. We performed experiments with ATP-sensing gates with longer blocking sequences (11 and 15 nucleotides) but found that in these cases the blocking sequence sequestered too much of the aptamer sequence, preventing the ligands from binding and activating the gate (**Figure A4.2**) These limitations are imposed by the particular aptamer sequence, and additional study of the kinetics and thermodynamics of aptamer activation will be necessary for further optimization of this design. It is also worth noting that these experiments were conducted using the P3 gate from **Section 6.2.2**, rather than the optimized P1, P2, L8 gate. Use of this new gate structure should result in a greatly improved response.

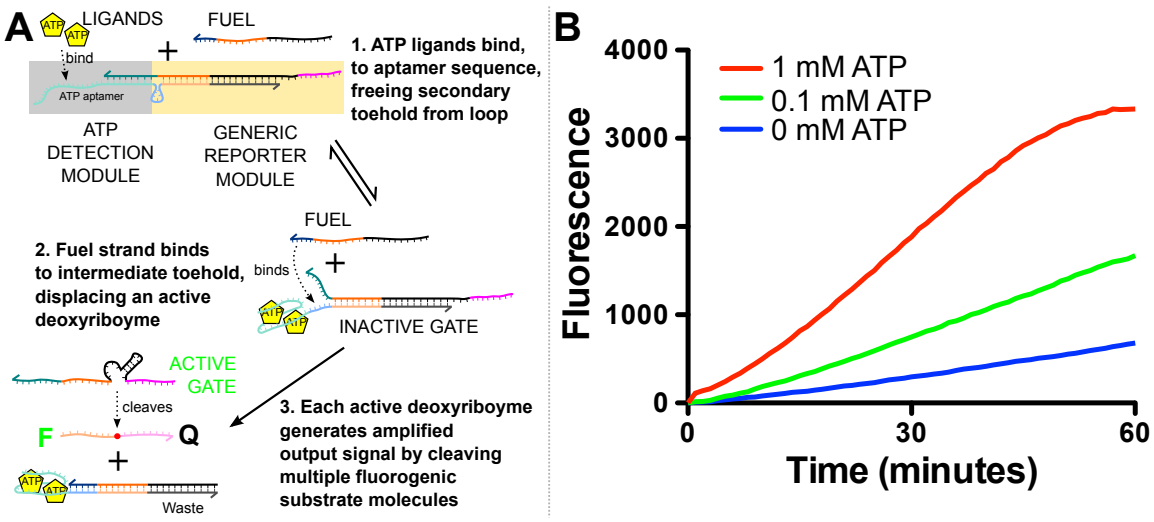

Figure 6.8 - Aptamer sensing using modular deoxyribozyme gates. a) Replacement of the target detection module with a partially blocked aptamer sequence allows the gate to be triggered by the binding of small molecules, in this case ATP. b) Kinetic traces showing gate response to various ATP concentrations.

6.2.4 Multiplexed analysis of STEC strains

We demonstrated the practical applicability of our modular sensor gates by designing six multiplexable sensor gates that each target a different STEC serotype (O26, O45, O103, O121, O145 or O157) and that each use a common reporter module. The target sequences were based on the PCR primers used for O-antigen STEC detection by Paddock *et al* [52]. These sequences are suitable for our purposes because PCR primer sequences are typically chosen to minimize secondary structure, which is also beneficial for TMSD reactions. We characterized each of the six STEC serotype-specific gates individually, as shown in **Figure 6.9B**. In each case, the sensor gates were prepared by annealing deoxyribozyme strands at 100 nM with the corresponding inhibitor strands at 125 nM, to account for variations in concentrations between stocks. Each experiment also used fuel strands at 500 nM and fluorogenic substrate at 250 nM. The target strands were single-stranded synthetic DNA oligonucleotides with the same sequences as a primer from the corresponding serotype. For the positive traces shown in **Figure 6.9B**, 50 nM of the target sequence of the particular serotype gate in question was added, and for the negative traces, 50 nM of *each* of the five incorrect serotype target sequences was added. In each case, we observed a significant response in the presence of the correct serotype target sequence, and a very low background in the presence of the other serotypes. This demonstrates that the sensor gates are highly sequence-specific. Variations in the activation rates between the detection gates for different STEC serotypes may be attributed to differences in the stability of the DNAzyme-

inhibitor complex caused by the different target sequences in the detection modules or by different levels of secondary structure in the target strands, which would impede the binding of the target to the toehold of the detection module.

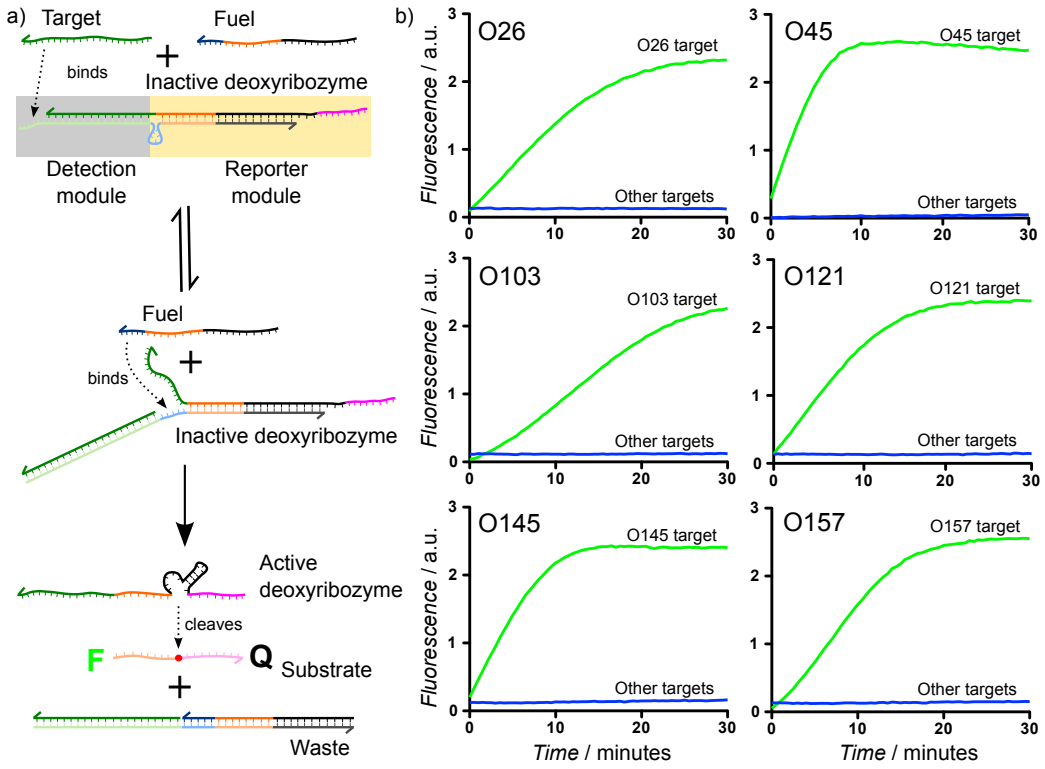

Figure 6.9 – Modular deoxyribozyme-based sensor gates and their application to detection of STEC target sequences. a) Basic mechanism of target detection and deoxyribozyme activation. Binding of the target strand to the detection module by toehold-mediated DNA strand displacement exposes the toehold in the reporter module, allowing the fuel strand to bind and complete displacement of the deoxyribozyme strand. The deoxyribozyme can then fold into a catalytically active conformation and generate an amplified fluorescent output by cleaving multiple substrate molecules labeled with a FRET pair. b) Detection of target sequences taken from the genomes of six STEC serotypes (O26, O45, O103, O121, O145 and O157), using sensor gates generated by fixing a reporter module and varying the input module. Each gate was characterized separately; the positive response in each case was obtained by adding the correct target sequence, and the negative controls were obtained by adding *all* of the other five target sequences, thereby demonstrating sequence specificity. The baseline response that occurs when the fuel is present in solution without the target strands was subtracted from each trace.

Since our STEC detection gates all use a common reporter module, we can multiplex detection of the six target sequences in a parallel gate array, with a single fluorescent readout via a common fluorogenic substrate molecule. In this case, a fluorescent signal is observed if *any* of the target sequences are present, giving a multiplexed detection circuit that implements "OR" logic. We demonstrated the multiplexed detection capabilities of our modular DNAzyme displacement gates by performing a six-way assay for the detection of *any* of the aforementioned six STEC serotype target sequences, as shown in **Figure 6.10**. Each STEC sensor gate was prepared as outlined above, and all six were combined along with the common fuel strands at 500 nM. This was replicated in seven wells; a single STEC target sequence at 50 nM was added to six of the wells, with the seventh serving as a negative control. After a 15 minute incubation at room temperature, the common substrate strand was added at 250 nM. Endpoint fluorescence values were measured after a further 30 minutes of incubation at room temperature. We observed a high signal to background ratio in all cases, with the relative heights of the bars for the positive traces corresponding roughly to the activation rates observed in the individual characterizations (**Figure 6.9B**). The background was slightly higher in this experiment, most likely due to the higher overall concentration of detector gates causing an increase in the spurious generation of signal. These data demonstrate that our system can simultaneously detect the STEC-specific sequences of interest in a mock bioassay scenario. A positive response from this six-way assay could be followed by a set of independent assays to determine the

identity of the strain present. We have also demonstrated that the sensor gates function well in the presence of randomized background DNA (**Figure A4.3**).

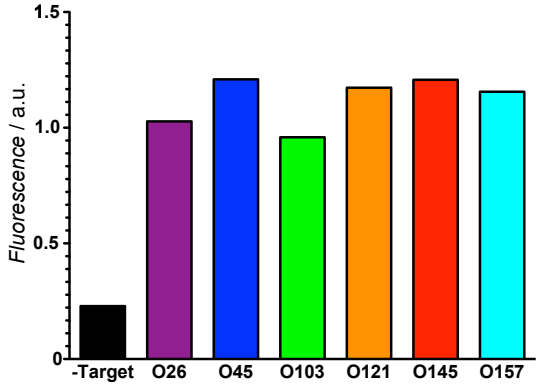

Figure 6.10 - Multiplexed detection of *any* of the six STEC serotypes of interest (O26, O45, O103, O121, O145 and O157). The six detectors from Figure 1b were used with a common fuel strand and a common substrate, and their response in the presence of each serotype signature was characterized. We observed high fluorescence values, which are similar for each of the target sequences, and a low response when no target is present.

Detection of low input concentrations is vital in pathogen detection applications. We investigated the limit of detection (LOD) of the O45 sensor gate by systematically diluting the entire system down to a 100 pM gate concentration. The left-hand axis of **Figure 6.11** shows kinetic traces of the O45 sensor gate in the presence of O45 target concentrations in the picomolar range. Lowering the gate concentrations reduced nonspecific signal generation in the absence of input while extending the timescale of the reactions. The right-hand axis of Figure 3 shows the LOD at various time points using the IUPAC definition of 3 standard deviations over background, to ensure statistically significant results [53]. We observe the optimal LOD of 7.4 pM at the four hour time point (see **Figure A4.4** for additional data). Thus our sensor gates are competitive with ELISA.

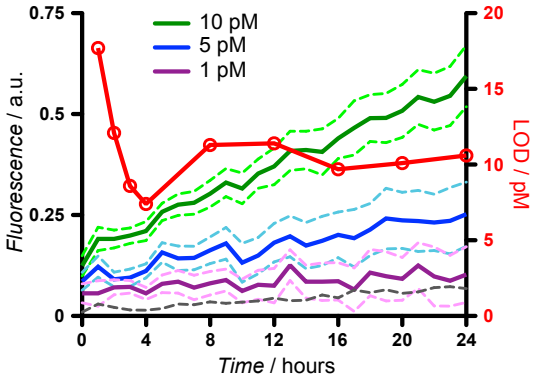

Figure 6.11 - Investigating the limit of detection of the O45 sensor gate. Colored line traces (left axis) show the response of the O45 detection circuit with various input concentrations: 10 pM, 5 pM and 1 pM. For these experiments the system was diluted to 100 pM gate concentration with 10 pM excess inhibitor, 100 pM fuel and 250 nM substrate. The background signal in the absence of input has been subtracted from all traces. Solid lines are average fluorescence values from 5 replicates, and dashed lines are one standard error above and below the mean in each case. Red data points (right axis) are detection limits at 3σ above the standard error of the background at various time points, calculated using the standard IUPAC definition[53].

6.2.5 Use of modular gates in DNAzyme cascades

Although characterization of modular gate behavior to this point has been focused on the optimization and direct performance as a biosensor, the design principles of DNAzyme cascading, outlined in detail in **Chapters 4 and 5**, should be directly applicable to developing modular gate cascades. However, modification of the gate structure from the original gates in **Chapter 3** altered several of the parameters used in the design of the original SCS. In particular, the availability of the toehold that binds to the SCS, now sequestered in a loop, is far more restricted in this system than its linear form in the original gates. Thus, use of the original SCS design (**Design 8**, **Figure 5.10**) resulted in almost no enzyme activation, likely due to the high structural stability and greater steric hindrance. This necessitated a reassessment of the SCS design that would be

most compatible with the modular gate structure. The increased protection of the toehold lessens the stringency with which the activator needed to be sequestered. Thus we revisited a previous SCS design (**Design 4, Figure 5.5**) that resulted in activation in our original gates, but where the leakage was too great to be considered a viable candidate. Here, the SCS is just a stem loop design, with a 7bp stem and a 22bp loop, with a 4bp toehold for the upstream DNAzyme. The mechanism of modular gate cascades remains quite similar, with the addition of both an input and a loop being necessary to release the upstream DNAzyme, which is then able to cleave the SCS. This cleavage product now acts as the fuel strand (functionally identical to the activator strand in the original DNAzyme gates) and along with $Input_1$, is sufficient to activate the downstream DNAzyme (**Figure 6.12**).

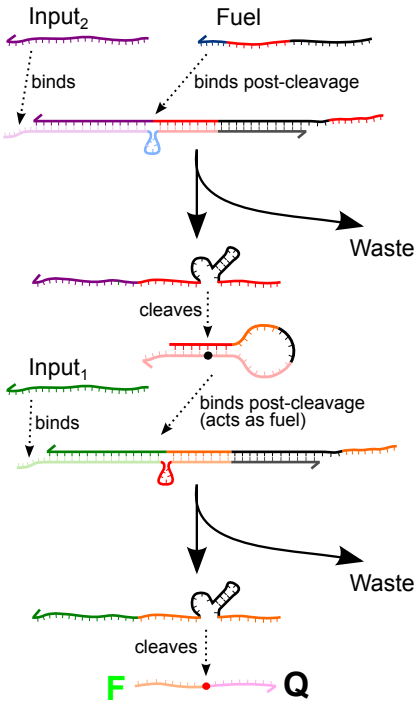

Figure 6.12 – Modular gate cascade scheme. Each enzyme requires its respective input, with the cascade initiated by the addition of upstream fuel.

Using the P3 gates, we designed a two layer cascade, which was activated by the addition of a fuel strand corresponding to a potential third layer SCS sequence. We achieved very good performance, comparable to our previous two-layer cascade design, with only a moderate reduction in activation rate. This is likely because the binding of its corresponding input sequence now regulates each enzyme layer, and the increased protection of the toehold ensures that the cleaved SCS product (Act) binds more slowly (**Figure 6.13**). Use of the third layer sequences as inputs indicates the viability of constructing a three-layer cascade, currently being implemented in the lab. This design also used a 2bp clamp, instituted by removing the two bases on the input closest to the inner loop toehold to prevent a full displacement of the DM, significantly reduced leakage in the cascades. The clamp sequence was also designed to mismatch to the activator toehold of the SCS, which resulted in reduced leakage in the cascades, in line with NUPACK predictions of relative binding stability at room temperature. A more detailed discussion of the design process of the P3 modular gate cascades can be found in **Section 6.2.6**. A prototypical two-layer cascade was also implemented using the P1, P2, L8 gate using a DNA-only SCS and ACT (representing a cleaved form of the SCS), to achieve a similar kinetic profile as the P3 gate, albeit with slightly less leakage, as would be expected from this gate. The RNA (cleavable) version of this cascade is also currently being implemented in the lab.

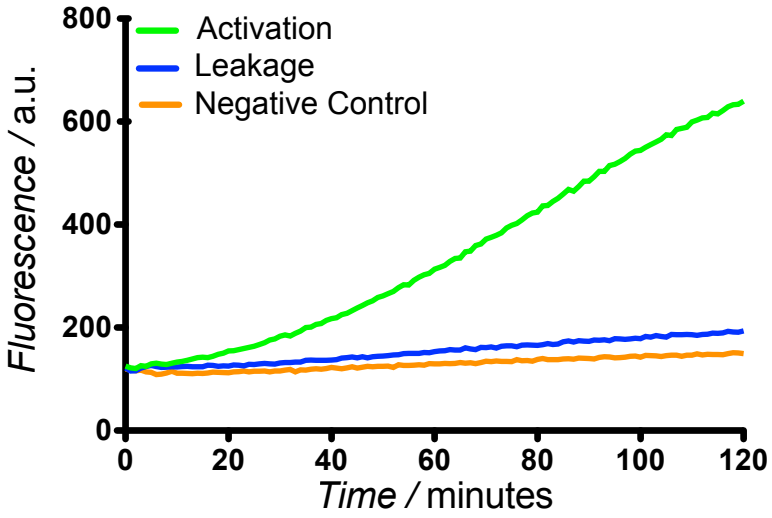

Figure 6.13 – Performance of a two layer modular gate cascade using the P3 gate design. In the negative control (blue), Dz1/Inh and Dz2/Inh, SCS, and Inputs 1&2 are present in solution. With the addition of a DNA-only SCS, coming from the third layer, as a sequestered fuel (brown), there is very little increase in circuit leakage. Addition of an activator (ACT) sequence, corresponding to the sequence derived from cleavage and dissociation of the waste product from an RNA SCS molecule, as a released fuel, we see a positive response (green).

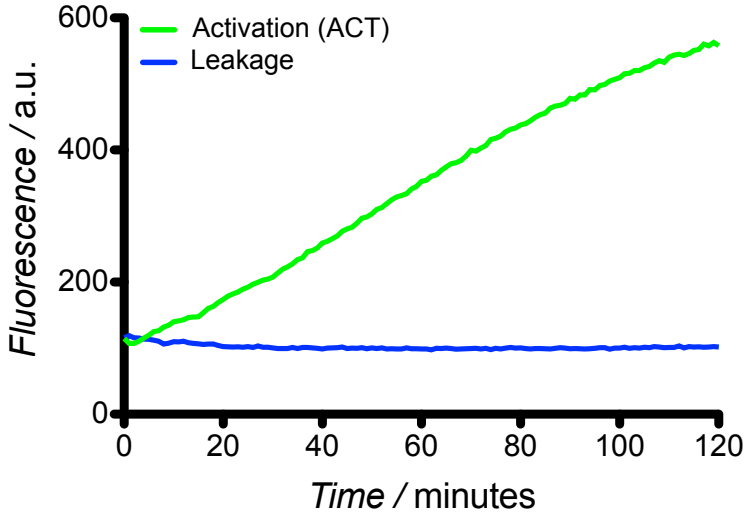

Figure 6.14 – Current development of a two layer cascade using the P1, P2, L8 gate design. The SCS and ACT are the DNA analog of the pre- and post-cleavage of the two layer RNA SCS molecule. The RNA version is currently being tested.

6.2.6 Rational design of SCS structure for modular gates

The reporter module of the modular gates is identical in sequence to the first layer of the DNAzyme displacement gates (alternatively designated as the U2 sequence); therefore we assumed that the SCS structure of **Design 8** (SCS-D8) would be sufficient to activate these gates. Here, the SCS activator would serve as the fuel strand, which are functionally equivalent. The addition of the input would release the target domain, and the cleaved SCS (ACT) sequence would displace the rest of the DNAzyme, resulting in signal. By using common DNAzyme sequences such as U2 for the basis of all experiments, different advances in cascading or gate structure would still be compatible with each other.

Unfortunately, the increased stability of the modular gate structure was incompatible with the pseudostable SCS-D8, and we observed no activation through cascading of the downstream modular gate using this SCS structure. This is not wholly unsurprising, as SCS-D8 was based on the availability of a free downstream toehold, and thus the activator required significant protection to ensure optimal sequestration. However, the introduction of mismatches to the fuel strand resulted in very little signal leakage of the gate, even in its single-stranded form. This likely resulted in a significant steric hindrance between the ACT and the loop toehold and overhanging DNAzyme detection domain strand. The retention of the inner stem of the SCS-D8 may have also contributed, making displacement even more unfavorable. With the failure of this structure, it became apparent that successful cascading no longer required a highly

protected fuel strand. Therefore, we revisited our earlier SCS design that were deemed nonviable due to high leakage, indicative of poor sequestration. As these designs still resulted in successful cascading, their weak protection of the activator strand was likely to be a good fit for the modular gate cascades.

As **Design 4 (Chapter 5)** was one of the most straightforward of the original designs, requiring a minimal secondary structure as it consists of merely a single stem and loop (**Figure 5.5**), we explored this design for modular gates cascades. Because the structure of the cascades is now primarily contained within the gate, the development of a viable stem loop structure was trivial compared to the design of the original SCS cascades. The design process focused around minor alterations of the toehold interactions, specifically the incorporation of mismatched bases similar to the linear fuel strand detailed in **Section 6.2.2**. By sequestering the respective fuels into the SCS Design 4 structure, we were able to generate productive SCS designs two different gate structures (P3 and P1, P2, L8). Performance of these gates is shown in **Section 6.2.5**.

6.2.7 Modular Gate Cascades for Advanced Circuit Behaviors

Although the development of modular gate cascades is still being explored and optimized, together, the performance of modular gates as biosensors and the regulation of each gate by its corresponding input hold significant potential for the utilization of modular gates in complex decision-making networks. We defined several advanced circuit behaviors to be conceptually implemented that would serve as individual components capable of being integrated into larger networks, with wet lab implementation occurring in the near future. The modularity of each

gate allows for the rapid development of such components, as each process can be designed independently. For example, the interaction of multiple enzymes through a single SCS intermediary represents the first step for complex network regulation and input integration. For this, the demonstration of both fan-in and fan-out capabilities is critical, each concept should be easily implemented using modular gates. Here, we can implement fan-in by altering the DM of the upstream DNAzymes while keeping the DM fixed (**Figure 6.15**). Because each upstream DNAzyme (Dz2 and Dz3) has the same DM, they are able to cleave the same SCS sequence, which releases the fuel (activator) for Dz1. Note that although the fuel for both Dz2 and Dz3 is shown as a single-stranded input, it can also be derived from another upstream SCS molecule as well. Additionally, the upstream regulation by separate inputs functionally acts as an OR gate, in which either $Input_2$ OR $Input_3$ is sufficient to activate the downstream DNAzyme.

Additionally, we can construct a fan-out circuit, in which a single upstream DNZzyme is able to activate two downstream DNAzymes (**Figure 6.16**). In the figure below, Dz3 cleaves a single SCS, which can in turn activate either Dz1 and/or Dz2, depending on the availability of Inputs 1 & 2. Although each downstream enzyme has one substrate binding arm in common, the second, unhybridized arm, is free to vary. We previously observed that this is sufficient to prevent cleavage of the same substrate sequences. Thus, we can split the initial signal into two DNAzymes, each able to cleave their respective substrates, which can be either FRET substrates labeled with different fluorophores or SCS molecules themselves, able to propagate the signal further.

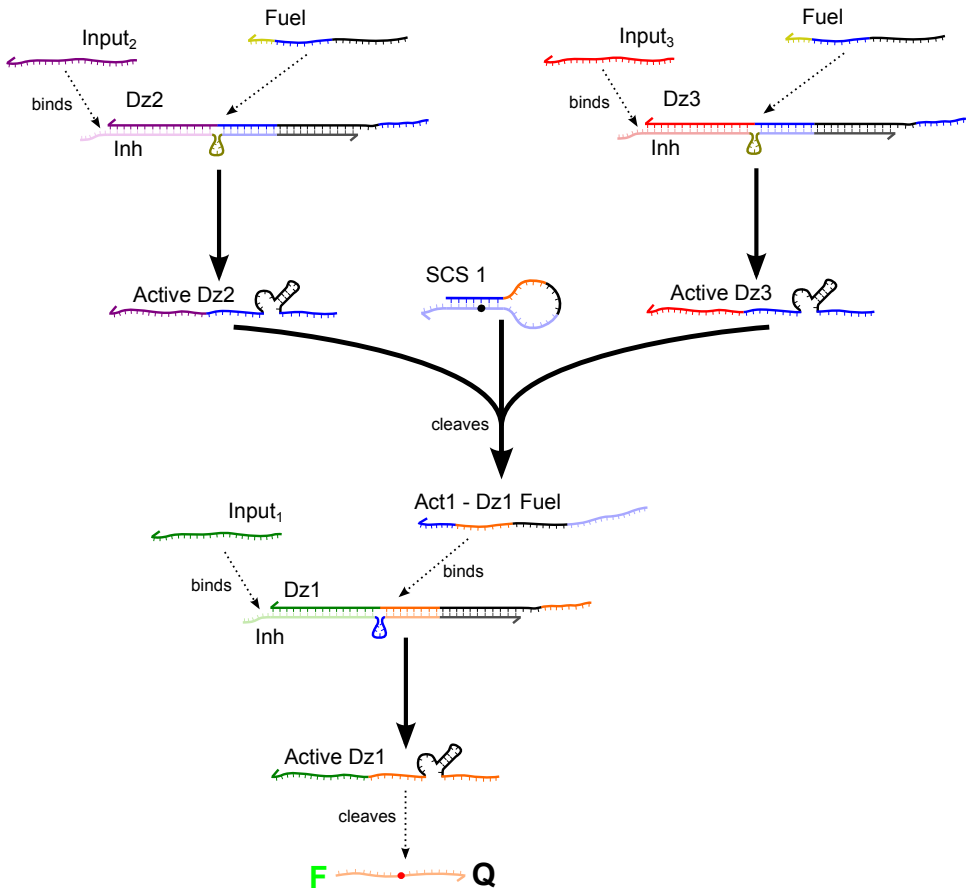

Figure 6.15 – Depiction of modular gate cascades executing fan-in. Two upstream DNAzymes (Dz2 and Dz3) contain the same reporter domain sequence (blue) and are therefore able to cleave the same SCS molecule after activation by their respective fuel and input. This acts as fuel for the downstream DNAzyme (Dz1) and activates in the presence of $Input_1$. Dz1 is then able to cleave a FRET substrate or another SCS to propagate the signal further.

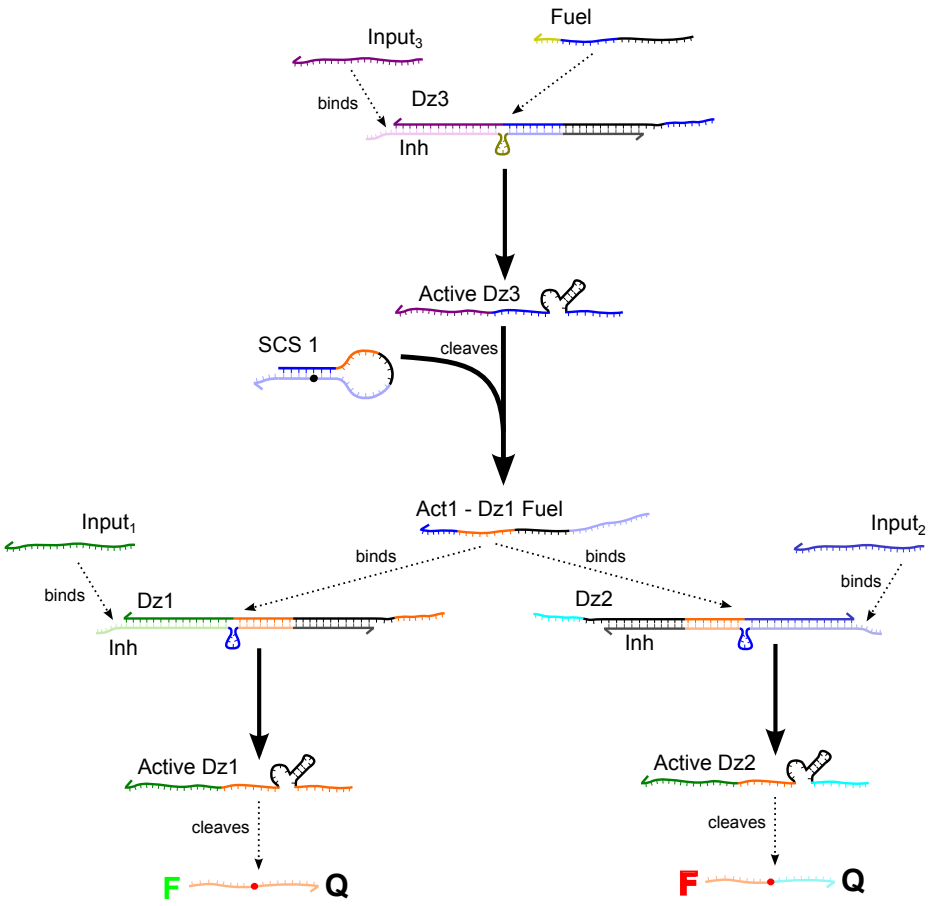

Figure 6.16 – Depiction of modular gate cascades executing fan-out. A single upstream DNAzyme (Dz3) cleaves the SCS molecule, after activation by its respective fuel and input. This SCS is now able to activate two DNAzymes (Dz1 and/or Dz2) depending on the availability of Inputs 1&2 in the system. By changing the second binding arm between DNAzymes 1&2, we can ensure each DNAzyme binds to a unique substrate. These two substrates can be FRET-labeled outputs or additional SCS molecules, to propagate the reaction further.

Although the integration and separation of signals is important for building more complex decision networks, the modular gate design also enables the execution of more advanced circuit behaviors, such as DNAzyme cycles as demonstrated in **Figure 6.17**. Here, we have two inhibited gates (Dz1 and Dz2), although with two corresponding SCS molecules. With the addition of a fuel strand to initiate the reaction, a small amount of Dz2 is released, capable of cleaving SCS1. This

acts as fuel for Dz1, which, upon cooperative binding with $Input_1$, is released.[54] The active Dz1 is then able to cleave SCS2, which releases more of Dz2 and the cycle continues. The activity of Dz1 can then be monitored by the addition of a substrate, which will be cleaved more rapidly in a cycle where increasing amounts of Dz1 can be generated. By matching the corresponding sequences, any two DNAzymes can be linked, enabling the engineering of more complex interactions such as three-DNAzyme cycles.

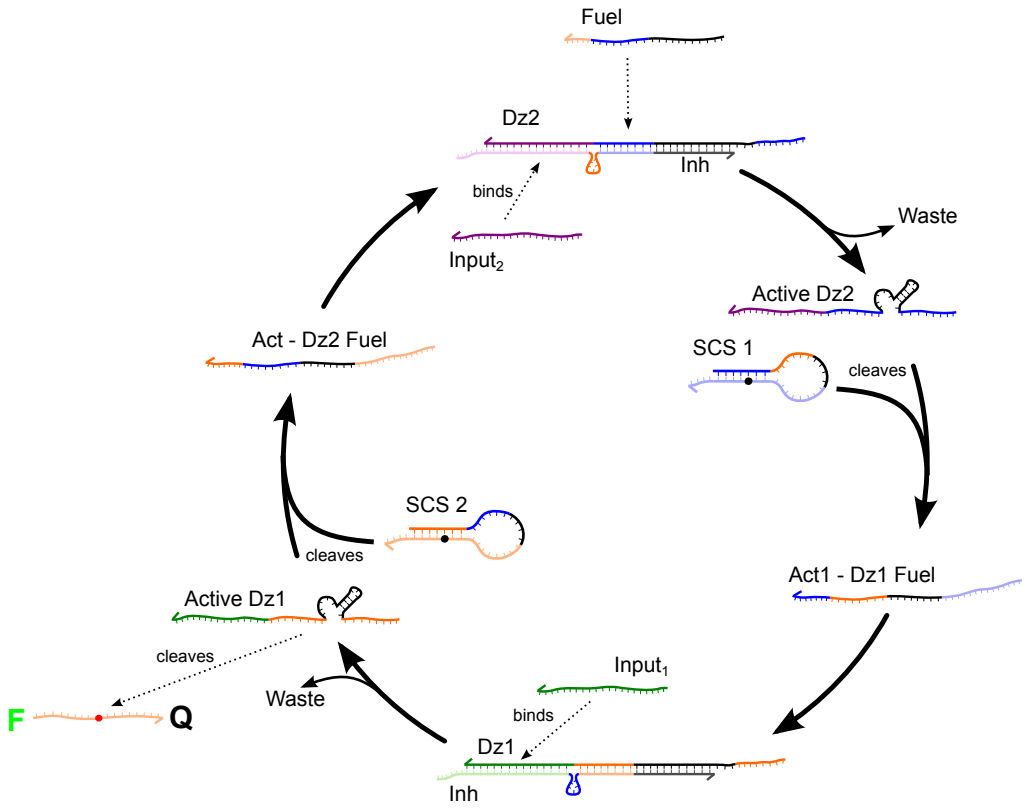

Figure 6.17 – Depiction of modular gate cascades executing a DNAzyme cycle. As Inputs 1&2 are available in the system, the addition of fuel to the upstream DNAzyme activates Dz2, which cleaves SCS1 to produce Act1. In conjunction with $Input_1$, Dz1 becomes active and cleaves SCS2 to produce Act2. As more fuel becomes available, each enzyme becomes more active, increasing the rate at which substrate is cleaved. The addition of a FRET substrate and its subsequent rate of cleavage is a relative indicator of the amount of DNAzyme present, which should be higher for positive feedback cycles such as this one.

6.3 Conclusions

The gate design presented herein is a major advance in modularity and versatility compared with the original DNAzyme displacement gates. By separating the reporter modular from the detection module, each gate is capable of detecting unique arbitrary sequences. The gate structure was designed specifically for biosensor performance, and optimized for fast activation with low leakage, even over extended time durations. The gate itself has been designed to detect a wide variety of single-stranded target DNA sequences, as well as RNA, small molecules, and plasmid DNA.

The purely nucleic acid-based approach to various target types and oligonucleotide sequences presented here has the advantage of simplicity and robustness compared with approaches for sequence specific detection based on protein enzymes, such as the polymerase chain reaction (PCR), strand displacement amplification[55], rolling circle amplification[56,57], or other isothermal nucleic acid amplification approaches[58,59]. Our approach is also distinct from other approaches that have used DNAzymes selected to provide fluorescent signals in response to bacteria based on affinity for the products excreted by the bacteria[60-62]. Our aim is to provide a simple, isothermal assay platform that is more sensitive, versatile, and cost-effective than alternative antibody-based approaches, such as the enzyme-linked immunosorbent assay (ELISA), and that can be performed with less expertise and less expensive lab equipment than PCR. Although PCR remains the gold standard for nucleic acid detection in terms of sensitivity, it is not isothermal and uses proteins, which require sophisticated

lab equipment and therefore limit its use in low cost field assays. All experiments were run with minimal purification of the strands and complexes, as this is important for reducing the cost of practical assay implementations. Sequestering the reporter module toehold in a bulge rather than in a duplex allowed us to implement a two-stage strand displacement reaction using a two-strand complex rather than the usual three, which further simplified the preparation of our sensors.

In addition, this work addresses the practical need for straightforward assays capable of identifying a large range of target strains with high sequence specificity. This versatile DNA-based sensor gate is well suited for isothermal, amplified detection and typing of multiple pathogen signatures. This allows new gates to be designed to target emerging pathogen strains easily and quickly. We demonstrated multi-strain detection capability using target sequences analogous to six different STEC strains, and we achieved a limit of detection of 7.4 pM after four hours. Target hybridization and strand displacement of the DM also confer high specificity, easily discriminating between gate specific targets and other targets sequences or a randomized DNA background.

The principles demonstrated here also show the expanded circuit design capabilities using the modular gate infrastructure. Having a true input control point for each DNAzyme improves both gate and circuit regulation, bringing this approach one step closer in approximating protein-based cascade functionality. The modular gate design lays the groundwork for truly complex decision trees and networks, capable of simultaneously integrating many different input

signatures. This capability is essential for the diagnosis of complex disorders and infections, the detection of multi-strain pathogens, hybridized biodetection and decision-making of pathogen serotyping, and the eventual construction of theranostic devices.

6.4 Future Directions

Although the O-antigen specific gates from **Section 6.2.4** performed well, each of these gates used the P3 mismatch version (as did the ATP gate from **Section 6.2.3**). This version was extensively characterized for leakage, mismatch response, sensitivity, and toehold size (**Appendix 4.2**, **Figures A4.5 and A4.6)**, and it showed an excellent signal to background ratio against single stranded targets. However, the P1, P2, L8 gate in **Figure A4.7** shows a reduced activation response from direct detection of plasmid DNA compared with the short, single-stranded target sequences. To detect this low signal over background, it is important to minimize gate leakage. While single-stranded targets are relatively small and diffuse easily and can provide a robust signal, plasmid DNA and genomic DNA targets are massive in comparison, meaning gate and target hybridization are likely diffusion limited. In addition, the formation of secondary structure and steric hindrance of these large targets may also have a effect on the rate of gate activation. Repeated testing of the O-antigen gates with the P1, P2, L8 structure is required. Although each individual gate generates a relatively low positive signal, the overall response may be improved by the addition of multiple gates, an option discussed above. In the specific case of STEC detection, this approach seems especially viable. Many target sequences have

been implicated in STEC infection (i.e. *Stx1, Stx2, EAE, HlyA*),[8,63,64] although there is some debate in the literature, especially in regards to the virulence plasmid,[64,66-70] as to the use of these sequences for positive STEC identification. Several of these sequences are inserted into the genome at multiple loci.[65] Thus, gates may be able to bind a single genome copy at multiple locations of a single sequence and across multiple sequences, and with each gate capable of multiple turnover, this may be an excellent approach for isothermal generation of a large output signal in the presence of low input concentrations. As additional gates also increase the overall leakage (**Figure 6.11**), testing of low leakage gate structures will be required if the approach of using multiple gates over long time durations is to be a viable option.

A complementary approach to improve gate performance in preparation for real-world biological sample detection is to remove excess inhibitor present in solution. As we are working under low-cost conditions, using minimal purification, we overcome strand impurities and concentration variation for each DNAzyme gate through the addition of excess inhibitor, typically on the order of 10-25%. This serves to effectively suppress any spurious DNAzyme signal, which may occur either through incomplete inhibition or through leakage reactions due to imperfect sequestration of the activator sequence in the SCS structure. However, the inhibitor strand also serves as the detection strand, binding directly to the target strand. Thus, the presence of excess inhibitor serves as a sink for the target strand, as it is thermodynamically more favorable for the target to bind free inhibitor than inhibitor complexed to the DNAzyme. To achieve the lowest

possible limit of detection, each target should bind and release a DNAzyme, but excess inhibitor effectively imposes a threshold on this lower limit. To overcome this, we are pursuing various methods of purifying out the excess inhibitor. One promising strategy appears to be the use of streptavidin-coated magnetic microspheres. Here, a capture strand is bound to the magnetic microspheres through a biotin-streptavidin conjugation, which contains the complementary sequence to the reporter domain of the inhibitor. As this region is unavailable when properly complexed with the DNAzyme, only the free inhibitior is able to bind the capture strand, and is subsequently removed when the beads are magnetically removed from solution (**Figure 6.18**)

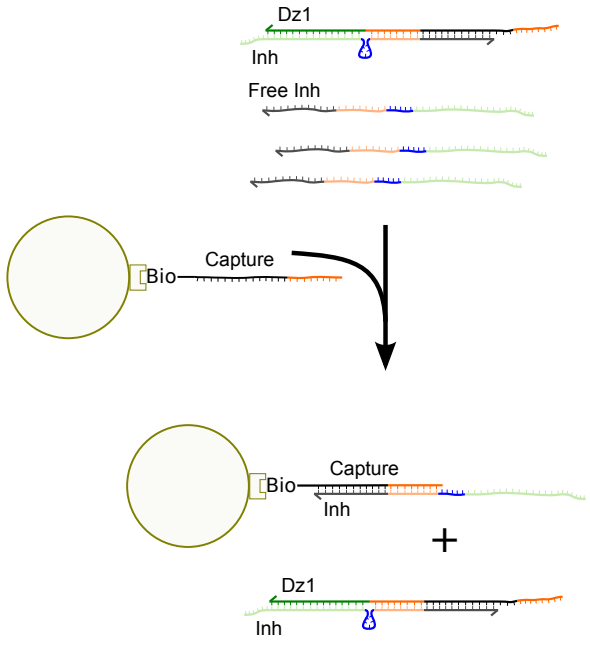

Figure 6.18 – Purification strategy to remove excess inhibitor from DNAzyme gates. Beads are streptavidin-coated magnetic microspheres. A biotinylated capture strand is added to the beads and the bead complex is mixed with the Dz/Inh complex. The capture strand is able to bind to the substrate binding arm and core of the inhibitor strand, which is only available with free inhibitor strands, as the DNAzyme for the bound inhibitor blocks that sequence. Magnetic removal of the beads ensures that most Dz/Inh complexes are recovered at the concentration is minimally altered.

Initial experiments using bead purification appear to have successfully removed excess inhibitor. Using the strand displacement DNAzyme/inhibitor FRET labeled complex from **Figure A2.3c**, we were able to test the response of this complex to increasing amounts of activator (**Figure 6.19**). By using a strand displacement (non-amplified) response, rather than a DNAzyme displacement, we were able to observe a more sensitive response to low input concentrations. Testing conditions of unpurified DNAzyme/inhibitor (25% excess), DNAzyme/inhibitor without capture, and DNAzyme/inhibitor with capture, we found that we achieved a robust response to low concentrations of activator in solution only with the captured inhibitor. We hope to further characterize this method for improved response to low concentrations of input, such as plasmid detection, and test how this purification can improve gate LOD. As excess inhibitor may also assist in rebinding spuriously activated DNAzymes to suppress leak, we will also characterize whether purification increases the signal from leakage over time.

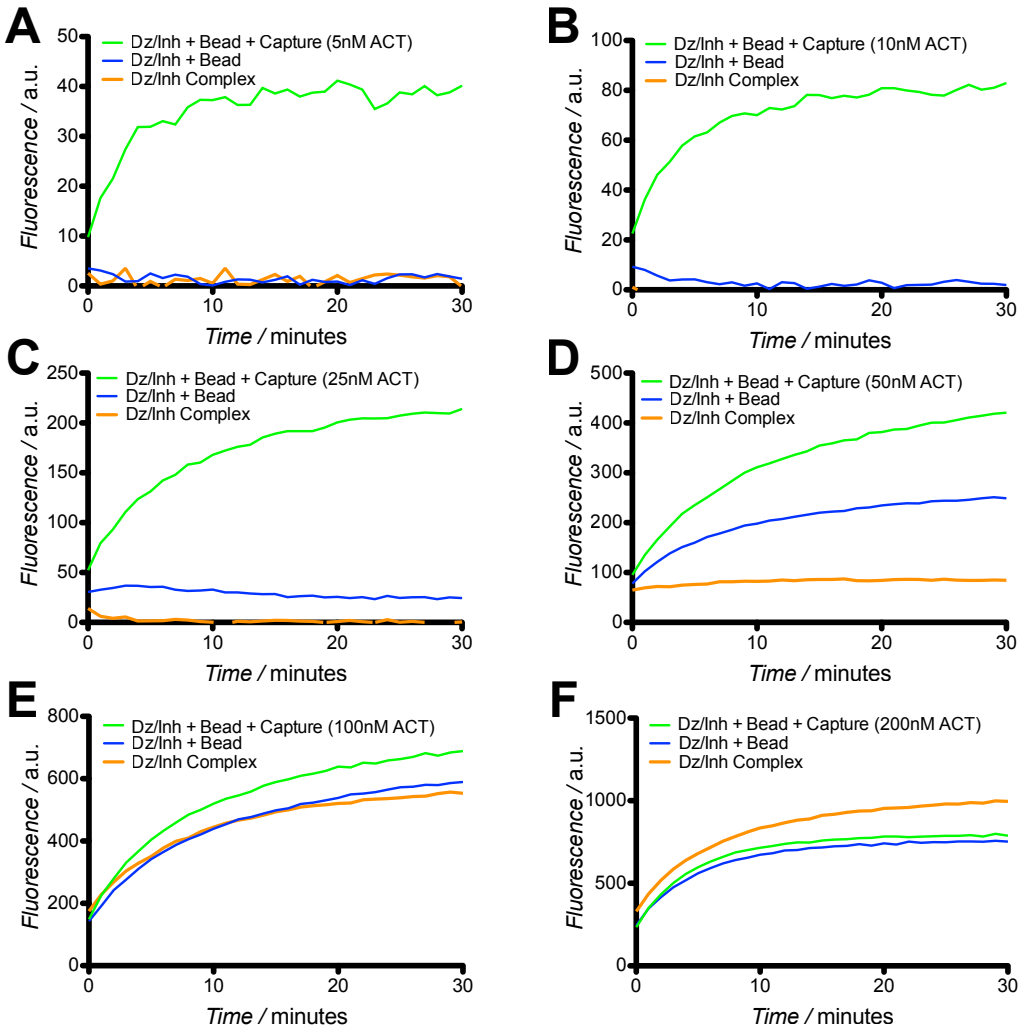

Figure 6.19 – Magnetic bead purification of FRET Dz/Inh complexes. Each complex was originally incubated with 25% excess inhibitor, and each strand of the Dz (FAM) and Inh (TAM) complex was labeled with a fluorophore. The negative control (black dash) had no purification steps, while the bead control (orange) incubated the Dz/Inh complex with the beads, but lacked the capture strand necessary to bind the free inhibitor. Black dash and orange traces should be similar in response. The purified sample (blue) incubated the beads and capture strand with the Dz/Inh complex for 30 min. 100 nM Dz/Inh complex was added to each well, and respective concentrations of activator strand was added. As activator binds to the inhibitor to form a waste complex, it removes FRET between the Dz and Inh strand, and the Dz strand can fluoresce. (A) 5 nM activator. (B) 10 nM activator. (C) 25 nM activator. (D) 50 nM activator. (E) 100 nM activator. (F) 200 nM activator. Purified complexes had increased signal at low concentrations of activator (A-C). The signal from the unpurified complexes approaches the purified response only at high concentrations of activator (F,G).

Another important future direction is the experimental implementation of alternate designs for more complex DNA circuits using modular gate cascades. Some work has already been done on the characterization of optimized gate design vs. SCS structure, but more work is still necessary. Although we have characterized the P3 and the P1, P2, L8 gates with cascades, other gate designs that had good performance should be tested as well. Gates such as P2 and P1, P2 gates also had reduced levels of leakage (**Figure 6.3**) compared to the P3 gate and should be investigated, as the design of an SCS-inhibitor interaction that contains three mismatches (including one G-T wobble) may impose design constraints on large circuits. Gates using this design may also have more variable activity profiles, which will make them unreliable in circuits. The use of modular gates for single nucleotide polymorphism (SNP) detection is detailed in **Appendix 4.3**. Alternative designs are detailed in **Appendix 4.4**.

A major advantage of the modular gate design is the ability to modify the reporter module to cleave a different substrate for each target sequence, a feature we have not yet explored in detail. If different fluorogenic readout substrates were conjugated to different fluorescently dyed beads, a large number of different strains could be detected in a single flow cytometry assay by gating on the bead fluorescence of the different bead populations. Finally, our sensor designs may also be integrated with a range of other readout technologies, such as lateral flow devices,[71,72] microfluidics,[73] and colorimetric readouts.[74,75]

6.5 Materials and Methods

6.5.1 Materials

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Substrate molecules (DNA-RNA chimeras) were purified using RNAse-free HPLC. Sequences for all oligonucleotides used herein are presented in Table S1 in the Supporting Information. ATP was purchased from Sigma-Aldrich (St. Louis, MO). DNA strands purified using standard desalting were resuspended in RNAse-free H_2O (Sigma-Aldrich, St. Louis, MO) at a stock concentration of 50 μM. These original stocks were diluted to working stocks of 2.5 μM, by diluting 50 μL stock DNA into 950 μL assay buffer (**Section 6.5.7**). RNAse-free HPLC strands were resuspended directly at 2.5 μM in RNAse-free H_2O.

6.5.2 Gate preparation

Typically, 60 μL of DNAzyme and 75 μL inhibitor (25% excess inhibitor) of 2.5 μM working stock solutions were added together and heated together at 95 °C for 3 minutes on a heat block, and subsequently annealed by cooling to room temperature over a minimum of 90 minutes. All other strands that required an initially hybridized state, including all SCS molecules, were also annealed using the same protocol.

6.5.3 Strand addition for modular gate experiments

General order of addition for each experiment is as follows: DNAzyme/inhibitor complex first, followed by SCS, fuel and input (for biosensing assays, the input is either a target strand, plasmid DNA, or ATP). The FRET substrate was added to

initiate the reaction. All experiments, with the exception of **Figure 6.10** was monitored and plotted as a kinetic trace. For **Figure 6.10**, input (here, the respective target strand) was added to initiate the reaction. After 15 minutes, substrate was added and monitored for 30 minutes. Final fluorescent values were reported as a bar graph.

6.5.4 Plasmid Extraction and Denaturation Protocol

pRSET emGFP (Life Technologies, Grand Island, NY) was transformed into SCS110 cells (Agilent Technologies, Santa Clara, CA). Cells were thawed on ice for roughly 30 min until liquid. One ng plasmid DNA was added to 100 uL cells, and incubated on ice for 30 min. Cells were then heat-shocked at 42°C for 45 seconds, and placed back on ice for 2 min. Cells were the incubated at 37°C for 10 min. After incubation, 20 uL cells were pipetted onto LB plates containing 50 ug/mL carbenicillin (Sigma Aldrich, St. Louis, MO) and incubated overnight at 37°C. Individual colonies were selected from this plate and grown in a 5 mL starter culture containing LB media + 50 ug/mL carbenicillin. After an overnight incubation, cells were added to 250 mL LB + 50 ug/mL carbenicillin and grown overnight. After incubation, cells were pelleted by centrifugation at 20,000 x g and lysed using a Maxiprep kit (Qiagen, Venlo, Limburg, Netherlands). Pelleted DNA was resuspended in RNAse free water. To denature the plasmid, 20 uL 1M NaOH was added to 90 uL plasmid DNA in RNAse free water. After 10 min, 20 uL 1M HCl was added to the solution to bring the pH back down.[76] Plasmid added to wells with DNAzyme gates comprised a maximum of 10% of the total volume.

6.5.5 Preparation of background DNA experiments

Herring sperm DNA (Promega, Madison, WI) was used as a control for a random DNA background. The herring sperm DNA was prepared several different ways. First, 1 µg/mL dsDNA was added to a well containing O157 P3 gate + substrate (**Table A4.7, Figure A4.3B**). The O157 gate was annealed separately. In **Figure A4.3C**, Dz O157, Inh O157, and herring sperm DNA were annealed together, that is, heated to 95°C for 3 min and cooled to room temperature over a minimum of 2 hours. In **Figure A4.3D**, herring sperm DNA was heated to 95°C and cooled rapidly on ice to prevent rehybridization. This was then added to a well containing O157 P3 gate + substrate.

6.5.6 Magnetic Bead Separation

Quantity of excess inhibitor of 135 µL Dz-Inh complexes (**Section 6.5.2**) was calculated to be 37.5 picomoles. Binding capacity of streptavidin-coated magnetic microspheres (Promega, Madison, WI) was determined from product information, which was estimated at 750-1250 pmol of biotinylated-oligo per mg of beads. Using the lower 750 pM capacity estimation, 66.6 µL beads (1mg/mL) corresponding to 50 pmol binding capacity were used. Beads were washed once by a 5 min magnetic separation and resuspended in assay buffer containing 200 pmol biotinylated capture strand. After incubation for 30 min on a rotator, beads were washed 2X in assay buffer. 135 µL Dz-Inh complex was added to the beads and incubated for 60 min with rotation. Beads were magnetically separated, and the purified Dz-Inh solution was removed. An activator strand was added to each well and fluorescence was monitored over 30 min.

6.5.7 Assay conditions and instrumentation

All assays were performed at room temperature (23 °C) in a buffer of 1M NaCl, 50 mM HEPES, 1 mM $ZnCl_2$, pH 7.0. Fluorescence was read on a Spectramax M2e fluorescent plate reader (Molecular Devices, Sunnyvale, CA) in a 200 µL reaction volume (492 nm excitation, 518 nm emission).

6.6 References

(1) Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N. *Nano Lett* **2006**, *6*, 2598.

(2) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.

(3) Qian, L.; Winfree, E.; Bruck, J. *Nature* **2011**, *475*, 368.

(4) Stojanovic, M. N.; Stefanovic, D. *Nat Biotechnol* **2003**, *21*, 1069.

(5) Chen, S. X.; Zhang, D. Y.; Seelig, G. *Nature Chemistry* **2013**, *2013*.

(6) Zhang, D. Y.; Chen, S. X.; Yin, P. *Nat Chem* **2012**, *4*, 208.

(7) Carlson, R. *Nat Biotechnol* **2009**, *27*, 1091.

(8) Paton, J. C.; Paton, A. W. *Clin Microbiol Rev* **1998**, *11*, 450.

(9) Brooks, J. T.; Sowers, E. G.; Wells, J. G.; Greene, K. D.; Griffin, P. M.; Hoekstra, R. M.; Strockbine, N. A. *J Infect Dis* **2005**, *192*, 1422.

(10) Kappeli, U.; Hachler, H.; Giezendanner, N.; Beutin, L.; Stephan, R. *Emerg Infect Dis* **2011**, *17*, 180.

(11) Johnson, K. E.; Thorpe, C. M.; Sears, C. L. *Clin Infect Dis* **2006**, *43*, 1587.

(12) Brown, C. W., III; Lakin, M. R.; Stefanovic, D.; Graves, S. W. *ChemBioChem* **2014**, *15*, 950.

(13) Zhang, D. Y.; Seelig, G. *Nat Chem* **2011**, *3*, 103.

(14) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. *Science* **2006**, *314*, 1585.

(15) Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E. *Science* **2007**, *318*, 1121.

(16) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2008**, *130*, 13921.

(17) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318.

(18) Li, B.; Ellington, A. D.; Chen, X. *Nucleic Acids Res* **2011**, *39*, e110.

(19) Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2013**, *110*, 5386.

(20) Li, B.; Jiang, Y.; Chen, X.; Ellington, A. D. *J Am Chem Soc* **2012**, *134*, 13918.

(21) Bath, J.; Green, S. J.; Allen, K. E.; Turberfield, A. J. *Small* **2009**, *5*, 1513.

(22) Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H. *Angew Chem Int Ed Engl* **2004**, *43*, 4906.

(23) Zhang, D. Y.; Winfree, E. *J Am Chem Soc* **2009**, *131*, 17303.

(24) Li, Y.; Breaker, R. R. *Curr Opin Struct Biol* **1999**, *9*, 315.

(25) Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611.

(26) Flynn-Charlebois, A.; Wang, Y.; Prior, T. K.; Rashid, I.; Hoadley, K. A.; Coppins, R. L.; Wolf, A. C.; Silverman, S. K. *J Am Chem Soc* **2003**, *125*, 2444.

(27) Lee, C. S.; Mui, T. P.; Silverman, S. K. *Nucleic Acids Res* **2011**, *39*, 269.

(28) Li, Y.; Breaker, R. R. *Proc Natl Acad Sci U S A* **1999**, *96*, 2746.

(29) Parker, D. J.; Xiao, Y.; Aguilar, J. M.; Silverman, S. K. *J Am Chem Soc* **2013**, *135*, 8472.
(30) Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. *J Am Chem Soc* **2005**, *127*, 13124.
(31) Walsh, S. M.; Sachdeva, A.; Silverman, S. K. *J Am Chem Soc* **2013**, *135*, 14928.
(32) Wang, Y.; Silverman, S. K. *Angew Chem Int Ed Engl* **2005**, *44*, 5863.
(33) Breaker, R. R.; Joyce, G. F. *Chem Biol* **1995**, *2*, 655.
(34) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.
(35) Stojanovic, M. N.; Stefanovic, D. *Journal of the American Chemical Society* **2003**, *125*, 6673.
(36) Stojanovic, M. N.; Mitchell, T. E.; Stefanovic, D. *J Am Chem Soc* **2002**, *124*, 3555.
(37) Lederman, H.; Macdonald, J.; Stefanovic, D.; Stojanovic, M. N. *Biochemistry* **2006**, *45*, 1194.
(38) Pei, R.; Matamoros, E.; Liu, M.; Stefanovic, D.; Stojanovic, M. N. *Nat Nanotechnol* **2010**, *5*, 773.
(39) Mokany, E.; Bone, S. M.; Young, P. E.; Doan, T. B.; Todd, A. V. *J Am Chem Soc* **2010**, *132*, 1051.
(40) Kolpashchikov, D. M. *Chem Rev* **2010**, *110*, 4709.
(41) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.
(42) Elbaz, J.; Wang, F.; Remacle, F.; Willner, I. *Nano Lett* **2012**, *12*, 6049.
(43) Kahan-Hanum, M.; Douek, Y.; Adar, R.; Shapiro, E. *Sci Rep* **2013**, *3*.
(44) Orbach, R.; Remacle, F.; Levine, R. D.; Willner, I. *Proc Natl Acad Sci U S A* **2012**, *109*, 21228.
(45) Wang, F.; Elbaz, J.; Orbach, R.; Magen, N.; Willner, I. *J Am Chem Soc* **2011**, *133*, 17149.
(46) Wang, F.; Elbaz, J.; Teller, C.; Willner, I. *Angew Chem Int Ed Engl* **2011**, *50*, 295.
(47) Wang, F.; Elbaz, J.; Willner, I. *J Am Chem Soc* **2012**, *134*, 5504.
(48) Lu, C. H.; Wang, F.; Willner, I. *J Am Chem Soc* **2012**, *134*, 10651.
(49) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.
(50) Achenbach, J. C.; Nutiu, R.; Li, Y. *Analytica Chimica Acta* **2005**, *534*, 41.
(51) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656.
(52) Paddock, Z.; Shi, X.; Bai, J.; Nagaraja, T. G. *Vet Microbiol* **2012**, *156*, 381.
(53) Long, G. L.; Winefordner, J. D. *Analytical Chemistry* **1983**, *55*.
(54) Zhang, D. Y. *J Am Chem Soc* **2011**, *133*, 1077.

(55) Walker, G. T.; Fraiser, M. S.; Schram, J. L.; Little, M. C.; Nadeau, J. G.; Malinowski, D. P. *Nucleic Acids Res* **1992**, *20*, 1691.
(56) Fire, A.; Xu, S. Q. *Proc Natl Acad Sci U S A* **1995**, *92*, 4641.
(57) Zhao, W.; Ali, M. M.; Brook, M. A.; Li, Y. *Angew Chem Int Ed Engl* **2008**, *47*, 6330.
(58) Kim, J.; Easley, C. J. *Bioanalysis* **2011**, *3*, 227.
(59) Vincent, M.; Xu, Y.; Kong, H. *EMBO Rep* **2004**, *5*, 795.
(60) Aguirre, S. D.; Monsur Ali, M.; Kanda, P.; Li, Y. *Journal of Visualized Experiments* **2012**, *63*, e3961.
(61) Aguirre, S. D.; Monsur Ali, M.; Salena, B. J.; Li, Y. *Biomolecules* **2013**, *3*, 563.
(62) Huang, P.-J. J.; Liu, M.; Liu, J. *Reviews in Analytical Chemistry* **2013**, *32*, 77.
(63) Fagan, P. K.; Hornitzky, M. A.; Bettelheim, K. A.; Djordjevic, S. P. *Appl Environ Microbiol* **1999**, *65*, 868.
(64) Johnson, T. J.; Nolan, L. K. *Microbiol Mol Biol Rev* **2009**, *73*, 750.
(65) Serra-Moreno, R.; Jofre, J.; Muniesa, M. *J Bacteriol* **2007**, *189*, 6645.
(66) Burland, V.; Shao, Y.; Perna, N. T.; Plunkett, G.; Sofia, H. J.; Blattner, F. R. *Nucleic Acids Res* **1998**, *26*, 4196.
(67) Muniesa, M.; Hammerl, J. A.; Hertwig, S.; Appel, B.; Brussow, H. *Appl Environ Microbiol* **2012**, *78*, 4065.
(68) Newton, H. J.; Sloan, J.; Bulach, D. M.; Seemann, T.; Allison, C. C.; Tauschek, M.; Robins-Browne, R. M.; Paton, J. C.; Whittam, T. S.; Paton, A. W.; Hartland, E. L. *Emerg Infect Dis* **2009**, *15*, 372.
(69) Paytubi, S.; Dietrich, M.; Queiroz, M. H.; Juarez, A. *Plasmid* **2013**, *70*, 52.
(70) Schmidt, H.; Kernbach, C.; Karch, H. *Microbiology* **1996**, *142 (Pt 4)*, 907.
(71) Allen, P. B.; Arshad, S. A.; Li, B.; Chen, X.; Ellington, A. D. *Lab Chip* **2012**, *12*, 2951.
(72) Martinez, A. W.; Phillips, S. T.; Butte, M. J.; Whitesides, G. M. *Angew Chem Int Ed Engl* **2007**, *46*, 1318.
(73) Hong, J. W.; Quake, S. R. *Nat Biotechnol* **2003**, *21*, 1179.
(74) Song, T.; Liang, H. *J Am Chem Soc* **2012**, *134*, 10803.
(75) Zhao, W.; Lam, J. C.; Chiuman, W.; Brook, M. A.; Li, Y. *Small* **2008**, *4*, 810.
(76) Oh, S. H.; Chater, K. F. *J Bacteriol* **1997**, *179*, 122.

Chapter 7.
High Throughput DNAzyme Screening

7.1 Introduction

Biological systems catalyze chemical reactions primarily through the use of protein enzymes. These enzymes are exceedingly efficient, selected by evolution over billions of years, and catalyze a vast and diverse range of reactions within the cell. The pressure of natural selection has ensured that protein structure, protein-protein interactions, and cofactor activation have developed to an optimal point, rendering further improvement using artificial selection methods infeasible. Though prior work on biocomputing devices using naturally occurring proteins shows promise[1-5], the complexity of protein folding, non-physiological stability, and the available repertoire of existing protein-protein interactions limits the use of proteins in the design of synthetic molecular computation circuits and devices[6].

The discovery of a group of catalytic RNA molecules, known as ribozymes, dramatically altered this landscape. The single-stranded nature of RNA enables it to function as a biological catalyst, and it has since been found to catalyze many reactions previously thought to be restricted to proteins. This discovery led to the “RNA world” hypothesis, in which RNA molecules catalyzed the chemical reactions in the prebiotic and protobiotic Earth[7,8], with the discovery of RNA self-replication providing a mechanism for Darwinian evolution. In addition to RNA-only catalysis, RNA can also complex with other molecules, such as in the ribosome, where mRNA, tRNA, and rRNA all interact within the protein scaffold.

Due to the biochemical similarity between RNA and DNA, it followed that DNA could potentially behave as a biological catalyst as well. Although DNA is almost exclusively double-stranded in nature, functioning as an information carrier in genomic or plasmid form, the advent of solid-state DNA synthesis and artificial selection processes such as SELEX (**S**ystematic **E**volution of **L**igands by **Ex**ponential Enrichment) enabled the discovery of DNA enzymes, also known as DNAzymes or deoxyribozymes[9-11]. After this initial finding, DNAzymes have been subsequently selected as catalyzed to a wide variety of chemical reactions, just like its RNA equivalent[12-17].

Despite the continually expanding capabilities of these nucleic acid enzymes, catalysis rates of individual reactions remain orders of magnitude lower than protein enzymes[18,19]. This is widely presumed to be the result of the larger biochemical diversity available to proteins through their more numerous and more chemically varied amino acid composition. In the case of RNA transesterification, the enzyme-catalyzed rate enhancement by 8-17 is achieved through the use of two catalytic mechanisms, which involves the deprotonation of the 2' hydroxyl group through base catalysis[20] and the arrangement of the structure to an in-line conformation (**Figure 7.1**)[21]. However, there are two alternative mechanisms, which involve the protonation of a non-bridging phosphate ion and the protonation of the 5' oxygen, both of which are specific acid catalysis[19]. While protein-based enzymes such as RNAse A likely engage all four of these mechanisms[22,23], it appears that many well-studied ribozymes and DNAzymes make use of only two of these. However, there does not appear to be

a hard limitation preventing DNAzymes from engaging the alternative transfer mechanisms[19]. By making use of these additional strategies, nucleic acid enzymes could achieve rate enhancements on the order of protein enzymes.

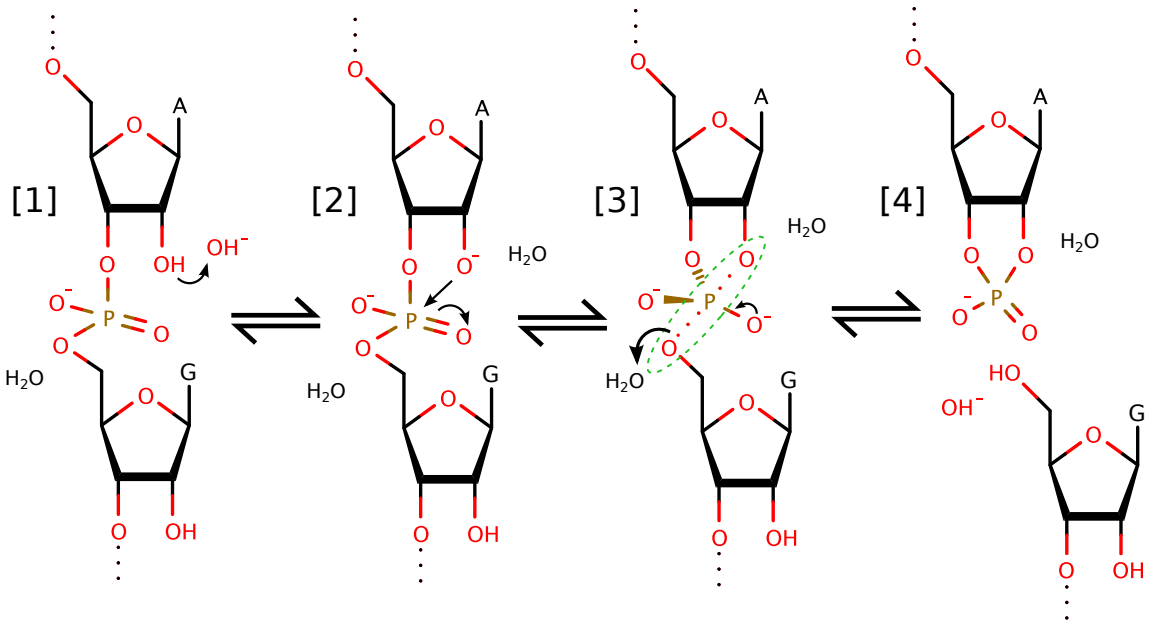

Figure 7.1 – Proposed RNA transesterification mechanism utilized by the 8-17 DNAzyme for RNA base hydrolysis. This DNAzyme appears to make near perfect use of two different catalytic strategies. First, the 2' hydroxyl group is deprotonated through base catalysis, shown in step 1. This group then participates in a nucleophilic attack on the phosphate center (step 2), resulting in a pentacoordinate species. The efficiency of this nucleophilic attack is also enhanced with an in-line orientation of the nucleophile, electrophile center and leaving group, which facilitates this process. This is highlighted by the green elliptical in step 3. After the leaving group is protonated, the reaction is complete. (Figure adapted from Emilsson, 2003)

Selection of catalytically active DNAzyme sequences is achieved through the SELEX process. Thus, enzymatic activity is determined by the specific setup and conditions of the selection procedure. The 8-17 DNAzyme, named for selection round (8) and the clone number (17), was originally selected under the following conditions: 1M NaCl, 10 mM $MgCl_2$, 50 mM Tris, pH 7.5[10]. Characterization of additional mutations by Li found a rate enhancement with 8-17 in the presence of $ZnCl_2$ instead of $MgCl_2$[24]. In both cases, selection conditions were fixed with limited reactants available for DNAzyme utilization. In light of the potential rate enhancements described by Emilsson and Breaker[18,19], DNAzymes appear to be intriguing candidates for small molecule high throughput screening.

While inhibitor compounds may be useful, our main interest lies in the discovery of activating compounds. The 8-17 DNAzyme was chosen for screening, due to its small size and high catalytic activity, as well as its extensive use in related experiments in our lab. The discovery of activators has numerous potential benefits. It has been shown the DNAzymes similar to 8-17 appear to have perfected the use of two different catalytic mechanisms[18]. Therefore, compounds able to increase the rate of 8-17 substrate cleavage would likely engage an additional transfer mechanism, such as acid catalysis, not yet observed in nucleic acid enzymes. This would have a substantial effect on our understanding of ribozyme and DNAzyme catalysis, potentially shedding light on the RNA world hypothesis and the origins of RNA enzyme complexes with peptides and other small molecules. Additionally, activators could also be of

significant importance in areas where DNAzymes are used[14,25-28], such as our DNAzyme displacement gates and cascades[29]. Improved performance of enzyme catalysis could speed up decision times by orders of magnitude, which could bring DNA circuit performance closer to electronic circuits. Improving decision times would also advance their capabilities as biosensor elements, making them more competitive with protein-based assays.

7.2 Results

7.2.1 Preparation of DNAzyme assay conditions for HTS compatibility

Development of a protocol for high throughput screening for the 8-17 DNAzyme required the refinement of assay parameters. These included a positive control, a negative control, and a "mid-control" to be quenched at the end of the experiment to assess the relative increase or decrease in enzyme activity in the presence effector compounds. The AllX library of compounds was chosen for this screening effort[30,31], due to its wide compound and scaffold variability. Each well contains an entire scaffold library, that is, a single base compound with all possible substitutions at each position on the compound. Scaffolds can be small molecule- or peptide-based. This library was tested in a 96 well format.

Several controls were necessary to determine compatibility between typical buffer conditions (1M NaCl, 50 mM HEPES, 1mM $ZnCl_2$, pH7) and the AllX scaffold library, which is stored in dimethylformamide (DMF), as well as to define assay parameters. As other potential libraries contained dimethylsulfoxide (DMSO) rather than DMF, we also examined the effect of DMSO on DNAzyme performance. (**Figure A5.1**) Parameters tested included buffer conditions,

enzyme purification, enzyme concentration, and EDTA quenching concentration (**Figure 7.2**).

Prior experiments in this dissertation focused on the use of DNAzymes for bioassay development, thus it was desirable to minimize additional purification steps to reduce cost. To ensure DNAzyme performance was optimized for performance to reliably assess compound interaction, we tested DNAzyme response through PAGE purification. We found that using PAGE purified DNAzyme strands resulted in a more robust enzyme performance, in line with the findings of other groups (**Figure 7.2A**)[32,33]. To characterize a suitable quencher for the mid-control reaction, we tested various concentrations of EDTA. Consistent with expected results for a buffer containing a 1mM metal cation concentration, the reaction was only quenched by the addition of at least 1mM EDTA (**Figure 7.2B**). As the concentration of EDTA did not affect the fluorescent response, a final concentration of 0.1M EDTA was used for the screening protocol.

After reaction initiation and quenching conditions were established. Initially, we explored the use of DNAse I for a positive control, to cleave the chimeric substrate strand, but that failed to produce the requisite fluorescent response (**Figure A5.2**). We found that the addition of a significant excess of 8-17 DNAzyme was sufficient for the positive control, rapidly cleaving all available substrate (**Figure 7.2C**). As with previous experiments, the negative control remained the reaction without input, in this case the DNAzyme strand. Here, this was the substrate by itself. To ensure the screening ran long enough to evaluate

compound effects, the enzyme concentration was chosen to produce 50% maximum fluorescence after 20 minutes. With the use of PAGE purified DNAzyme strands, which were already shown to have increased activity compared to standard desalting, the optimal DNAzyme concentration was found to be 15 nM (**Figure 7.2C**). All three controls (positive, negative, and mid), conducted with 8 replicates and accompanying error bars. These replicates were essential to ensure consistency across a 96 well plate, and our results showed exceptional reproducibility. **Figure 7.2D** demonstrates the mid control with quenching by EDTA. Final experimental conditions and assay parameters are detailed in the materials and methods (**Section 7.4**). Additional parameters were also examined and are reported in **Appendix 5**.

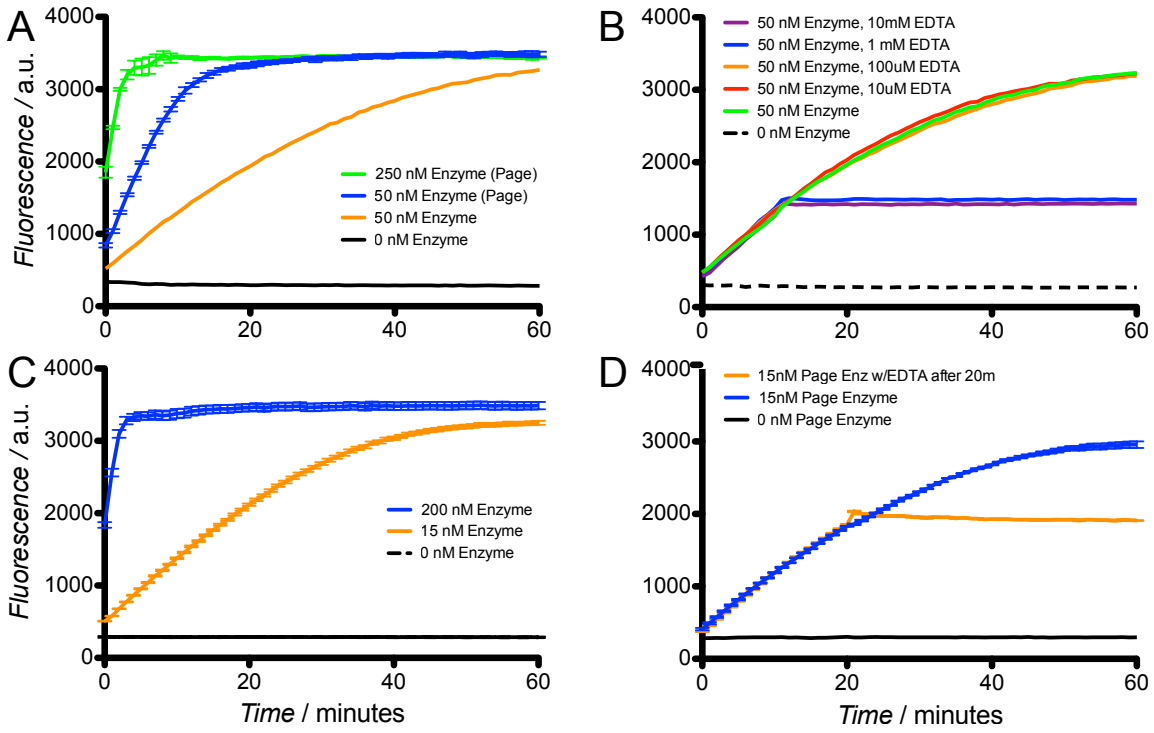

Figure 7.2 – Determination of scaffold library screening parameters. (A) The use of PAGE purification resulted in an increased DNAzyme performance (n=3). (B) DNAzyme activity quenching via EDTA after 10 min, requiring a minimum of 1 mM EDTA for complete quenching. (C) Relative performance at the final concentrations of DNAzyme for negative (0 nM), mid (15 nM), and positive (200 nM) controls (n=8). (D) Mid control performance with and without quenching (n=3).

		1 mg/ml, 10% DMF/water					0.5 mg/ml, 5% DMF/water					
Plate Layout	1	2	3	4	5	6	7	8	9	10	11	12
A	loCntrl	1169	1324	1420	1481	882	1169	1324	1420	1481	882	midCntrl
B	loCntrl	1170	1343	1421	1433	531	1170	1343	1421	1433	531	midCntrl
C	loCntrl	1171	1344	1422	1295	914	1171	1344	1422	1295	914	midCntrl
D	loCntrl	1172	1345	923	1277	506	1172	1345	923	1277	506	midCntrl
E	hiCntrl	1174	1346	924	1387	886	1174	1346	924	1387	886	midCntrl
F	hiCntrl	1275	1347	1455	1409	sc_1mg/ml	1275	1347	1455	1409	sc_0.5mg/ml	midCntrl
G	hiCntrl	1276	1418	1456	1509	sc_1mg/ml	1276	1418	1456	1509	sc_0.5mg/ml	midCntrl
H	hiCntrl	1319	1419	1477	1002	sc_1mg/ml	1319	1419	1477	1002	sc_0.5mg/ml	midCntrl

	1 mg/ml, 10% DMF/water						0.5 mg/ml, 5% DMF/water					
% of midCntrl	1	2	3	4	5	6	7	8	9	10	11	12
A	0.4	67.4	62.1	72.6	65.0	88.4	73.5	65.8	71.4	59.2	85.8	95.4
B	-0.1	83.9	62.4	71.7	15.0	58.2	84.5	68.1	71.9	14.8	59.5	98.6
C	-0.1	83.1	84.6	48.1	84.6	82.7	90.4	85.7	54.4	89.2	82.2	103.9
D	-0.2	74.4	69.4	54.6	79.1	64.1	80.4	77.0	59.2	83.4	63.4	92.3
E	216.0	86.4	63.9	78.4	86.1	83.9	92.9	69.8	83.7	86.7	90.6	98.4
F	202.0	63.8	70.2	68.0	85.5	88.8	69.1	77.6	69.3	90.0	92.3	102.1
G	218.7	76.4	56.4	81.3	90.9	93.3	84.0	63.6	90.6	91.4	94.5	110.0
H	215.0	66.0	57.7	80.7	77.2	89.4	75.7	63.5	86.7	85.7	91.0	99.3

Table 7.1 – Results from AllX scaffold library screening against the 8-17 DNAzyme. (A) Well layout, including all four controls. High control (4X, Column 1, Rows E-H) contains 200 nM DNAzyme. Mid control (8X, Column 12) contains 15 nM DNAzyme. Lo control (4X, Column 1, Rows A-d) contains 0 nM DNAzyme. Solvent control (3X, Column 6, Rows F-H) contains buffer with DMF, with no compounds added. Compound reference numbers are listed in respective wells. (B) Relative fluorescent values, based off of the averaged results of all 8 mid control replicates establishing a baseline of 100% activity. No activators were found. The top two inhibitors are highlighted in yellow and green.

Upon determination of final assay parameters, the AllX library was successfully screened. **Table 7.1A** shows the well layout of the library screen on a 96 well plate. Scaffold libraries were screened at two different concentrations, 1 mg/mL listed under the yellow column and 0.5 mg/mL listed under the blue column. **Table 7.1B** contains the relative fluorescent response of each well displayed as a percent of activation compared to the mid-control samples (15 nM DNAzyme, no compound, EDTA quenched after 20 minutes). Although no activator compounds were discovered, we found two promising inhibitors - a nitrosamine-based compound being a very strong inhibitor (85% inhibition) and a N-methyltriamine being a decent inhibitor (52% inhibition), (**Figure 7.3**).

Figure 7.3 – Structures of two inhibitors of 8-17 DNAzyme activity. R groups indicate positions of modification. (A) Nitrosamine (B) N-methyltriamine.

7.3 Conclusions

In the work presented here, we screened the 8-17 DNAzyme against the AllX scaffold library. We discovered one scaffold based on nitrosamine that resulted in excellent inhibition (~85%) and several others that resulted in ~50% inhibition, the best one being N-methyltriamine. The inhibition mechanism for these compounds is currently unknown and would require further experimentation. Although no activators were found in this study, the work here presents a path for additional screening, either using different compound libraries such as the Torrey

Pines (TPIMS) combinatorial library or different DNAzymes such as 10-23. Newly discovered DNAzymes, such as the DNA-cleaving I-R3, may employ alternative hydrolysis strategies and would therefore be exciting candidates. Additionally, DNAzymes that catalyze reactions other than RNA transesterification may also be screening candidates for their respective chemistries, allowing them to be more kinetically competitive with RNA-cleaving DNAzymes.

7.4 Materials and Methods

7.4.1 Oligonucleotide Sequences

Oligonucleotide sequences for the enzyme and substrate are listed 5'-3', and substrate binding regions are color-coded. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). DNAzymes, inhibitors, and input strands were ordered purified with standard desalting. DNA/RNA chimeric FRET reporter substrates were ordered purified using RNase-free HPLC. All sequences are listed in **Appendix 5**, along with their respective concentrations in each experiment. Oligonucleotides were resuspended in RNase-free H_2O (Sigma-Aldrich) in accordance with the manufacturer-provided specifications at a stock concentration (50 μM). Working stocks were made by adding the resuspended oligonucleotide solution (50 μL) into buffer (950 μL). All reactions were run in a buffer of NaCl (1M), HEPES (50 mM), and $ZnCl_2$ (1mM), at pH 7.0, unless otherwise noted (through the addition of DMSO, DMF, or the quenching by EDTA).

7.4.2 Parameter Control Experiments

DNAzyme sequence #1 (Dz1) and its corresponding substrate (Sub1) were used for the actual screening assay, and **Figures 7.2 and A5.2-A5.4** controls for activity and fluorescent response. **Figure A5.1 and A5.5** demonstrate the fluorescent response of Dz2 and Sub2 in the presence of DMSO and DMF and varying pH, although these sequences were ultimately not used in the screening. In each case, the reaction was initiated through the addition of the corresponding DNAzyme. Enzyme performance was characterized on a PTI (Birmingham, NJ) Quantamaster-40 fluorimeter at an excitation wavelength of 492 nm and an emission wavelength of 518 nm for Sub1 and an excitation wavelength of 648 nm and an emission wavelength of 668 nm for Sub2.

7.4.3 Scaffold Library Screening

Screening was performed against the AllX scaffold library. Compounds were initially at a stock concentration of 20 ug/mL and 10 ug/mL in DMF; addition to DNAzyme and substrate-containing wells diluted these compounds to a final concentration of 1 ug/mL and 0.5 ug/mL, respectively. Each control contains only enzyme and substrate, the low control containing 0 nM enzyme, the mid control containing 15 nM enzyme, and the high control containing 200 nM enzyme, as determined in **Figure 7.2**. The reaction was initiated with the addition of substrate, to a final concentration of 500 nM. The reaction was quenched by 10 mM EDTA after 20 min and immediately read out on a 1420-040 Victor 3V multilabel fluorescent plate reader (Perkin-Elmer).

7.5 References

(1) *Biomolecular Information Processing - From Logic Systems to Smart Sensors and Actuators*; Katz, E., Ed.; Wiley-VCH: Weinheim, Germany, 2012.

(2) Katz, E.; Privman, V. *Chem Soc Rev* **2010**, *39*, 1835.

(3) Katz, E.; Wang, J.; Privman, M.; Halamek, J. *Anal Chem* **2012**, *84*, 5463.

(4) Wang, J.; Katz, E. *Anal Bioanal Chem* **2010**, *398*, 1591.

(5) Zhou, J.; Arugula, M. A.; Halamek, J.; Pita, M.; Katz, E. *J Phys Chem B* **2009**, *113*, 16065.

(6) Martin, C. H.; Nielsen, D. R.; Solomon, K. V.; Prather, K. L. *Chem Biol* **2009**, *16*, 277.

(7) Gilbert, W. *Nature* **1986**, *319*.

(8) Joyce, G. F. *Nature* **2002**, *418*, 214.

(9) Breaker, R. R.; Joyce, G. F. *Chem Biol* **1995**, *2*, 655.

(10) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.

(11) Santoro, S. W.; Joyce, G. F. *Biochemistry* **1998**, *37*, 13330.

(12) Li, Y.; Breaker, R. R. *Curr Opin Struct Biol* **1999**, *9*, 315.

(13) Li, Y.; Lu, Y. *Functional Nucleic Acids for Analytical Applications*; Springer, 2009.

(14) Baum, D. A.; Silverman, S. K. *Cell Mol Life Sci* **2008**, *65*, 2156.

(15) Silverman, S. K. *Acc Chem Res* **2009**, *42*, 1521.

(16) Silverman, S. K.; Baum, D. A. *Methods Enzymol* **2009**, *469*, 95.

(17) Walsh, S. M.; Sachdeva, A.; Silverman, S. K. *J Am Chem Soc* **2013**, *135*, 14928.

(18) Breaker, R. R.; Emilsson, G. M.; Lazarev, D.; Nakamura, S.; Puskarz, I. J.; Roth, A.; Sudarsan, N. *RNA* **2003**, *9*, 949.

(19) Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. *RNA* **2003**, *9*, 907.

(20) Li, Y.; Breaker, R. R. *Journal of the American Chemical Society* **1999**, *121*, 5364.

(21) Soukup, G. A.; Breaker, R. R. *RNA* **1999**, *5*, 1308.

(22) Adams, R. L. P.; Knowler, J. T.; Leader, D. P. *The biochemistry of the nucleic acids*; Chapman and Hall, 1992.

(23) Raines, R. T. *Chemical reviews* **1998**, *98*, 1045.

(24) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. *Nucleic Acids Res* **2000**, *28*, 481.

(25) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.

(26) Dass, C. R. *Trends Pharmacol Sci* **2004**, *25*, 395.

(27) Dass, C. R.; Choong, P. F.; Khachigian, L. M. *Mol Cancer Ther* **2008**, *7*, 243.

(28) Willner, I.; Shlyahovsky, B.; Zayats, M.; Willner, B. *Chem Soc Rev* **2008**, *37*, 1153.

(29) Brown, C. W., III; Lakin, M. R.; Stefanovic, D.; Graves, S. W. *ChemBioChem* **2014**, *15*, 950.
(30) Houghten, R. A.; Pinilla, C.; Giulianotti, M. A.; Appel, J. R.; Dooley, C. T.; Nefzi, A.; Ostresh, J. M.; Yu, Y.; Maggiora, G. M.; Medina-Franco, J. L.; Brunner, D.; Schneider, J. *J Comb Chem* **2008**, *10*, 3.
(31) Santos, R. G.; Appel, J. R.; Giulianotti, M. A.; Edwards, B. S.; Sklar, L. A.; Houghten, R. A.; Pinilla, C. *Molecules* **2013**, *18*, 6408.
(32) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.
(33) Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D. *Proc Natl Acad Sci U S A* **2013**, *110*, 5386.

Chapter 8.
Surface-based Conjugation for DNA and Protein Compartmentalization

8.1 Introduction

Solution-based DNA computation has generated significant interest, and many groups have developed molecular logic gates based on DNA. The principles have been utilized to construct circuits of impressive scale. However, the continued development of complex computational interactions will likely require a more efficient method for scaling up while minimizing spurious interactions. Although significant work has gone into inhibition and sequestration of active sequences in solution-based assays, including our own efforts in DNAzyme cascading, the lack of physical separation of circuit components ensures that leakage through cross-reactivity will continue to hinder circuit development and growth. Work on the physical separation of circuit components has been limited; to our knowledge, only two instances of direct conjugation of DNA logical elements to microspheres have been reported to date[1,2]. This represents a field with significant future potential.

As flow cytometry excels in the analysis of many separate populations, it appears ideally suited for the development of multi-analyte computational analysis. There are many different modalities available for the coupling of computational elements to microspheres suitable for analysis by flow cytometry, each with application-specific advantages. Given the nascence of this field, characterization of different attachment methods to microspheres would fill an important need.

Direct attachment of DNA strands and logic gates to microspheres has been used in prior studies[1,2]. This method has obvious advantages for design simplicity and preparation and fulfills one of the key constraints for use in biosensor devices. For certain applications, this approach is highly beneficial and warrants further characterization. In other cases, such as DNA-substrate interactions, immobilization of circuit elements may be problematic. To address this issue, the exploration of alternative modalities that can provide element mobility to facilitate component interactions, such as biological membranes, would be highly beneficial.

Membranes have been demonstrated in several new areas of DNA nanotechnology, including lipid vesicles[3,4], planar bilayers[5,6], and mammalian cell lines[7]. The latter case was an important demonstration both for the performance of DNA computation on a cell surface as well as the use of flow cytometry analysis of computational systems. Although direct cellular analysis is of high interest, cells are complex systems composed of many different molecular species and cell culture work can be expensive and time-consuming. These limitations indicate the necessity of a model system capable of both DNA and protein interaction and amenable to flow cytometry analysis. One promising platform is supported lipid bilayers (SLB)[8], in which a pure lipid bilayer is coated on the surface of silica microspheres, or “cushioned” bilayers for polystyrene particles[9]. These SLBs have been shown to retain natural lipid fluidity, ensuring that elements associated with the SLB are able to freely diffuse in a 2D plane. This approach has been demonstrated for the sequestration and release of

dye[10], chemotherapeutic compounds[11] and interfering oligonucleotides[12], along with integration of transmembrane proteins[13].

The use of SLBs may be of use for DNA computational architectures that include proteins, potentially enabling the discovery of new circuit behaviors. This system may also have use in the continued development and characterization of synthetic transcriptional circuits[14-16]. For these reasons, it is important to characterize the use of microspheres and SLBs for both proteins as well as DNA, which may be of use to the wide variety of DNA computation implementations currently being explored. One of the aspects we explored in depth was the use of supported lipid bilayers for the stable and specific attachment of hexahistidine-tagged proteins. This is a commonly used protein tag for the purification of proteins, and would be highly desirable for use in bioassay development and the 2D constraint of proteins on the surface of SLBs, which has important implications for surface-based synthetic circuit designs. However, reports on the use of histidine-tag immobilization of proteins on microspheres have been mixed.

8.2 Background on histidine tags for assay development

Immobilized metal affinity chromatography (IMAC) is one of the most widely used technologies for protein purification. The small hexa- or deca-histidine tags have significant advantages over other affinity tags, due to their lack of steric hindrance and ease of use. These benefits have made histidine tags attractive targets for use in protein-based biological assays. The histidine tag has a relatively weak (~μM K_d) monovalent binding affinity for any of its

receptors[17]. This results in a relatively rapid equilibration of bound and unbound states, which is useful for protein chromatography.

However, the weak binding affinity of his-tags for their known partner receptors has made the development of biomimetic surfaces, which require stable and directional binding of proteins, problematic. While several groups have reported stable his-tag binding on supported surfaces[17-20], others have shown that his-tag binding is insufficient for biological assays demonstrated discrepancies between binding on different surfaces[21-23]. Previous studies on large, planar systems have proposed two mechanisms for achieving stable protein binding: rebinding and avidity. The model of rebinding proposes that upon dissociation from its receptor, a ligand will instead rebind to the surface with a different receptor, rather than diffuse from the surface[24,25]. Thus, rebinding is highly dependent on surface receptor density. Rebinding effects were first described for the Ni-NTA system using surface plasmon resonance to measure the binding of polyhistidine tagged proteins a high density of Ni-NTA groups in a porous dextran layer[17]. In this study, the high receptor density found on the dextran layer was postulated to increase the apparent affinity of his-tagged proteins. This model was also the proposed mechanism of another low affinity tag (glutathione-S-transferase) that was effective for protein immobilization on the surface of microspheres with high receptor density, despite its low monovalent affinity for its surface bound receptor[26]. Microsphere studies with Ni-NTA have reported his-tagged binding to be insufficient for use in assay development[21].

Multivalency of his-tags has also been proposed as a mechanism to achieve stable protein binding via avidity effects derived from simultaneous molecular binding interactions with multiple Ni-NTA groups on a surface[27]. The use of mobile receptors in supported lipid bilayers[28], can improve avidity. Mobility allows for receptor rearrangement, increasing the apparent affinity of his-tagged proteins. Supported lipid bilayers have been shown to provide an excellent platform for directionally oriented protein binding via histidine chelator lipids[29,30], especially at high densities[31,32]. Stable binding has been reported using 1% Ni-NTA incorporated into lipid bilayers supported on a silica chip[27]. Similar attempts were made to transfer such bilayers to microsphere surfaces, however, the bilayer proved to be unstable in these efforts[21].

8.3 Results

8.3.1 Direct Attachment to Microspheres

The standard of DNA computation on microspheres has been the use of direct non-covalent attachment through conjugation via biotin-streptavidin. To provide a reference point for our surface-based DNAzyme characterizations, we sought to implement direct attachment to observe DNAzyme cleavage off of streptavidin-coated polystyrene microspheres (Spherotech, Lake Forest, IL). Using a biotinylated chimeric substrate labeled with FAM (a fluorescein derivative), we attached the substrate to the surface of the microsphere (**Figure 8.1A**). With the addition of the E6 DNAzyme, we observed a loss of fluorescence from the bead surface over the course of 24 hours (**Figure 8.1B**). Although this demonstrated the feasibility of direct attachment in agreement with earlier studies, this differed

from the implementation by Stojanovic[2], which also attached the DNAzyme to the surface of the microsphere. In our characterization, the E6 DNAzyme was added to the solution and could freely diffuse to the bead surface. Thus, this effort was a first step in the binding and substrate cleavage with a microsphere/flow cytometry-based format; additional circuit element development for the implementation of DNA logic and computation is currently underway.

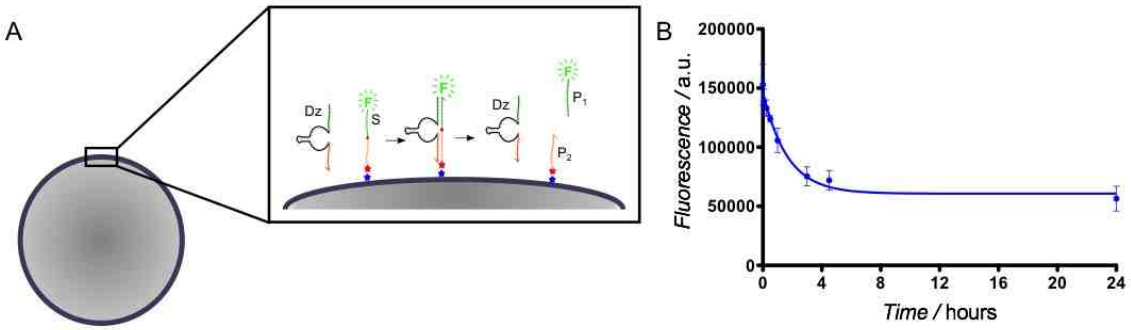

Figure 8.1 – Chimeric substrate cleavage from a microsphere with an E6 DNAzyme, measured by a loss of fluorescence.

8.3.2 Covalent attachment of DNA to SLBs

Although the direct, non-covalent attachment of DNA directly to microspheres is of interest due to its simplified set up, many elements of DNA logic and computation could benefit from a mobile yet 2D constrained interaction. Although circuit elements would be topologically constrained in the same manner and orientation as with direct attachment, they would have a significant advantage in their ability to interact with other circuit elements attached to the same surface outside of their local attachment area. To implement this mobility, we attached the chimeric DNAzyme substrate to supported lipid bilayers. The fluidity of the lipid leaflets ensures that the DNA elements would have the mobility to interact

with multiple circuit elements. While the biotin-streptavidin attachment method has potential use, it also may be problematic. For SLBs, this approach requires a biotinylated lipid, with the attachment of the streptavidin followed by the attachment of a biotinylated DNA oligonucleotide. Although possible using tetrameric streptavidin, this can result in ostrich quenching of the fluorescein labeled DNA substrate strand[33], in which fluorescein can associate with the biotin-binding pocket of streptavidin, resulting in altered fluorescence. This may prevent access of the DNAzyme to its substrate. Given the flexibility of single-stranded DNA, this interaction should be avoided. To avoid this self-quenching, we explored an alternative method of conjugation, adapting a previously established method for coupling a thiolated DNA strand to a maleimide lipid (Avanti Polar Lipids, Alabaster, Alabama) through a thioether linkage reaction[34,35]. This results in a covalent attachment of DNA to the surface, which can then be read out on the flow cytometer (**Figure 8.2**).

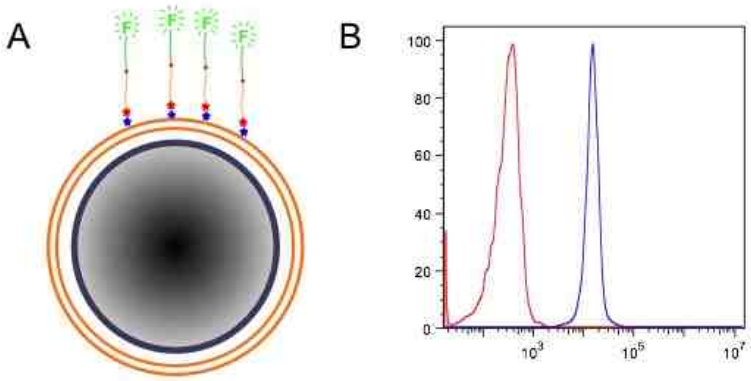

Figure 8.2 – Coupling of a chimeric substrate to supported lipid bilayers through a covalent thioether conjugation. (A) Cartoon representation (B) Flow cytometry analysis of labeled (blue) vs. unlabeled (red) beads.

After the demonstration of successful coupling, free E6 DNAzyme was added to the solution and allowed to freely diffuse to the bead surface where it cleaved substrate. This cleavage was measured as a loss of fluorescence (**Figure 8.3**). This is the first successful demonstration of DNAzyme substrate cleavage from supported lipid bilayers.

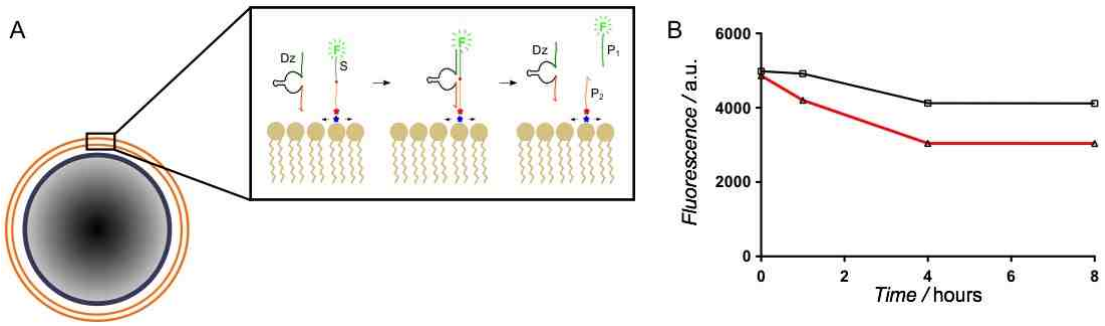

Figure 8.3 – (A) Attachment of a chimeric substrate to supported lipid bilayers and cleavage using an E6 DNAzyme. (B) Loss of fluorescence measured by flow cytometry, indicative of substrate cleavage (red) from the negative control (black) without the addition of E6.

8.3.3 Non-covalent attachment of proteins to SLBs

The attachment of proteins to supported lipid bilayers is also of high importance, due to their use in various synthetic DNA circuits[14-16,36]. One of the most commonly used tags is the 6xHis tag, a repeat of six consecutive histidine amino acids that chelates nickel ions complexed with nitrilotriacetic acid (Ni-NTA), due to its small size and simple purification techniques. We explored the use of a mobile lipid bilayer for the rearrangement for Ni-NTA receptors for the development of stable his-tagged proteins to microspheres, which has not been previously demonstrated in the literature[21]. Using high concentrations of Ni-NTA added to an EggPC lipid bilayer, we were able to demonstrate stable attachment

of GFP to microspheres, even over several days (**Figure 8.4A**), and elution with imidazole resulting in <1% non-specific binding (**Figure 8.4B**).

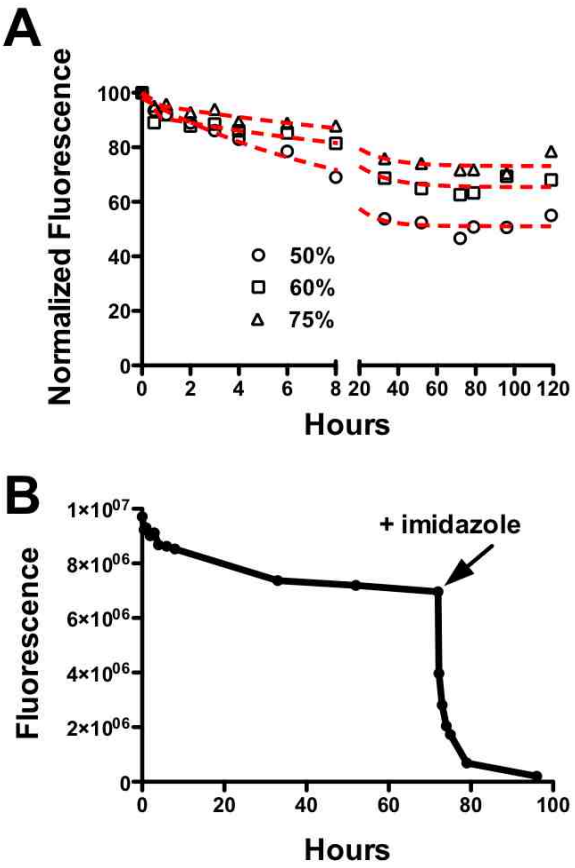

Figure 8.4 – High percentages and long pre-incubation times yield low dissociation rates and highly specific binding. (A) High percentages of Ni-NTA MSLBs demonstrate a biphasic dissociation rate and retain stably bound his-tagged GFP for 120 hours. (B) Addition of 250 mM imidazole after 72 hours shows non-specific binding is less than 1%.

To ensure that this attachment strategy would be sufficient to retain multiple populations, we attached GFP to a single population of beads under various conditions (**Figure 8.5**) and mixed this with an unlabeled population. Several time points were taken to monitor whether the GFP dissociated from the surface of the bound population and associated with the unlabeled population. Optimal results

occurred under a 50% labeling density after a 20 hour incubation with GFP, which showed that populations remained distinct, even after a day (20 hours).

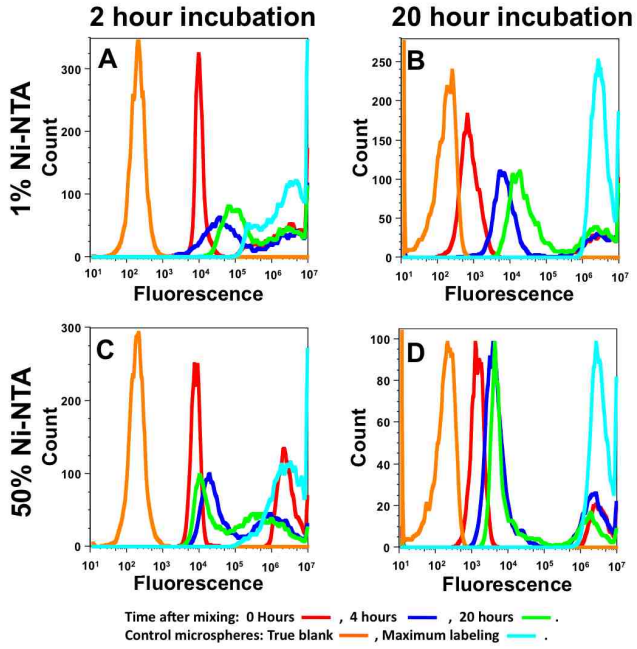

Figure 8.5 - Fluorescence transfer between unlabeled and GFP labeled microsphere populations. Low receptor density incubation times yields populations that rapidly converge, while increasing receptor density and incubation time significant stabilizes bead fluorescence, reducing crosstalk. (A) 1% Ni-NTA labeling density, 2hr incubation time (B) 1% Ni-NTA labeling density, 20hr incubation time (C) 50% Ni-NTA labeling density, 2hr incubation time (D) 50% Ni-NTA labeling density, 20hr incubation time.

8.3.4 Direct peptide insertion into SLBs

As transmembrane signaling is a critical mechanism for the cellular compartmentalization of signaling, we sought to characterize the use of supported lipid bilayers for the direct insertion of a transmembrane peptide and the interaction with a synthetic protease construct[37]. After solubilizing the bilayers using a mild detergent, the peptide was added to supported lipid bilayers and the hydrophobic sequence spontaneously inserted into the bilayer. After removing the detergent, the fluorescent tag was cleaved from the peptide upon addition of

the protease (**Figure 8.6**). This demonstrates the viability of monitoring enzymatic reactions on the surface of supported lipid bilayers.

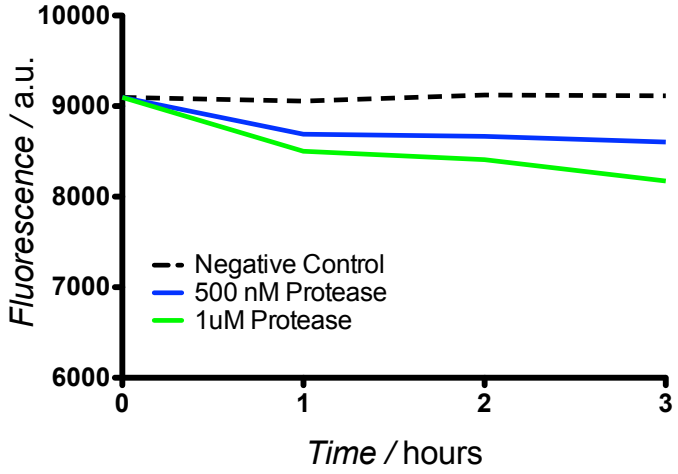

Figure 8.6 – Cleavage of a transmembrane peptide from the surface of a silica microsphere supported lipid bilayer. Loss of fluorescence was observed with the addition of 500 nM (blue) or 1 μM (green) protease.

8.3.5 Multiplex analysis of SLBs

The use of flow cytometry for monitoring surface compartmentalization of DNA computational elements holds considerable promise for monitoring the interactions between multiple bead populations simultaneously. While this is easily done with hard-dyed polystyrene beads, the use of silica microspheres for supported lipid bilayers makes multiplexing a significant challenge. Here, we explored the development of multiplexable supported lipid bilayers with the use of multiple fluorescence lipids. The use of a pyrene-labeled lipid (Ex: 350, Em: 405) along with a carboxyfluorescein-labeled lipid (Ex: 488, Em: 530) allowed the discrimination of multiple separate populations, which can be gated on each fluorescent channel respective to the concentration of lipid added to each bead population (**Figure 8.7**). With the use of different fluorophores, such as one in the

far red channel, biochemical reactions can be analyzed off of a multiplexed supported lipid bilayer platform.

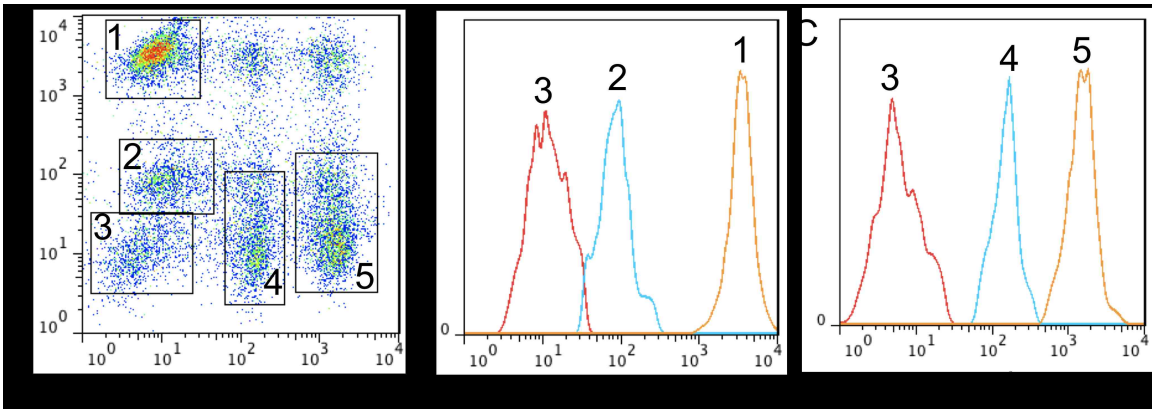

Figure 8.7 – Multiplexed supported lipid bilayers with the addition of pyrene (x-axis) and carboxyfluorescein (y-axis). Flow cytometry analysis (using a Mo-Flo cytometer) was able to discern separate populations of beads with a low (0%), medium (0.01%), and high (1%) of labeled lipids.

8.4 Conclusions

The work demonstrated here shows the feasibility of surface-based attachment methods for potential use in DNA computation. We have repeated the demonstration of direct microsphere attachment and DNAzyme activity shown previously[2]. We have also demonstrated a significant amount of work towards the characterization of supported lipid bilayers for attachment of elements that could be used in DNA computing circuits, including DNA, proteins, and peptides. SLBs offer a wide variety of attachment chemistries, including covalent, non-covalent, and direct insertion mechanisms. The use of multiple fluorescent lipids provide a mechanism for the multiplexing of supported lipid bilayers, ensuring that this technology can still be used to monitor multiple populations simultaneously.

8.5 Future Directions

Successful demonstration of multiple attachment strategies opens the prospect of coupling and interaction of multiple DNA computing elements to form a communicating network. Work is underway in our laboratory to develop strand displacement and DNAzyme displacement for direct microsphere attachment as well as for supported lipid bilayers. Although we saw promising results for monitoring cleavage reactions, both DNA and protease, on the surface of SLBs, optimization to improve these reactions would be essential for the generation of robust and responsive DNA circuits. Additional coupling strategies, such as click chemistry,[38] should also be explored.

8.6 Materials and Methods

8.6.1 General experimental conditions

Unless otherwise noted, all experiments were conducted on an Accuri C6 flow cytometer (Becton-Dickinson, San Jose, CA) using a 488 nm laser for excitation and 530/30 filter for emission.

E6 sequence: CTCTTCAGCGATGGCGAAGCCCACCATGTTAGTGA

E6 substrate sequence (Direct attachment): /5-BiosG/ TTT TCA CTA TrAG GAA GAG TTT /36-FAM/

E6 substrate sequence (Thiol attachment): /5ThioMC6-D/ TCA CTA TrAG GAA GAG /iSpC3 /36-FAM/

8.6.2 Synthesis of supported lipid bilayers

The microspheres were prepared as previously described[39]. Briefly, 1 mM Ni-NTA lipids were added in a mol/mol% ratio to generic EPC lipids (Avanti Polar Lipids). The chloroform solution was dried out and reconstituted in PBS buffer, and the lipid solution was then extruded or sonicated until clear to form vesicles. Vesicles were vortexed for 45 min with 10 um nanoporous silica microspheres with 100 Å pores (GFS Chemicals), surface treated to be hydrophilic. The microsphere supported lipid bilayers (MSLB) were then washed three times in buffer.

8.6.3 DNA-SLB covalent attachment

Vesicles contained 10% MBP-PE (Avanti Polar Lipids, Alabaster, Alabama) in DOPC were prepared in VP buffer (20 mM citric acid, 35 mM Na_2PO_4, 108 mM NaCl, 1 mM EDTA, pH 4.5) and sonicated until clear using a bath sonicator. Reducing buffer (20 mM citric acid, 35 mM Na_2PO_4, 108 mM NaCl, 1 mM EDTA, 20 mM DTT, pH 5.0) was degassed under argon for 90 min. 5 uL of 1 mM thiolated DNA was added to 0.5 mL reducing buffer for a 10 μM final concentration. This solution was eluted on a Nap-5 column after equilibration with VP buffer. DNA was added in a 1000X excess of MBP concentration and mixed under a constant argon stream for up to 8hrs with stirring.

8.6.4 GFP experiments

Microspheres were incubated for various times with a thermostable GFP protein containing a C-terminal hexahistidine tag, expressed from the pET-CK3-sfGFP[40]. This allowed for all protein assays to be run at room temperature to monitor bilayer stability while minimizing protein denaturation.

8.7 References

(1) Frezza, B. M.; Cockroft, S. L.; Ghadiri, M. R. *J Am Chem Soc* **2007**, *129*, 14875.

(2) Yashin, R.; Rudchenko, S.; Stojanovic, M. N. *J Am Chem Soc* **2007**, *129*, 15581.

(3) Perrault, S. D.; Shih, W. M. *ACS nano* **2014**.

(4) Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H.; Simmel, F. C. *Science* **2012**, *338*, 932.

(5) Burns, J. R.; Gopfrich, K.; Wood, J. W.; Thacker, V. V.; Stulz, E.; Keyser, U. F.; Howorka, S. *Angew Chem Int Ed Engl* **2013**.

(6) Burns, J. R.; Stulz, E.; Howorka, S. *Nano Lett* **2013**, *13*, 2351.

(7) Rudchenko, M.; Taylor, S.; Pallavi, P.; Dechkovskaia, A.; Khan, S.; Butler, V. P., Jr.; Rudchenko, S.; Stojanovic, M. N. *Nat Nanotechnol* **2013**, *8*, 580.

(8) Bayerl, T. M.; Bloom, M. *Biophysical journal* **1990**, *58*, 357.

(9) Renner, L.; Pompe, T.; Lemaitre, R.; Drechsel, D.; Werner, C. *Soft Matter* **2010**, *6*, 5382.

(10) Piyasena, M. E.; Zeineldin, R.; Fenton, K.; Buranda, T.; Lopez, G. P. *Biointerphases* **2008**, *3*, 38.

(11) Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Padilla, D.; Durfee, P. N.; Brown, P. A.; Hanna, T. N.; Liu, J.; Phillips, B.; Carter, M. B.; Carroll, N. J.; Jiang, X.; Dunphy, D. R.; Willman, C. L.; Petsev, D. N.; Evans, D. G.; Parikh, A. N.; Chackerian, B.; Wharton, W.; Peabody, D. S.; Brinker, C. J. *Nat Mater* **2011**, *10*, 389.

(12) Ashley, C. E.; Carnes, E. C.; Epler, K. E.; Padilla, D. P.; Phillips, G. K.; Castillo, R. E.; Wilkinson, D. C.; Wilkinson, B. S.; Burgard, C. A.; Kalinich, R. M.; Townson, J. L.; Chackerian, B.; Willman, C. L.; Peabody, D. S.; Wharton, W.; Brinker, C. J. *ACS Nano* **2012**, *6*, 2174.

(13) Buranda, T.; Huang, J.; Ramarao, G.; Ista, L. K.; Larson, R. S.; Ward, T. L.; Sklar, L. A.; Lopez, G. P. *Langmuir* **2003**, *19*, 1654.

(14) Elowitz, M. B.; Leibler, S. *Nature* **2000**, *403*, 335.

(15) Kim, J.; White, K. S.; Winfree, E. *Mol Syst Biol* **2006**, *2*, 68.

(16) Kim, J.; Winfree, E. *Mol Syst Biol* **2011**, *7*, 465.

(17) Nieba, L.; Nieba-Axmann, S. E.; Persson, A.; Hamalainen, M.; Edebratt, F.; Hansson, A.; Lidholm, J.; Magnusson, K.; Karlsson, A. F.; Pluckthun, A. *Anal Biochem* **1997**, *252*, 217.

(18) Gershon, P. D.; Khilko, S. *J Immunol Methods* **1995**, *183*, 65.

(19) Dietrich, C.; Boscheinen, O.; Scharf, K. D.; Schmitt, L.; Tampe, R. *Biochemistry* **1996**, *35*, 1100.

(20) Dietrich, C.; Schmitt, L.; Tampe, R. *Proc Natl Acad Sci U S A* **1995**, *92*, 9014.

(21) Lauer, S. A.; Nolan, J. P. *Cytometry* **2002**, *48*, 136.

(22) Sigal, G. B.; Bamdad, C.; Barberis, A.; Strominger, J.; Whitesides, G. M. *Anal Chem* **1996**, *68*, 490.

(23) Dorn, I. T.; Pawlitschko, K.; Pettinger, S. C.; Tampé, R. *Biological chemistry* **1998**, *379*, 1151.
(24) Erickson, J.; Goldstein, B.; Holowka, D.; Baird, B. *Biophysical journal* **1987**, *52*, 657.
(25) Lagerholm, B. C.; Thompson, N. L. *The Journal of Physical Chemistry B* **2000**, *104*, 863.
(26) Tessema, M.; Simons, P. C.; Cimino, D. F.; Sanchez, L.; Waller, A.; Posner, R. G.; Wandinger-Ness, A.; Prossnitz, E. R.; Sklar, L. A. *Cytometry Part A* **2006**, *69*, 326.
(27) Nye, J. A.; Groves, J. T. *Langmuir* **2008**, *24*, 4145.
(28) Vaz, W. L.; Goodsaid-Zalduondo, F.; Jacobson, K. *FEBS letters* **1984**, *174*, 199.
(29) Dorn, I. T.; Eschrich, R.; Seemüller, E.; Guckenberger, R.; Tampé, R. *Journal of molecular biology* **1999**, *288*, 1027.
(30) Trammell, S. A.; Wang, L.; Zullo, J. M.; Shashidhar, R.; Lebedev, N. *Biosensors and Bioelectronics* **2004**, *19*, 1649.
(31) Thess, A.; Hutschenreiter, S.; Hofmann, M.; Tampé, R.; Baumeister, W.; Guckenberger, R. *Journal of Biological Chemistry* **2002**, *277*, 36321.
(32) Somersalo, K.; Anikeeva, N.; Sims, T. N.; Thomas, V. K.; Strong, R. K.; Spies, T.; Lebedeva, T.; Sykulev, Y.; Dustin, M. L. *Journal of Clinical Investigation* **2004**, *113*, 49.
(33) Wu, Y.; Simons, P. C.; Lopez, G. P.; Sklar, L. A.; Buranda, T. *Analytical biochemistry* **2005**, *342*, 221.
(34) Martin, F. J.; Hubbell, W. L.; Papahadjopoulos, D. *Biochemistry* **1981**, *20*, 4229.
(35) Martin, F. J.; Papahadjopoulos, D. *Journal of Biological Chemistry* **1982**, *257*, 286.
(36) Xie, Z.; Wroblewska, L.; Prochazka, L.; Weiss, R.; Benenson, Y. *Science* **2011**, *333*, 1307.
(37) D'Arcy, A.; Chaillet, M.; Schiering, N.; Villard, F.; Lim, S. P.; Lefeuvre, P.; Erbel, P. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2006**, *62*, 157.
(38) Kolb, H. C.; Finn, M.; Sharpless, K. B. *Angewandte Chemie International Edition* **2001**, *40*, 2004.
(39) Davis, R. W.; Flores, A.; Barrick, T. A.; Cox, J. M.; Brozik, S. M.; Lopez, G. P.; Brozik, J. A. *Langmuir* **2007**, *23*, 3864.
(40) Kiss, C.; Fisher, H.; Pesavento, E.; Dai, M.; Valero, R.; Ovecka, M.; Nolan, R.; Phipps, M. L.; Velappan, N.; Chasteen, L. *Nucleic acids research* **2006**, *34*, e132.

Chapter 9.
Conclusions and Future Directions

9.1 Conclusions

9.1.1 Development of DNAzyme displacement gates

We have demonstrated the successful regulation of DNAzyme activity by strand displacement for the construction of molecular logic gates. This combines the advantages of both technologies: the use of DNAzymes can provide DNA circuits with a rich diversity of biochemical reactions and innate catalysis without the need for additional strands and hybridization reactions, while strand displacement offers a sensitive and specific mechanism for programming molecular interactions. Each has been extensively but independently characterized, and unifying these biochemical concepts can have far-reaching implications for the potential development of DNA circuits and computers. We also demonstrated the introduction of mismatches for the release of a DNAzyme, which furthers the understanding of strand displacement interactions and capabilities. This represents an important first effort in creating robustly regulated DNAzymes for molecular logic and for the development of hybrid logic circuits.

9.1.2 Multi-layer DNAzyme cascades

Using the DNAzyme displacement logic gates, we developed a structured chimeric substrate (SCS) that enabled a DNA signal to be propagated from one DNAzyme logic gate to another. This was a major advance for the development of DNAzyme cascades; there had only been minor success in this area[1,2], which had not been broadly applicable to the DNA computational community at large. The rational design process for the SCS structure was described in detail,

demonstrating both the wide variety of structural changes and the large effects they can have on signal propagation. The success of the SCS demonstrates the efficacy of this approach for the design of dynamic DNA nanostructures; however, the granularity at which alterations were made implies a large potential structure-space, even for the design of sequences that contain multiple design constraints.

The SCS molecule served as an effective communication module between two DNAzyme displacement gates, which enabled the construction of a five layer DNAzyme cascade, the longest to date. This molecule can interact with molecular logic gates outside of the 8-17 DNAzyme displacement gates used in the cascade, yet another step in the creation of hybrid DNA circuits. These cascades are also amenable to use in bioassays, as they are resistant to background and do not require additional purification. The potential of this system was demonstrated in a representative three-input Dengue serotyping assay, in which the circuit responded with a positive output response only in the presence of all three targets. This has implications both for molecular logic and computation as well as the understanding and development of synthetic enzyme cascades, which may aid in the understanding of complex signal propagation and transduction networks.

9.1.3 A versatile, modular DNA biosensor

We have designed a new DNA biosensor structure that couples each reporter DNAzyme to an input recognition domain, which we call the detection domain. These domains are orthogonal to each other and therefore can vary

independently, which gives this sensor a broad flexibility both for the input as well as the output. We have extensively characterized this biosensor structure, optimizing it for low leakage while retaining high activation even in the continuous presence of the fuel strand. As with the cascades, this biosensor is highly specific and performs well in a random DNA background under minimal purification. This biosensor was shown to perform well with the detection of many different targets of a multiplex assay, with a limit of detection below 10 pM. Given its modularity, we altered the detection domain to show response against a variety of inputs, including RNA, ssDNA, dsDNA, and small molecules such as ATP. We also demonstrated initial success to in cascading this new gate structure. This has significant implications for the combination of biosensing with molecular logic for higher order computation and the development of intelligent biosensors.

9.1.4 High-throughput screen of the 8-17 DNAzyme

The 8-17 DNAzyme was screened against a scaffold library to look for molecular cofactors that could lead to catalytic rate enhancement. As DNAzymes are artificially selected under specific conditions, the use of high-throughput screens for testing large numbers of randomized compounds has the potential to uncover new and unique catalytic behaviors. Although no activating compounds were discovered, two strongly inhibitory scaffolds were identified. These compounds could have potential for establishing a threshold for *in vitro* assays or regulating the *in vivo* activity of therapeutic DNAzymes.

9.1.5 Surface-based DNA and protein attachment

We have established a baseline characterization of several different modalities for surface-based DNA computation. The compartmentalization of DNA circuit units provides a straightforward method to standardizing circuit components via physical separation, which is a critical step towards rapid scaling of circuit size and complexity, allowing circuits to integrate more inputs and leading to the development of intelligent biosensors. Due to their ease of use, we have demonstrated the attachment and cleavage of chimeric substrates from polystyrene beads, which holds promise for the development of hybrid bead-solution interactions. However, to overcome the limitations of previous work using direct attachment to beads[3,4], we have characterized the binding of both DNA and proteins to supported lipid bilayers[5], which have the potential to create mobile DNA logic elements constrained on bead surfaces. We have explored several different attachment strategies, such as covalent thioether attachment of DNA, non-covalent 6xHis-tag/Ni-NTA of proteins, and direct peptide insertion. Although further characterization is necessary, each of these shows considerable promise for the potential use of supported lipid bilayers for surface-based interactions.

9.2 Future Directions

Taken as a whole, the work presented in this dissertation expands the understanding of the underlying principles of DNA computation for use in biodetection applications. These principles establish a comprehensive framework with which the continued development of more sophisticated computational tools

and circuits can be carried out. Outlined below are some of the most promising directions that can now feasibly be achieved through the knowledge gained through this work. Conceptualization and implementation for many of these ideas are currently underway.

9.2.1 Modular gate cascades

Our laboratory is currently in the process of continuing work and development on the modular gate cascades. Successful implementation of a two-layer cascade using the modular gate framework has been demonstrated in **Section 6.2.5**. This gate design has already shown considerable promise as a direct biosensor, with a low limit of detection (LOD) and high specificity, sensitivity, and modularity. The development of cascades using the modular gate designs offers an apparent method for the construction of multi-layer DNAzyme cascades capable of performing molecular logic. Most importantly, the design ensures each DNAzyme is regulated by a corresponding input; thus, the inclusion of multiple modular gates into a single circuit also ensures the integration of multiple inputs, a key parameter for the development of complex computation. This can give rise to advanced circuit behaviors, examples of which are described in **Section 6.2.7**. Given the performance of individual gates as biosensors, the addition of complex logic will be a crucial next step for the development of a next-generation biosensor and ultimately, intelligent biosensors.

9.2.2 Optimization of modular gate design

Although much improvement has already been made regarding the performance of the modular gates through the introduction of mismatches into the loop toehold

for fuel strand displacement, further optimizations are highly desirable. In particular, the rate of activation against double-stranded DNA targets such as whole plasmid DNA is very slow (full activation is achieved after several days of incubation). Improving this activation rate while retaining low leakage will be a major advancement in the performance of a biosensor and open up many new target opportunities currently out of reach due to the long sensing times.

Optimization of modular gate performance will likely also target the removal of excess inhibitor, as its presence currently acts as a sink for target inputs, negatively affecting the attainable limit-of-detection of the gate. While new data indicates that magnetic bead purification hold promise for fast, isothermal removal, the excess inhibitor may not be benign as currently presumed, but may play a role in the suppression of gate leakage or the reinhibition of activated DNAzymes. Thus, while the removal of inhibitor may result in a lower detection limit, it may also increase spurious gate activation. Further characterization of these effects is warranted.

9.2.3 Detection of biological samples

Although much effort has been made toward the development of DNAzyme displacement gates, cascades, and modular gates for biological applications, including their performance in a random DNA background, their performance against real world biological sample targets has not yet been demonstrated. Efforts are currently underway to explore their use against multi-strain targets such as Dengue (a single-stranded RNA target) and Shiga-toxin producing *E. coli* (a double-stranded DNA target). One of the major concerns is the low

concentration of nucleic acid targets, likely in the low picomolar to femtomolar range, which is currently below our characterized LOD for a ssDNA target input.

9.2.4 Additional screening targets for DNAzymes

Although the screen of the 8-17 DNAzyme was unsuccessful in finding catalytic cofactors, the demonstration of the first successful high throughput screen of a DNAzyme holds considerable promise for the potential success of future targets. Although the AllX scaffold library contained a significant amount of diverse compounds, the screening of other small molecule libraries may yet yield activating compounds. Additionally, there are DNAzymes known to contain alternative mechanisms of RNA-cleavage outside of exclusive αγ catalysis, such as 10-23 or X-motif[6]. The difference in RNA catalysis may yield activation in the presence of certain compounds that could amplify or accelerate these mechanisms. Additionally, newly discovered DNA-cleaving DNAzymes[7] are of significant interest due to the use of an all-DNA substrate, and their selection under physiologically relevant conditions would make the transition to *in vivo* use a realistic goal.

9.2.5 Development of theranostic circuits

Although the work in this dissertation focused predominately on the development of the input/detection methods, the demonstration of DNAzyme cascading was a major advance for the extension of a signal cascade through the output of a DNA strand. This approach may offer a straightforward method for the development of theranostic circuits, in which the release of a DNA strand could be an active DNAzyme capable of cleaving RNA strands *in vivo*[8-10]. In this way, the

mechanism of interpreting inputs of a logic circuit developed herein could be coupled to the release of an output capable of affecting cellular functionality. Alternatively, there has been significant work and interest in the use of supported lipid bilayers for the targeted delivery of therapeutic compounds or nucleic acids to cancer cells[11-14]. If the release of these compounds could be triggered by the computational decisions of a DNA logic circuit, this would be a major step towards the development of theranostic devices.

9.2.6 In vivo *DNA computation*

The promising long-term future of DNA computation lies with the continued progress toward an *in vivo* biological computer. While there have been several key demonstrations of this concept, overall the sophistication of *in vivo* DNA-based computation is still in its infancy. Although some progress has been made as to the performance of strand displacement logic gates in cells[15], the location and availability of these gates outside endosomal compartments remains unclear. The use of DNAzymes *in vivo* has a similarly conflicted scientific record. While there have been several promising studies[8-10,16], there are also concerns about the efficacy and performance in cells[17-21] and these systems did not employ any molecular logic. Taken as a whole, one of the most anticipated advances in the field DNA computation will be the improved design and performance of autonomous, multi-component circuits under non-pristine conditions, particularly *in vivo* diagnostics.

9.3 References

(1) Gerasimova, Y. V.; Cornett, E. M.; Edwards, E.; Su, X.; Rohde, K. H.; Kolpashchikov, D. M. *ChemBioChem* **2013**, *14*, 2087.

(2) Elbaz, J.; Lioubashevski, O.; Wang, F.; Remacle, F.; Levine, R. D.; Willner, I. *Nat Nanotechnol* **2010**, *5*, 417.

(3) Yashin, R.; Rudchenko, S.; Stojanovic, M. N. *J Am Chem Soc* **2007**, *129*, 15581.

(4) Frezza, B. M.; Cockroft, S. L.; Ghadiri, M. R. *J Am Chem Soc* **2007**, *129*, 14875.

(5) Bayerl, T. M.; Bloom, M. *Biophysical journal* **1990**, *58*, 357.

(6) Breaker, R. R.; Emilsson, G. M.; Lazarev, D.; Nakamura, S.; Puskarz, I. J.; Roth, A.; Sudarsan, N. *RNA* **2003**, *9*, 949.

(7) Gu, H.; Furukawa, K.; Weinberg, Z.; Berenson, D. F.; Breaker, R. R. *J Am Chem Soc* **2013**, *135*, 9121.

(8) Dass, C. R. *Trends Pharmacol Sci* **2004**, *25*, 395.

(9) Dass, C. R.; Choong, P. F.; Khachigian, L. M. *Mol Cancer Ther* **2008**, *7*, 243.

(10) Dass, C. R.; Saravolac, E. G.; Li, Y.; Sun, L. Q. *Antisense Nucleic Acid Drug Dev* **2002**, *12*, 289.

(11) Ashley, C. E.; Carnes, E. C.; Epler, K. E.; Padilla, D. P.; Phillips, G. K.; Castillo, R. E.; Wilkinson, D. C.; Wilkinson, B. S.; Burgard, C. A.; Kalinich, R. M.; Townson, J. L.; Chackerian, B.; Willman, C. L.; Peabody, D. S.; Wharton, W.; Brinker, C. J. *ACS Nano* **2012**, *6*, 2174.

(12) Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Padilla, D.; Durfee, P. N.; Brown, P. A.; Hanna, T. N.; Liu, J.; Phillips, B.; Carter, M. B.; Carroll, N. J.; Jiang, X.; Dunphy, D. R.; Willman, C. L.; Petsev, D. N.; Evans, D. G.; Parikh, A. N.; Chackerian, B.; Wharton, W.; Peabody, D. S.; Brinker, C. J. *Nat Mater* **2011**, *10*, 389.

(13) Epler, K.; Padilla, D.; Phillips, G.; Crowder, P.; Castillo, R.; Wilkinson, D.; Wilkinson, B.; Burgard, C.; Kalinich, R.; Townson, J.; Chackerian, B.; Willman, C.; Peabody, D.; Wharton, W.; Brinker, C. J.; Ashley, C.; Carnes, E. *Adv Healthc Mater* **2012**, *1*, 348.

(14) Pascal, J.; Ashley, C. E.; Wang, Z.; Brocato, T. A.; Butner, J. D.; Carnes, E. C.; Koay, E. J.; Brinker, C. J.; Cristini, V. *ACS Nano* **2013**, *7*, 11174.

(15) Bonnet, J.; Yin, P.; Ortiz, M. E.; Subsoontorn, P.; Endy, D. *Science* **2013**.

(16) Kahan-Hanum, M.; Douek, Y.; Adar, R.; Shapiro, E. *Sci Rep* **2013**, *3*.

(17) Young, D. D.; Lively, M. O.; Deiters, A. *J Am Chem Soc* **2010**, *132*, 6183.

(18) Fokina, A. A.; Meschaninova, M. I.; Durfort, T.; Venyaminova, A. G.; Francois, J. C. *Biochemistry* **2012**, *51*, 2181.

(19) Peracchi, A. *Rev Med Virol* **2004**, *14*, 47.

(20) Achenbach, J. C.; Chiuman, W.; Cruz, R. P.; Li, Y. *Curr Pharm Biotechnol* **2004**, *5*, 321.

(21) Baum, D. A.; Silverman, S. K. *Cell Mol Life Sci* **2008**, *65*, 2156.

Appendix 1
Section A1.1. Oligonucleotide sequences and concentrations

All oligonucleotide sequences are listed 5' to 3'. Functional domains have been color-coded to match the corresponding domains in the figures, domain junctions are indicated by a space, and strand names have been annotated with the corresponding labels from the figures. The dinucleotide junctions that are cleaved in the substrate strands have been highlighted using a yellow background, and mismatched bases in AND gate inhibitors are shown as single red letters. The RNA base at the cleavage site in each substrate strand is represented as rA, and the fluorophore (fluorescein) and quencher (TAMRA) are represented as FAM and TAM respectively.

Table A1.1. Oligonucleotide sequences and concentrations for **Figures 3.2A** and **A1.2**.

Strand	Sequence	Conc. (nM)
Input	GCCGGTCGAA AACTAAGA TACAT	100
DNAzyme (Dz)	GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Inhibitor (Inh)	ATGTA TCTTAGTT TTCGACCGGC	125
Substrate	FAM-TCTTAGTT rAG GATAGTTC AT-TAM	250

Table A1.2. Oligonucleotide sequences and concentrations for **Figure 3.2B**.

Strand	Sequence	Conc. (nM)
Input	ATGTA TCTTAGTT TTCGACCGGC	125
DNAzyme (Dz)	GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Substrate	FAM-TCTTAGTT rAG GATAGTTC AT-TAM	250

Table A1.3. Oligonucleotide sequences and concentrations for **Figures 3.2C** and **A1.1**.

Strand	Sequence	Conc. (nM)
$Input_1$	CGGTCGAA AACTAAGA TGGAG	100
$Input_2$	GACCT GAACTATC TCCGAGC	100
DNAzyme (Dz)	GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Inhibitor (Inh)	CTCCA TCTTAGTT TTCGACCGGCTCGGA GATAGTTC AGGTC	125
Substrate	FAM-TCTTAGTT rAG GATAGTTC AT-TAM	250

Table A1.4. Oligonucleotide sequences and concentrations for **Figure 3.3**.

Strand	Sequence	Conc. (nM)
$Input_1$	AACTAAGA TGATGTGGAG	100
$Input_2$	GAGGTTGATG GAACTATC	100
DNAzyme (Dz)	GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100

Inhibitor (Inh), *n*=0	CTCCACATCA TCTTAGTT TTCGACCGGCTCGGA GATAGTTC CATCAACCTC	125
Inhibitor (Inh), *n*=1	CTCCACATCA TCTTAGTT TTCGACCAGCTCGGA GATAGTTC CATCAACCTC	125
Inhibitor (Inh), *n*=2	CTCCACATCA TCTTAGTT TTCAACCGGCTAGGA GATAGTTC CATCAACCTC	125
Inhibitor (Inh), *n*=3	CTCCACATCA TCTTAGTT TTCAACCAGCTAGGA GATAGTTC CATCAACCTC	125
Substrate	FAM-TCTTAGTT rAG GATAGTTC AT-TAM	50

Table A1.5. Oligonucleotide sequences for **Figure 3.4**.

Strand	**Sequence**	**Conc. (nM)**
$Input_1$	AACTAAGA TGATGTGGAG	100
$Input_2$	GAGGTTGATG GAACTATC	100
DNAzyme AND gate strand	GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Inhibitor for AND gate	CTCCACATCA TCTTAGTT TTCAACCAGCTAGGA GATAGTTC CATCAACCTC	125
Substrate for AND gate	FAM-TCTTAGTT rAG GATAGTTC AT-TAM	50
$Input_3$	CGGCTCGGA TCTATCCA CATTC	125
DNAzyme NOT gate strand	TGGATAGA TCCGAGCCGGTCGAA AACTAAGA	100
Substrate for NOT gate	FAM-TCTTAGTT rAG TCTATCCA AT-TAM	50

Section A1.2. Further investigation of AND gate controls

In **Figure 3.2C** we observed that our DNAzyme displacement AND gate showed slightly different behavior upon the addition of $Input_2$. While addition of $Input_1$ showed no increase in signal compared to the control comprising the inhibited gate complex and the substrate, the addition of $Input_2$ lead to a reproducible initial increase in fluorescence, as shown in **Figure A1.1A** (which is a reproduction of **Figure 3.2C**, included here for ease of comparison).

We investigated this effect by running additional controls with just the substrate and the individual input strands, as shown in Figure S1B. We observed an identical response to that seen in the presence of the inhibited gate complex. We conclude that $Input_2$ interacts directly with the reporter substrate, which has a direct effect on the observed fluorescence. This is likely due to additional undesired complementarity between $Input_2$ and the 3' end of the substrate

molecule, near the quencher. Binding of the activator this close to the quencher may interfere with the FRET efficiency. Alternatively, activator binding to the substrate strand may result in a straightened helix, which could increase the distance between the FRET pair, again reducing quencher efficiency. This effect also may sequester a significant fraction of the second input that, when combined with the slower kinetics of cooperative hybridization, yields the significantly slower rates than was observed for the YES gates.

Therefore the different behavior of the input control can be attributed not to incomplete inhibition of the DNAzyme (which makes sense given that the fluorescence trace does not rise following the initial "burst"), but to undesired interactions between $Input_2$ and the substrate. We believe that further sequence optimization could enable us to eliminate this effect, producing a similar response in the presence of just $Input_2$ as was seen in the presence of just $Input_1$.

The initial burst phase seen in **Figure A1.1B** was not observed in **Figure A1.1A**, likely due to different dead time between the two experiments. Concentrations are given in **Table A1.3**.

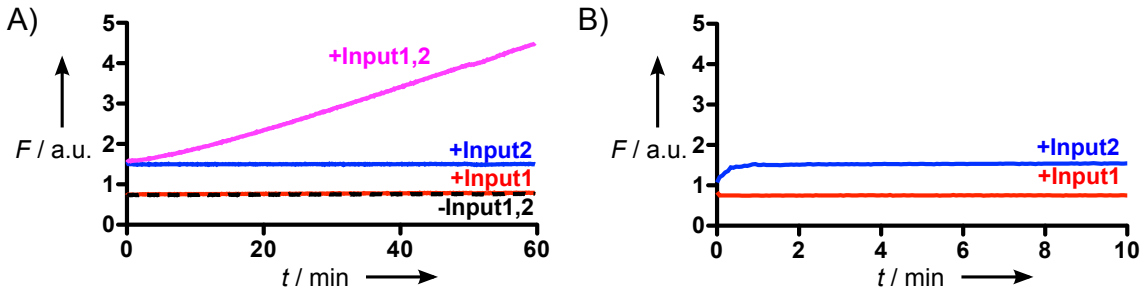

Figure A1.1. Investigation of activator-substrate interactions in the DNAzyme displacement AND gate. A) Data from main text illustrating action of DNAzyme displacement AND gate. B) Control experiments consisting of just activator strands and substrate in solution, in the absence of DNAzymes.

Section A1.3. Investigation of concentration effects and input thresholding by excess inhibitor strands

Preparation of DNAzyme-inhibitor complexes is a crucial part of our experimental procedure. Any DNAzymes that are not inhibited will be free to cleave substrates regardless of whether input strands are present, producing an unwanted leakage signal that gradually rises over time due to the multiple-turnover action of the DNAzyme. Therefore, it is important to ensure that all DNAzymes are fully inhibited.

One way to achieve this is to purify DNAzyme-inhibitor complexes after annealing by gel electrophoresis; however, this is both time- and labor-intensive, and therefore costly. Therefore we investigated the effect of annealing the DNAzyme-inhibitor complexes in the presence of a stoichiometric excess of inhibitor, to account for concentration and purity variations between stocks. We used different concentrations of activator and excess inhibitor in the YES gate example from **Figure 3.2A**, and the results are presented in **Figure A1.2**. We found that adding inhibitor strands in 10-20% excess relative to the concentration of the DNAzyme strand produced optimal performance. We observed lower leakage with increasing amounts of excess inhibitor, but this reduced the system's sensitivity and increased the response time for lower input concentrations. This is most likely due to a thresholding effect, whereby the excess inhibitor absorbs some (or all) of the input strands because the inhibitor strands bind more rapidly to the input strands than to the DNAzyme-inhibitor complexes, because the inhibitor contain a longer sequence complementary to the input. Similar effects are observed in other systems based on strand

displacement systems[1-3]. We observed a clear signal over background down to 1 nM input concentrations in certain conditions, which shows that the low input concentrations are amplified into a larger concentration of cleaved substrate due to the isothermal signal amplification capability of DNAzymes.

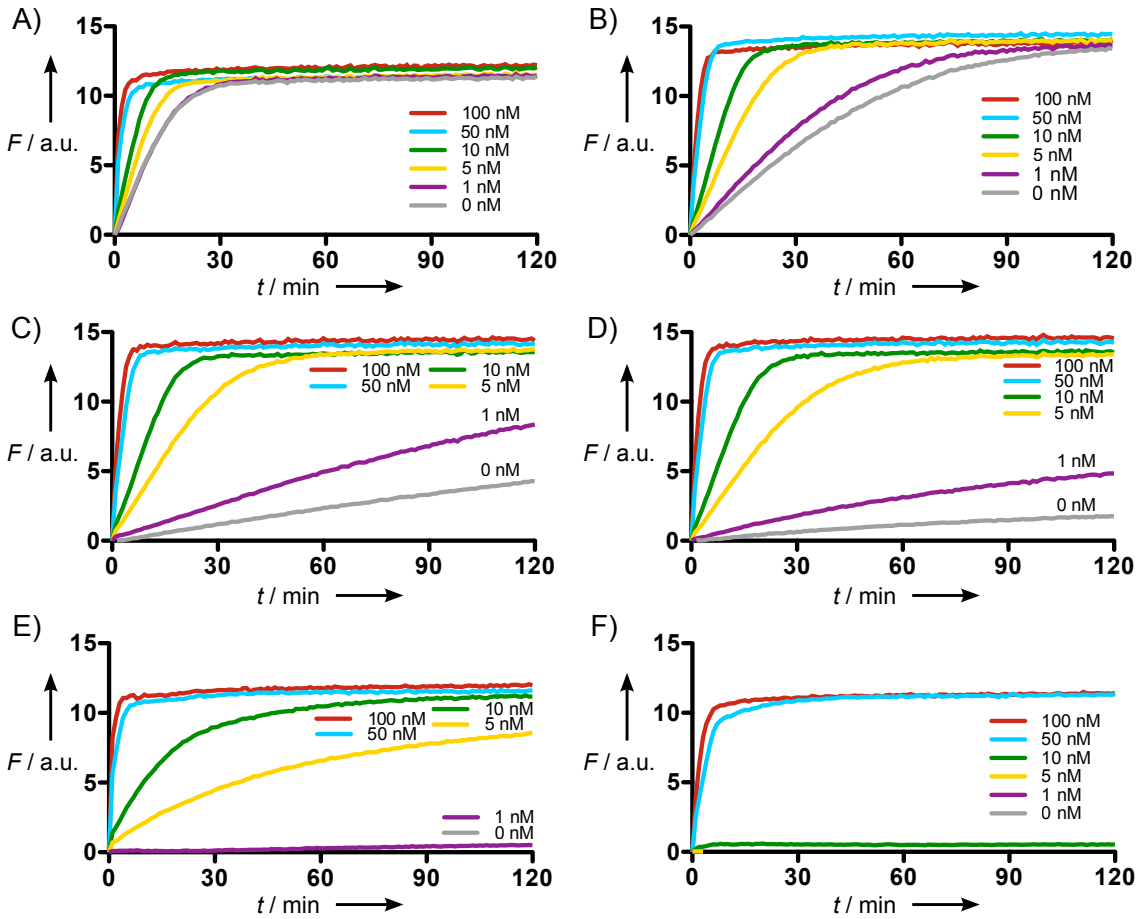

Figure A1.2. Kinetic traces of DNAzyme displacement YES gate output for different concentrations of input and excess inhibitor. In each panel, the concentration of the input strand (labeled "Input" in Figure 2A) varies as labeled. The concentration of excess inhibitor in solution varies between panels, as follows: A) 0 nM, B) 5 nM, C) 10 nM, D) 15 nM, E) 20 nM, F) 25 nM. In all cases, the concentration of inhibited DNAzymes (labeled Dz-Inh in Figure 2A) was 100 nM and the concentration of substrate was 50 nM. As the inhibitor concentration increases, it produces lower leakage but at the cost of increasing the limit of detection and slowing the response to lower concentrations of input.

Section A1.4. Thermodynamic investigation of destabilization of AND gate inhibitor by mismatched bases in the catalytic core

In **Figure 3.3**, the use of three mismatches in the AND gate inhibitor produced the optimal response in the presence of both activators, compared to the inhibitors with 0, 1, and 2 mismatches. To investigate this further, we used NUPACK[4,5] to generate predicted structures and complexes, along with their corresponding minimal free energies (MFEs) and the relative probability of formation. The results of these calculations are summarized in **Tables A1.6** and **A1.7**.

Columns (a) and (b) of Table S6 show a trend of increasing minimum free energy (MFE) with increasing numbers *n* of mismatches, for both the DNAzyme-inhibitor complex (Dz-I) and the DNAzyme-inhibitor complex with both inputs bound (Dz-Inh-$Input_1$-$Input_2$). The latter is formed when both inputs are added, and maybe be viewed as an unwanted structure because the both activators have bound to the DNAzyme-inhibitor complex, but the active DNAzyme has not been released into solution. The concentration of free DNAzyme strands in solution corresponds to the formation of an inhibitor-activator complex (Inh-$Input_1$-$Input_2$) without the DNAzyme strand still bound, since the formation of this structure indicates the displacement of the DNAzyme strand.

Column (c) of **Table A1.6** presents the equilibrium concentrations of the Dz-Inh-$Input_1$-$Input_2$ complex (corresponding to unsuccessful displacement of the DNAzyme strand by both activators) and of free DNAzyme (Dz) in solution (corresponding to successful displacement). These concentrations do not add up to 100 nM because of additional complexes involving the Dz strand, for instance

the initial Dz-Inh complex and those complexes where only a single activator is bound (Dz-Inh-Input$_1$ and Dz-I-Input$_2$). For clarity, we do not tabulate these concentrations here. We see that the concentration of free DNAzyme in solution is very low for *n*=0,1,2 but increases dramatically for *n*=3. We observe a similar, but less pronounced, trend in column (d), where we included an additional 25 nM of inhibitor to more accurately reflect our experimental conditions. In this case, the additional inhibitor is able to rebind some of the displaced Dz strands, which accounts almost exactly for the discrepancy in the concentration of free Dz (68 nM versus 97 nM). In any case, these calculations concur with the trend from our experimental data from **Figure 3.3B**, where we saw low activation for *n*=0,1,2 but a fast kinetic trace for *n*=3. This is likely due to the fact that three mismatches ensures that the core region of the DNAzyme remains bound to the inhibitor only by several short duplexes, which decreases the relative stability of any single duplex. In the absence of each activator, the binding arm duplexes are sufficient to maintain a stable complex, even in the presence of one activator (**Figure 3.3C**). However, with both activators present, in the catalytic core region only duplexes of 6, 3, 3, and 6 bases remain, and DNA breathing likely accounts for the rapid dissociation of the DNAzyme from this complex.

We observed that the MFE structure for the Dz-Inh-Input$_1$-Input$_2$ complexes was not quite the idealized structure that we depicted in **Figure 3.3A**. As shown in **Figure A1.3**, the MFE structure actually has incomplete binding of the two input strands, which presumably leads to a more stable complex overall by extending the length of the duplex on either side of the mismatches bases. We computed

free energies for these idealized structures, which are listed in **Table A1.8**. The results are shown in column (a) of **Table A1.7**, and we observed a similar trend of increasing free energy with increasing *n*.

Finally, we investigated the stability of the binding between the catalytic core and the (potentially mismatched) inhibitor strand in isolation. We used just the catalytic core portion of the DNAzyme strand as one input sequence, and the corresponding portion of the inhibitor strand (for *n*=0,1,2,3) as a second input sequence. As shown in columns (b) and (c) of **Table A1.7**, we observed a similar trend of almost complete binding of the inhibitor to the catalytic core for *n*=0,1 and no binding at all for *n*=3. Interestingly, the binding percentage for *n*=2 in this experiment was lower than observed when modeling the entire complex, which suggests that the additional flanking duplex between the DNAzyme and inhibitor strands (see **Figure A1.3**) may play a role in stabilizing the complex in this case.

It is important to bear in mind the limitations of thermodynamic modeling. In particular, NUPACK models all structures at equilibrium conditions, which may not accurately reflect the dynamic behavior of our DNAzyme displacement devices, which are inherently reliant on transient interactions between catalytically active DNAzyme strands and their substrates. While we have largely concerned ourselves with the MFE structured and energies, there will be many additional variants of each structure present in solution. Additionally, the direct applicability of NUPACK's structural predictions to our experiments is limited by the fact that the tables of thermodynamic parameters that serve as the basis of the NUPACK prediction algorithm are only really valid within a certain range of

salt concentrations. In particular, our reactions require Zn^{2+} ions in the buffer to serve as cofactors for the DNAzyme cleavage reaction, and the effects of these ions on the relative stability of the various DNA structures (and, indeed, on the folding of the DNAzyme strand) are a subject of ongoing research[6,7].

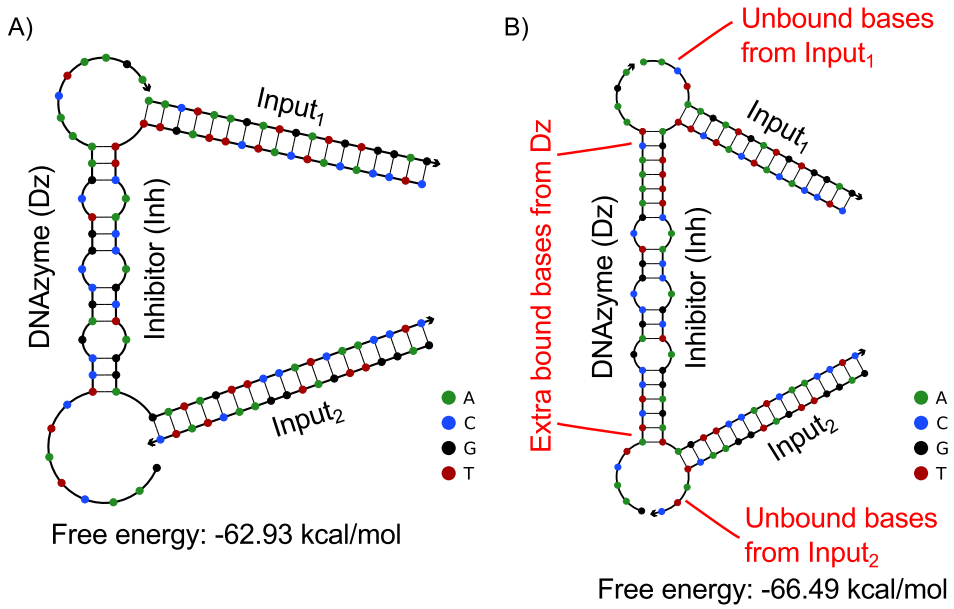

Figure A1.3. (A) Idealized structure of the Dz-Inh-$Input_1$-$Input_2$ complex with *n*=3 mismatches. (B) Minimum free energy structure of the Dz-Inh-$Input_1$-$Input_2$ complex for *n*=3. Note that, in the most stable structure, the two input strands are not completely bound to the inhibitor: rather, some bases are left overhanging. Presumably this produces a more stable structure overall by increasing the length of the bound duplex on either side of the outermost mismatched base on the inhibitor.

Table A1.6. Results of thermodynamic calculations using NUPACK to assess the stability of AND gates with various numbers n of mismatched bases in the part of the inhibitor strand that is complementary to the catalytic core. All NUPACK calculations where performed in DNA mode at 25°C, with a maximum complex size of 4 and using the default advanced options (dangles="some", Na^{+}=1.0M, Mg^{++}=0.0M). All sequences are presented in **Table A1.4** – the sequences for the DNAzyme strand (Dz) and the two input strands ($Input_1$ and $Input_2$) are the same throughout, and the sequence of the inhibitor strand (Inh) changes with varying *n* as shown in Table S4. In all cases the concentrations of Dz, $Input_1$ and $Input_2$ were each 100 nM. Calculations were performed with concentrations of Inh at both 100 nM and 125 nM, to assess the effect of excess inhibitor on the predictions. In columns (c) and (d), the concentrations of Dz, Inh, $Input_1$ and $Input_2$ were all 100 nM. In columns (e) and (f), the concentrations of Dz, $Input_1$ and $Input_2$ were 100 nM and the concentration of Inh was 125 nM.

Number of mismatches in inhibitor, *n*	**(a) MFE of Dz-Inh complex (kcal/mol)**	**(b) MFE of Dz-Inh-$Input_1$-$Input_2$ complex (kcal/mol)**	**100 nM inhibitor**		**125 nM inhibitor**	
			(c) Conc. of Dz-Inh-$Input_1$-$Input_2$ (nm)	**(d) Conc. of free Dz (nM)**	**(e) Conc. of Dz-Inh-$Input_1$-$Input_2$ (nm)**	**(f) Conc. of free Dz (nM)**
0	-50.49	-83.52	93	0	56	0
1	-44.34	-77.36	93	0	56	0
2	-39.61	-72.64	89	4.2	55	0.4
3	-33.46	-66.49	2	97	1.3	68

Table A1.7. Results of additional thermodynamic calculations using NUPACK. Parameters, settings and sequences were as described in Table S6. (a) Free energies of the "idealized" structures of the Dz-Inh-Input$_1$-Input$_2$ complexes. In each case, the idealized structure assumes complete binding of the two inputs to the inhibitor strand, which completely displaces the two substrate binding arms of the DNAzyme from the complex. These structured are presented explicitly in **Table A1.8** below. (b) and (c) Hybridization of just the core sequence of the DNAzyme strand (the black bases in **Table A1.**4) were used as one input strand, with the corresponding part of the inhibitor strand (black bases and red mismatches in **Table A1.4**). The maximum complex size was 2. In column (b), the concentrations of core strand and inhibitor strand were both 100 nM, and in column (c) they were 100 and 125 nM respectively.

Number of mismatches in inhibitor, *n*	**(a) Free energy of idealized Dz-Inh-Input$_1$-Input$_2$ complex structure (kcal/mol)**	**(b) Percentage of core sequence (100 nM) bound to inhibitor (100 nM)**	**(c) Percentage of core sequence (100 nM) bound to inhibitor (125 nM)**
0	-79.96	100%	100%
1	-73.81	97%	100%
2	-69.08	48%	55%
3	-62.93	0%	0%

Table A1.8. Idealized structures for Dz-Inh-Input$_1$-Input$_2$ complexes for different numbers *n* of mismatched bases in the inhibitor. The structures are represented in the dot-paren-plus format, which is a standard notation for the secondary structures of multi-strand complexes. Parentheses denote paired bases, dots denote unpaired bases, and plus signs denote strand breaks.

Number of mismatches in inhibitor, *n*	Dot-paren-plus representation for idealized structure of Dz-Inh-Input$_1$-Input$_2$ complex
0	((+))+.....)))))))))))))))).....+))))))))))))))))))))
1	((((((((((((((((((((((((.((((((((((((((((((((((((((+))+.....)))))))).)))))))).....+))))))))))))))))))))
2	((((((((((((((((((((.(((((((.((((((((((((((((((((((((((+))+.....))).)))))))).)))).....+))))))))))))))))))))
3	((((((((((((((((((((.(((.(((.((((((((((((((((((((((((((+))+.....))).))).))).))).....+))))))))))))))))))))

Appendix 2

A2.1.1 Oligonucleotide sequences

Oligonucleotide sequences are presented in **Tables 4.1-4.8**. All sequences are listed 5' to 3'. The cleavage sites in the substrates have been highlighted in red and the catalytic cores of DNAzymes are highlighted in bold. The RNA base at the cleavage site in each substrate (including SCS) strand is represented as rA. Fluorescein fluorophores and TAMRA quenchers are represented as /FAM/ and /TAM/ respectively.

Table A2.1. Sequences from multi-layer cascade experiments (**Figure 4.2b**), concentration profile of two-layer DNAzyme signaling cascade (**Figure A2.1**), and two-layer cascade experiment in DNA background (**Figure A2.6**).

Strand	**Sequence**
5th layer DNAzyme	GGGAGCCGTCCGAGCCGGTCGAAACTGTGGT
SCS_5	GCCGCTATACAAAGGTCGAAATATTTGTACCACAGTrAGCGGCTCCC
4th layer DNAzyme	GGTAGCGC**TCCGAGCCGGTCGAA**ATATTTGT
4th layer inhibitor	GTGGTACAAATATTTCGACCGGC
SCS_4	GCGCCTATTCCCCGGTCGAAACAGGGGAACAAATATrAGGCGCTACC
3rd layer DNAzyme	ACATGCCG**TCCGAGCCGGTCGAA**ACAGGGGA
3rd layer inhibitor	TTTGTTCCCCTGTTTCGACCGGC
SCS_3	GCCGCTAATACATGGTCGAAAGTATGTATCCCCTGTrAGCGGCATGT
2nd layer DNAzyme	ATCACGCC**TCCGAGCCGGTCGAA**AGTATGTA
2nd layer inhibitor	GGGGATACATACTTTCGACCGGC
SCS_2	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG
1st layer DNAzyme	GAACTATC**TCCGAGCCGGTCGAA**AACTAAGA
1st layer inhibitor	ATGTATCTTAGTTTTCGACCGGC
1st layer reporter substrate	/FAM/-TCTTAGTTrAGGATAGTTCAT-/TAM/

Table A2.2. Sequences from characterization of two-layer dengue serotyping circuits **(Figure 4.3b and Figure A2.4).**

Strand	Sequence
DEN-1 target	ACCAACAACAAACACCAAA
DEN-1 upstream DNAzyme	ACACCAAA**TCCGAGCCGGTCGAA**CATCATTC
DEN-1 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGATTTGGTGTTTGTTGTTGGT
DEN-1 SCS	CAAACTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGTTTGGTGT
DEN-2 target	ACTGCTCTTAACATCCTC
DEN-2 upstream DNAzyme	ACATCCTC**TCCGAGCCGGTCGAA**CATCATTC
DEN-2 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGAGAGGATGTTAAGAGCAGT
DEN-2 SCS	CCTCCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGGAGGATGT
DEN-3 target	GTGTGCCAGTCTTCAAGC
DEN-3 upstream DNAzyme	CTTCAAGC**TCCGAGCCGGTCGAA**CATCATTC
DEN-3 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGAGCTTGAAGACTGGCACAC
DEN-3 SCS	AAGCCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGGCTTGAAG
DEN-4 target	TATTGAAGTCAGGCCACT
DEN-4 upstream DNAzyme	AGGCCACT**TCCGAGCCGGTCGAA**CATCATTC
DEN-4 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGAAGTGGCCTGACTTCAATA
DEN-4 SCS	CACTCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGAGTGGCCT
DengueA target	CATCATTCCAGGCACAGA
DengueB target	CATGGGCTACTGGATAGA
Downstream DNAzyme	TGGATAGA**TCCGAGCCGGTCGAA**AACTAAGA
Downstream inhibitor	CATTCTCTTAGTTTTCGACCAGCTAGGATCTATCCAGTAGCCCATG
Downstream reporter substrate	/FAM/-TCTTAGTTrAGTCTATCCAAT-/TAM/

Table A2.3. Sequences from demonstration that SCS cleavage is necessary for signal propagation **(Figure A2.2).**

Strand	Sequence
Upstream DNAzyme	ATCACGCC**TCCGAGCCGGTCGAA**AGTATGTA
Uncleavable SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTAGGGCGTGATG
SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG
Downstream DNAzyme	GAACTATC**TCCGAGCCGGTCGAA**AACTAAGA
Downstream inhibitor	ATGTATCTTAGTTTTCGACCGGC
Downstream reporter substrate	/FAM/-TCTTAGTTrAGGATAGTTCAT-/TAM/

Table A2.4. Sequences from demonstration of first SCS input-output combination **(Figure A2.3a).**

Strand	Sequence
Upstream DNAzyme	ATCACGCC**TCCGAGCCGGTCGAA**AGTATGTA
Upstream inhibitor	CTCCTGTGCATACATACTTTCAACCAGCTAGGAGGCGTGATGATGAGTTTG
Input 1	AGTATGTATGCACAGGAG
Input 2	CAAACTCATCATCACGCC
SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG
Downstream DNAzyme	GAACTATC**TCCGAGCCGGTCGAA**AACTAAGA
Downstream inhibitor	ATGTATCTTAGTTTTCGACCGGC
Downstream reporter substrate	/FAM/-TCTTAGTTrAGGATAGTTCAT-/TAM/

Table A2.5. Sequences from demonstration of second SCS input-output combination **(Figure A2.3b).**

Strand	Sequence
Upstream DNAzyme	ATCACGCC**TCCGAGCCGGTCGAA**AGTATGTA
Upstream inhibitor	AAACATACATACTTTCGACCGGC
Input	GGTCGAAAGTATGTATGTTT
SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG
Downstream DNAzyme	TGATAGTTCATGTATCTTAGTTTTCGGAACTATC**AGCGATGACTGTTTTCAGTCCACCCATGT**AACTAAGA
Downstream reporter substrate	/FAM/-TCTTAGTTrAGGATAGTTCAT-/TAM/

Table A2.6. Sequences from demonstration of third SCS input-output combination **(Figure A2.3c).**

Strand	Sequence
Upstream DNAzyme	ATCACGCC**TCCGAGCCGGTCGAA**AGTATGTA
Upstream inhibitor	AAACATACATACTTTCGACCGGC
Input	GGTCGAAAGTATGTATGTTT
SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG
Downstream fluorophore strand	/FAM/-GCCGGTCGAAAACTAAGA
Downstream quencher strand	ATGTATCTTAGTTTTCGACC-/TAM/

Table A2.7. Sequences from optimization of DEN-3 serotyping circuit **(Figure A2.5b).**

Strand	Sequence
Initial DEN-3 target	TGCCAGTCTTCAAGCATG
Initial DEN-3 upstream DNAzyme	CAAGCATG**TCCGAGCCGGTCGAA**CATCATTC
Initial DEN-3 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGACATGCTTGAAGACTGGCA
Initial DEN-3 SCS	CATGCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGCATGCTTG
Optimized DEN-3 target	GTGTGCCAGTCTTCAAGC
Optimized DEN-3 upstream DNAzyme	CTTCAAGC**TCCGAGCCGGTCGAA**CATCATTC
Optimized DEN-3 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGAGCTTGAAGACTGGCACAC
Optimized DEN-3 SCS	AAGCCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGGCTTGAAG
DengueA target	CATCATTCCAGGCACAGA
DengueB target	CATGGGCTACTGGATAGA
Downstream DNAzyme	TGGATAGA**TCCGAGCCGGTCGAA**AACTAAGA
Downstream inhibitor	CATTCTCTTAGTTTTCGACCAGCTAGGATCTATCCAGTAGCCCATG
Downstream reporter substrate	/FAM/-TCTTAGTTrAGTCTATCCAAT-/TAM/

Table A2.8. Sequences from optimization of DEN-2 serotyping circuit **(Figure A2.5d).**

Strand	Sequence
Initial DEN-2 target	CTCTTAACATCCTCACAG
Initial DEN-2 upstream DNAzyme	CCTCACAG**TCCGAGCCGGTCGAA**CATCATTC
Initial DEN-2 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGACTGTGAGGATGTTAAGAG
Initial DEN-2 SCS	ACAGCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGCTGTGAGG
Optimized DEN-2 target	ACTGCTCTTAACATCCTC
Optimized DEN-2 upstream DNAzyme	ACATCCTC**TCCGAGCCGGTCGAA**CATCATTC
Optimized DEN-2 upstream inhibitor	TCTGTGCCTGGAATGATGTTCAACCAGCTAGGAGAGGATGTTAAGAGCAGT
Optimized DEN-2 SCS	CCTCCTCCTCTTAGGTCGAAAACTAAGAGAATGATGrAGGAGGATGT
DengueA target	CATCATTCCAGGCACAGA
DengueB target	CATGGGCTACTGGATAGA
Downstream DNAzyme	TGGATAGA**TCCGAGCCGGTCGAA**AACTAAGA
Downstream inhibitor	CATTCTCTTAGTTTTCGACCAGCTAGGATCTATCCAGTAGCCCATG
Downstream reporter substrate	/FAM/-TCTTAGTTrAGTCTATCCAAT-/TAM/

A2.2.2 DNA sequence design

The high-level structures of the strands and complexes were designed based on biophysical expectations of the stability of the complexes and their dynamic interactions with the other components of the system. The conserved sequences of the catalytic cores of the 8-17 and E6 DNAzymes were obtained from the literature [8,9].

Sequence design for SCS molecules was performed using a custom Python script that uses the NUPACK secondary structure prediction algorithm [10] and the ISO numeric representation of nucleic acid secondary structure [11] to find suitable domain assignments for the SCS sequence. Randomly generated sequences were tested using NUPACK to assess their equilibrium binding to the downstream DNAzyme and inhibitor strands in both the pre-cleavage state (to estimate leak rates) and the post-cleavage state (to estimate activation rates). Sequences that passed these tests were assessed for unwanted secondary structure using NUPACK and ISO, and candidate sequences were manually checked and optimized. Sequences for loop-inhibited DNAzyme logic gates were derived from the sequences of the DNAzyme displacement logic gates in the two-layer cascade via ensemble defect optimization using the NUPACK design tool [12].

For the dengue serotyping bioassays, we first performed a ClustalW sequence alignment on the genomes of all four dengue serotypes. Conserved and unconserved regions were identified manually and candidate target

sequences were selected from these regions. These were then tested for secondary structure using NUPACK and optimized by hand as necessary.

It is worth noting that NUPACK only models systems at thermodynamic equilibrium, and because the SCS participates in highly dynamic, transient interactions we can only draw limited conclusions about the behavior of our circuits from NUPACK predictions. We were forced to approximate the ribose base at the cleavage site by a deoxyribose base, because the available thermodynamic tables that serve as the basis of the NUPACK structure prediction algorithm [13] do not include parameters for DNA-RNA hybrids. Furthermore, the thermodynamic tables are only strictly valid within a certain range of salt concentrations. In particular, our reactions require Zn^{2+} ions in the buffer to serve as cofactors for the DNAzyme cleavage reaction, and the effects of these ions on DNA folding and on the relative stability of the various DNA structures are subjects of ongoing research [6,7,14-16].

A2.2.3 Multi-layer cascade experiments (Figure 4.2b,c)

Sequences are listed in **Table A2.1**. Concentrations for **Figure 4.2b**: 100 nM DNAzyme per layer, 125 nM inhibitor per layer (except the top layer), 100 nM SCS per layer, 250 nM fluorescent reporter substrate. Concentrations for Figure 2c: 100 nM 1st layer DNAzyme, 75 nM 2nd Layer DNAzyme, 50 nM 3rd layer DNAzyme, 25 nM 4th layer DNAzyme, 25% excess inhibitor and equimolar SCS per layer relative to DNAzyme concentration, 250 nM fluorescent reporter substrate. Pre-annealed DNAzyme-inhibitor complexes were added to buffer first, then pre-annealed SCS molecules, then fluorescent reporter substrate. Input

(active DNAzyme in the top layer) was added last to initiate the reaction. Loss of FRET was observed over two hours. Each trace was baseline-corrected by subtracting the initial value for that trace from each time point in that trace.

A2.2.4 Characterization of two-layer dengue serotyping circuits (Figure 4.3b, Figure A2.4) and secondary structure optimization in dengue serotyping circuits (Figure A2.5b,d)

Sequences for **Figure 4.3b** and **Figure A2.4** are listed in **Table A2.2**. Sequences for **Figure A2.5b** are listed in **Table A2.7**. Sequences for **Figure A2.5d** are listed in **Table A2.8**. Concentrations: 100 nM DNAzyme (upstream & downstream), 125 nM inhibitor (upstream & downstream), 100 nM inputs (DengueA, DengueB, DEN-k for k=1,2,3,4 as appropriate), 250 nM fluorescent reporter substrate. In **Figure A2.4**, experiments using multiple serotype-specific input strands were run using 100 nM of *each* serotype-specific input.

Pre-annealed DNAzyme-inhibitor complexes were added to buffer first, then pre-annealed SCS molecules, then inputs. The system was incubated at room temperature for 2 hours, then fluorescent reporter substrate was added, and the endpoint fluorescence value was observed after incubation at room temperature for a further 6 hours (2 hours for Figure **A2.5b,d**). All endpoint fluorescence values were baseline-corrected relative to the corresponding fluorescence value at the time of substrate addition. In **Figure 4.3b** and **Figure A2.4**, the baseline-corrected fluorescence values were normalized to the endpoint fluorescence of the positive trace, so that values between 0 and 1 could be reported. In **Figure A2.5b,d**, the baseline-corrected fluorescence values were plotted with no further data processing.

A2.2.5 Concentration profile of two-layer DNAzyme signaling cascade (Figure A2.1)

Sequences are listed in **Table A2.1**. DNAzyme concentrations (upstream & downstream) and SCS molecules and inputs varied according to the figure legend. In each case, downstream inhibitor was used in 25% excess relative to the concentration of the downstream DNAzyme. 250 nM reporter substrate was used. Pre-annealed DNAzyme-inhibitor complexes were added to buffer first, then pre-annealed SCS molecules, then fluorescent reporter substrate. Inputs were added last to initiate the reaction. Loss of FRET was observed over 30 minutes. Raw fluorescence values are plotted.

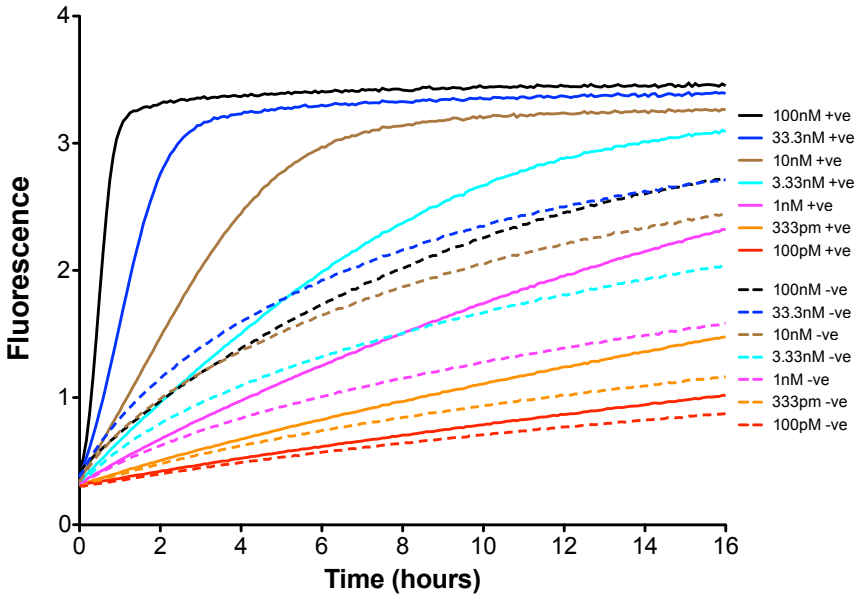

Figure A2.1: Concentration profile of two-layer DNAzyme signaling cascade. Broken lines indicate response in absence of active upstream DNAzymes (-ve control) and solid lines indicate response in presence of active upstream DNAzymes (+ve control). Concentrations of DNAzymes, inputs and SCS molecules vary as shown in the legend; inhibitor concentrations were also varied to ensure a 25% excess of inhibitor with respect to the DNAzyme concentration in each case. Concentration of the readout substrate was the same in all cases. Reducing the concentrations of circuit elements reduced both leakage rates and activation rates, as expected.

A2.2.6 Demonstration that SCS cleavage is necessary for signal propagation (Figure A2.1)

Sequences are listed in **Table A2.3**. Concentrations: 100 nM DNAzymes (upstream & downstream), 125 nM inhibitor (downstream), 100 nM SCS (cleavable or uncleavable, as appropriate), 250 nM fluorescent reporter substrate.

Pre-annealed DNAzyme-inhibitor complexes were added to buffer first, then pre-annealed SCS molecules, then fluorescent reporter substrate. Active upstream DNAzyme was added last to initiate the reaction. Loss of FRET was observed over two hours. Raw fluorescence values were plotted with no additional data processing.

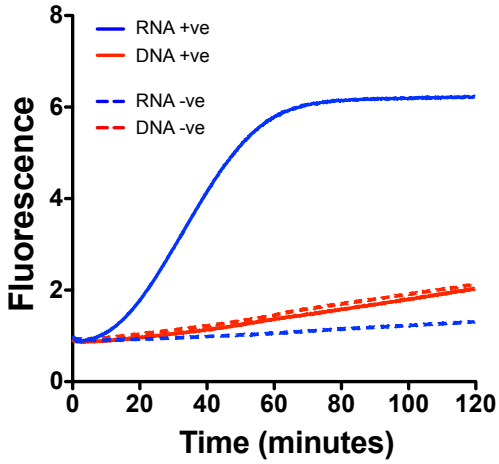

Figure A2.2: Demonstration that cleavage is required for signal propagation by the SCS in a two-layer cascade. Red traces are the response of an SCS molecule with the rA (ribose) base at the cleavage replaced by a dA (deoxyribose) base, both in the presence (+ve control, solid line) and absence (-ve control, broken line) of the upstream active DNAzyme. The substitution of dA for rA at the cleavage site makes the SCS molecule uncleavable. For comparison, the blue traces are the response of a cleavable SCS in the two layer cascade, both in the presence (solid line) and absence (broken line) of the upstream active DNAzyme. Addition of the upstream active DNAzyme does not produce any additional leakage in the case of the uncleavable SCS, which demonstrates that simply opening the outer stem by strand displacement does not produce downstream signal propagation. Hence, the cleavage step is required.

A2.2.7 Demonstration of SCS input-output combinations (Figure A2.3a-c)

Sequences are listed in **Tables A2.4-6**. Concentrations: (a) 100 nM DNAzymes (upstream & downstream), 125 nM inhibitor (upstream & downstream), 100 nM SCS, 50 nM reporter substrate, 100 nM input 1, 100 nM input 2. (b) 100 nM DNAzymes (upstream & downstream), 125 nM inhibitor (upstream), 100 nM SCS, 50 nM reporter substrate, 100 nM input. (c) 100 nM DNAzyme (upstream), 125 DNAzyme inhibitor (upstream), 100 nM SCS, 100 nM input, 100 nM fluorescent reporter strand, 125 nM downstream inhibitor labeled with quencher.

Inhibited DNAzymes (either pre-annealed DNAzyme-inhibitor complexes or annealed loop-inhibited DNAzyme strands) were added to buffer first, then pre-annealed SCS molecules, then inputs. The system was incubated for 2 hours at room temperature, then the reporter (either a fluorescent reporter substrate or a strand displacement reporter complex) was added, and the endpoint fluorescence value was observed after a further 30 minutes incubation at room temperature. Each endpoint fluorescence value was baseline-corrected relative to the corresponding fluorescence value at substrate addition.

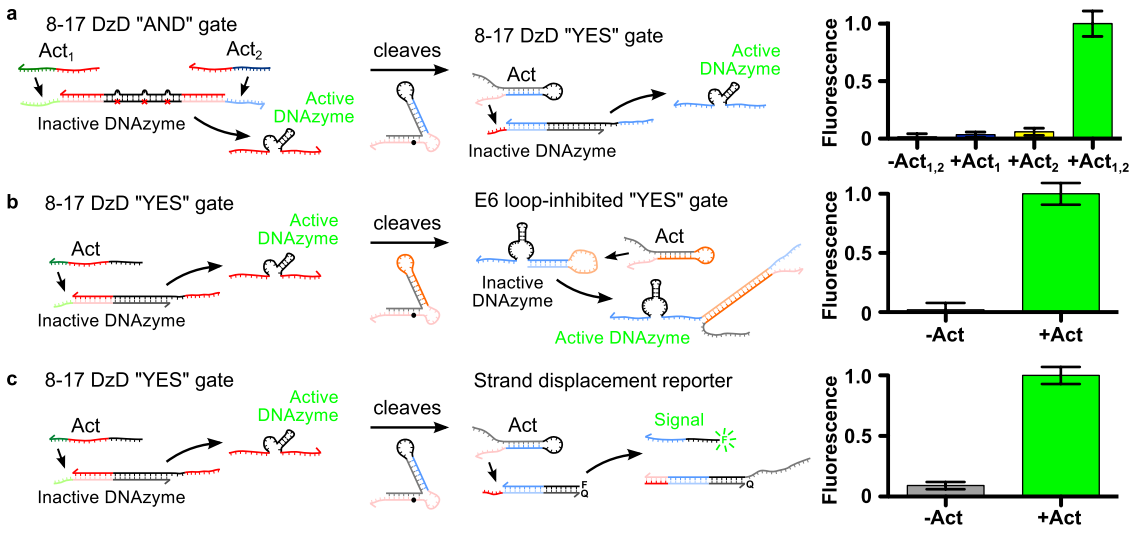

Figure A2.3. Application of the SCS as a generic interface molecule. An active DNAzyme from the input module cleaves the SCS, releasing an activator for the output module. This shows that our SCS design enables interoperability between different architectures, which is an important goal for future development of DNA logic circuits. Error bars on bar charts show the 95% confidence interval from triplicate runs of each experiment. a) Input module is a previously reported 8-17 DNAzyme displacement (DzD) "AND" gate with mismatched bases in the catalytic core portion of the inhibitor [17], which is activated by two inputs in a cooperative strand displacement reaction [18]. Output module is an 8-17 DzD "YES" gate. b) Input module is an 8-17 DzD "YES" gate, output module is a loop-inhibited "YES" gate based on the E6 catalytic motif [8]. Since the E6 DNAzyme cleaves the same dinucleotide junction as the 8-17 DNAzyme, we can use the same fluorescent reporter substrate in this case. c) Input module is an 8-17 DzD "YES" gate, output module is a DNA strand displacement reporter gate in which the activator released by cleavage of the SCS simply displaces a fluorescently-labeled strand from the reporter complex. The advantage of using a strand displacement gate as the reporter is that it does not amplify leakage, which might be preferable for certain applications. More generally, this reaction demonstrates that the SCS design could be used to interface DNAzymes with arbitrary strand displacement circuits and alternative DNAzyme catalytic motifs such as the 10-23 RNA-cleaving DNAzyme [9] and DNA-cleaving DNAzymes [19-21].

A2.2.8 – Dengue Biosensor Specificity

Sequences can be found in **Table A2.2.** Concentrations can be found in **Section A2.2.4**.

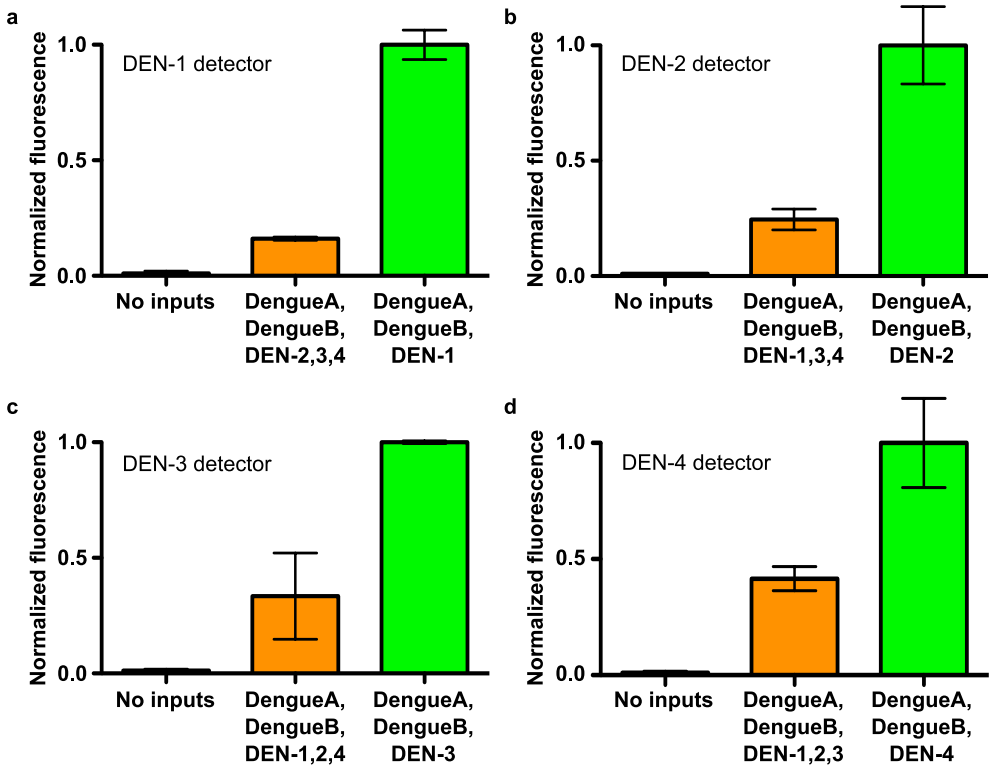

Figure A2.4: Demonstration of serotype-specific response from dengue serotyping circuits, which were presented in Figure 4.3. In each case, the negative control (grey) is the response in the absence of all three inputs, and the positive control (green) is the response in the presence of the two conserved inputs and the correct serotype-specific input (DEN-1, DEN-2, DEN-3 or DEN-4). The orange bar is the response in the presence of the two conserved inputs and all three incorrect serotype-specific sequences. **a**, Serotype specificity of DEN-1 detection circuit. **b**, Serotype specificity of DEN-2 detection circuit. **c**, Serotype specificity of DEN-3 detection circuit. **d**, Serotype specificity of DEN-4 detection circuit. In all cases, we observe a significantly reduced response when the incorrect serotype-specific sequences are present. In fact, the magnitudes of the non-specific responses to the incorrect serotype-sequences correlate with the background activations observed in the presence of the downstream DengueB input sequence in **Figure 4.3**, so it is likely that the non-specific activation seen in the presence of the incorrect serotype-specific sequences is in fact largely caused by the presence of DengueB. Hence we conclude that our four dengue detection circuits are in fact serotype-specific. Error bars represent the 95% confidence interval from three replicate experiments.

A2.2.9 – Dengue Biosensor Structural Optimization

Sequences can be found in **Tables A2.2**, **A2.7**, and **A2.8**. Concentrations can be found in **Section A2.2.4**.

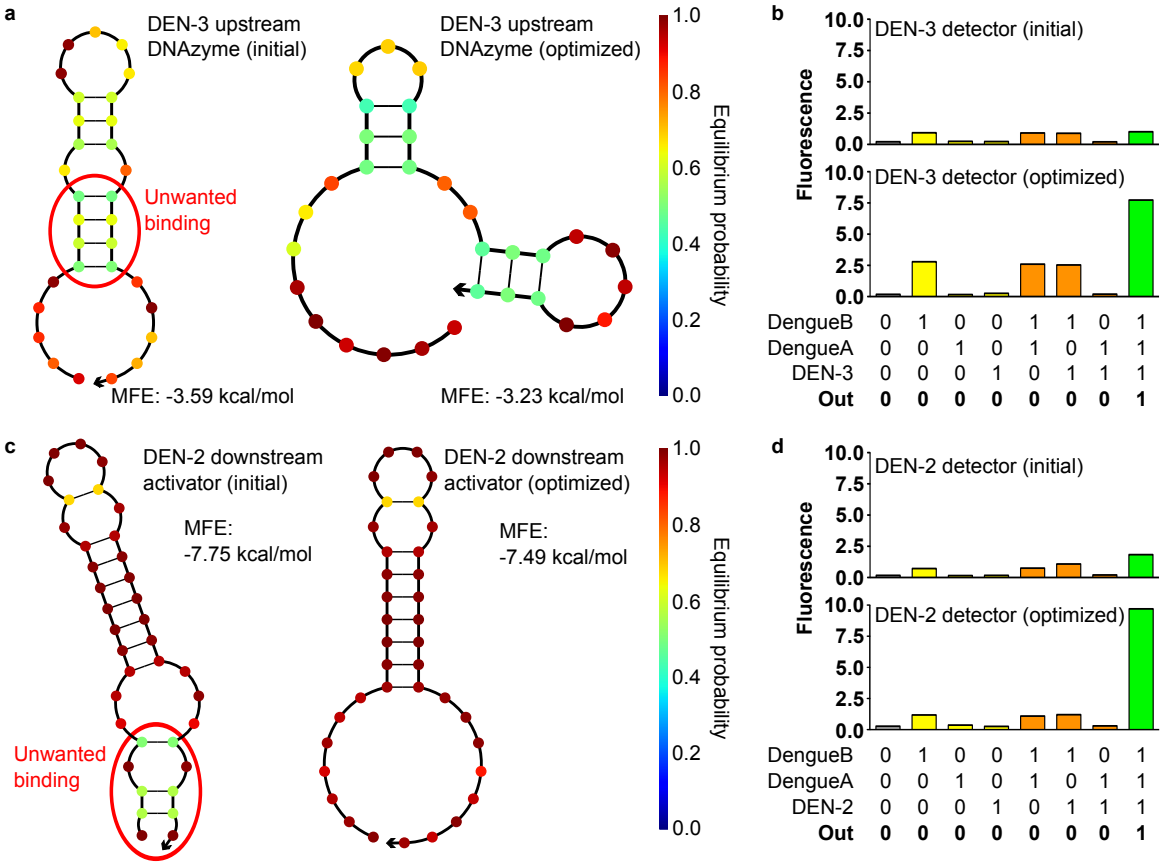

Figure A2.5: Performance improvements from secondary structure optimization of dengue serotyping circuit components. a) The upstream DNAzyme derived from the initial DEN-3 target sequence was found to have unwanted secondary structure, highlighted in red. We hypothesized that this would cause difficulty binding to the SCS molecule in the dengue serotyping circuit, because the two substrate binding arms of the DNAzyme were hybridized to each other. We switched to a different DEN-3 target sequence that removed this unwanted binding in the DNAzyme strand. **b**) Performance of the DEN-3 detection circuit using the initial and the optimized target sequences (and associated upstream DNAzymes and SCS molecules). The initial circuit produced no activation above background in the presence of all three inputs, whereas the optimized circuit produced a significant response in this case. **c**) The SCS cleavage product derived from the initial DEN-2 target sequence was found to have unwanted secondary structure, highlighted in red. We hypothesized that this would sequester the toehold of the downstream activator strand even after the SCS was cleaved, leading to low activation of the circuit. We switch to a different DEN-2 target sequence that removed this unwanted

binding in the activator strand. **d**) Performance of the DEN-2 detection circuit using the initial and the optimized target sequences (and associated upstream DNAzymes and SCS molecules). The initial circuit produced little activation above background in the presence of all three inputs, whereas the optimized circuit produced a significant response in this case. This highlights the importance of design optimization to prevent the formation of unwanted structures. The bar charts are representative data that illustrate the performance difference between the initial and optimized versions of the circuits.

A2.2.10 Two-layer cascade experiment in DNA background (Figure A2.6)

Sequences are listed in **Table A2.1**. Concentrations: 100 nM DNAzyme (2nd and 1st layers), 125 nM inhibitor (1st layer), 100 nM SCS (SCS_2), 50 nM fluorescent reporter substrate (1st layer). Herring sperm DNA (Promega, Madison, WI) was annealed (as described previously) and various amounts were added to 96 well plates containing buffer. Pre-annealed downstream DNAzyme-inhibitor complexes were added first, then pre-annealed SCS molecules, then fluorescent reporter substrate. Input (active upstream DNAzyme) was added last to initiate the reaction. Loss of FRET was observed over 30 minutes. Each positive kinetic trace was baseline-corrected by subtracting each time point observed from a negative control (run in the same experimental conditions but no active upstream DNAzyme present) from the corresponding time point in each positive trace. None of the negative controls showed a significant increase in fluorescence.

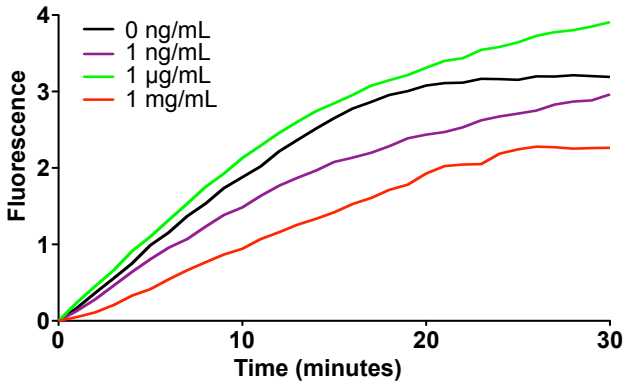

Figure A2.6. Operation of DNAzyme signaling cascades in the presence of background DNA. The two-layer cascade experiment was repeated in the presence of the labeled concentrations of denatured herring sperm DNA, covering six orders of magnitude from 1 ng/mL to 1 mg/mL. We observed no systematic loss of performance caused by the presence of background DNA. We hypothesize that the single-stranded SCS design allows rapid refolding following interactions with the background DNA, preventing spurious activation.

A2.2.11 - Discussion of spurious interactions in cascades

We have identified the following potential mechanisms for spurious activation in our DNAzyme signaling cascades.

1. Unwanted activation of one of the DNAzymes in the cascade by spontaneous dissociation of its inhibitor. The spuriously activated DNAzyme can then proceed to cleave its substrate, which may be an SCS molecule or a fluorescently labeled readout substrate.
2. Unwanted binding of an inhibited upstream DNAzyme to the SCS molecule. This may open the outer stem even in the absence of a cleavage reaction, providing a window of opportunity for the downstream effector sequence to interact with the downstream, inhibited DNAzyme.
3. Direct binding of the downstream inhibitor toehold to the sequestered toehold in the outer loop, leading to activation of the downstream DNAzyme by a toehold-mediated strand displacement (TMSD) reaction.
4. Spontaneous dissociation of one or both of the duplex stems of the SCS molecule, which reduces the topological constraint on the downstream toehold, enabling it to more easily activate the downstream DNAzyme by TMSD.

Our design work on DNAzyme displacement logic gates has shown that the effect of mechanism #1 can be reduced by extending the length of the inhibitor strands and, if necessary, introducing additional inhibitor strands free in solution to bias the equilibrium towards continued deactivation of the DNAzymes.

We addressed mechanism #2 by running additional controls with an SCS molecule in which the RNA base at the cleavage site is replaced by the corresponding DNA base. Since the 8-17 DNAzyme is RNA-cleaving, this single base substitution suffices to make the SCS molecule uncleavable by the upstream DNAzyme (Figure S3). These results demonstrated that stem opening by the binding of the upstream DNAzyme is not a significant leak mechanism.

Previous studies of hairpin opening kinetics [22] have shown that the rate constant for opening a hairpin by TMSD is several orders of magnitude slower when the toehold is contained within the hairpin (an "internal toehold") as opposed to outside (an "external toehold"). This is directly relevant to the kinetics of unwanted binding of the downstream, inhibited DNAzyme to the sequestered our SCS molecule, as in mechanism #3, which can be thought of as opening a loop via an internal toehold. Hence we conclude that mechanism #3 is probably not the dominant leakage mechanism.

Therefore, mechanism #4 is most likely to be the primary source of leakage in our cascades. The most obvious way to reduce such spurious dissociation of duplexes in the SCS molecule would be to extend these duplexes, in order to increase their melting temperature. However, the desire to retain multiple turnover in the cleavage of SCS molecules by upstream DNAzymes restricts the length of the DNAzymes' substrate binding arms to ~8-10 nucleotides each, so that the product strands unbind rapidly from the DNAzyme after cleavage. This, in turn, places upper limits on the lengths of the duplexes in the SCS structure, since these must be displaced by the substrate binding arms

via TMSD reactions. One of the substrate binding arms must also bind to the SCS toehold, so this also had to be taken into consideration when designing the basic SCS structure. A 4nt toehold was chosen to strike a balance between speed of toehold binding and subsequent strand displacement, and leaving enough of the substrate binding arm left so that the stem could be sufficiently long to reduce leakage as far as possible.

A2.2.12 - Discussion of rate-limiting steps in DNAzyme signaling cascades

Cleavage of an SCS molecule by an active upstream DNAzyme is a complex reaction with a number of steps that could be rate-limiting. In particular:

1. In order to cleave the SCS molecule, the upstream DNAzyme must initiate a TMSD reaction to open the outer stem and then nucleate a second binding event with the outer loop, so that both substrate binding arms are bound to the SCS molecule and the catalytic core is correctly positioned opposite the cleavage site. The second binding event is similar to a "remote toehold" strand displacement reaction [23], in that an internal diffusion step is required, which slows down the reaction considerably. Furthermore, the cleavage reaction must compete with the attempts of the SCS molecule to reform its dual-stem structure. This could cause the upstream DNAzyme to be displaced from the complex before cleavage takes place.
2. Post-cleavage, activation of the downstream DNAzyme involves a TMSD reaction in which the invading strand has some secondary structure (the short hairpin comprising the inner stem and inner loop from Figure 1a), as

opposed to being a single strand with no secondary structure. It is well known that secondary structure in the invader strand reduces the rate of strand displacement reactions [24].

A2.2.13 - Discussion of SCS design parameters and circuit leakage

The design of the SCS balances a number of constraints. To minimize the rate of leakage, that is, unwanted downstream activation caused by uncleaved SCS molecules interacting with DNAzyme gates, stems were made as long as possible. However, for efficient catalytic turnover the substrate binding arms of the DNAzymes should each be at most 8 nucleotides long, constraining the length of the stems.[25] To maximize the length of the outer stem and toehold while respecting this constraint, we place the cleavage site within the outer stem. Finally, to reduce leakage we make the SCS as compact as possible, by using the outer loop (part of the input module) to also serve as the toehold for the downstream activator. Minimizing loop size makes them more difficult to invade when in the sequestered state, reducing the potential for unwanted circuit element interactions. This approach does not constrain the recognition sequences of the different DNAzymes in the cascade. Overall, our SCS design exploits the predictable, sequence-specific folding of DNA to program a favorable reaction pathway directly into the structure of the substrate molecule.

Despite these efforts, the primary limiting factor for DNAzyme signaling cascades using the SCS remains the rate of circuit leakage. Although constraints on the SCS limit design space to a certain extent, a number of alternative strategies offer the potential to overcome this challenge. These include the physical

separation of circuit components by attaching them to the surfaces of microspheres,[26] and the rational introduction of mismatched bases[27] in the SCS design. Such strategies may be used in conjunction with one another, as physical separation should reduce the number of interactions between inactivate circuit components, while mismatches should reduce the conditional probability of unwanted signal generation given that an interaction has taken place. A number of such design alternatives are currently being explored. Additionally, while our work has advanced DNA logic for bioassay development, several challenges still remain. Most notable is the use of the fluorogenic chimeric substrate, which is expensive and requires RNAse free conditions for optimal response. However, recent work on DNA cleaving DNAzymes [19-21] may offer a path forward using cheaper pure DNA substrates instead of chimeric DNA/RNA substrates.

A2.2.14 - Discussion of SCS sequence effect on kinetic rates

As stated in **Section 5.2.1**, one of the primary objectives in the SCS design process was to optimize kinetic rates of activation vs. leakage. The modularity of the SCS design allowed the sequence to be altered while retaining the structure of the molecule. Each individual SCS sequence behaves identically from a mechanistic standpoint; however, the variation between layers of the five-layer cascade (**Figure 4.2B**) suggested that they do not all respond with the same kinetic profile. Although global kinetic modeling would be required to assign quantitative values to each of these rates, we may still be able to achieve a qualitative understanding of these differences through a careful look into the sequence selection process.

In **Section 5.2.3**, we discussed the mechanism for choosing each sequence. Briefly, we used NUPACK to explore the structure space of each DNAzyme pair, keeping the downstream DNAzyme sequence fixed while allowing the downstream inhibitor toehold, upstream DNAzyme, and corresponding SCS upstream substrate sequences to vary freely. To assess the potential performance of a given sequence, we compared the equilibrium binding probabilities between the pre-cleavage SCS molecule (SCS) and the post-cleavage SCS molecule (ACT) in the presence of the downstream DNAzyme and inhibitor. In this way, we would expect a low probability of the SCS binding with the downstream inhibitor, thus favoring a low percentage of activate downstream DNAzyme (referred to as the leak parameter). However after cleavage, the interaction between the downstream inhibitor and ACT would be favored, increasing the percentage of active DNAzyme (referred to as the activation parameter). Optimally, we look to achieve the highest activation/leak ratio, while ensuring good structural formation. In the graph below, the activation and leak parameters are listed for each SCS used in the five-layer cascade.

Table A2.9 – Activation vs. Leak percentages for each SCS sequence

Strand	% Leak	% Activation
SCS_2	**~0**	**6.7**
SCS_3	**0.24**	**71.55**
SCS_4	**0.29**	**71.57**
SCS_5	**0.28**	**50**

Note that the SCS_2 design was a misfolded structure from the original NUPACK script it was coded from, which may explain the low activation percentage. Additionally, the custom Python script used to search for optimal sequences was

not used for the first SCS sequence, but was used for all three subsequent sequences. There are two striking features from this table. First, it seems likely that despite the good performance of the SCS_2 (**Design 8**) demonstrated in **Figure 5.10**, this structure may indeed be the rate-limiting step of the cascades. Second, despite the third and fourth SCS sequences being nearly identical in both leak and activation parameters, they result in vastly different kinetic responses. While the lag phase between the two-layer and three-layer cascades is only slightly offset, there is a significant increase in the lag phase between the three-layer and the fourth-layer.

It remains unclear what structural differences contribute to the large differences in activation rate, and why similar activation/leak profiles (SCS_3 and SCS_4) result in very different kinetic traces in a multi-layered cascade. There are no obvious differences with downstream DNAzyme, Dz-Inh complex, SCS structure, or ACT-Inh complex. In **Figure A2.7**, we can observe the relative differences in structural stability of the ACT structure, which is generate from the SCS after cleavage by the upstream DNAzyme. It appears that the ACT structure from SCS_2 may have the most stable stem structure, in which all seven bases bind at a high probability. This may offer a possible explanation as to why this is the lowest activation rate (a stable ACT structure may slow the rate of strand displacement of the downstream Dz-Inh complex). However, this line of reasoning may be problematic in that the next most stable structures (SCS_3, which retains a seven base stem, albeit at a lower probability of remaining bound; SCS_4, which retains a six base pair stem at a high probability) also yield

the highest rate of activation. Finally, the weakest ACT structure is derived from SCS_5, having only a moderate probability of retaining a five base pair stem yet this sequence produces only a medium activation percentage, well below seemingly more stable structures. It is possible that this activation is slowed from the many of the unpaired bases exhibiting a lower probability, indicating that this structure may fold up on itself to some degree.

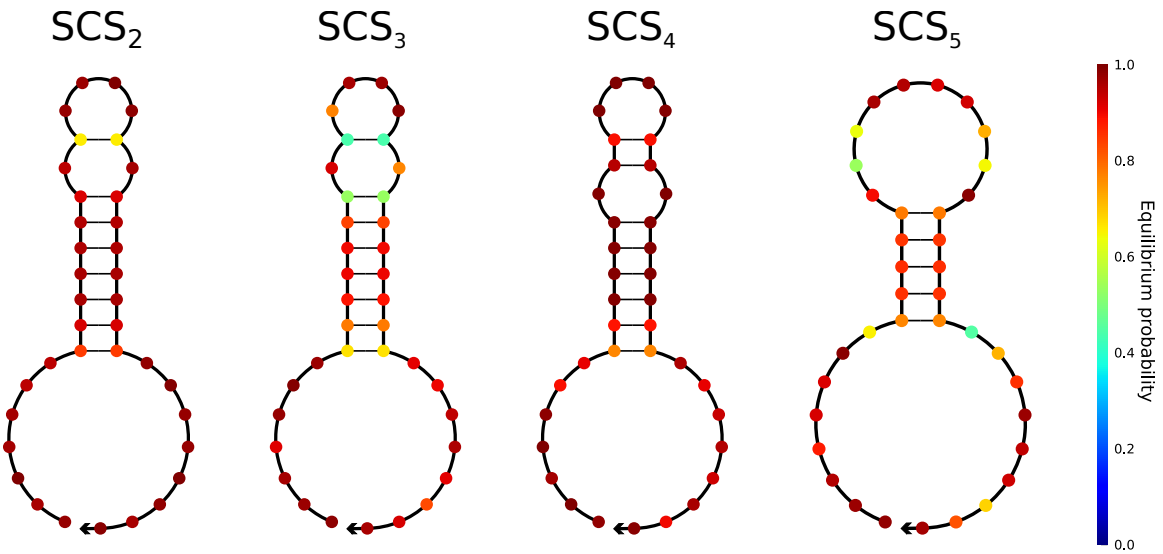

Figure A2.7 – Comparison of ACT (post-cleavage) structural probabilities for the four respective SCS structures.

Absent of any defining factors that could readily explain the differences in equilibrium predictions and observed kinetic rates, it is clear that more work needs to be done in this area. Kinetic modeling and the determination of kinetic rates to each of the individual reactions may provide significant insight into this mechanism. It may also be beneficial to run a global sequence interaction program to predict cross-reactivity between layers, something that has not yet been explored.

Appendix 3

A3.1 Oligonucleotide sequences and concentrations

Section A3.1.1 Oligonucleotide sequences and concentrations for each variant of SCS **Design 1** for DNAzyme cascades. The variant used in **Figure 5.2** is highlighted in cyan.

Table A3.1

Strand (D1v1)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTTAGTTGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.2

Strand (D1v2)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTCGTTTTCGACCGGCT	120
SCS	CTCCATCTTCGTTGGGTATTrAGGCGGACAGCCGGTCGAAAACGAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.3

Strand (D1v3)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTTAGTTTGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.4

Strand (D1v4)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTTAGTTTTGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.5

Strand (D1v5)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTTAGTTTTCGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.6

Strand (D1v6)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTAAGTTTTCGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.7

Strand (D1v7)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTTAGTTTTCGAGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100

Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.8

Strand (D1v8)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	CTCCATCTAAGTTTTCGAGGGTATTrAGGCGGACAGCCGGTCGAAAACTAAGATGGAG	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.9

Strand (D1v4)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	AGCCGGTCGAAAACTAAGATGGAGGGGTATTrAGGCGGACTAGTTTTCGACCGGCT	100
UE	GTCCGCTCCGAGCCGGTCGAAAATACCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.2 Oligonucleotide sequences and concentrations for each variant of SCS **Design 2** for DNAzyme cascades. The variant used in **Figure 5.3** is highlighted in cyan. NUPACK code for this design is in **Section A3.2.1**.

Table A3.10

Strand (D2v1)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	AGCCGGTCGAAAACTAAGACGTGAGGGTATTrAGGCGGACTCACG	100
UE	GAGTCCGCTCCGAGCCGGTCGAAAATACCCT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.11

Strand (D2v2)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGCT	120
SCS	AGCCGGTCGAAAACTAAGACGCCCAGGGTATTrAGGCGGACTGGGCG	100
UE	CAGTCCGCTCCGAGCCGGTCGAAAATACCCT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.3 Oligonucleotide sequences and concentrations for SCS **Design 3** for DNAzyme cascades (**Figure 5.4**).

Table A3.12

Strand (D3)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CGGGTTCTTAGTTTTCGACC	120
SCS	AGCCGGTCGAAAACTAAGACGCCCAGGGTATTrAGGCGGACTGGGCG	100
UE	GTTTATGCTCCGAGCCGGTCGAAACCCGTTTCT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.4 Oligonucleotide sequences and concentrations for SCS **Design 4** for DNAzyme cascades (**Figure 5.5**). NUPACK code for this design can be found in **Section A3.2.2**.

Table A3.13

Strand (D4)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	GAAGTTCTTAGTTTTCGACC	120
SCS	GGGATGTGAAGTrAGGATGGGACGGTCGAAAACTAAGAACTTCAC	100
UE	GTCCCATCTCCGAGCCGGTCGAAACTTCACATCCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.14 Proposed variants of the SCS **Design 4**. Bold bases were targeted for variation, as described by the strand name. This is an example of the granularity of the rational design process. M2 indicates the mutation of two bases, either on the 3' end or the 5' end of the activator. The corresponding sequences for each SCS Design 4 variant is found in **Table A3.12**. These sequences were not explicitly tested but such targeted locations for optimization were used in later variants.

Strand (D4v1)	Sequence	Conc (nM)
M2-5'	GGGATGTGAAGTrAGGATGGGAC**AA**TCGAAAACTAAGAACTTCAC	100
M2-3'	GGGATGTGAAGTrAGGATGGGACGGTCGAAAACTAA**AG**ACTTCAC	100
M2-5'3'	GGGATGTGAAGTrAGGATGGGAC**AA**TCGAAAACTAA**AG**ACTTCAC	100
S9	GGGAT**GA**GTGAAGTrAGGATGGGACGGTCGAAAACTAAGAACTTCAC**TC**	100
7GC	GGGATGTG**CC**GTrAGGATGGGACGGTCGAAAACTAAGAAC**GG**CAC	100
9GC	GGGAT**GA**GTG**CC**GTrAGGATGGGACGGTCGAAAACTAAGAAC**GG**CAC**TC**	100

Section A3.1.5 Oligonucleotide sequences and concentrations for each variant of SCS **Design 5** for DNAzyme cascades. The variant used in **Figure 5.6** is highlighted in cyan.

Table A3.15 - NUPACK code for this design can be found in **Section A3.2.3.**

Strand (D5v1)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CGTATTCTTAGTTTTCGACC	120
SCS	GGTCGAAAACTAAGAATACGGGACTACAGTTAGTAGTrAGCGTATGAGGG	100
UE	CCCTCATACGCTCCGAGCCGGTCGAAACTACTAACT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.16 - NUPACK code for this design can be found in **Section A3.2.4.**

Strand (D5v2)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	GCCACTCTTAGTTTTCGACC	120
SCS	GGTCGAAAACTAAGAGTGGCACCAGACTTAGGCCACTCATAAA	100
UE	TTTATGAGTGGCTCCGAGCCGGTCGAAAGTCTGGT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.6 Oligonucleotide sequences and concentrations for SCS **Design 6** for DNAzyme cascades (**Figure 5.7**).

Table A3.17 - NUPACK code for this design can be found in **Section A3.2.5.**

Strand (D6)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CGACCCGTCTTAGTTTTCGACCGGC	120
SCS	GGTCGAAAACTAAGACGTACTAGTACTACTACTAGTACGGGAA	100
SCS ACT	GGTCGAAAACTAAGACGTACTAGTACTACTACTA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.7 Oligonucleotide sequences and concentrations for each variant of SCS **Design 7** for DNAzyme cascades. The variant used in **Figure 5.8** is highlighted in cyan.

Table A3.18 - NUPACK code for this design can be found in **Section A3.2.6.**

Strand (D7v1)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	GTAGCTCTTAGTTTTCGACC	120
SCS	CACGCGTAGCGGTCGAAAACTAAGAGCTACAATrAGGCGTGAGG	100
UE	CCTCACGCTCCGAGCCGGTCGAAATTGTAGC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.19 - NUPACK code for this design can be found in **Section A3.2.7.**

Strand (D7v2)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	TTTACTCTTAGTTTTCGACC	120
SCS	CCCTACGACTTTACGGTCGAAAACTAAGAGTAAAGTGCAATrAGCGTAGGGATGAA	100
UE	TTCATCCCTACGTCCGAGCCGGTCGAAATTGCACTTTAC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.20 - NUPACK code for this design can be found in **Section A3.2.8.**

Strand (D7v3)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	TCTGATCTTAGTTTTCGACC	120
SCS	AAAGCCGTGATCGGTCGAAAACTAAGATCAGATACATrAGCGGCTTTAAC	100
UE	GTTAAAGCCGTCCGAGCCGGTCGAAATGTATCTGA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.21 - NUPACK code for this design can be found in **Section A3.2.9.**

Strand (D7v4)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	TCCAATCTTAGTTTTCGACC	120
SCS	CAAACGCTCCAATCGGTCGAAAACTAAGATTGGATAACTrAGGCGTTTGATG	100
UE	CATCAAACGCTCCGAGCCGGTCGAAAGTTATCCAA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Section A3.1.8 Oligonucleotide sequences and concentrations for each variant of SCS **Design 8** for DNAzyme cascades. The variant used in **Figure 5.9** is highlighted in cyan. NUPACK code for this design can be found in **Section A3.2.10.** Manual adjustments were made between the original variant (D8v1) and the final variant (v4).

Table A3.22

Strand (D8v1)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	ATGTATCTTAGTTTTCGACC	120
SCS	CACGCCTATCTTAGGTCGAAAACTAAGATTCATTTACTrAGGGCGTGATTAG	100
SCS ACT	CACGCCTATCTTAGGTCGAAAACTAAGATTCATTTACTA	100
UE 11-10	CTAATCACGCCTCCGAGCCGGTCGAAAGTAAATGAA	100
UE 11-8	CTAATCACGCCTCCGAGCCGGTCGAAAGTAAATG	100
UE 10-8	TAATCACGCCTCCGAGCCGGTCGAAAGTAAATG	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.23

Strand (D8v2)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	TCCAATCTTAGTTTTCGACC	120
SCS	CACGCCTGTCTTAGGTCGAAAACTAAGATTCATTTACTAGGGCGTGATTAG	100
SCS ACT	CACGCCTGTCTTAGGTCGAAAACTAAGATTCATTTACTA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.24

Strand (D8v3)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	ATGTATCTTAGTTTTCGACC	120
SCS	ACGCCCTATCTTAGGTCGAAAACTAAGATTCATTACTrAGGGCGTGATT	100
UE	AATCACGCCCTCCGAGCCGGTCGAAAGTAATGAA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A3.25 – This variant is used in **Figure 5.10**.

Strand (D8v4)	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	ATGTATCTTAGTTTTCGACCGGC	120
SCS	CGCCCTAATCTTAGGTCGAAAACTAAGATACATACTrAGGGCGTGATG	100
UE	ATCACGCCTCCGAGCCGGTCGAAAGTATGTA	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

A3.2 NUPACK Codes

A3.2.1 NUPACK Code for Design 2

```
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options for
# material, sodium, magnesium, and dangles
#
material = dna    temperature[C] = 25.0    # optional units: C (default) or K
trials = 2
sodium[M] = 1.0    # optional units: M (default), mM, uM, nM, pM
dangles = some
#
# target structure using DU+ notation
#
structure stickfigure = U5D5(U25)
#
# sequence domains
#
domain core = GGTCGAA
domain cleave = TAGG
domain sub = AACTAAGA
domain U1 = N9
domain U2 = N7
domain TH = N5
#
# thread sequence domains onto target structures
#
stickfigure.seq = U1 cleave U2 core sub TH
#
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
stickfigure.stop = 1.0

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

A3.2.2 NUPACK Code for Design 4

```
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options for
# material, sodium, magnesium, and dangles
#
material = dna
temperature[C] = 25.0    # optional units: C (default) or K
trials = 2
sodium[M] = 1.0    # optional units: M (default), mM, uM, nM, pM
dangles = some

#
# target structure using DU+ notation
#
structure stickfigure = U5D7(U25)
structure enzyme = U9 D3 ( U3 ) U17

#
# sequence domains
#
domain core = GGTCGAA
domain cleave = AG
domain sub = AACTAAGA
domain U1 = N11 T
domain U2 = G N7
domain TH = N7
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.

#
# thread sequence domains onto target structures
#
stickfigure.seq = U1 cleave U2 core sub TH
enzyme.seq = U2* corea1 corea2 coreb U1*

#
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
stickfigure.stop = 1.0

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

A3.2.3 NUPACK Code for Design 5

```
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options for
# material, sodium, magnesium, and dangles
#
material = dna
temperature[C] = 23.0    # optional units: C (default) or K
trials = 2
sodium[M] = 1.0    # optional units: M (default), mM, uM, nM, pM
dangles = some

#
# target structure using DU+ notation
#
structure stickfigure = U15D5(U2D5(U5)U2)U5

#
# sequence domains
#
domain core = GGTCGAAAACTAAGA
domain a = N5
domain b = N2
domain c = N5
domain d = N5
domain e = N4 T
domain f= AG
domain g = N5
domain h = N5

#
# thread sequence domains onto target structures
#
stickfigure.seq = core a b c d e f g h

#
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
stickfigure.stop = 1.0

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

A3.2.4 NUPACK Code for Design 5v2

```
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options for
# material, sodium, magnesium, and dangles
#
material = dna
temperature[C] = 25.0    # optional units: C (default) or K
trials = 10
sodium[M] = 1.0    # optional units: M (default), mM, uM, nM, pM
dangles = some

#
# target structure using DU+ notation
#
structure stickfigure = U15D5(U10)U5

#
# sequence domains
#
domain core = GGTCGAA
domain cleave = TAGG
domain sub = AACTAAGA
domain U1 = N9
domain U2 = N7
domain TH = N5

#
# thread sequence domains onto target structures
#
stickfigure.seq = core sub TH U2 cleave U1

#
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
stickfigure.stop = 1.0

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY
```

A3.2.5 NUPACK Code for Design 6

This variant has a 2nt clamp added to the downstream enzyme, and the inhibitor has 3nt extra at the core end which is not displaced by the activator. This variant has a 4nt loop.

```
##############################################################################
################
#
#% downstream_name = U2
#% upstream_name = R
#% link_type = ASL3.2-4Loop_5prime
#% num_iterations = 1500
#% pyxis_tests = asl:50
#% max_leak_pct = 15
#% min_act_pct = 5
#% temp = 25
#% test_conc = 1e-7
#
##############################################################################
################

#
# sequence domains
#
domain u1a = N4
domain u1b = SSN
domain u1c = N
domain u2a = AN
domain u2c = N4
domain clva = A
domain clvb = G
domain d1a = GAAC
domain d1b = TATC
domain d2a = AACTAA
domain d2b = GA
domain corea1 = TCC
domain corea2 = GA
domain coreb1 = GCC
domain coreb2a = GGT
domain coreb2b = CGAA
domain clamp = SSS

# Upstream enzyme
structure u = U9 D3 ( U3 ) U14
u.seq = u1a u1b u1c corea1 corea2 coreb1 coreb2a coreb2b u2a clamp u2c

# Downstream enzyme
structure d = U12 D3 ( U3 ) U13
d.seq = clamp d1a d1b corea1 corea2 coreb1 coreb2a coreb2b d2a d2b

# Downstream inhibitor with clamp and core extension
structure d_inhibitor_clamp_extn = U25
d_inhibitor_clamp_extn.seq = coreb2a* coreb1* corea2* corea1* d1b* d1a* clamp* u2a* clva clvb
u1c*
```

```
# Downstream substrate
structure d_sub = U18
d_sub.seq = d2b* d2a* clva clvb d1b* d1a*

# Communicator
structure asl = U4 D11 ( U4 ) U16
asl.seq = u2c* clamp* u2a* clva clvb u1c* u1b* u1a* u1b u1c clvb* clva* u2a clamp d1a d1b corea1 corea2 coreb1

# Waste strand from cleavage
structure waste = U10
waste.seq = u2c* clamp* u2a* clva

# Downstream activator
structure d_activator = U36 # there will be some secondary structure due to weak hairpin and 8-17 core
d_activator.seq = clvb u1c* u1b* u1a* u1b u1c clvb* clva* u2a clamp d1a d1b corea1 corea2 coreb1

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
dangles = some
trials = 10
```

A3.2.6 NUPACK Code for Design 7v1

This variant has 8nt arms on both upstream (U) and downstream (D) enzymes. The SCS element has each binding arm split into a 3nt toehold and 5nt stem. The downstream enzyme is constrained to our existing 8-17 sequence (U2) whereas the upstream enzyme can be freely varied, but without any G in the substrate binding arms. There is a sequence constraint between the upstream enzyme sequence and the toehold used on the downstream inhibitor strand.

```
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
trials = 5
#
# sequence domains
#
domain u1a = H3
domain u1b = H5 # NB: u1 = 5'-u1a-u1b-3', 8nt total
# We must subdivide corea like this so that NUPACK accepts the design.
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.
domain u2a = A H2
domain u2b = H5 # NB: u2 = 5'-u2a-u2b-3', 8nt total
domain clv = AG
domain d1 = GAACTATC # 5' substrate binding arm of U2 enzyme
domain d2 = AACTAAGA # 3' substrate binding arm of U2 enzyme
#
# target structures, list from 5' to 3'
#
# NB: "canonical" 8-17 core secondary structure is U1 D3 ( U3 ) U5
# So for substrate binding arms of X nt each, the whole thing is U[X+1] D3 ( U3 ) U[X+5]
structure asl = D5 ( D5 ( U15 ) U5 ) U3
structure u = U9 D3 ( U3 ) U13
#structure u_bound_to_asl = D8 ( U1 D3 ( U3 ) U5 D8 ( + U18 ) U2 )
structure d_activator = U33 # there will be some secondary structure due to weak hairpin and 8-17 core
#structure d = U9 D3 ( U3 ) U13
#structure d_inhibitor = U21
# thread sequence domains onto target structures, list from 5' to 3'
asl.seq = u1b u2b coreb d2 u2b* u2a* clv u1b* u1a*
u.seq = u1a u1b corea1 corea2 coreb u2a u2b
#u_bound_to_asl.seq = u1a u1b corea1 corea2 coreb u2a u2b u1b u2b d1 u2b* u2a* clv u1b* u1a*
d_activator.seq = u1b u2b coreb d2 u2b* u2a*
#d.seq = d1 corea1 corea2 coreb d2
#d_inhibitor.seq = u2b d1* coreb*
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
#asl.stop = 1.0
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

A3.2.7 NUPACK Code for Design 7v2

This variant has 12nt arms on both upstream (U) and downstream (D) enzymes. The SCS element has each binding arm split into a 5nt toehold and 7nt stem. The downstream enzyme is constrained to our existing U2 sequence whereas the upstream enzyme can be freely varied. There is a sequence constraint between the upstream enzyme sequence and the toehold used on the downstream inhibitor strand.

```
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
trials = 5
#
# sequence domains
#
domain u1a = H5
domain u1b = H7 # NB: u1 = 5’-u1a-u1b-3’, 12nt total
# We must subdivide corea like this so that NUPACK accepts the design.
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.
domain u2a = A H4
domain u2b = H7 # NB: u2 = 5’-u2a-u2b-3’, 12nt total
domain clv = AG
domain d1 = GAACTATC # 5’ substrate binding arm of U2 enzyme
domain d2 = AACTAAGA # 3’ substrate binding arm of U2 enzyme
# target structures, list from 5’ to 3’
#
# NB: “canonical” 8-17 core secondary structure is U1 D3 ( U3 ) U5
# So for substrate binding arms of X nt each, the whole thing is U[X+1] D3 ( U3 ) U[X+5]
structure asl = D7 ( D7 ( U15 ) U7 ) U5
structure u = U13 D3 ( U3 ) U17
#structure u_bound_to_asl = D8 ( U1 D3 ( U3 ) U5 D8 ( + U18 ) U2 )
structure d_activator = U41 # there will be some secondary structure due to weak hairpin and 8-17 core
#structure d = U9 D3 ( U3 ) U13
#structure d_inhibitor = U21
#
# thread sequence domains onto target structures, list from 5’ to 3’
#
asl.seq = u1b u2b coreb d2 u2b* u2a* clv u1b* u1a*
u.seq = u1a u1b corea1 corea2 coreb u2a u2b
#u_bound_to_asl.seq = u1a u1b corea1 corea2 coreb u2a u2b u1b u2b d1 u2b* u2a* clv u1b* u1a*
d_activator.seq = u1b u2b coreb d2 u2b* u2a*
#d.seq = d1 corea1 corea2 coreb d2
#d_inhibitor.seq = u2b d1* coreb*
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#asl.stop = 1.0
#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

A3.2.8 NUPACK Code for Design 7v3

This variant has a 2nt clamp for arm end of activator, 7nt outer stem and 5nt inner stem. The downstream inhibitor toehold is 5bp.

```
#
# NUPACK answers (using a 3' activator/inhibitor combo):
# ASL strand: AAAGCCGTGATCGGTCGAAAACTAAGATCAGATACATAGCGGCTTTAAC (-10.66 kcal/mol)
# D activator strand: AAAGCCGTGATCGGTCGAAAACTAAGATCAGATACAT
# D inhibitor strand: TGATCTTAGTTTTCGACC
# D strand (codename U2): GAACTATCTCCGAGCCGGTCGAAAACTAAGA
# U strand (codename V): GTTAAAGCCGTCCGAGCCGGTCGAAATGTATCTGA
#
# Equimolar DAct, DInh and D give ~89% free D
# Equimolar ASL, DInh and D gives ~5.5% free D

################################################################################
################

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
trials = 5

#
# sequence domains
#
domain u1a = N3
domain u1b = S2 N3 S2 # NB: u1 = 5'-u1a-u1b-3', 10nt total
# We must subdivide corea like this so that NUPACK accepts the design.
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.
domain u2a = A N4
domain u2c = N2
domain u2b = N3 # NB: u2 = 5'-u2a-u2b-3', 10nt total
domain clv = AG
domain d1 = GAACTATC # 5' substrate binding arm of U2 enzyme
# We must divide d2 like this so that NUPACK accepts the design.
domain d2a = AACTAA
domain d2b = GA # 3' substrate binding arm of U2 enzyme. NB: d2 = 5'-d2a-d2b-3', 8nt total
domain clamp = N2

# NB: "canonical" 8-17 core secondary structure is U1 D3 ( U3 ) U5
# So for substrate binding arms of X nt each, the whole thing is U[X+1] D3 ( U3 ) U[X+5]
#
# NNB: all structures and sequences are listed from 5' to 3'

# Upstream enzyme
structure u = U11 D3 ( U3 ) U15
u.seq = u1a u1b corea1 corea2 coreb u2a u2c u2b
```

```
# ASL
structure asl = D7 ( D5 ( U13 ) U9 ) U3
asl.seq = u1b u2b clamp coreb d2a d2b u2b* u2c* u2a* clv u1b* u1a*
#asl.stop = 1.0 # (1.0 is the default)

# Downstream activator
structure d_activator = U37 # there will be some secondary structure due to weak hairpin and 8-17 core
d_activator.seq = u1b u2b clamp coreb d2a d2b u2b* u2c* u2a*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 coreb d2a d2b

# Downstream inhibitor
structure d_inhibitor = U20
d_inhibitor.seq = u2b u2c d2b* d2a* coreb*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

A3.2.9 NUPACK Code for Design 7v4

This variant has a 2nt clamp for arm end of activator, 7nt outer stem and 7nt inner stem.

```
#
# NUPACK suggestions:
# ASL strand: CAAACGCTCCCATCGGTCGAAAACTAAGATTGGATAACTAGGCGTTTGATG (-
10.31 kcal/mol)
# D activator strand: CAAACGCTCCCATCGGTCGAAAACTAAGATTGGATAACT
# D inhibitor strand: TCCCATCTTAGTTTTCGACC
# D strand: GAACTATCTCCGAGCCGGTCGAAAACTAAGA
# U strand: CATCAAACGCTCCGAGCCGGTCGAAAGTTATCCCA
#
# Equimolar DAct, DInh and D gives ~87% free D.
# Equimolar ASL, DInh and D gives ~8% free D.

# NUPACK suggestions, modified for a better-looking ASL structure:
# ASL strand: CAAACGCTCCAATCGGTCGAAAACTAAGATTGGATAACTAGGCGTTTGATG (-
14.62 kcal/mol)
# D activator strand: CAAACGCTCCAATCGGTCGAAAACTAAGATTGGATAACT
# D inhibitor strand: TCCAATCTTAGTTTTCGACC
# D strand: GAACTATCTCCGAGCCGGTCGAAAACTAAGA
# U strand: CATCAAACGCTCCGAGCCGGTCGAAAGTTATCCAA
#
# Equimolar Dact, DInh and D gives ~93% free D.
# Equimolar ASL, DInh and D gives ~8% free D.

##############################################################################
################=
#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
trials = 5

#
# sequence domains
#
domain u1a = N3
domain u1b = N7 # NB: u1 = 5'-u1a-u1b-3', 10nt total
# We must subdivide corea like this so that NUPACK accepts the design.
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.
domain u2a = A N4
domain u2b = N5 # NB: u2 = 5'-u2a-u2b-3', 10nt total
domain clv = AG
domain d1 = GAACTATC # 5' substrate binding arm of U2 enzyme
# We must divide d2 like this so that NUPACK accepts the design.
domain d2a = AACTAA
domain d2b = GA # 3' substrate binding arm of U2 enzyme. NB: d2 = 5'-d2a-d2b-3', 8nt total
domain clamp = N2

# NB: "canonical" 8-17 core secondary structure is U1 D3 ( U3 ) U5
```

```
# So for substrate binding arms of X nt each, the whole thing is U[X+1] D3 ( U3 ) U[X+5]
#
# NNB: all structures and sequences are listed from 5' to 3'

# Upstream enzyme
structure u = U11 D3 ( U3 ) U15
u.seq = u1a u1b corea1 corea2 coreb u2a u2b

# ASL
structure asl = D7 ( D7 ( U13 ) U7 ) U3
asl.seq = u1b u2b clamp coreb d2a d2b u2b* u2a* clv u1b* u1a*
#asl.stop = 1.0 # (1.0 is the default)

# Downstream activator
structure d_activator = U39 # there will be some secondary structure due to weak hairpin and 8-17 core
d_activator.seq = u1b u2b clamp coreb d2a d2b u2b* u2a*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 coreb d2a d2b

# Downstream inhibitor
structure d_inhibitor = U20
d_inhibitor.seq = u2b d2b* d2a* coreb*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

A3.2.10 NUPACK Code for Design 8

This is a hybrid variant, derived from a combination of several of the earlier versions. Here we focus on retaining 8bp binding arms for the upstream enzyme. The inner stem is 6bp and the outer stem is 7bp, with a 4bp upstream toehold and a 5bp downstream toehold (one of which is in the inner stem). This was also the first design to move the cleavage site into the outer stem.

```
##############################################################################
################
#
# ASL5.8 - trying to use upstream enzymes with 8bp binding arms
#
##############################################################################
################

#
# sequence domains
#
domain u1a = N4
domain u1b = S N2 S # NB: u1 = 5'-u1a-u1b-3', 10nt total
domain u2a = A N4
domain u2b = N3 # NB: u2 = 5'-u2a-u2b-3', 10nt total
# We subdivide clv like this so we can use the bases in the cleavage product strands.
domain clva = G
domain clvb = A # NB: clv = 5'-clvb-clva-3', 2nt total
# We must subdivide corea like this so that NUPACK accepts the designs for the enzyme strands.
domain corea1 = TCC
domain corea2 = GAGCC
domain coreb = GGTCGAA # NB: core = 5'-corea1-corea2-coreb-3', 15nt total.
domain dt1 = N
domain dt2 = N
domain d1 = GAACTATC # 5' substrate binding arm of U2 enzyme
# We must divide d2 like this so that NUPACK accepts the design.
domain d2a = AA
domain d2b = CTAAGA # 3' substrate binding arm of U2 enzyme. NB: d2 = 5'-d2a-d2b-3', 8nt
total
domain clamp = N5

# NB: "canonical" 8-17 core secondary structure is U1 D3 ( U3 ) U5
# So for substrate binding arms of X nt at the 5' end and Y nt at the 3' end, the whole thing is
U[X+1] D3 ( U3 ) U[Y+5]
#
# NNB: all structures and sequences are listed from 5' to 3'

# Upstream enzyme
structure u = U9 D3 ( U3 ) U13
u.seq = u1a u1b corea1 corea2 coreb u2a u2b

# ASL
structure asl = D6 ( D7 ( U9 ) U9 ) U4
asl.seq = u1b clva* clvb* dt2* d2b* coreb d2a d2b dt2 dt1 u2b* u2a* clvb clva u1b* u1a*

# Downstream activator
structure d_activator = U39 # there will be some secondary structure due to weak hairpin and 8-
17 core
```

```
d_activator.seq = u1b clva* clvb* dt2* d2b* coreb d2a d2b dt2 dt1 u2b* u2a* clvb

# Waste strand from cleavage
structure waste = U9
waste.seq = clva u1b* u1a*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 coreb d2a d2b

# Downstream inhibitor
structure d_inhibitor = U20
d_inhibitor.seq = u2b* dt1* dt2* d2b* d2a* coreb*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
dangles = some
trials = 10
```

A3.2.11 NUPACK Code for Design 8 – Third Layer

```
##############################################################################
################
#
# ASL5.8 - trying to use upstream enzymes with 8bp binding arms
#
#% downstream_name = A
#% upstream_name = B
#% link_type = ASL8-8_3Prime_Inh3
#% num_iterations = 1000
#% pyxis_tests = asl:50
#% max_leak_pct = 10
#% min_act_pct = 30
#% temp = 25
#% test_conc = 1e-6
#
##############################################################################
################

#
# sequence domains
#
domain u1a = N4
domain u1b = N4
domain clva = G
domain clvb = A
domain u2a = T
domain u2b = N2
domain u2c = N4
domain u2d = N
domain d1 = ATCACGCC
domain d2a = AG
domain d2b = T
domain d2c = ATGTA
domain clamp = N5
domain corea1 = TCC
domain corea2 = GA
domain corea3 = GCC
domain coreb = GGTCGAA

# Communicator
structure asl = D7 ( D6 ( U10 ) U6 ) U4
asl.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb clva u1b u1a

# Downstream activator
structure d_activator = U37 # there will be some secondary structure due to weak hairpin and 8-
17 core
d_activator.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb

# Waste strand from cleavage
structure waste = U9
waste.seq = clva u1b u1a

# Upstream enzyme
```

```
structure u = U9 D3 ( U3 ) U13
u.seq = u1a* u1b* corea1 corea2 corea3 coreb u2a* u2b* u2c* u2d*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 corea3 coreb  d2a d2b d2c

# Downstream inhibitor
structure d_inhibitor = U23
d_inhibitor.seq = u2c* u2d* d2c* d2b* d2a* coreb* corea3*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
dangles = some
trials = 1
f
```

A3.2.12 NUPACK Code for Design 8 – Fourth Layer

```
##########################################################################################
#
# ASL5.8 - trying to use upstream enzymes with 8bp binding arms
#
#% downstream_name = B5
#% upstream_name = C
#% link_type = ASL8-8_3Prime
#% num_iterations = 1500
#% pyxis_tests = asl:50
#% max_leak_pct = 10
#% min_act_pct = 30
#% temp = 25
#% test_conc = 1e-6
#
##########################################################################################

#
# sequence domains
#
domain u1a = N4
domain u1b = N4
domain clva = G
domain clvb = A
domain u2a = T
domain u2b = N2
domain u2c = N4
domain u2d = N
domain d1 = ACATGCCG
domain d2a = AC
domain d2b = A
domain d2c = GGGGA
domain clamp = N5
domain corea1 = TCC
domain corea2 = GA
domain corea3 = GCC
domain coreb = GGTCGAA

# Communicator
structure asl = D7 ( D6 ( U10 ) U6 ) U4
asl.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb clva u1b u1a

# Downstream activator
structure d_activator = U37 # there will be some secondary structure due to weak hairpin and 8-
17 core
d_activator.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb

# Waste strand from cleavage
structure waste = U9
waste.seq = clva u1b u1a

# Upstream enzyme
```

```
structure u = U9 D3 ( U3 ) U13
u.seq = u1a* u1b* corea1 corea2 corea3 coreb u2a* u2b* u2c* u2d*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 corea3 coreb  d2a d2b d2c

# Downstream inhibitor
structure d_inhibitor = U23
d_inhibitor.seq = u2c* u2d* d2c* d2b* d2a* coreb* corea3*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
dangles = some
trials = 1
```

A3.2.13 NUPACK Code for Design 8 – Fifth Layer

```
##############################################################################
################
#
# ASL5.8 - trying to use upstream enzymes with 8bp binding arms
#
#% downstream_name = C4
#% upstream_name = D
#% link_type = ASL8-8_3Prime
#% num_iterations = 1500
#% pyxis_tests = asl:50
#% max_leak_pct = 10
#% min_act_pct = 30
#% temp = 25
#% test_conc = 1e-6
#
##############################################################################
################

#
# sequence domains
#
domain u1a = N4
domain u1b = N4
domain clva = G
domain clvb = A
domain u2a = T
domain u2b = N2
domain u2c = N4
domain u2d = N
domain d1 = GGTAGCGC
domain d2a = AT
domain d2b = A
domain d2c = TTTGT
domain clamp = N5
domain corea1 = TCC
domain corea2 = GA
domain corea3 = GCC
domain coreb = GGTCGAA

# Communicator
structure asl = D7 ( D6 ( U10 ) U6 ) U4
asl.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb clva u1b u1a

# Downstream activator
structure d_activator = U37 # there will be some secondary structure due to weak hairpin and 8-
17 core
d_activator.seq = u1b* clva* clvb* u2a* u2d* d2c* coreb d2a d2b d2c u2d u2c u2b u2a clvb

# Waste strand from cleavage
structure waste = U9
waste.seq = clva u1b u1a

# Upstream enzyme
```

```
structure u = U9 D3 ( U3 ) U13
u.seq = u1a* u1b* corea1 corea2 corea3 coreb u2a* u2b* u2c* u2d*

# Downstream enzyme
structure d = U9 D3 ( U3 ) U13
d.seq = d1 corea1 corea2 corea3 coreb  d2a d2b d2c

# Downstream inhibitor
structure d_inhibitor = U23
d_inhibitor.seq = u2c* u2d* d2c* d2b* d2a* coreb* corea3*

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS,
WWWWWW, YYYYYY

#
# design material, temperature (C), and trials
#
material = dna
temperature = 25.0
dangles = some
trials = 1
```

Appendix 4

A4.1 Oligonucleotide sequences and concentrations

All oligonucleotide sequences are listed 5' to 3'. Functional domains have been color-coded to match the corresponding domains in the figures, and strand names have been annotated with the corresponding labels from the figures. The dinucleotide junctions that are cleaved in the substrate strands have been highlighted using a yellow background and mismatched bases are underlined. The RNA base at the cleavage site in each substrate strand is represented as rA, and the fluorophore (fluorescein) and quencher (TAMRA) are represented as FAM and TAM respectively.

Table A4.1 Oligonucleotide sequences and concentrations for **Figure 6.2**

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGAAACAACACTC	100
INH	ATAGGGTTGAGTGTTGTT CTCCA TCTTAGTT TTCGACCGGC	125
INH (P1)	ATAGGGTTGAGTGTTGTT CTCCT TCTTAGTT TTCGACCGGC	125
INH (P2)	ATAGGGTTGAGTGTTGTT CTCGA TCTTAGTT TTCGACCGGC	125
INH (P3)	ATAGGGTTGAGTGTTGTT CTGCA TCTTAGTT TTCGACCGGC	125
INH (P1,2)	ATAGGGTTGAGTGTTGTT CTCGC TCTTAGTT TTCGACCGGC	125
INH (P1,2,L8)	ATAGGGTTGAGTGTTGTT CATCTCGC TCTTAGTT TTCGACCGGC	125
Fuel	GGTCGAAAACTAAGATGGAG	100
Fuel (L8)	GGTCGAAAACTAAGATGGAGATG	100
Target	AACAACACTCAACCCTAT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.2 Oligonucleotide sequences and concentrations for **Figure 6.3/6.4**

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGAAACAACACTC	100
INH	ATAGGGTTGAGTGTTGTT CTCCA TCTTAGTT TTCGACCGGC	125/50
INH (P2)	ATAGGGTTGAGTGTTGTT CTCGA TCTTAGTT TTCGACCGGC	125/50
INH (P3)	ATAGGGTTGAGTGTTGTT CTGCA TCTTAGTT TTCGACCGGC	125/50
INH (P1,2)	ATAGGGTTGAGTGTTGTT CTCGC TCTTAGTT TTCGACCGGC	125/50
INH (P1,2,L8)	ATAGGGTTGAGTGTTGTT CATCTCGC TCTTAGTT TTCGACCGGC	125/50
Fuel	GGTCGAAAACTAAGATGGAG	100/40
Fuel (L8)	GGTCGAAAACTAAGATGGAGATG	100/40
Target	AACAACACTCAACCCTAT	100/40
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	250

Table A4.3 Oligonucleotide sequences and concentrations for **Figure 6.5**

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGAAACAACACTC	40
INH	ATAGGGTTGAGTGTTGTT CTCCA TCTTAGTT TTCGACCGGC	50
INH (P1)	ATAGGGTTGAGTGTTGTT CTCCT TCTTAGTT TTCGACCGGC	50
INH (P2)	ATAGGGTTGAGTGTTGTT CTCGA TCTTAGTT TTCGACCGGC	50
INH (P3)	ATAGGGTTGAGTGTTGTT CTGCA TCTTAGTT TTCGACCGGC	50
INH (P1,2)	ATAGGGTTGAGTGTTGTT CTCGC TCTTAGTT TTCGACCGGC	50
INH (P1,2,L8)	ATAGGGTTGAGTGTTGTT CATCTCGC TCTTAGTT TTCGACCGGC	50
CT Fuel	GGTCGAAAACTAAGCTGGAG	40
GT Fuel	GGTCGAAAACTAAGGTGGAG	40
CT Fuel (L8)	GGTCGAAAACTAAGCTGGAGATG	40
GT Fuel (L8)	GGTCGAAAACTAAGGTGGAGATG	40
Target	AACAACACTCAACCCTAT	40
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	250

Table A4.4 Oligonucleotide sequences and concentrations for **Figure 6.6. Figure 6.6E** also contains 100 and 50 nM fuel, as indicated on the graph.

Strand	Sequence	Conc (nM)
Dz R1	GAACTATC TCCGAGCCGGTCGAA AACTAAGA AACTACAACA	100
INH R1	CTTGTGGCTGTTGTAGTTCTCCATCTTAGTTTTCGACCGGC	125
Target R1	AACTACAACAGCCACAAG	100
Dz R2	GAACTATC TCCGAGCCGGTCGAA AACTAAGA AACAACACTC	100
INH R2	ATAGGGTTGAGTGTTGTT CTCCA TCTTAGTT TTCGACCGGC	125
Target R2	AACAACACTCAACCCTAT	100
rTarget R2	rArArCrArArCrArCrTrCrArArCrCrCrTrArT	100
Dz R3	GAACTATC TCCGAGCCGGTCGAA AACTAAGA TGGATGAACG	100
INH R3	GTCTATTTCGTTCATCCAGTTGTTCTTAGTTTTCGACCGGC	125
Target R3	TGGATGAACGAAATAGAC	100
Dz R4	GAACTATC TCCGAGCCGGTCGAA AACTAAGA GAAGGGAGAA	100
INH R4	GTCCGCCTTTCTCCCTTCGTTGTTCTTAGTTTTCGACCGGC	125
Target R4	GAAGGGAGAAAGGCGGAC	100
Fuel R1,R2	GGTCGAAAACTAAGATGCAG	500
Fuel R3,R4	GGTCGAAAACTAAGAACTAC	500
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.5 Oligonucleotide sequences and concentrations for **Figure 6.7.**

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGAAACAACACTC	40
INH (P1,2,L8)	ATAGGGTTGAGTGTTGTT CATCTCGC TCTTAGTT TTCGACCGGC	50
GT Fuel (L8)	GGTCGAAAACTAAGGTGGAGATG	100
Target	pRSET emGFP	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	250

Table A4.6 Oligonucleotide sequences and concentrations for **Figure 6.8** and **Figure 6.14.**

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGAACCTTCCT	100
INH (P3)	ACCTGGGGGAGTATGTGCGGAGGAAGGTCTCCATCTTAGTTTTCGACCGGC	125
Fuel	GGTCGAAAACTAAGATGCAG	100
Target	ATP	0.1,1 mM
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	250

Table A4.7 Oligonucleotide sequences and concentrations for **Figures 6.9** and **6.10.**

Strand	Sequence	Conc (nM)
Dz O26	GAACTATC TCCGAGCCGGTCGAA AACTAAGA GATACTTTGAACCTT	100
INH O26	GGATAT AAGGTTCAAAGTATC CTCCATCTTAGTTTTCGACCGGC	125
Target O26	GATACTTTGAACCTTATATCCCAATATAGT	50
Dz O45	GAACTATC TCCGAGCCGGTCGAA AACTAAGA GCCAAACCAACTATG	100
INH O45	GACAGT TCATAGTTGGTTTGGC CTCCATCTTAGTTTTCGACCGGC	125
Target O45	GCCAAACCAACTATGAACTGTC	50
Dz O103	GAACTATC TCCGAGCCGGTCGAA AACTAAGA ATTTTACTGGAAAAA	100
INH O103	GGTGCT TTTTTCCAGTAAAAT CTCCATCTTAGTTTTCGACCGGC	125
Target O103	TCTGATATTTTACTGGAAAAAAGCACC	50
Dz 121	GAACTATC TCCGAGCCGGTCGAA AACTAAGA AGTATAACCTTTTAC	100
INH O121	ATGAAA GTAAAAGGTTATACT CTCCATCTTAGTTTTCGACCGGC	125
Target O121	AGTATAACCTTTTACTTTCATGACAGGA	50
Dz O145	GAACTATC TCCGAGCCGGTCGAA AACTAAGA CATACACTCCTAAAT	100
INH O145	CAACAG ATTTAGGAGTGTATG CTCCATCTTAGTTTTCGACCGGC	125
Target O145	CATACACTCCTAAATCTGTTGATGGTA	50
Dz O157	GAACTATC TCCGAGCCGGTCGAA AACTAAGA TGTCATTCGTGACAA	100
INH O157	GAATGG TTGTCACGAATGACA CTCCATCTTAGTTTTCGACCGGC	125
Target O157	TGTCATTCGTGACAACCATTC	50
Fuel	GGTCGAAAACTAAGATGCAG	500
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.8 Oligonucleotide sequences and concentrations for **Figure 6.11.** Gate concentrations for this figure are listed in pM.

Strand	Sequence	Conc (pM)
Dz O45	GAACTATC TCCGAGCCGGTCGAA AACTAAGA GCCAAACCAACTATG	100
INH O45	GACAGT TCATAGTTGGTTTGGC CTCCATCTTAGTTTTCGACCGGC	125
Target O45	GCCAAACCAACTATGAACTGTC	1/5/10
Fuel	GGTCGAAAACTAAGATGCAG	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	250000

Table A4.9 Oligonucleotide sequences and concentrations for **Figure 6.13.**

Strand	Sequence	Conc (nM)
Dz1	GAACTATCTCCGAGCCGGTCGAAAACTAAGACGTCACCTAC	100
INH1	GGTAGGTAGGTGACGGTTGTTCTTAGTTTTCGACCGGC	125
$Input_1$	TCACCTACCTACCTACATC	100
SCS1	GCTGGCGTrAGTACAAATGGGTCGAAAACTAAGAACTACGC	100
Dz2	CATTTGTATCCGAGCCGGTCGAAACGCCAGCCCATAACCCT	100
INH2	GGGTGAGGGTTATGGATTGTGCTGGCGTTTCGACCGGC	125
$Input_2$	CCATAACCCTCACCCCATCTA	100
ACT2	GTAACATAAGGTCGAAACGCCAGCACTATCC	100
SCS2	GACGGGATAGTAACATAAGGTCGAAACGCCAGCACTATCC	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.10 Oligonucleotide sequences and concentrations for **Figure 6.14.**

Strand	Sequence	Conc (nM)
Dz1	GAACTATCTCCGAGCCGGTCGAAAACTAAGACGTCACCTAC	100
INH1	GGTAGGTAGGTGACGCGCGTACCTCTTAGTTTTCGACCGGC	125
$Input_1$	TCACCTACCTACCTACATC	100
SCS1	GCTGGCGTAGTACAAATGGGTCGAAAACTAAGGACTACGCG	100
ACT1	GTACAAATGGGTCGAAAACTAAGGACTACGCG	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.11 - Oligonucleotide sequences and concentrations for **Figure A4.6**.

Strand	Sequence	Conc (nM)
Dz4.0	GGGAGGCGTAATAAT GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Dz4.1	GGGAGGCGTAATAAT A GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Dz4.2	GGGAGGCGTAATAAT TA GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Dz4.3	GGGAGGCGTAATAAT CTA GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Dz4.4	GGGAGGCGTAATAAT TCTA GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Dz4.5	GGGAGGCGTAATAAT TTCTA GAACTATC TCCGAGCCGGTCGAA AACTAAGA	100
Inh	CCG GCTCGGA GATAGTTC TAGAA ATTATTACGCCTCCC ATTATTGGCG	125
Fuel	TTCTAGAACTATCTCCGAGCC	100
Target	CGCCAATAAT GGGAGGCGTAATAAT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.12 Oligonucleotide sequences and concentrations for SNP detection (**Figure A4.8 and A4.9**).

Strand	Sequence	Conc (nM)
Dz R2	GAACTATC TCCGAGCCGGTCGAA AACTAAGA AACAACACTC	100
INH P1, P2, L8	ATA GG GTT GAGTGTTGTT CATCTCGCTCTTAGTT TTCGACCGGC	125
Target	AACAACACTC AAC CC TAT	100
Fuel	GGTCGAAAACTAAGGTGGAGATG	100
SNP2	AACAACACTC AAC CC TCT	100
SNP3	AACAACACTC AAC CC GAT	100
SNP4	AACAACACTC AAC CA TAT	100
SNP5	AACAACACTC AAC GC	100
SNP6	AACAACACTC AAA CC TAT	100
SNP2,3	AACAACACTC AAC CC GCT	100
SNP2,4	AACAACACTC AAC CA TCT	100
SNP3,4	AACAACACTC AAC CG GAT	100
SNP4,5	AACAACACTC AAC GA TAT	100
SNP4,6	AACAACACTC AAA CA TAT	100
SNP5,6	AACAACACTC AAA GC TAT	100
SNP6,7	AACAACACTC ATA CC TAT	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.13 Oligonucleotide sequences and concentrations for DSL detection (**Figure A4.10**)

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGC	120
Target	TCCCACCAACAGCAGGGATATTAAA	100
DSL	AACAGCAGGGATATCTGCAGCCGGTCGAAAACTAAGATGGAG ATATCCCTGCTGTTGGTGGGA	varies
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

Table A4.14 Oligonucleotide sequences and concentrations for VSD detection (**Figure A4.11**)

Strand	Sequence	Conc (nM)
Dz	GAACTATCTCCGAGCCGGTCGAAAACTAAGA	100
INH	CTCCATCTTAGTTTTCGACCGGC	120
Target	AACTAAGATGGAGGGGATATTAAA	100
VSD-Inh5	AATATCCCCTCCATCTTAGTT	100
VSD-Inh8	TTTAATATCCCCTCCATCTTAGTT	100
VSD-Act	AGCCGGTCGAAAACTAAGATGGAGGGG	100
Substrate	FAM-TCTTAGTTrAGGATAGTTCAT-TAM	50

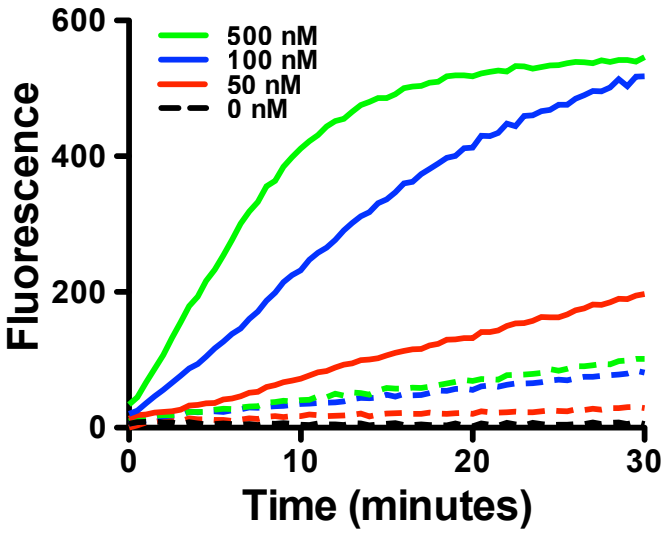

Figure A4.1 – Leak dependency on fuel concentration. Using the P3 gate structure (sequence is O45 DNAzyme and inhibitor, **Table A4.7**), gate response and leakage is directly related to the amount of fuel strand present in solution. Reducing fuel concentration decreases rate of both the positive response and the leakage. Concentrations were 100 nM O157 gate, 25 nM excess inhibitor, 50 nM substrate, 50 nM target, varied concentrations of fuel with 1 mismatch.

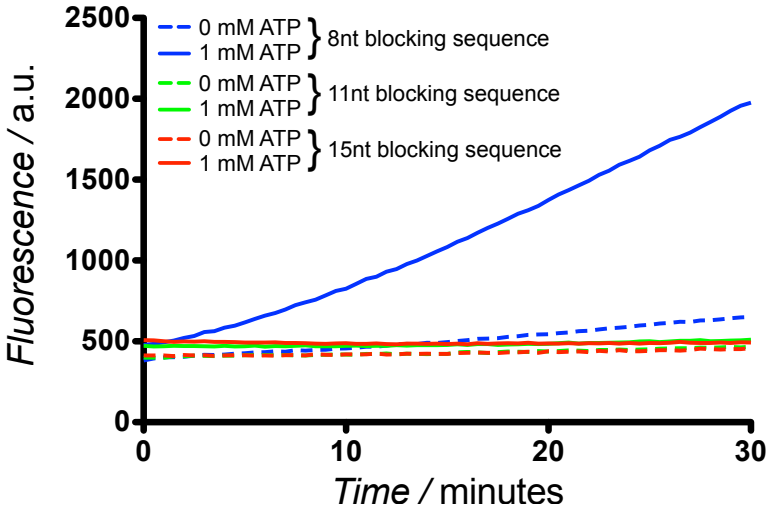

Figure A4.2 – Effect of longer blocking sequences on ATP aptamer activation. Only the 8bp blocking sequence is sufficient for activation (blue). Longer sequences, such as the 11bp (green) or 15 bp (red) lengths, blocked the ATP from displacing the detection domain, resulting in no gate activation.

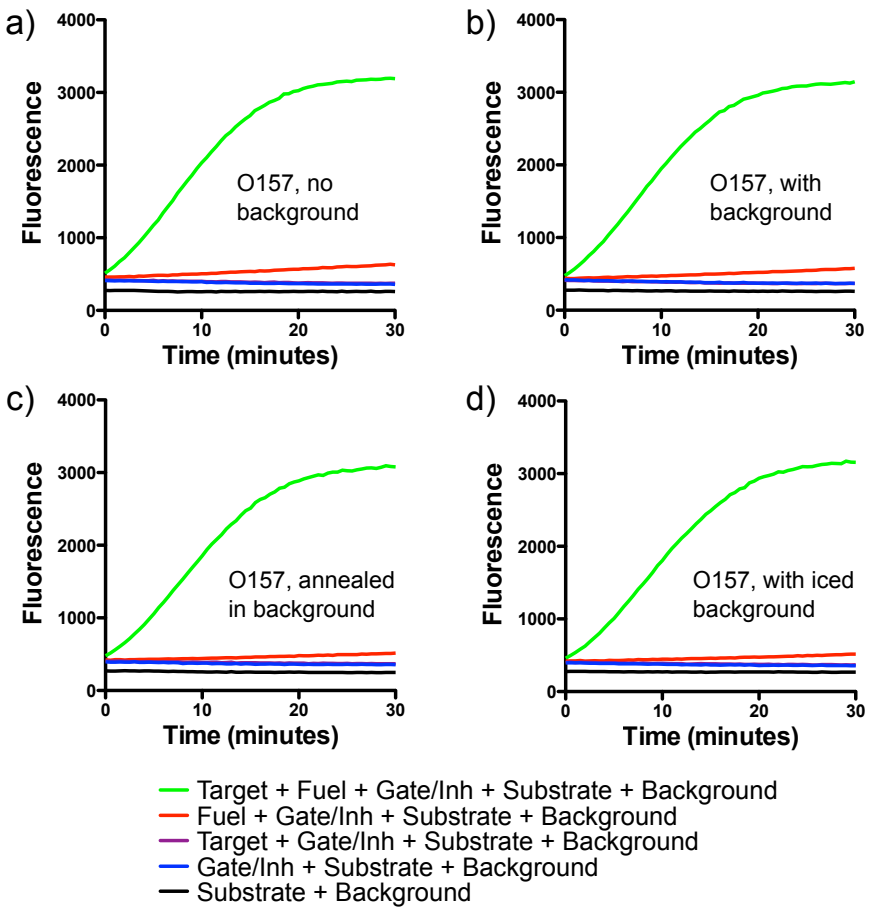

Figure A4.3 – Characterization of the O157 gate (P3 mismatch) (**Table A4.7**) in the presence of various preparation of a random DNA background (**Section 6.5.5**). a) Control experiment with no background. b) Characterization in presence of background DNA. c) DNAzyme-inhibitor complexes were annealed in the presence of herring sperm DNA, allowing the possibility of gate misfolding due to interactions with the background. d) DNAzyme-inhibitor gate complexes and the background DNA were heated separately, and the background was quenched on ice before adding to the solution, to prevent rebinding of the background DNA. All experiments used 100 nM O157 gate, 25 nM excess inhibitor, 50 nM target strand, 500 nM fuel, 250 nM substrate. Experiments in background also contained 1 μM herring sperm DNA (Promega, Madison, WI).

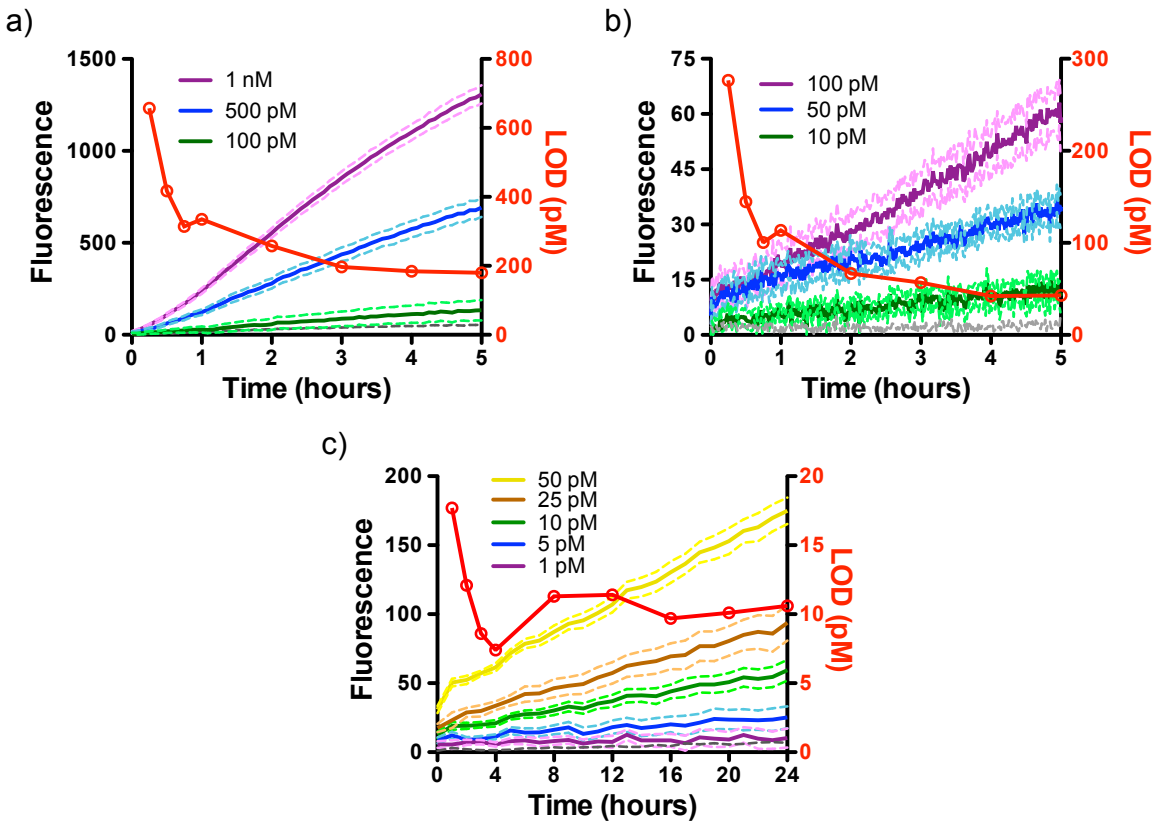

Figure A4.4 – Additional investigation of detection limits of O45 sensor gate. Serial dilutions of circuit components were used to reduce the concentration of the entire circuit (substrate was held constant at 250 nM). Colored line traces (left axis) show response with various input concentrations. The background signal in absence of input has been subtracted from all traces. Solid lines are average fluorescence values from 5 replicates, and dashed lines are one standard error above and below the mean in each case. Red data points (right axis) are detection limits at 3σ above the standard error of the background at various time points, calculated using the standard IUPAC definition. a) 10 nM O45 gate, 1 nM excess inhibitor, 10 nM fuel with 1 mismatch, 250 nM substrate, 1 ug/mL background, varied target concentrations. b) 1 nM O45 gate, 100 pM excess inhibitor, 1 nM fuel with 1 mismatch, 250 nM substrate, 1 ug/mL background, varied target concentrations (4 replicates). Note that a higher sampling frequency was used for the fluorescence measurements in this reaction, causing the traces to appear more jagged. c) 100 pM O45 gate, 10 pM excess inhibitor, 100 pM fuel with 1 mismatch, 250 nM substrate, varied target concentrations.

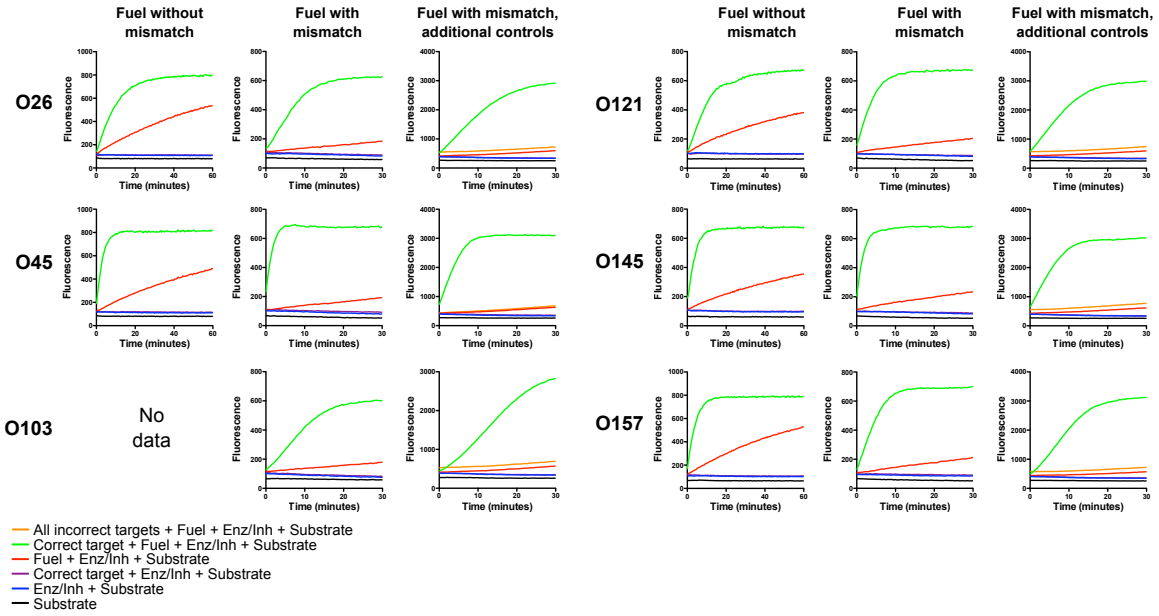

Figure A4.5 – Full characterization of STEC detection gates with additional controls. The inclusion of a mismatched base in the fuel strands significantly reduces the response in the absence of the target strand by reducing unwanted binding of the fuel strand to the deoxyribozyme-inhibitor complex. Experiments with only the incorrect target strands demonstrate sequence specificity as well as highlighting that unwanted fuel binding is the dominant source of leakage in these reactions. Experiments in columns 1 & 2 were performed using 50 nM substrate, column 3 was performed with 250 nM substrate (this explains the comparatively lower leak in column 3). All experiments used 100 nM deoxyribozyme-inhibitor complexes, 25 nM excess inhibitor, 50 nM of each corresponding target strand used in the reaction, and 500 nM fuel.

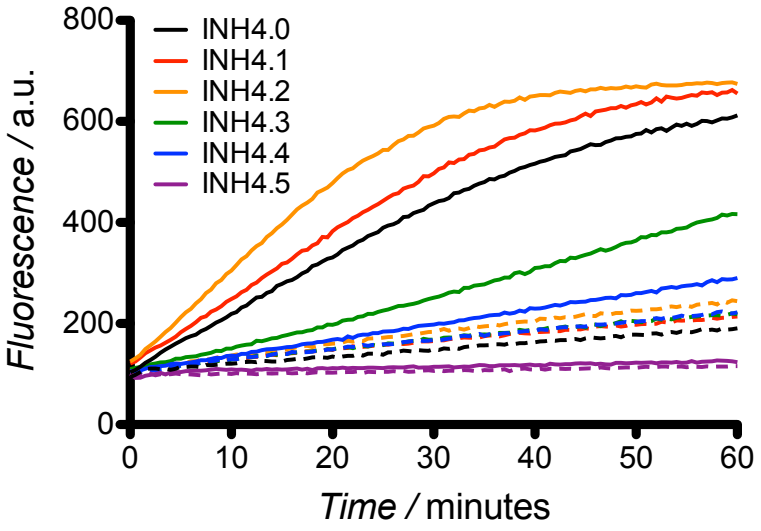

Figure A4.6 – Examination of loop toehold length on activation and leakage. The activation (solid line) and leakage (dash line) for each looped toehold length was tested. Starting from inhibitor 0 (black), bases were added to the enzyme strand to hybridize to the inhibitor, decreasing the available size of the toehold. Interestingly, it appeared that binding one (red) or two (orange) bases serves to increase both activation and leakage. Binding of all five bases (purple) eliminated binding of fuel and prevented gate activation.

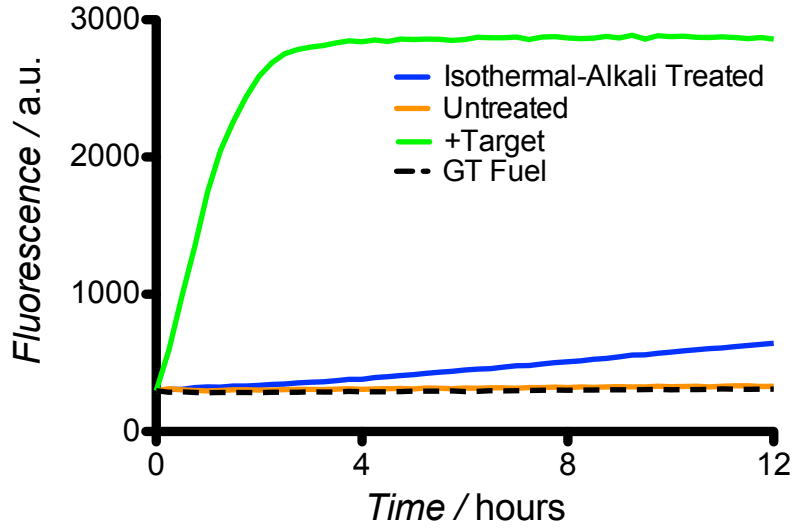

Figure A4.7 – Kinetic profile of P1, P2, L8 gate with various detection targets. Gate and substrate only (black dash) was nearly identical to the untreated (dsDNA) plasmid target. The single-stranded target (green) activated the gate quickly whereas direct plasmid detection was much slower, indicating the significant effect of diffusion and steric hindrance on gate activation.

A4.2 Discussion of P3 Leak Profiles

Additional controls (**Figure A4.5**) showed that the DNAzyme-inhibitor complexes are highly stable in the absence of the fuel strand, even in the presence of the target sequence. The fuel strand is the major source of non-specific background signal, which varies in a concentration dependent manner (**Figures A4.1 and A4.5**). We surmise that imperfections in the sensor gate complex and breathing of the duplexes near to the secondary toehold bulge may allow strand invasion via the secondary toehold in the absence of the target sequence, allowing the fuel strand to displace a catalytically active deoxyribozyme even though the inhibitor in the detection module remains bound. To address this, we used fuel strands with a rationally introduced mismatched base in the toehold domain, which considerably reduced the rate of spurious activation due to the fuel strand binding to the toehold in the bulge (**Figure A4.5**). The use of excess inhibitor also helps to inhibit the deoxyribozymes more efficiently in the face of concentration variations. However, the excess inhibitor may hinder circuit operation by binding to target strands, preventing them from activating the gates, and by rebinding activated deoxyribozyme strands to deactivate them, thereby reducing assay sensitivity. We also strengthened the inhibitor binding to the gate by extending the inhibitor by 3 nt beyond the length of the fuel strand, and used an excess of substrate relative to the deoxyribozyme concentration to enable a higher maximum signal-to-background ratio through multiple-turnover amplification.

A4.3 Use of modular gates for SNP detection

Using the P1, P2, L8 gate design and R2 detection sequence, we hypothesized that the addition of mismatches to simulate a single nucleotide polymorphism would result in a significant change in the kinetic response of the gate. We tested this under various shortened toehold lengths of 5bp and 3bp. We instituted a single mismatch between the target strand and the inhibitor toehold (detection domain), corresponding to the location of the SNP on the target strand in relation to the toehold. This was followed by the addition of a second mismatch between the target and the detection domain, where the location of the SNP could be at either position. The SNP targets are labeled for the location the SNP was placed, relative to the 3' end of the target strand and the 5' side of the inhibitor (**Table A4.11**). SNP locations were chosen at positions either in the middle of the toehold (3bp – P2; 5bp – P4), the last base of the single stranded toehold (3bp – P3; 5bp – P5), the first base of the double stranded detection domain (3bp – P4; 5bp – P6) or the second base of the double stranded detection domain (5bp – P7).

Overall, the best results were using two mismatches with a 5bp toehold (**Figure A4.8**). However, the results were somewhat curious from both the 5bp toehold as well as the 3bp toehold (**Figure A4.9**). In the case of the 5bp toehold, the position of the a single mismatch or two mismatches had very little effect on the kinetic rates, a result in striking contrast to our earlier optimization of gate structure (**Section 6.2.2**). Although these results are not directly comparable, as the aforementioned optimization was with the fuel strand binding, rather than the

target strand, it was surprising we did not see a larger difference in rate. This is especially true with the 3bp toehold, where the addition of two mismatches in the 2&3 position would have left only a single base toehold on the 5' end, yet still performed similarly to the target strand with no mismatch (T2). While the addition of a mismatch in the P4 position did decrease the gate response, the rate was still far higher than would be expected given a 1-X-1 (X) toehold (in the SNP2,4 example) or a 2-X (X) toehold (in the SNP 3,4 example), where X indicates the position of the mismatch in the 3bp toehold and first position of the detection domain.

Although this mechanism shows promise, more work must be done before to further increase the signal difference between SNP and non-SNP detection. Although many of the obvious SNP locations have been tested, further experiment to verify kinetic response is necessary. It is unclear the cause of the high fluorescent response with multiple mismatches and whether this can be overcome through a more judicious choice of SNP location or if a structural alteration in gate design is necessary. It may be possible to sequester the outer toehold, currently in a linear form, to a looped form like the inner toehold, and whether such a change would result in a more pronounced SNP detection response.

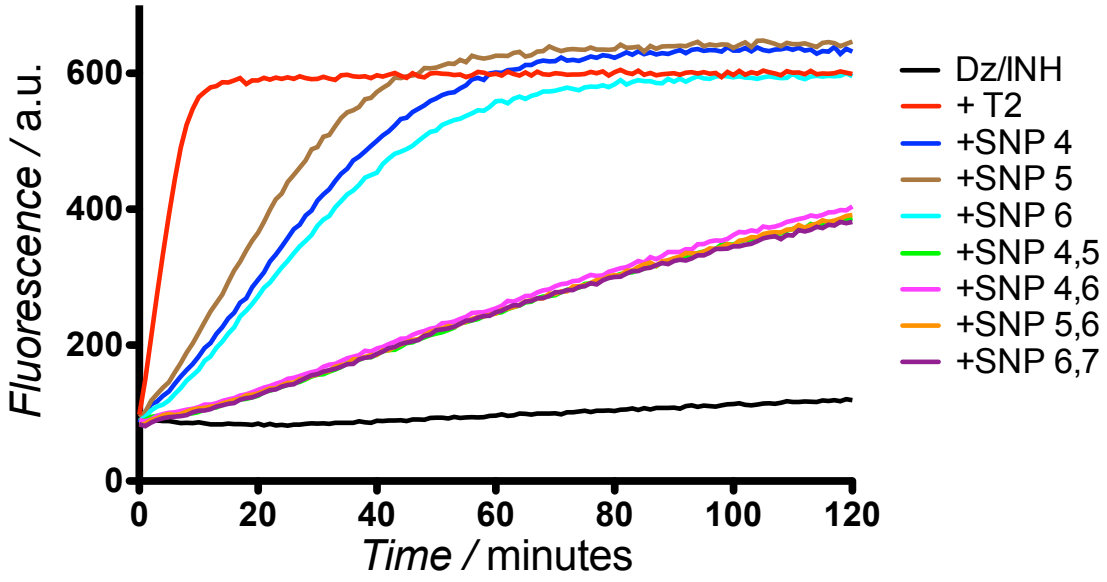

Figure A4.8 – SNP detection with a 5bp toehold. SNP names indicate the position of the SNP relative to the 3' end of the target strand. SNPs with a single mismatch are relative to the T2 target, which contains no mismatches. SNPs with two mismatches can have the SNP location in either position and kinetic traces are relative to targets with only a single mismatch.

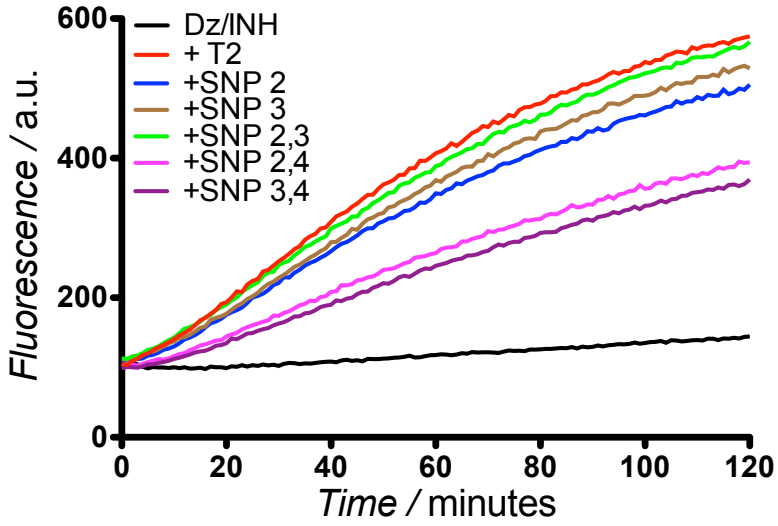

Figure A4.9 – SNP detection with a 3bp toehold. SNP names indicate the position of the SNP relative to the 3' end of the target strand. SNPs with a single mismatch are relative to the T2 target, which contains no mismatches. SNPs with two mismatches can have the SNP location in either position and kinetic traces are relative to targets with only a single mismatch.

A4.4 Alternative biosensor designs

Several different attempts were made to develop an arbitrary sequence biosensor compatible with the DNAzyme displacement gates from **Chapter 3**. The two most interesting were “detection stem loop” (DSL, **Figure A4.10**), and a “viral strand displacement” (VSD, **Figure A4.11**). For the DSL, the target strand would bind the toehold and displace the stem. The activator, sequestered in the DSL loop, would then become single-stranded and able to displace the inhibitor from the Dz-Inh complex. This stem loop was fairly sensitive, generating a signal even at 10 pM (**Figure A4.10C**), but still resulted in significant leakage (black dash, **Figure A4.10B**). For the VSD, the target displaced the activator from the VSD complex, which could then displace the inhibitor from the Dz-Inh complex. However, to shorten the complex, the toehold and substrate binding arm of the DNAzyme were made to match the target strand. Although this also resulted in a robust response, the sequence overlap may constrain the design of cascades and require a new substrate for each new enzyme. Even still, the performance of this complex may be worth revisiting in the context of the new cascade designs.

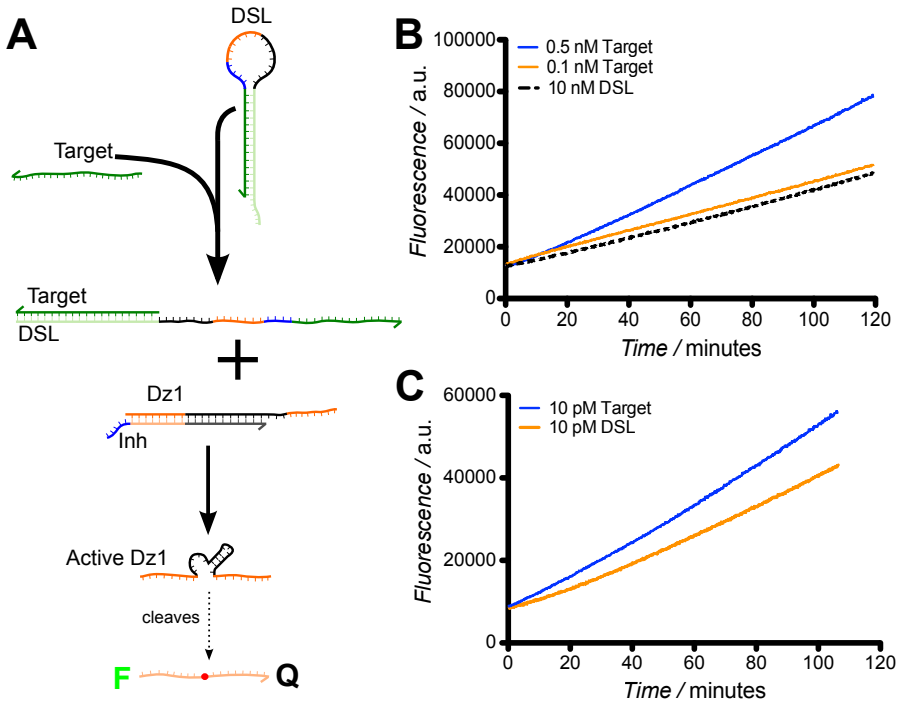

Figure A4.10 – Detection stem loop for arbitrary sequence detection. (A) Mechanism of DSL and target interaction. Target binds to the toehold on the DSL and displaces the stem. This opens the loop, which contains a sequestered activator. This activator then binds to the gate and removes the inhibitor, producing an active DNAzyme. (B) 100 nM Dz-Inh gate with 10 nM DSL. (C) 100 nM Dz-Inh gate with 10 pM DSL.

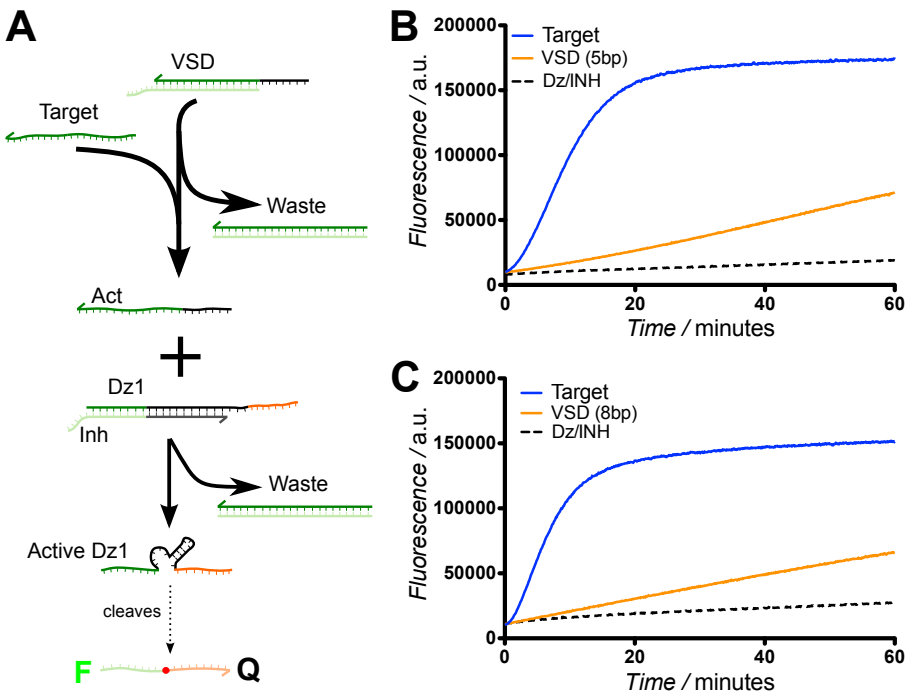

Figure A4.11 – Viral strand displacement for arbitrary sequence detection. (A) Mechanism of VSD and target interaction. Binding of the target strand to the VSD complex releases the activator, which binds to the Dz-Inh complex to produce active DNAzyme. (B) 100 nM gate with 100 nM VSD and target using a 5bp toehold. (C) 100 nM gate with 100 nM VSD and target using an 8bp toehold.

Appendix 5

A5.1. Oligonucleotide sequences

All oligonucleotide sequences are listed 5' to 3'. Functional domains (substrate binding arms and cleavage sites) have been color-coded to match with the corresponding domains. The dinucleotide junctions that are cleaved in the substrate strands have been highlighted using a yellow background, and mismatched bases are highlighted in red text. The RNA base at the cleavage site in each substrate strand is represented as rA. For Sub1, the fluorophore (fluorescein) and quencher (TAMRA) are represented as FAM and TAM respectively. For Sub2, the fluorophore (Cy5) and quencher (Black Hole Quencher-2) are represented as Cy5 and BHQ_2 respectively. Concentrations are listed in regards to each experiment.

Table A5.1 – Corresponding sequences for each DNAzyme and substrate pair. Dz1 and Sub1 were sequences used for all screening experiments.

Strand	Sequence
DNAzyme 1 (Dz1)	GAACTATC TCCGAGCCGGTCGAA AACTAAGA
Substrate 1 (Sub1)	FAM-TCTTAGTT rAG GATAGTTC AT-TAM
DNAzyme 2 (Dz2)	CTGTCCGC TCCGAGCCGGTCGAA AATACCCAT
Substrate 2 (Sub2)	Cy5-CATGGGTATT rAG GCGGACAGG-BHQ_2

A5.2 Reaction Concentrations

Table A5.2 – Concentrations of DNA strands and EDTA, as well as relative percentages of DMF and DMSO in solution.

Experiment	Dz	Sub	% DMF/DMSO	EDTA
Figure 7.2A	0/50/250	250	--	--
Figure 7.2B	0/50	250	--	10uM, 100uM, 1mM, 10mM
Figure 7.2C	0/15/200	250	--	--
Figure 7.2D	0/15	250	--	100mM
Figure 7.4	100	50	1	--
Figure 7.5	0/25/50/100/200	250	--	--
Figure 7.6	0/15	250	0/0.5/1/1.5/2/5/10	--
Figure 7.7	0/50	250	--	--
Figure 7.8	100	50	--	--

A5.3 Additional screening data

[redacted]

	1 mg/ml, 10% DMF/water						0.5 mg/ml, 5% DMF/water					
Plate Layout	1	2	3	4	5	6	7	8	9	10	11	12
A	loCntrl	1169	1324	1420	1481	882	1169	1324	1420	1481	882	midCntrl
B	loCntrl	1170	1343	1421	1433	531	1170	1343	1421	1433	531	midCntrl
C	loCntrl	1171	1344	1422	1295	914	1171	1344	1422	1295	914	midCntrl
D	loCntrl	1172	1345	923	1277	506	1172	1345	923	1277	506	midCntrl
E	hiCntrl	1174	1346	924	1387	886	1174	1346	924	1387	886	midCntrl
F	hiCntrl	1275	1347	1455	1409	sc_1mg/ml	1275	1347	1455	1409	sc_0.5mg/ml	midCntrl
G	hiCntrl	1276	1418	1456	1509	sc_1mg/ml	1276	1418	1456	1509	sc_0.5mg/ml	midCntrl
H	hiCntrl	1319	1419	1477	1002	sc_1mg/ml	1319	1419	1477	1002	sc_0.5mg/ml	midCntrl

[redacted]

	1 mg/ml, 10% DMF/water						0.5 mg/ml, 5% DMF/water					
Fluorescence	1	2	3	4	5	6	7	8	9	10	11	12
A	346398	1160956	1097086	1224929	1132013	1416276	1235443	1142315	1209988	1061143	1385696	1501695
B	340073	1362236	1100808	1214068	523985	1048850	1369578	1169752	1216460	520746	1064992	1540379
C	339824	1352385	1371194	926771	1370252	1348071	1441687	1383984	1002828	1426392	1341250	1605777
D	338329	1246371	1186005	1005065	1303852	1120997	1319852	1277979	1062103	1355520	1112626	1464218
E	2969002	1392252	1119119	1294964	1388781	1362617	1471431	1190146	1360120	1395822	1443368	1538851
F	2798717	1117242	1194886	1168509	1381455	1421102	1182522	1285468	1184075	1436120	1464722	1583422
G	3002459	1270741	1027913	1330354	1447554	1476005	1362775	1115585	1443986	1453416	1490734	1679572
H	2957693	1144201	1043022	1323712	1280154	1429603	1261858	1114149	1395628	1383977	1448728	1549960

[redacted]

	1 mg/ml, 10% DMF/water						0.5 mg/ml, 5% DMF/water					
% of sc	1	2	3	4	5	6	7	8	9	10	11	12
A	0	74	69	80	72	98	79	71	77	64	93	105
B	0	93	69	79	17	64	91	74	78	16	64	109
C	0	92	94	53	93	91	98	93	59	96	89	115
D	0	82	77	60	87	71	87	83	64	90	68	102
E	239	95	71	87	95	93	100	75	90	94	98	109
F	223	70	78	75	94	98	75	84	75	97	100	113
G	242	84	62	90	100	103	91	69	98	99	102	122
H	238	73	64	89	85	99	82	69	94	93	98	110

Table A5.3 – (A) Well layout, as described in **Table 7.1**. (B) Raw fluorescent values (C) Relative fluorescent values, based off of the average of three solvent control replicates.

Table A5.4 – Statistical values of the various screening controls.

	Mean	SD	CV
loCntrl	341156	3579	1.0
hiCntrl	2931968	90844	3.1
midCntrl	1557984	65917	4.2
sc_1mg/ml	1442237	29552	2.0
sc_0.5mg/ml	1468061	21201	1.4

A5.4 Assay screening parameter controls

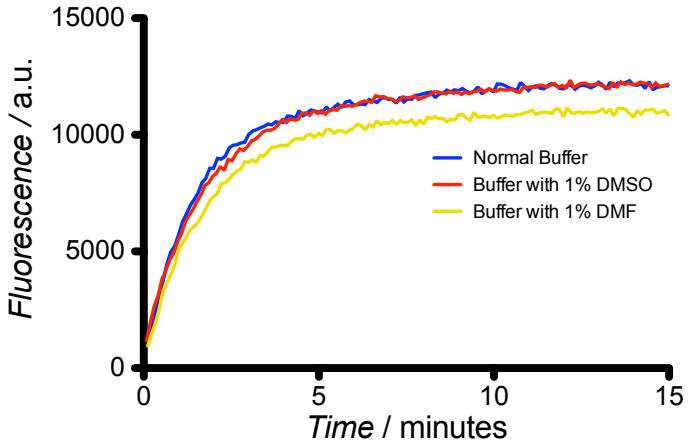

Figure A5.1 – Comparison of typical buffer conditions (1M NaCl, 50 mM HEPES, 1 mM $ZnCl_2$, pH7), with the addition of either 1% DMSO (red trace) or 1% DMF (yellow trace).

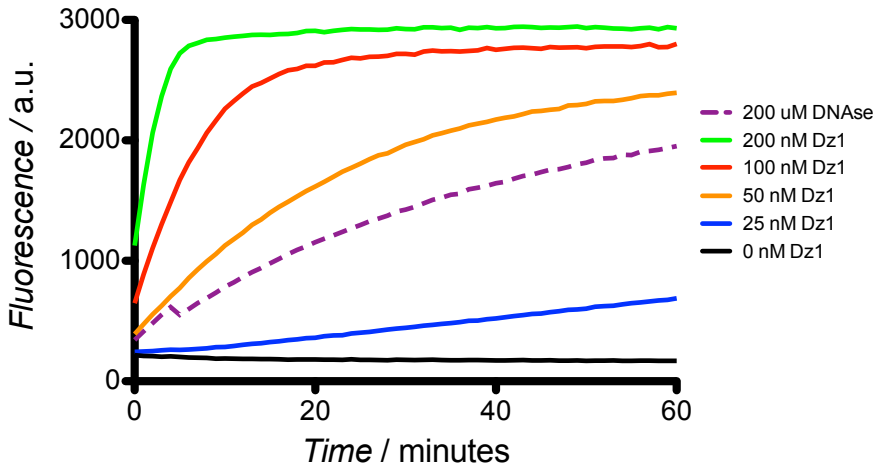

Figure A5.2 – Comparison of various concentrations of the 8-17 DNAzyme (Dz1) kinetic profiles vs. DNAse I, cleaving the substrate (Sub1). Performance of DNase was significantly below DNAzyme, making it an ineffective control for substrate cleavage.

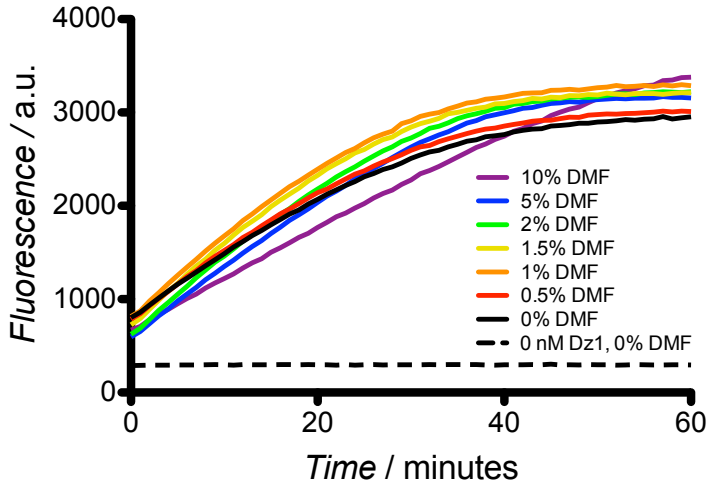

Figure A5.3 – Assessment of the Dz1/Sub1 performance in the presence of increasing percentages of DMF in solution. Only the 10% DMF solution appears to have any effect on DNAzyme activity.

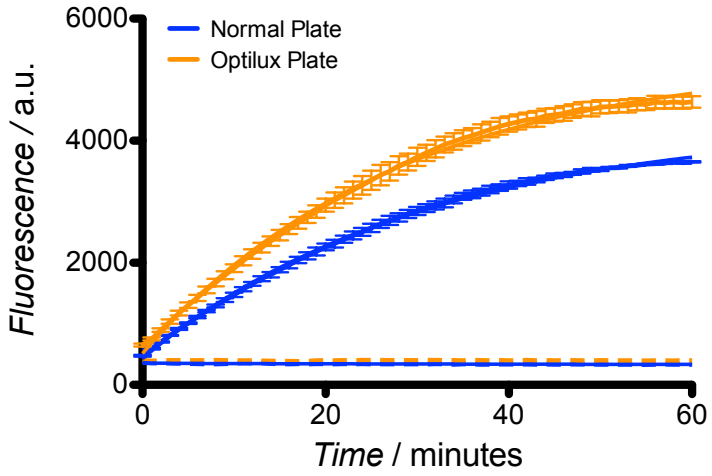

Figure A5.4 – Optilux plate usage has a moderate effect on fluorescent signal obtained on the plate reader when compared to a non-optical 96 well plate. This demonstrates the necessity of using standardized products for accurate comparison of fluorescent values.

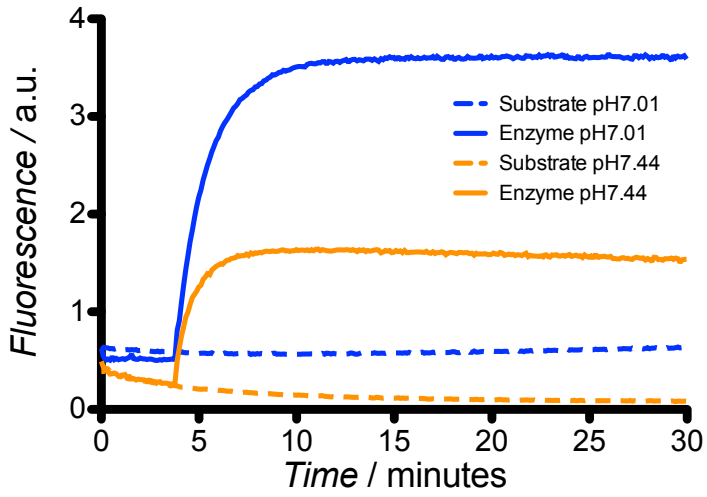

Figure A5.5 – Use of a slightly lower pH buffer resulted in a more prominent signal from the Cy5-labeled Sub2 cleavage. It also appeared to stabilize the negative control, which trended downward over time. It is unknown what caused this effect, and whether the pH7 buffer removed this effect due to a different pH or just being a newly made buffer.

Appendix 6 – Appendix References

(1) Qian, L.; Winfree, E. *J R Soc Interface* **2011**, *8*, 1281.
(2) Qian, L.; Winfree, E. *Science* **2011**, *332*, 1196.
(3) Qian, L.; Winfree, E.; Bruck, J. *Nature* **2011**, *475*, 368.
(4) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *Siam Review* **2007**, *49*, 65.
(5) Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 170.
(6) Faulhammer, D.; Famulok, M. *J Mol Biol* **1997**, *269*, 188.
(7) Okumoto, Y.; Sugimoto, N. *J Inorg Biochem* **2000**, *82*, 189.
(8) Breaker, R. R.; Joyce, G. F. *Chem Biol* **1995**, *2*, 655.
(9) Santoro, S. W.; Joyce, G. F. *Proc Natl Acad Sci U S A* **1997**, *94*, 4262.
(10) Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. *SIAM Rev* **2007**, *49*, 65.
(11) Fanning, M. L.; Macdonald, J.; Stefanovic, D. In *ACM-BCB 2011*; ACM: 2011.
(12) Zadeh, J. N.; Wolfe, B. R.; Pierce, N. A. *J Comput Chem* **2011**, *32*, 439.
(13) SantaLucia, J., Jr. *Proc Natl Acad Sci U S A* **1998**, *95*, 1460.
(14) Kim, H.-K.; Rasnik, I.; Liu, J.; Ha, T.; Lu, Y. *Nat Chem Biol* **2007**, *3*, 763.
(15) Mazumdar, D.; Nagraj, N.; Kim, H.-K.; Meng, X.; Brown, A. K.; Sun, Q.; Li, W.; Lu, Y. *J Am Chem Soc* **2009**, *131*, 5506.
(16) Kim, H. K.; Li, J.; Nagraj, N.; Lu, Y. *Chem Eur J* **2008**, *14*, 8696.
(17) Brown, C. W., III; Lakin, M. R.; Stefanovic, D.; Graves, S. W. *ChemBioChem* **2014**, *15*, 950.
(18) Zhang, D. Y. *J Am Chem Soc* **2011**, *133*, 1077.
(19) Chandra, M.; Sachdeva, A.; Silverman, S. K. *Nat Chem Biol* **2009**, *5*, 718.
(20) Gu, H.; Furukawa, K.; Weinberg, Z.; Berenson, D. F.; Breaker, R. R. *J Am Chem Soc* **2013**, *135*, 9121.
(21) Xiao, Y.; Wehrmann, R. J.; Ibrahim, N. A.; Silverman, S. K. *Nucleic Acids Res* **2012**, *40*, 1778.
(22) Green, S. J.; Lubrich, D.; Turberfield, A. J. *Biophys J* **2006**, *91*, 2966.
(23) Genot, A. J.; Zhang, D. Y.; Bath, J.; Turberfield, A. J. *J Am Chem Soc* **2011**, *133*, 2177.
(24) Gao, Y.; Wolf, L. K.; Georgiadis, R. M. *Nucleic Acids Res* **2006**, *34*, 3370.
(25) Santoro, S. W.; Joyce, G. F. *Biochemistry* **1998**, *37*, 13330.
(26) Yashin, R.; Rudchenko, S.; Stojanovic, M. N. *J Am Chem Soc* **2007**, *129*, 15581.
(27) Jiang, Y. S.; Bhadra, S.; Li, B.; Ellington, A. D. *Angew Chem Int Ed* **2014**, *53*, 1845.