Engineering DNA Computation with Unnatural Nucleotides and Protein Function with Unnatural Amino Acids

by

Alexander Michael Prokup

B.S. Chemical Engineering, University of Illinois at Urbana-Champaign, 2010

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Chemistry

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This thesis was presented

by

Alexander Michael Prokup

It was defended on

July 2nd, 2015

and approved by

Seth Horne, PhD, Chemistry Department

Kabirul Islam, PhD, Chemistry Department

Marcel Bruchez, PhD, Chemistry Department and Biological Sciences Department

Committee Chair: Alexander Deiters, PhD, Chemistry Department

ENGINEERING DNA COMPUTATION WITH UNNATURAL NUCLEOTIDES AND PROTEIN FUNCTION WITH UNNATURAL AMINO ACIDS

Alexander Michael Prokup, PhD

University of Pittsburgh, 2015

Synthetic engineering methods, such as DNA computation and unnatural amino acid mutagenesis, have provided a route to improve the control of DNA and proteins. DNA computation encompasses a broad field that attempts to build computational devices from DNA structures. Logic gates are a fundamental component of any larger computational network, and have been constructed from purely DNA frameworks. Operation is determined by strict rules, which allow for the predictable creation of complex circuits. Described herein are methods to alter DNA logic gates with photochemical caging groups, interface logic gates with protein output, and optically control DNA amplification cycles. These methods have enabled precise temporal and spatial control, as well as merged the interface between DNA circuits and biological systems. Unnatural amino acid mutagenesis enables the site specific alteration of protein residues, through the insertion of a non-canonical amino acid. Incorporation of these unnatural residues has greatly expanded the function of proteins, with the introduction of new chemical functionalities. These chemical handles have enabled applications such as the study of abasic bypass in DNA polymerases and protein-RNA crosslinking.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

There are many people who have helped me throughout my time in graduate school. I'd like to thank my advisor, Dr. Alex Deiters, for his support and scientific guidance. I know I am a much better scientist due to his help and instruction. I'd like to extend a special thanks to Dr. Kabirul Islam, Dr. Seth Horne, and Dr. Marcel Bruchez for serving on my defense committee.

To past and present group members, specifically James Hemphill, Kalyn Brown, Yuta Naro, Jessica Torres-Kolbus, and Chunjung Chou, for teaching me lab techniques, providing feedback on papers and presentations, and their friendship. Thank you to my Mom, Sara and Christopher, all of my family, and all of my friends, for your loving support and continually encouraging me to pursue my dreams. Also, to Danielle Conte, who has motivated me to be the best I can be and provided reassurance during difficult times.

1.0 DNA COMPUTATION

The field of DNA computation has grown significantly since Adleman used DNA to solve the Hamiltonian path problem in 1994.[1] DNA-based devices have been designed to perform mathematical operations,[2,3] simulate human memory,[4] simulate games,[5] and perform logic operations.[6,7] Although DNA computational devices act similarly to silicon-based computers, they are not necessarily direct competitors. DNA enables greater information storage density than electrical systems[8] and facilitates highly parallelized computation with sequence-specific strands. However, susceptibility to environmental factors such as pH, temperature, salt concentrations, long operating times, and accumulation of background signal[9] are limiting factors of DNA-based platforms. Additionally, retrieval of information from these systems ultimately requires the use of electrical instrumentation, e.g., plate readers or DNA sequencers. Yet, the greatest advantage of DNA-based circuitry is the ability to directly interact with other biomolecules in a cellular environment.[7] Therefore, a major goal of DNA computation is to mimic the operating processes of electrical systems using devices composed of biological molecules.

DNA is an excellent material for designing and building computational devices. The Watson-Crick base pairing and double helical structure is well understood and allows for precise and predictable hybridization between multiple DNA strands. Devices

have been designed to incorporate aptamers,[10] deoxyribozyme activity,[11] G-quadruplexes[12], and molecular beacons[13] demonstrating the variety of functions and interactions that can be incorporated into DNA computation. Synthetic modifications, such as quenchers, fluorophores, photocleavable caging groups, unique nucleobases, and different backbones greatly expand the capabilities of DNA beyond its natural functions.

Operation of the devices typically relies on toe-hold mediated strand exchange (Figure 1.1).[14] A toe-hold is a single-stranded overhang connected to a DNA duplex. The purpose of a toe-hold (typically about six nucleotides long) is to facilitate strand displacement and hybridization by bringing two strands within close proximity. In the absence of a toe-hold, strand displacement is kinetically slow and offers little or no thermodynamic benefit (Figure 1.1A). If a toe-hold is available, the incoming strand is able to bind and displace the hybridized strand completely (Figure 1.1B). Without an exposed complementary region for the displaced DNA, the exchange is unidirectional. If a toe-hold exists for both the incoming and outgoing strands, then the exchange is reversible (Figure 1.1C). Reaction pathways can be precisely controlled by selecting a unique toe-hold sequence for each subsequent device. Multiple devices can be connected serially, generally by designing the outgoing strand of one computation event to serve as the incoming strand for the next event. To prevent premature activation of the consecutive displacement reaction, the toe-hold can be masked in a double stranded DNA structure. Serial connections allow for the creation of larger and more complex circuits from smaller modular units.[15] Toe-holds are an integral design

component of nearly all DNA-based circuits and enable extraordinary control of strand exchange reactions.

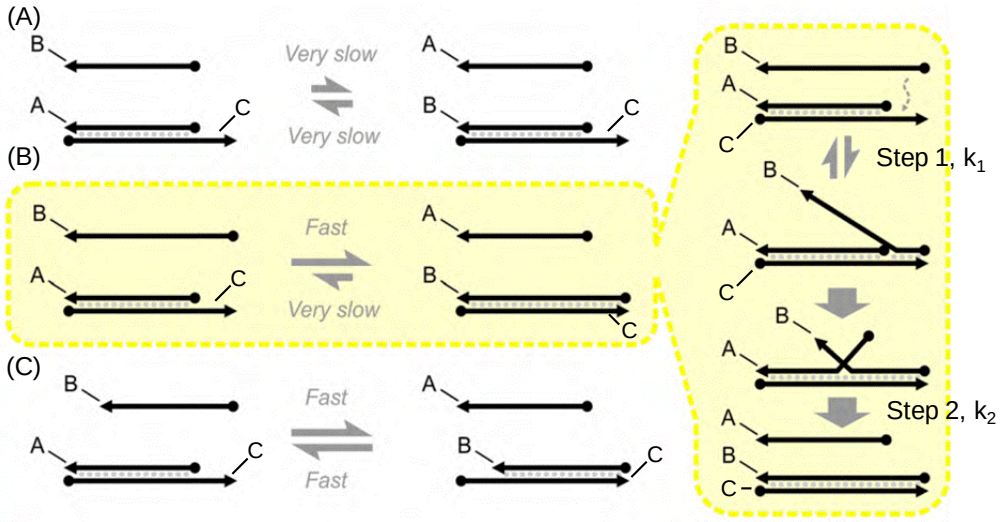


Figure 1.1. Toe-hold mediated strand displacement schematic. Toe-holds are shown as ssDNA overhangs connected to the DNA duplexes. (A) If strand B cannot bind to the gate toe-hold, the displacement of A is unfavorable. (B) If a toe-hold on the gate duplex is only available for strand B, but not A, the displacement of strand A is quick and essentially irreversible. The strand exchange reaction outlined in yellow shows the intermediate steps of B binding and removing strand A. Toe-hold binding (Step 1) and strand displacement (Step 2) are also indicated. Corresponding rate constants are 10^6 $M^{-1}s^{-1}$ for the toe-hold binding (Step 1, k_1) and 1.0 s^{-1} for strand displacement (Step 2, k_2). (C) When toe-holds are present in the gate duplex for strands B and C, the hybridization is reversible. Figure adapted with permission from Chen *et al. Curr. Opin. Biotechnol.* 2010, *21* (4), 392-400.

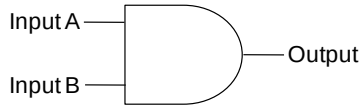
Winfree investigated the kinetics and thermodynamics of toe-hold reactions in great detail.[14, 16] The rate constant k_1 for binding of the toe-hold region to a complementary domain was around 10^6 $M^{-1}s^{-1}$, whereas the rate constant k_2 for strand

displacement was 1.0 s^{-1} (Figure 1.1B). However, these calculated rate constants were only reasonably accurate when the concentrations of DNA strands were below a low nanomolar critical concentration. Above this concentration, the toe-hold binding step is no longer the rate limiting step. These rate constants are also anticipated to change as the sequence of each domain is altered. As expected, when the toe-hold length was increased, the overall rate of strand displacement increased. For example, increasing the toe-hold length from 3 to 5 nucleotides increased the rate from 10^3 to 10^6 $M^{-1}s^{-1}$.[16] For a typical six nucleotide toe-hold, the average binding energy was −8.3 kcal/mol. Strengthening toe-hold binding through incorporation of more G and C nucleotides increased the binding energy to −12.1 kcal/mol, whereas weakening the toe-hold binding by adding more A and T nucleotides decreased the binding energy to −5.3 kcal/mol.[14] Even a single basepair mismatch can lead to noticeable effects on overall strand displacement.[9] To further aid in the design of circuits, a software program called Visual DSD (DNA strand displacement) was created.[17] The program enables easy visualization of reaction networks and has been used in the creation of logic gates[2] and larger circuits, such as a neural network.[4]

One of the most basic computational devices is the logic gate, which accepts two or more inputs and produces a single output. The operation of a logic gate is outlined by each individual truth table, which displays all possible combinations of inputs. For example, an AND gate will only yield an output if both inputs A and B are present (Figure 1.2). Logic is displayed in Boolean format, where a “1” represents the presence of an input and “0” represents the absence of an input. To create larger and more complex circuits, logic gates can be connected in series or parallel. In DNA

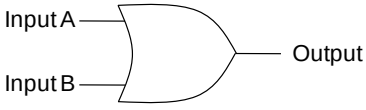
computation, the inputs are strands of DNA instead of voltages used in electronic gates.[6,7,18] Outputs are also DNA strands, and are commonly detected by fluorescence emission of a covalently attached fluorophore. A major limitation of current DNA gate designs is the stoichiometric dependence of the output on the amount of input. Generally, one input strand cannot generate more than one output strand. Therefore, DNA devices have been designed that self-amplify[15] or rely on amplification processes[19,20] to boost output signals. Amplification cycles are discussed in detail in Chapter 2.0 .

(A)



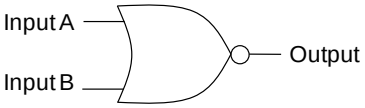
Input A	Input B	Output (luminescence)
0	0	0
1	0	0
0	1	0
1	1	1

(B)



Input A	Input B	Output (luminescence)
0	0	0
1	0	1
0	1	1
1	1	1

(C)



Input A	Input B	Output (luminescence)
0	0	1
1	0	0
0	1	0
1	1	0

Figure 1.2 Logic gate symbols and corresponding truth tables. (A) The AND logic gate generates an output only in the presence of both inputs A AND B. (B) An OR will produce an output if A OR B is present. (C) The NOR gate, also known as a negated OR gate, operates as if an inverter was placed after the OR gate. Therefore, output generation is prevented in the presence of input A OR B. Inputs and outputs are shown in Boolean format (1 or 0).

Ghadiri has created three logic gates from a similar DNA scaffold.[21] The AND gate was composed of a DNA strand modified with a 3' carboxy fluorescein. An output

was produced only in the presence of both inputs: a complementary strand of DNA and a duplex-binding fluorescent dye, Hoechst 33342. Fluorescence resonance energy transfer (FRET) between the two dyes enabled output in the form of fluorescein fluorescence at 520 nm, from excitation of the Hoechst dye at 350 nm. A negated AND gate, or NAND gate, was constructed by using ethidium bromide as the second input instead of the Hoechst dye. Excitation of the 3' fluorescein (490 nm) produced fluorescence output at 520 nm. The output was continually produced until addition of both inputs. Formation of the duplex from the complementary DNA input strand enabled intercalation of ethidium bromide. The fluorescence produced by fluorescein was then quenched by ethidium bromide. An INHIBIT gate combined the aspects of both the AND and NAND gates. When a third input was added, an INHIBIT gate prevented any output generation. Three inputs were necessary for the INHIBIT gate: the complementary DNA strand, the Hoechst dye, and the ethidium bromide dye. The INHIBIT gate would function exactly as the AND gate in the absence of ethidium bromide. When added, the ethidium bromide quenched any fluorescein fluorescence, thus inhibiting output production. These three designs demonstrate how logic gate frameworks can be easily outfitted with light-responsive elements. Creation of new gates was as simple as exchanging fluorophores. However, these gates could not be non-invasively spatially or temporally controlled.

Ghadiri has also created logic gates based on the principles of toe-hold mediated strand exchange (Figure 1.3).[22] These logic gates were constructed on solid supports in order to physically separate gates to prevent premature activation. With the gates separated, the solution phase enabled communication between the gates. A simple

modification of sequences allowed for construction of AND, OR, and AND-NOT (NAND) gates. In order to observe gate activity, all final output strands from the gates interacted with a reporter gate to produce a fluorescent signal. Different sequence domains appended to the output strands allowed for directionality of information traveling through a network. For example, the addition of the B domains on the output strand in Figure 1.3 directed the output strand to the next appropriate gate. Multi-level circuits were created using these directional domains. Notably, a XOR gate was fashioned from the connection of the AND, OR, and AND-NOT gates. Overall, premature gate activation was minimized by using solid-support gate structures and the modular nature and uniform construction allowed for multi-level networks to be created. However, the necessity of solid supports limits the application of logic gates to *in vitro* environments only. Additionally, the scale of networks may be limited by the solution-phase diffusion of outputs and inputs between gates.

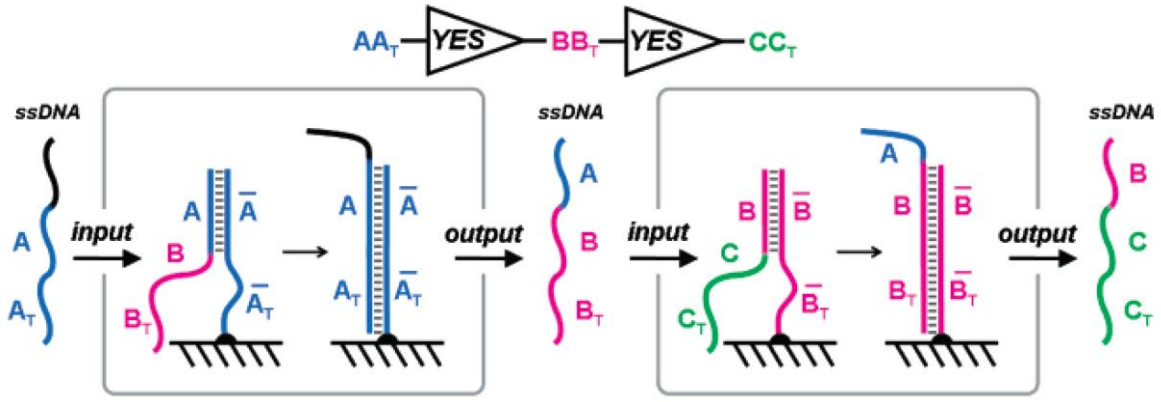


Figure 1.3. Multi-level networks were created from solid-support DNA logic gates. The input strand A_TA can interact with the first YES gate to release an intermediate output strand ABB_T through a toe-hold mediated strand exchange. This output can then act as an input for the second YES gate in a similar manner to displace the output strand BCC_T. Domains are abbreviated by capital letters (e.g., A, B, and C). Complementary domains are indicated with a line above the letter. Adapted with permission from Frezza *et al. J. Am. Chem. Soc.* 2007, *129* (48), 14875-14879. Copyright 2007 American Chemical Society.

One of the first interactive circuits composed entirely of nucleic acids was the molecular automaton MAYA.[5, 23] The purpose of MAYA was to play a game of tic-tac-toe against a human player. Furthermore, MAYA demonstrated how DNA-based circuits could be programmed to think and respond, similar to how a computer program acts. Tic-tac-toe was chosen as the model game due to its straightforward gameplay. The moves were simplified by always starting with MAYA choosing the center. Symmetry of the game board then dictated that all future human moves would either be to a corner or side space, further streamlining the programming process. Deoxyribozyme logic gates based on the E6 core were used in all wells of the 3x3 game board (Figure 1.4A). A combination of AND, YES, and NOT gates was constructed to interact with DNA (Figure 1.4B). The deoxyribozyme RNase activity could be activated or inactivated based on the binding of the inputs. Similar to a toe-hold mediated strand exchange, the inputs would bind to open ssDNA loops and cause an opening of the hairpin duplex structure. A fluorescence output was generated by the cleavage of fluorophore and quencher labeled DNA substrates at an internal ribonucleotide modification. Each move was signaled by the release of a fluorescent signal. In over 100 games, MAYA did not make an incorrect move. Although MAYA would never be able to compete with silicon-based

computing in terms of complexity or speed, the ability to program responses into DNA could eventually be used for controlling molecular-based therapeutic and diagnostic devices.

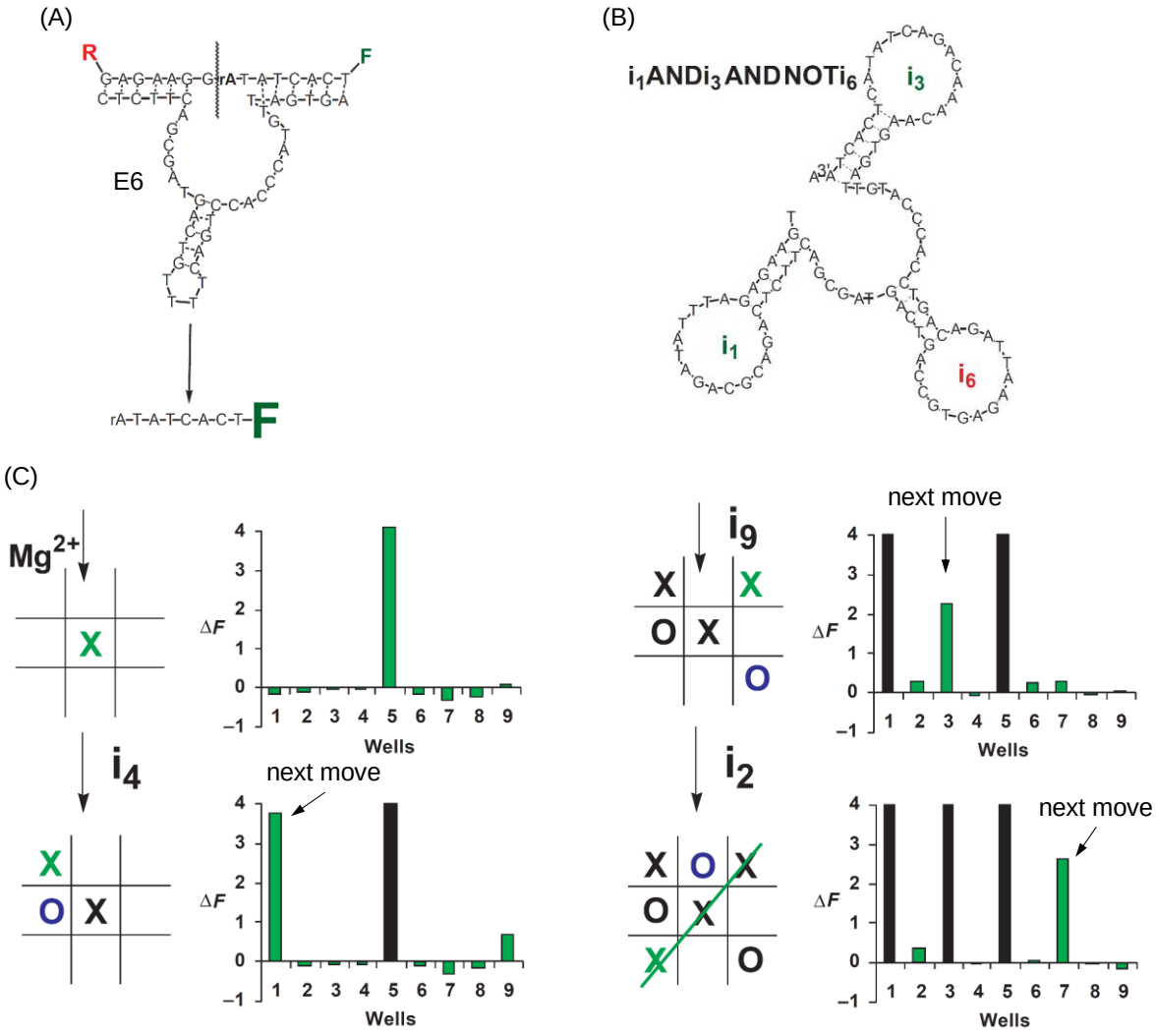


Figure 1.4. Molecular automaton MAYA playing tic-tac-toe. (A) The deoxyribozyme E6 was used as the base structure for all further logic gate designs. A DNA strand containing quencher (R) and fluorophore (F) moieties was cleaved in the middle of the strand at the ribonucleotide rA. (B) An example of an i1ANDi3NOTi6 logic gate. The binding of i1 AND i3 would activate the deoxyribozyme. However, binding of i6 would inactivate the ribozyme. (C) A game of tic-tac-toe was played with MAYA (X markers) against the human (O makers). Green and blue markers indicated new respective moves by MAYA and the human. The fluorescence column graphs to the right of the game board show the next selected move by MAYA in green. Previous moves are indicated by a black column. Adapted with permission from Stojanovic and Stefanovic. *Nat. Biotechnol.* 2003, *21* (9), 1069-1074.

Winfree used modular DNA logic gates to create a larger circuit.18 Three basic logic gate designs were created for AND, OR, and NOT gates, based on toehold-mediated strand exchange reactions. The central component is a complex of three hybridized DNA strands. Alone, the complex functioned as an AND gate (Figure 1.5A). Fluorophore and quencher modifications were made to adjacent 3' and 5' ends to enable fluorescence emission as a quantifiable output (Figure 1.5B). To create an OR gate, a translator gate was added upstream of the gate.18 The translator gate essentially acted as a one-input YES gate, releasing an output of a different sequence in the presence of another input. Toehold-binding regions of the translator output strands were buried in duplexes to prevent background activity in the absence of the translator input. For the OR gate, two input strands were translated into the same output sequence. Thus, either input one OR input two would activate the gate. A NOT gate functioned according to a similar premise, except the input strand would compete for the input to the translator gate (Figure 1.5C). The input (let-7c) and translator input (K) were complementary, thus binding prevented triggering of the translator gate. This

competition culminated in the inverse of a YES gate, by preventing output in the presence of input. In addition to DNA inputs, the gates were also successfully operated using RNA inputs. Scalability was demonstrated by combining the gates into a larger circuit containing up to twelve gates total in order to compute a pattern of six inputs. However, as more layers were added to the network of gates, the processing time also increased. Adding two extra layers increased processing time from 2 hours to 10 hours. Overall, the creation of DNA-based logic gates enabled modular and simplified designs that could be easily scaled into larger networks because of the uniform nature of all inputs and outputs.

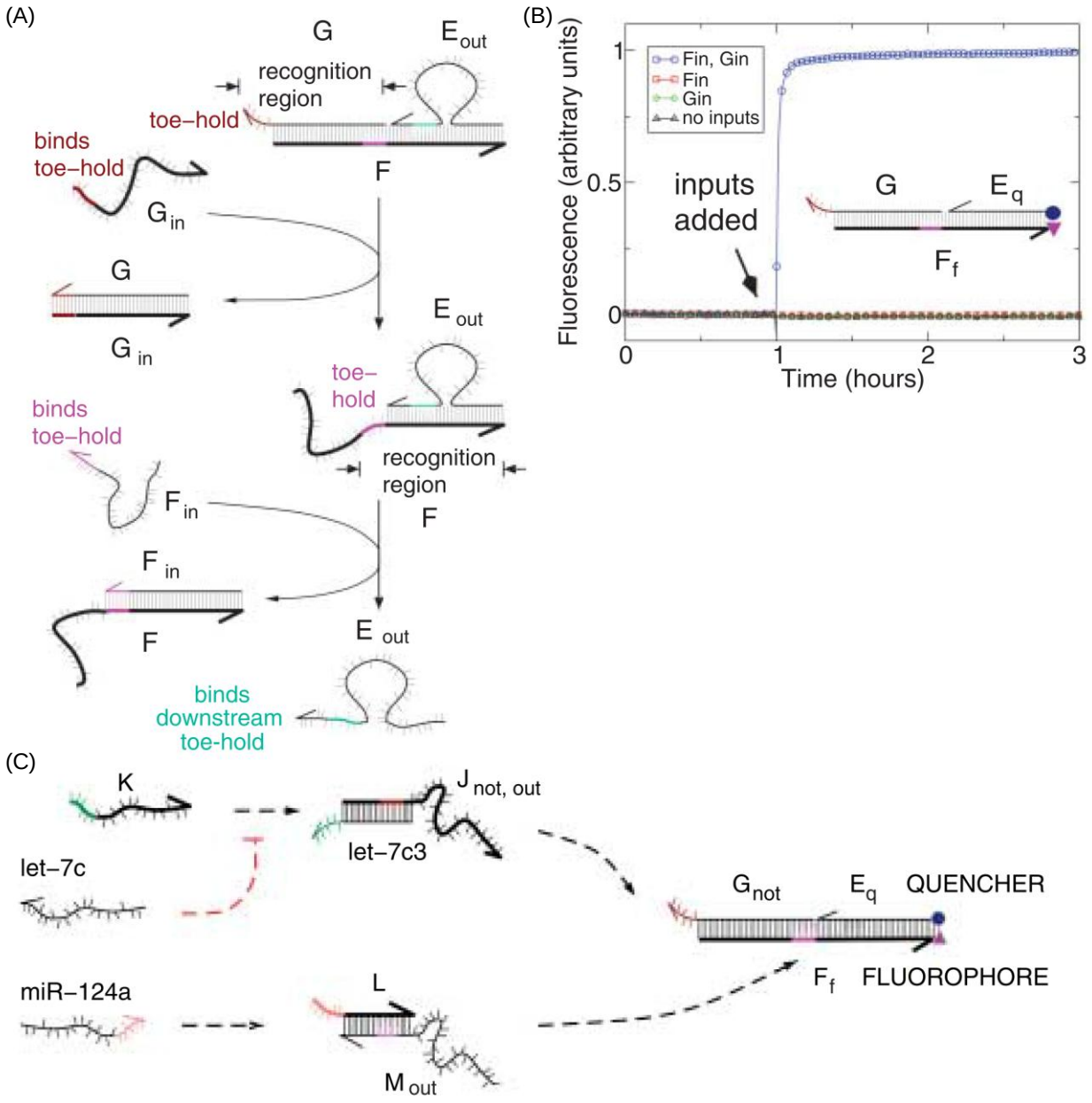

(A)
G
E out
recognition region
toe–hold
binds toe–hold
F
G in
G
E out
G in
toe–hold
binds toe–hold
recognition region
F in
F
F in
F
E out
binds downstream toe-hold
(B)
Fluorescence (arbitrary units)
Time (hours)
Fin, Gin
Fin
Gin
no inputs
inputs added
G
E q
F f
(C)
K
J not, out
let–7c3
let–7c
G not
E q
QUENCHER
F f
FLUOROPHORE
miR–124a
L
M out

Figure 1.5. DNA logic gates designed to operate based on toehold-mediated strand exchange. (A) The AND gate was composed of three DNA strands (G, E_{out}, and F). The first input (G_{in}) can bind and remove strand G to reveal a toe-hold for the second input F_{in}. Displacement of the output strand E_{out} completes the mechanism. (B) When fluorophore and quencher moieties were added to the 5' end of E_{out} and 3' end of F, respectively, the reaction could be quantitatively monitored. Only in the presence of both inputs G_{in} and F_{in} would the quencher and fluorophore moieties be separated. (C) The (NOT let-7c) AND miR-124a gate functioned similarly to the AND gate in part (B). If the let-7c input was present, it would bind to K, preventing the activation of the translator gate. In the absence of let-7c, the translator input K will release strand $J_{not,out}$ from the translator gate, which can act as the first input to the AND gate. Similarly, miR-124a will displace strand L from the bottom translator gate. Adapted with permission from Seelig *et al. Science* 2006, *314* 5805, 1585-1588.

While many different DNA computation devices have been constructed, there still remain many challenges and limitations to practical applications. However, some of these limitations are being solved through the use of optical control. Light represents a powerful input with a wide range of advantages over chemical or biological inputs.[24,25,26,27] The addition of photocleavable caging groups to DNA logic gates has enabled photocontrol over the temporal and spatial activation of DNA computation events.[6] Since photons can be emitted and received by electrical circuits, the addition of photocleavable caging groups to DNA logic gates also enables the interfacing of biological and electrical circuits. Closing the gap between biological and electrical systems expands upon the ability to program biological systems using *in silico* methods, and will be discussed further in this chapter. Light-cleavable caging groups have also enabled precise spatial and temporal control of DNA-based amplification systems.[28] Previously, spatial control of DNA computation systems has relied on pre-formed

biological structures.[29] When incorporated into a fuel-catalyst cycle, a spatially activated signal could be generated in any customizable pattern. Further modification of the cycle allowed for propagation of signal through a semi-solid, similar to a chemical wire. This lays the groundwork for creating larger spatially organized DNA networks, and will be discussed further in Chapter 2.0 . Typically, once a DNA network has been activated, signal propagation through the circuit is irreversible, and the gates cannot be used a second time. A reversible NAND logic gate was designed from a DNA structure modified with a spiropyran fluorophore. Light was used to reversibly switch the spiropyran to merocyanine, thereby activating the gate. However, full reversibility of the system also required a change in pH to alter the secondary structure of the DNA gate. Optically reversible logic gates can be reset quickly to allow multiple uses from a single gate, greatly reducing the time and cost of materials to build new gates.

Despite the limitations of DNA computational devices, there have been several “real-world” applications. For example, the mRNA localization in whole zebrafish embryos was spatially probed using the hybridization chain reaction.[29] This paves the way for the detection of other nucleic acids and nucleic acid-binding biomolecules in more complex organisms using simple and relatively inexpensive devices. Genetic logic circuits have also been investigated for their potential in creating “smart” therapeutics.[30] These circuits are able to identify mRNA disease indicators and release a drug only into the affected cells. Thus, precise treatment occurred only where the drug was needed and may help to eliminate off-target effects. Looking forward, DNA computational circuits may be useful in an industrial setting to monitor the production of a bio-based product. Bacteria and other microorganisms are increasingly finding applications as

mini-bioreactors.[31] However, industrial control systems are only designed for exogenous variables, such as temperature and pressure. To optimize and control for internal conditions, DNA computational circuits could be programmed to prevent against product degradation or apoptosis.

1.1 PHOTOCHEMICALLY CONTROLLED AND GATE

1.1.1 Results and Discussion

In order to control hybridization of oligonucleotides, DNA strands can be modified with photocleavable organic protecting group, known as a caging group. Irradiation of the caged DNA strand with UV light initiates a photochemical reaction that results in restoration of the natural nucleotide structure. The use of chemical inputs introduces variables, such as cellular uptake, processing, and diffusion that reduce the reliability of a logic gate to be controlled in a biological environment. A system in which the logic gate machinery is preassembled and later activated with light provides enhanced control and specificity. Caged nucleic acids allow for light-activation of DNA hybridization in a precise manner that other research tools cannot accommodate. Previous examples have shown that photocaged oligonucleotides can be used for recombinant DNA manipulation,[32,33,34] DNA aptamer activation,[35] ribozyme and deoxyribozyme regulation,[36,37,38] and control of gene expression through antisense technology[39,40,41,42,43] as well as RNA interference mechanisms.[44,45,46] However, photocaged nucleic acids have not been used in cellular computation or the

development of DNA based logic gates. Photochemical control of logic gate function can be achieved by employing caging groups on DNA strands responsible for toe-hold displacement. The photochemical triggering of a functional logic gate allows for spatial and temporal activation which may be used to enhance control over signaling cascades of complex DNA computation circuits.[4]

Shown in Figure 1.6 is a light-triggered AND gate based on a gate complex[18] as well as a caged (A_4) and a noncaged (B_0) DNA strand. The AND logic gate will only deliver an output signal if both photochemical input signals of different wavelengths (I_1 and I_2) are present. The gate complex is composed of three ssDNA oligomers: a fluorophore strand G_F, a quencher strand G_Q, and a toe-hold containing strand G_T. The fluorophore and quencher moieties are in close proximity preventing fluorescence. In order to activate the gate, A_4 and B_0 need to induce a toe-hold displacement cascade resulting in the removal of G_Q from G_F. The A_4 strand binds to the toe-hold of G_T separating the gate complex, allowing B_0 to bind to the toe-hold exposed on G_F. This event releases G_Q, permitting emission of the excited fluorophore. It was hypothesized that caging groups installed on select thymidine bases of the A_4 strand will prevent hybridization and thus prevent strand exchange. Therefore, without the proper light inputs for decaging (input I_1 = 365 nm) and excitation (input I_2 = 532 nm) no output signal will be observed. Thus, step 1 involves UV irradiation at 365 nm for decaging of the nucleotides. After caging group removal, complementary regions are exposed, enabling DNA:DNA hybridization. In step 2, A_4 will dislodge G_T via a toe-hold mediated strand displacement mechanism.[2] Following step 2, the gate complex consists of only fluorophore and quencher strands. Step 3 occurs spontaneously because a second toe-

hold region is exposed on the gate complex after the G_T strand was expelled by the A_4 strand. During step 4, quencher and fluorophore strands are separated by a second toe-hold mediated exchange with the strand B_0. In step 5, irradiation at 532 nm now leads to excitation of the fluorophore and observation of fluorescence emission as the output signal.

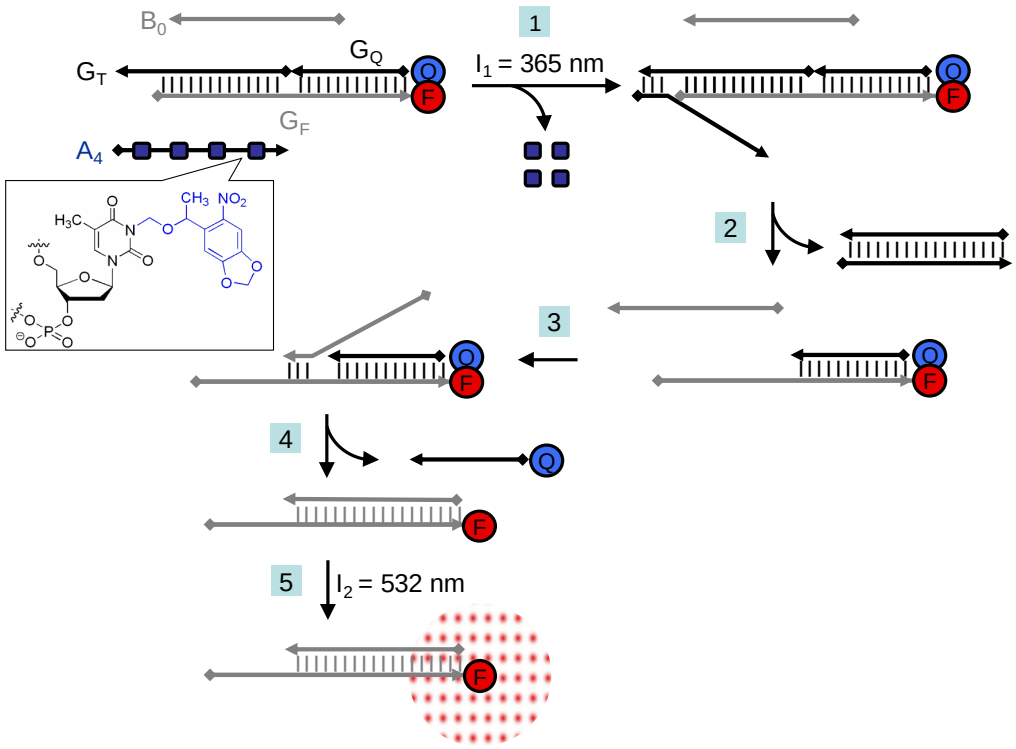


Figure 1.6. Light-triggered DNA-based AND gate using irradiation at 365 and 532 nm as input signals I_1 and I_2, respectively, and fluorescence as the output signal. The NPOM (6-nitropiperonyloxymethylene) caging group installed on thymidine nucleotides is represented by a blue square.[33] Quencher Q = Iowa Black RQ. Fluorophore F = tetramethylrhodamine (TAMRA). Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

Due to the impact of toe-holds on strand displacement kinetics,[14] toe-holds were initially selected as the primary targets for the installation of nucleotide caging groups. If needed, additional caging groups were added as evenly as possible throughout the remainder of the oligonucleotide. Thus, a set of caged and noncaged DNA oligonucleotides consisting of input strands or the gate itself were synthesized on the basis of previous reports[18,20] using standard oligonucleotide polymerization chemistry, in order to develop and investigate a light-triggered AND gate (Table 1.1).

Table 1.1. Sequences of caged and non-caged oligomers used in the light-triggered AND gate. Toehold regions are underlined and NPOM-caged thymidines are highlighted as T*. Q = Iowa Black RQ quencher. F = tetramethylrhodamine (TAMRA) fluorophore.

Strand	Sequence (5′ → 3′)
G_Q	Q-GTTAGATGTTAGTTTCACGAAGACAATGAT
G_F	TGTTTATGTGTTCCCTGATCTTTAGCCTTAATCATTGTCTTC GTGAAACTAACATCTAAC-F
B_0	GTTAGATGTTAGTTTCACGAAGACAATGATTAAGGC
B_4	GTTAGAT*GTTAGTTT*CACGAAGACAAT*GATT*AAGGC
G_T	TAAGGCTAAAGATCAGGGAACACATAAACAACCATA
G_{T1}	TAAGGCTAAAGATCAGGGAACACATAAACAACCAT*A
A_0	TATGGTTGTTTATGTGTTCCCTGATCTTTAGCCTTA
A_1	TAT*GGTTGTTTATGTGTTCCCTGATCTTTAGCCTTA
A_2	TAT*GGT*TGTTTATGTGTTCCCTGATCTTTAGCCTTA
A_3	TAT*GGT*TGTTTATGT*GTTCCCTGATCTTTAGCCTTA
A_4	TATGGT*TGTTTATGT*GTTCCCT*GATCTTT*AGCCTTA

In order to determine the effect of caging groups for the photochemical control of an AND gate, caging groups were initially added to the A_0 strand. A set of four

oligonucleotides, A_1–A_4, bearing 1–4 caging groups was synthesized and individually tested for function to study the design requirements for suppression of strand displacement. As the number of caging groups was increased, the fluorescence output of the gate linearly decreased in the absence of UV irradiation with I_1 = 365 nm (Figure 1.7). Optimal suppression of the output signal was observed with A_4, which contained four NPOM caging groups evenly distributed throughout the DNA strand and displayed no activity. Thus, the presence of only one or two caging groups in the six nucleotide toe-hold region was not sufficient to prevent initiation of the strand displacement reaction and subsequent gate function.

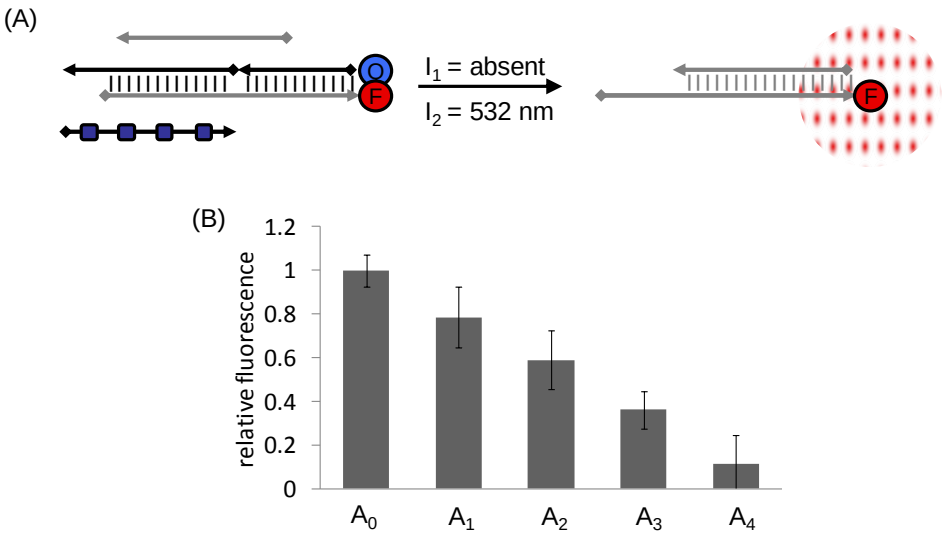


Figure 1.7. Investigation of the number of caging groups on strand A_0 that is required to inhibit gate function. (A) Simplified schematic of the light-activation of the AND gate. (B) The logic gate was not irradiated with 365 nm light in order to keep all caging groups in place, but only with 532 nm light as I_2. A linear decrease in fluorescence was observed with increasing numbers of caging groups. Four caged thymidines on strand A_4 produced optimal suppression of fluorescent signal. An average of three independent experiments is shown, and error bars represent standard deviations. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

An optimization of the UV irradiation time for decaging was conducted, and a time course was performed with the A_4 strand. Maximum fluorescence was observed after 15 min of UV irradiation at I_1 = 365 nm followed by a brief excitation at I_2 = 532 nm (Figure 1.8). Longer I_1 irradiation times lead to a decrease in fluorescence, possibly due to photobleaching of the fluorophore.[47] Activation of logic gates using noninvasive UV irradiation as an input signal shows that a DNA-based light switch can be generated, which holds promise for developing new applications of externally regulated DNA computation devices.

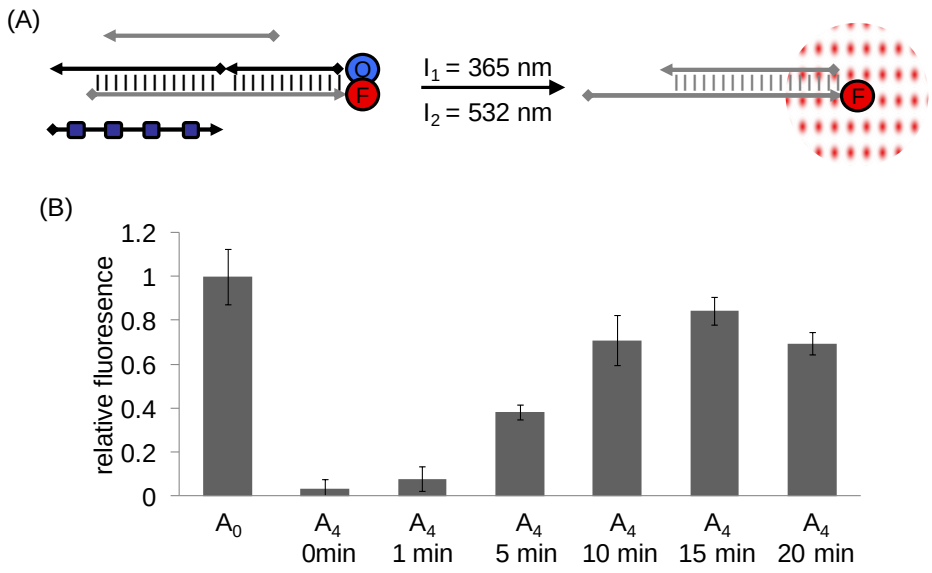


Figure 1.8. Time course of UV irradiation of the gate complex and the caged A_4 strand. (A) Simplified schematic of the light-activation of the AND gate. (B) A maximum fluorescence output signal is obtained with a 15 min irradiation at I_1 = 365 nm. An average of three independent experiments is shown, and error bars represent standard deviations. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

Although the AND gate was successfully activated by irradiation with UV light, the order of caging group removal was never investigated. To determine the which caging groups are removed from strand A_4 after irradiation with UV light, a primer partially complementary to strand A_4 could be added in addition to a DNA polymerase. In the absence of UV light, the caging group will not be removed, and the polymerase should not extend the primer past the location of the caging group.[48] Irradiation with UV light will remove the caging group and the primer will be extended until the polymerase reaches the end of strand A_4 or another caging group. Primer extension can be

analyzed by native-PAGE and an increase in primer length would indicate a caging group was removed. A restriction endonuclease digest can also be used to determine if a specific caging group was removed.[34] After irradiation with UV light for a specified time interval, the complementary strand to A_4 can be added with the restriction enzyme. Restriction enzymes would be selected such that one of the caged thymidines is present within the recognition site. Therefore, digestion of a specific sequence can only occur if the caging group has been removed. Analysis of the digestion products can be performed by native-PAGE. Appearance of short oligonucleotide fragments would indicate successful removal of a specific caging group by UV light.

After the successful light-triggering of gate activity using the caged strand A_4, four NPOM-caged thymidine nucleotides were introduced into the B_0 strand (B_4) in order to test if the AND gate can also be photochemically controlled by caging B_0. Strands containing fewer than four caging groups were dismissed on the basis of the results from testing $A_{1–4}$ (Figure 1.7). Here, the output signal was completely suppressed in the presence of B_4 and was only observed after irradiation with the input wavelengths of I_1 = 365 and I_2 = 532 nm (Figure 1.9). Thus, caged B_4 was also successful for photochemical control of the light-triggered AND gate.

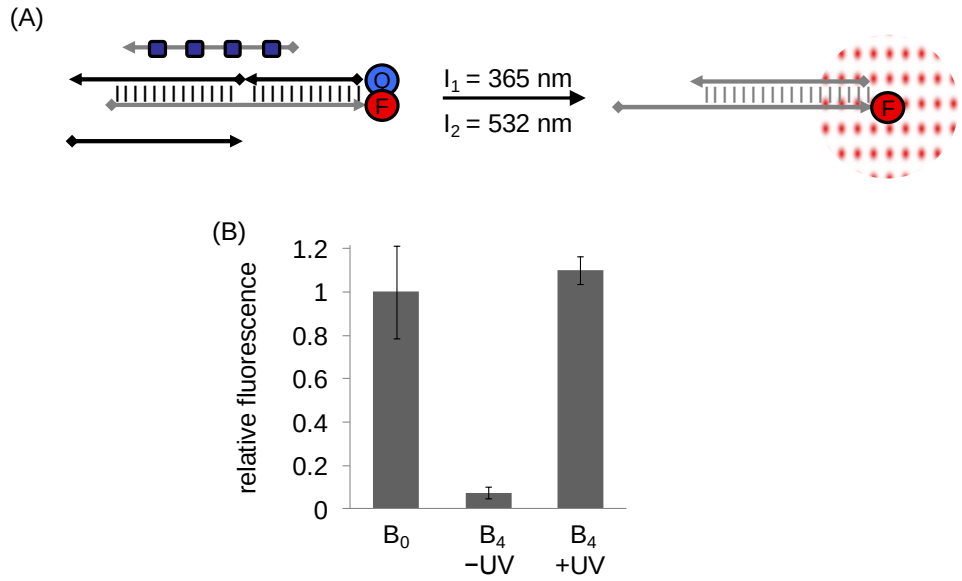


Figure 1.9. Gate activation through irradiation of a caged B_0 strand. Four caging groups on B_4 fully suppress AND gate function in the absence of UV irradiation, and decaging at I_1 = 365 nm (15 min) led to full restoration of DNA gate activity. (A) Simplified schematic of a light-triggered AND gate using the caged DNA strand B_4. (B) The logic gate was irradiated at I_2 = 532 nm with and without prior irradiation at I_1 = 365 nm. An average of three independent experiments is shown, and error bars represent standard deviations. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

In addition, caging of a thymidine located in the toe-hold region of the gate strand (G_{T1}) was investigated in order to ascertain if caging of the initiator toe-hold of the AND gate would allow for photochemical control. When noncaged A_0 and B_0 were added to a caged gate complex containing G_{T1}, a signal comparable to the noncaged G_T strand was observed indicating full function of the gate despite the presence of the NPOM caging group (Figure 1.10). Thus, the output signal was not repressed through the

addition of a single caging group to the gate complex toe-hold. The caged gate complex containing G_{T1} was also investigated with strands A_1 and A_2, which also contain caging groups in the toe-hold region of the complementary strand. Surprisingly, no complete suppression of logic gate function was observed by combining caging groups in the toe-holds of strands $A_{1–2}$ with G_{T1}. Thus, in agreement with caged A_0 strands (Figure 1.7B), caging just the toe-hold regions of interacting strands is not sufficiently effective at suppressing the strand displacement reaction. In conclusion, caging groups need to be evenly distributed throughout the sequence, including toe-hold and body regions, in order to enable photochemical control of DNA logic gate operations.

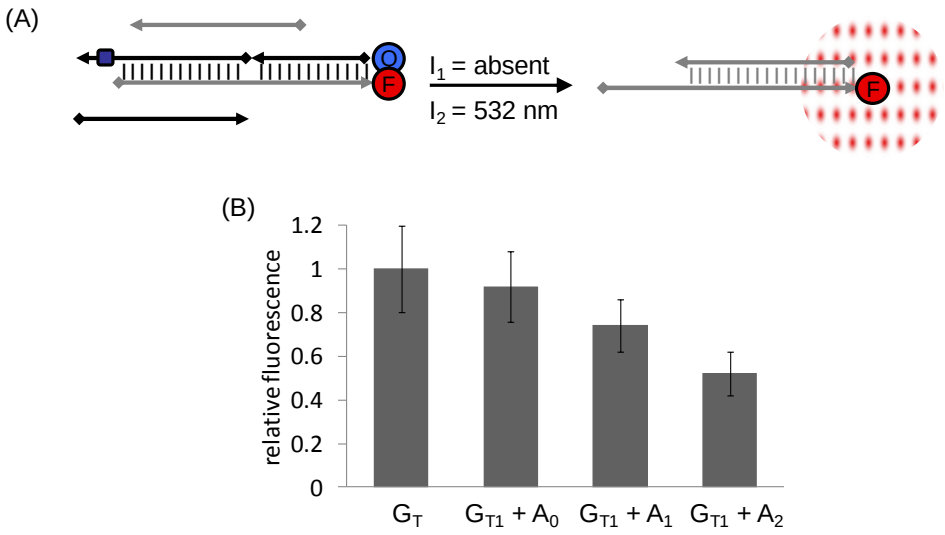


Figure 1.10. Caging of the toehold region of strands G_T. (A) Simple schematic of the caged AND gate containing caged G_{T1}. (B) The logic gate was not irradiated with 365 nm light, but only with 532 nm light. The addition of a caging group on strand G_T alone does not decrease fluorescence output. G_{T1} was analyzed with strands A_1 and A_2 to determine the effects of combining caged toe-hold regions on hybridizing strands. Full suppression of the gate output signal was not achieved, revealing an insufficient deactivation of gate function using caged toe-hold regions exclusively. An average of three independent experiments is shown, and error bars represent standard deviations. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

Unexpectedly, the AND gate could not be successfully controlled by installing caging groups only within the toe-hold or toe-hold binding sequences. At the end of the gate duplex near the toe-hold, the majority of nucleotides are A or T. These A/T base pairs are weaker than G/C base pairs and may not prevent against the natural temporary dissociation of gate duplexes. Therefore, adding clamp regions composed of G/C base pairs to the ends of gate duplexes may enable stronger duplex binding and gate activation can be controlled with fewer caging groups located in the toe-hold and toe-hold binding regions. In order to quickly and cost-effectively determine the optimal number of caging groups required to control gate activity, the caging groups can be simulated by replacing caged nucleotides with mismatched base pairs. The mismatches would prevent DNA:DNA hybridization similar to a caging group. However, the strands containing mismatches are commercially available and do not require specialized phosphoramidites.

To investigate whether the light-triggered AND gate could be controlled with temporal resolution using UV light as an input, three separate sets of experiments with A_4 were conducted and logic gates were irradiated at different time points (Figure 1.11A). The obtained fluorescent signal was only observed after UV irradiation at I_1 = 365 nm, but not before. Thus, temporal control over the light-triggered AND gate was achieved. Moreover, a step response of the gate was elicited through subsequent UV irradiations in two intervals (Figure 1.11B). The tunable nature of the step response displays a unique feature to control output intensity of a DNA-based AND gate using subsequent input stimuli I_1. Achieving a tunable step-response allows light-triggered DNA logic gates to be used as molecular controllers that can be adapted to enhance circuit cascades. As discussed by Ellington,[49] a disadvantage of current DNA logic gate technology is a lack of real-time response to changes in the environment. Using photochemical activation to achieve temporal control allows for the advancement of DNA based computation by overcoming this hurdle and enhances time dependent computation applications. These factors demonstrate the improvements upon existing DNA logic gates through temporal activation with light input signals.

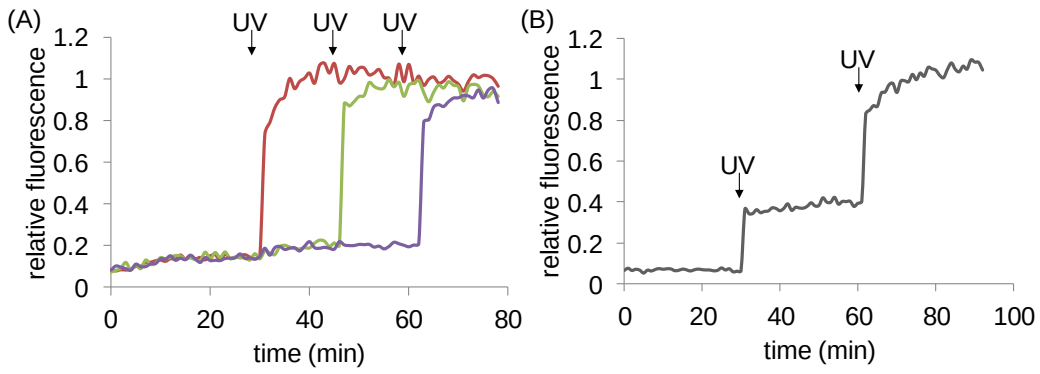


Figure 1.11. UV irradiation of the gate complex containing caged A_4 at different time points in order to demonstrate temporal control over DNA computation. (A) Baseline fluorescence was measured for 30 min, and three individual gates were irradiated with I_1 = 365 nm light at 30 (red), 45 (green), and 60 min (purple). (B) A single logic gate was irradiated for two intervals resulting in a step-like response. The output signal of the caged AND gate is dependent upon the irradiation interval and increases with additional UV exposure. Graphs represent an average of three independent experiments. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

In order to demonstrate spatial control of DNA computation via locally restricted light irradiation, the AND gate complex and the caged strand A_4 were embedded into a low-melt agarose gel. To ensure that the AND gate still functioned correctly in an agarose gel, a truth table was first completed (Figure 1.12A). The gel was either kept in the dark or spot irradiated with I_1 = 365 nm UV light, followed by imaging of the gel via excitation at I_2 = 532 nm. A distinct signal was obtained, and no fluorescence was observed in the absence of I_1 or I_2, demonstrating the ability to apply the developed light-triggered AND gate in spatially controlled DNA computation. DNA computation in a spatially restricted fashion was achieved through patterned UV irradiation using two

different masks (Figure 1.12B). A fluorescent output was only observed in irradiated areas that performed an AND logic operation (I_1 = 365 nm, I_2 = 532 nm) but not in areas where one input was missing (I_1 = absent, I_2 = 532 nm). This demonstrates that logic gate operation can be performed in semisolid structures and is not limited to solution based applications. Electronic systems depend on solid structures and spatially separated devices. Identification and recognition of spatially separated signals allows organization of objects and circuits and creates an important link between nonelectronic and electronic computational systems.

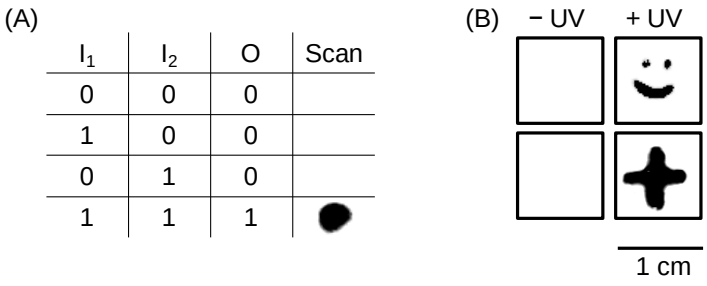


Figure 1.12. Spatial control of DNA logic gate function. (A) Truth table of the AND gate. Low-melt agarose containing the gate complex, the A_4 strand, and the B_0 strand was tested using all combinations of I_1 (365 nm) and I_2 (532 nm), providing the expected AND gate result. (B) Spatially restricted activation of the DNA logic gate through patterned UV irradiation using masks. Gel imaging revealed patterned fluorescence only in areas previously irradiated with I_1 = 365 nm light, followed by scanning of the entire gel with I_2 = 532 nm. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

In conclusion, a photochemically controlled AND gate was developed through the incorporation of caged thymidine nucleotides in a DNA-based logic gate. Strands of DNA were synthesized using specialized phosphoramidites, which enabled the use of specific wavelengths of light as inputs for a DNA-based AND gate. Many DNA-based computation methods rely on toe-hold mediated strand displacement. Thus, the design of caged oligomers was primarily focused on controlling gate activity by caging toe-hold regions. However, these experiments showed that exclusive caging of the toe-hold regions and the introduction of fewer than four evenly spaced caged nucleotides per 36 bases is ineffective for the photochemical control of strand displacement and DNA computation. Temporal control over DNA computation was achieved through introducing four caging groups and activating separate gate complexes at different time points, displaying fundamental properties of a light-switch for molecular circuits. When a single gate complex was irradiated at two intervals, a steplike response in the output signal was observed, suggesting that the phototriggered AND gate can act as a tunable DNA-based circuit. Integration of a light-activated AND gate for purposes of a step response could allow the gate to function as a manual feedback controller. Within a cascade of gates, the light-triggered AND gate can operate as a switch or controller and will allow for more complex and better controlled circuit designs. Moreover, photochemical activation enabled DNA-based logic operations in a spatially localized fashion. This was demonstrated by light-triggered pattern formation in a semisolid substrate, where DNA computation events were only observed in areas that received irradiation with both input wavelengths. Design rules were established that enabled light-activation of the gate and will be applicable to further developments, e.g., the

generation of other light-triggered logic gates. The use of light to control a DNA-based logic gate creates a new paradigm of inputs that will be beneficial when used in a biological context. Light allows for spatial and temporal control with high specificity, while overcoming the downfalls of chemical based inputs such as diffusion and delivery kinetics. Photochemical inputs also shorten the gap between DNA computation and silicon-based electrical circuitry, since light waves can be directly converted into electrical output signals and vice versa. This connection is supremely important for further developing the interface of DNA logic gates and electronic devices and, thus, the interface of biological systems with electrical circuits. Thus, the photochemical control demonstrated here lays the foundation for the programming of complex, DNA-based computation devices with unprecedented spatial and temporal resolution.

1.1.2 Experimental

Caged DNA Synthesis Protocol

DNA synthesis was performed using an Applied Biosystems (Foster City, CA) model 394 automated DNA/RNA synthesizer and standard β-cyanoethyl phosphoramidite chemistry by James Hemphill (Deiters Lab). The caged oligonucleotides were synthesized on a 40 nmol scale, with solid-phase supports obtained from Glen Research (Sterling, VA). Reagents for automated DNA synthesis were also obtained from Glen Research. Standard synthesis cycles provided by Applied Biosystems were used for all normal bases with 25 s coupling times. The coupling time was increased to 2 min for incorporation of caged deoxythymidine modified phosphoramidite. The NPOM-

caged deoxythymidine phosphoramidite was resuspended in anhydrous acetonitrile to a concentration of 0.1 M.

Preparation of the Logic Gate

Logic gates were assembled and quantified according to protocols developed by Winfree.[18] Noncaged strands G_Q, G_T, A_0, and B_0 were purchased from Integrated DNA Technologies (IDT), and strand G_F was purchased from Alpha DNA. A 100 μM stock of all oligonucleotides was made by adding the appropriate amount of nuclease-free deionized water. The logic gates were assembled in a PCR tube by adding 30 μL of each strand (e.g., G_T, G_Q, and G_F) to TEA/Mg^{2+} buffer (10 μL of a 10X stock, 0.4 M tris-acetate, 10 mM ethylenediaminetetraacetic acid, and 125 mM magnesium acetate), so that the final volume was 100 μL. This DNA was slow annealed in a PCR thermocycler (BioRad T100 Thermal Cycler) by heating to 95 °C and cooling to 12 °C with intermediate temperatures of 85, 75, 65, 55, 45, 35, 25, and 15 °C held for 1 min. After annealing, 25 μL DNA loading buffer (50% glycerol, 50% water, bromophenol blue) was added. The sample was loaded onto a 16% native-gel (wells of the gel were removed by manual scraping with a metal spatula to create one large flat-bottomed well). The gel was run at 100 V for 45 min in native-PAGE running buffer (200 mM glycine, 25 mM tris base). Afterwards, the gel was removed, placed on a TLC plate (covered with clear plastic), and illuminated briefly with a handheld UV light. For caged DNA, only a small edge of the gel was illuminated with UV light to determine where the bands were with UV shadow. The small illuminated part of the gel was not excised. The top-most band was excised, cut into smaller pieces, and added to a dialysis membrane tubing (6-8 kDa MWCO, Fisherbrand). The dialysis membrane containing gel pieces was placed in the

middle of a Bio-Rad Mini-PROTEAN vertical gel electrophoresis apparatus. A single cell was arranged with a plastic buffer dam and native-gel. The native-gel was necessary to complete the electrical circuit. The electroelution occurred overnight in a cold room (100 V). Next day, the liquid inside the dialysis tube was collected and concentrated by centrifugation in a 30 kDa MWCO spin-filter (Millipore). The concentrated AND gate (typically between 1-10 μM) was then stored at −20 °C and used as-is for further experiments.

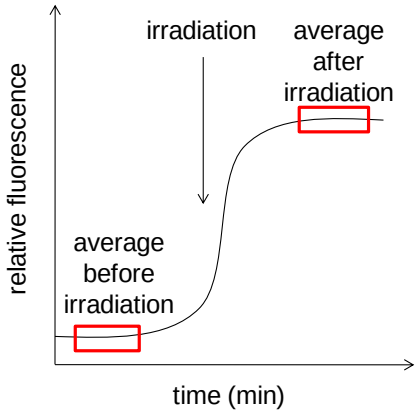
AND Gate Photoactivation

Fluorescence was measured on a BioTek Synergy 4 plate reader using an excitation wavelength of 532 nm and emission wavelength of 576 nm. Samples were prepared in triplicate TAE/Mg^{2+} by adding the appropriate reagents into a black walled and clear bottom 96-well plate. To each well, gate (2 μL of a 9.83 μM stock, 200 nM), strand B_0 or B_4 (8 μL of a 10 μM stock, 800 nM), and strands A_0-A_4 (8 μL of a 10 μM stock, 800 nM) were added to TAE/Mg^{2+} buffer (10 μL of a 10X stock, 0.4 M tris-acetate, 10 mM ethylenediaminetetraacetic acid, and 125 mM magnesium acetate) and water (72 μL). Relative fluorescence represents fluorescence of each sample relative to the positive control (set as 1.0). Error was calculated as the relative standard deviation. To decage the DNA strands, the plate was placed on the bench top and a handheld UV lamp (365 nm) was set on top of the wells. Foil was used to block UV irradiation for all other wells.

Investigation of Optimal Caging Group Number and Localization on Strand A

Samples were prepared in triplicate as described above (see AND Gate Photoactivation). Initial fluorescence (532 nm) was measured for 20 min. Caged oligomers A_1–A_4 were added, and fluorescence was measured for 20 min.

Bar graphs were created from fluorescence time courses, at the time-point listed in each experimental section (Figure 1.13). In case of each experiment presented below, an average of the data points before UV irradiation was subtracted from the average of data points after UV irradiation, in order to generate the corresponding bar graphs in the main text. This averaging of data before or after irradiation was done in order to minimize fluctuations and bias in data that might have been otherwise caused by selecting a single representative data/time point. Each fluorescence readout was then normalized to the readout of the non-caged logic gate.



final value = (average after irradiation) – (average before irradiation)

Figure 1.13. Schematic outlining the method used to determine the fluorescence values before and after UV irradiation for all solution-based experiments. Adapted with permission from Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J. Am. Chem. Soc.* 2012, *134* (8), 3810-5. Copyright 2012 American Chemical Society.

Irradiation Time Course

Samples were prepared in triplicate for each UV exposure experiment (0, 1, 5, 10, 15, and 20 min) as described above (see AND Gate Photoactivation Experiments). Initial fluorescence (532 nm) was measured for 30 min and the average before irradiation was calculated between 0-30 min. Wells were irradiated at 365 nm for the indicated time followed by fluorescence measurements for 30 min and the average after irradiation was calculated between 30-60 min.

Light-Activation of Strand B

Samples were prepared in triplicate as described above (see AND Gate Photoactivation. Fluorescence was measured for 30 min. The wells containing B_4 were irradiated at 365 nm for 15 min followed by fluorescence measurement for 30 min and the average before irradiation was calculated at 0-30 min. Wells were irradiated at 365 nm for the indicated time followed by fluorescence measurements for 30 min and the average after irradiation was calculated at 30-60 min.

Investigation of Toe-Hold Caging

Samples were prepared in triplicate as described above (see AND Gate Photoactivation) Fluorescence was measured for 30 min and the average before irradiation was calculated at 0-30 min. Wells were irradiated at 365 nm for the indicated

time followed by fluorescence measurements for 30 min and the average after irradiation was calculated at 30-60 min.

Spatial Control of Gate Function

A 100 µL 1.5% agarose solution in TAE/Mg^{2+} buffer was heated in a household microwave oven until all agarose was dissolved (15 sec). Before the agarose solidified (still warm to the touch), the gate complex (4 µL of a 9.83 µM stock, 200 nM), A_4 (16 µL of a 10 µM stock, 800 nM), and B_0 (16 µL of a 10 µM stock, 800 nM) were added to TAE/Mg^{2+} buffer (10 µL of a 10X stock), water (54 µL), and the agarose solution (100 µL). The gel containing the gate and caged strand was spread on a glass slide and allowed to solidify for 20 min in the dark. The gel was imaged on a General Electric Typhoon FLA 7000 phosphorimager for background fluorescence with an excitation wavelength of 532 nm and a 580 nm emission filter. The gel was then irradiated on a UVP high performance UV transilluminator with 365 nm for 15 min light passing through patterned aluminum foil. After UV irradiation, the gel was again imaged (5 min post irradiation).

1.2 INTERFACING SYNTHETIC DNA LOGIC OPERATIONS WITH PROTEIN OUTPUTS

1.2.1 Results and Discussion

A limitation in the current DNA-based gate designs is a lack of direct interfacing with biological components other than oligonucleotides. Reported herein is the first direct integration of DNA logic gates with protein outputs. Integration of protein components into DNA computation devices expands upon the capabilities of interfacing DNA-based circuits with biological systems through the incorporation of proteins as inputs or outputs. Protein-based DNA logic gates have been applied in antibody activation,[50] targeted cancer therapy,[51] and protein detection.[10] Aptamers are used frequently to directly interface DNA logic gates with proteins due to their high specificity for a target protein. Binding of a protein to the aptamer can initiate a toe-hold-like mediated strand displacement, which is a fundamental element of any DNA-based circuit. Use of the appropriate aptamer enables a DNA-based logic gate to detect a specific protein input.

A DNA logic gate was developed by Janssen *et al.* to activate an antibody after toe-hold mediated strand displacement.[50] An IgG_1-type monoclonal antibody that recognized the hemagglutinin (HA) epitope was used. The antibody was blocked by the addition of a bivalent peptide-DNA conjugate composed of two peptides bound to complementary DNA strands, which formed a duplex (Figure 1.14A). Toe-hold regions were extended from the duplex DNA to enable toe-hold mediated strand exchange.

When both peptides were bound to the peptide-DNA conjugate, the antibody was inactive and could not bind to the HA epitope. To ensure that the antibody was fully blocked, the peptide-DNA conjugate was added in slight excess to the antibody. Yeast cells displaying the HA epitope were added to the blocked yellow fluorescent protein-tagged antibodies and sorted by flow cytommetry. No fluorescently labeled cells were detected for blocked antibodies, showing that the peptide-DNA conjugate fully blocked epitope recognition. Addition of a displacement strand of 20-35 bases with an eight nucleotide toe-hold binding region removed the peptide-DNA conjugate from the antibody. The flexibility in displacement strand length enabled flexibility in the design of the DNA sequence. An AND logic gate was created from a peptide-DNA conjugate containing two 10 nucleotide long toe-hold regions (Figure 1.14B). The antibody activity was only restored in the presence of two oligonucleotide inputs, which displaced the conjugate through toe-hold mediated strand exchange. Full displacement was also possible from a single longer DNA strands, such that addition of either input activated the antibody. Thus, the same AND gate could be transformed into an OR gate by changing the length of the ssDNA inputs (Figure 1.14C).

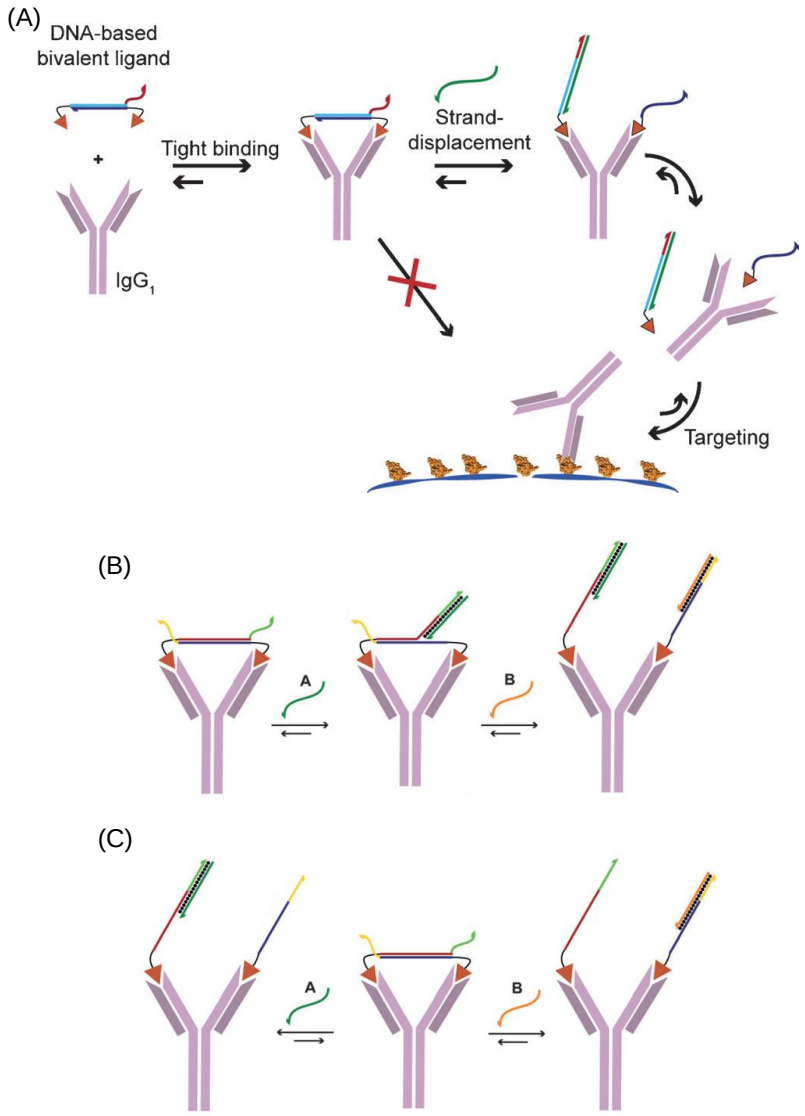


Figure 1.14. Antibody activation by toe-hold mediated strand displacement. (A) A peptide-DNA conjugate blocks the recognition of the HA epitope by an antibody. Addition of a displacement strand to the blocked antibody separates the two DNA strands of the conjugate, activating the antibody. (B) An AND gate was generated from a peptide-DNA conjugate containing two toe-hold regions. The conjugate was only removed after addition of both inputs. (C) An OR gate was constructed from the same structure as the AND gate. However, addition of either input, which were longer than the AND gate inputs, activated the antibody. Adapted with permission from Janssen *et al. Angew. Chem. Int. Ed.* 2015, *54* (8), 2530-2533.

DNA logic gates were also interfaced with cell-surface proteins by You *et al.*[51] The first component of the logic gates was a ssDNA containing an aptamer specific to a cancer cell-remained different stages of cancer within the same population. First, the aptamer strands were added to the cells and excess aptamer strands not bound to cell-surface proteins were washed away. A second component of the logic gates was a duplex with one strand containing a complementary toe-hold to the aptamer strand while the other strand was bound to a drug or dye. In the presence of both components (AND logic), the strand containing the drug or dye would be released by a toe-hold mediated strand exchange. Using this design, an AND gate was created that recognized two cancer cell-surface proteins, Sgc4f and Sgc8c (Figure 1.15A). Two separate aptamer strands were used to specifically bind the Sgc4f or Sgc8c protein. A duplex modified with a fluorescent moiety was added that contained a toe-hold for the Sgc8c aptamer strand. After a toe-hold meditated strand displacement, the fluorophore-modified strand was released and could bind to the Sgc4f aptamer strand. The fluorescently labeled strand could only remain bound to the cell surface through binding to the aptamer strand if both proteins were present. Four cell lines were tested, including CEM, HeLa, Ramos, and K562. The CEM cell line overexpressed both Sgc8c and Sgc4f and was expected to be the only cell line capable of triggering the AND gate. However, background fluorescence was observed with the HeLa cell line due to a low expression level of Sgc4f (Figure 1.15B). The K562 and Ramos cell lines did not express one or both of the proteins, therefore both aptamer strands could not bind to proteins on the cell surface, and no output was produced by the AND gate. Although high background was observed with HeLa cells, fluorescence measurements

demonstrated AND gate activation was 1.5-fold higher for CEM cells. This example shows how modular DNA circuits can be used to recognize cancer biomarkers and target therapy only to intended cell populations. Further optimization of conditions may be necessary to apply these DNA-based logic gates to the routine detection of cell-surface oncoproteins. More complex networks were also created using the AND gate and two similar logic gates (OR and NOT gates).

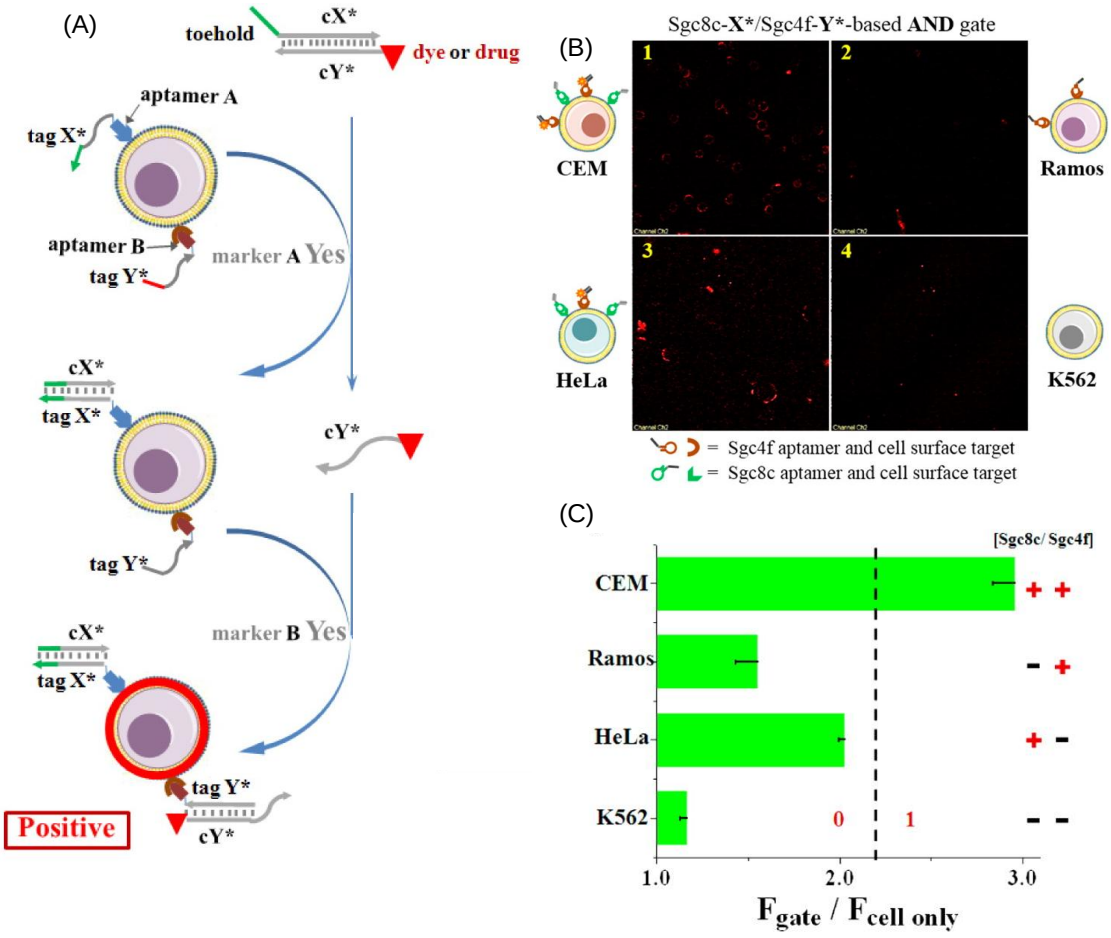


Figure 1.15. Recognition of cancer cell-surface biomarkers. (A) The AND gate aptamer strands bind the specific cell-surface proteins (blue and red). Exposed toe-hold binding regions from the aptamer strands (tagX* and tagY*) interact with the cX*:cY* duplex and displace the cY* strand, which is connected to a dye or drug (red triangle). Once displaced, the strand can bind the second free aptamer strand (tagY*) to induce fluorescence (if red triangle is a fluorophore) or deliver a therapeutic agent (if red triangle is a drug). (B) The AND gate recognizing the proteins Sgc8c and Sgc4f was tested with multiple cell lines. Maximum fluorescence was observed for CEM cell line, which overexpresses both proteins. Mild fluorescence was observed in HeLa cells due to the downregulated, but still present, Sgc4f protein. Fluorescence values were only 1.5-fold higher for CEM cells than HeLa cells. (C) Fluorescence values were measured by flow cytometry and are given for each cell type tested. Together, these 4 cell types simulate all possible combinations for the AND gate based on cell-specific expression of Sgc8c and Sgc4f. Adapted with permission from You *et al. J. Am. Chem. Soc.* 2015, *137* (2), 667-674. Copyright 2015 American Chemical Society.

Instead of exclusively using DNA strands as inputs, DNA logic gates that respond to small molecules or proteins as inputs have been created.[10] The main gate structures contained long single-stranded DNA aptamer domains that could bind specifically to adenosine or thrombin. An AND gate was constructed from three DNA strands, with the aptamer domains exposed on either side (Figure 1.16A). Binding of adenosine AND thrombin to their respective aptamers displaced a partially complementary strand modified with a quencher moiety. Removal of the quencher from the proximity of the fluorophore allowed output in the form of fluorescence emission. The structure of the OR gate was a modification of the AND gate (Figure 1.16B). Both of the aptamer domains were used, but one of the quencher moieties was replaced by a fluorophore. The fluorophore remained quenched until addition of adenosine OR thrombin. Although these gates could detect adenosine at 1 mM and thrombin at 1 μM, the protein was not

directly affected. The gates could not switch protein activity ON or OFF. The gates were also activated by DNA input strands complementary to the aptamer regions. Fluorescence was much higher than with the aptamer ligands, which was possibly due to a difference in the fluorophore environment after activation.

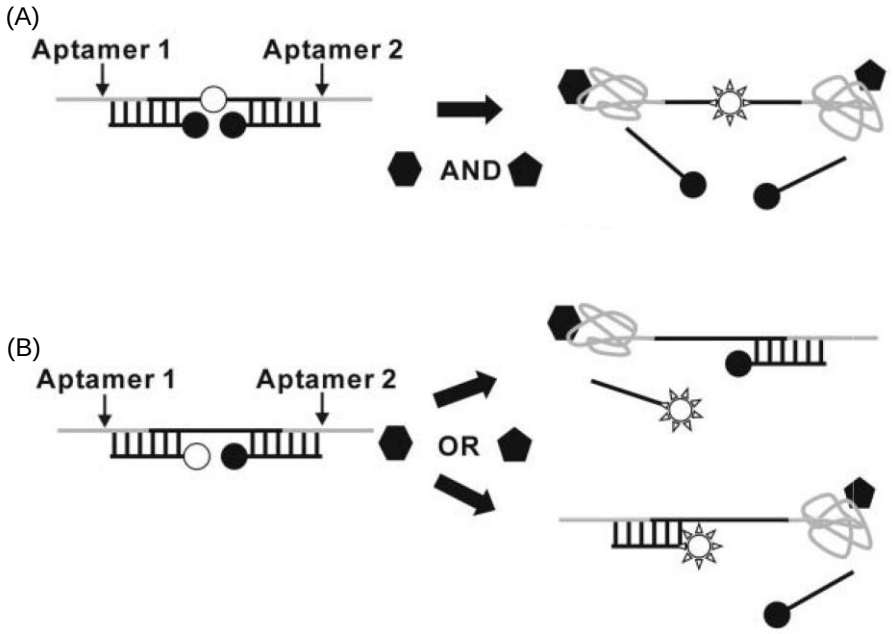


Figure 1.16. Aptamer logic gates designed to detect adenosine and thrombin. (A) An AND gate enabled fluorescence emission only after adenosine AND thrombin were present to remove the quencher strands. (B) The OR gate design was based on the AND gate structure. However, removal of either the adenosine OR thrombin aptamer strand enabled fluorescence output. The fluorophore is represented by an open circle, quenchers are represented by filled-in circles, adenosine is represented by a hexagon, and thrombin is represented by a pentagon. Reproduced from Ref. 10 with permission from The Royal Society of Chemistry.

An approach described herein to directly interface logic gate with protein outputs makes use of zinc-finger proteins since they are able to bind DNA without disrupting the structure of the logic gate and can be easily fused to split-protein components.[52] Zinc-finger proteins are naturally occurring, sequence-specific DNA binders that have been modified with various effector domains,[53] for example, to regulate transcription,[54] manipulate mitochondrial DNA,[55] or inhibit viral replication.[56] A single zinc finger will recognize an arrangement of three nucleotides. Generally, six fingers are fused together to create a larger protein that can identify a unique 18-nucleotide sequence. The most common group of zinc fingers is based on the Cys2His2 motif[57] and includes the proteins AaRT and E2C. The majority of zinc-finger proteins, such as E2C[58] recognize guanine-rich DNA sequences, however, the protein AaRT was specifically designed to bind adenine-rich sequences.[59] Changing a single nucleotide in the binding site has been shown to negatively affect the binding of zinc fingers to DNA by increasing the K_D value more than 100-fold.[60] The utilization of two unique zinc-finger proteins, AaRT and E2C, in the logic-gate design ensures sequence-specific protein activation. In addition to the zinc-finger component, the developed logic gates contain a split-luciferase enzyme to generate a luminescence readout, which gives a highly sensitive biosensor design.[61,62,63,64,65] However, the function of a wide range of other split-proteins, including green fluorescent protein,[66] beta-galactosidase[67] and TEV protease,[68] could be triggered as well. A split-protein consists of two halves of the original protein. Separately, each half is inactive and does not exhibit enzymatic activity. When the two complementary halves are brought into close proximity, the enzyme structure is reconstituted and activity is restored. In the DNA logic-gate design, the N-terminal half

of luciferase is fused to AaRT while the C-terminal half is fused to E2C (Figure 1.17 and Figure 1.18). Although each gate contains the zinc finger and split-luciferase fusion protein, the operation of each gate is inherently unique.

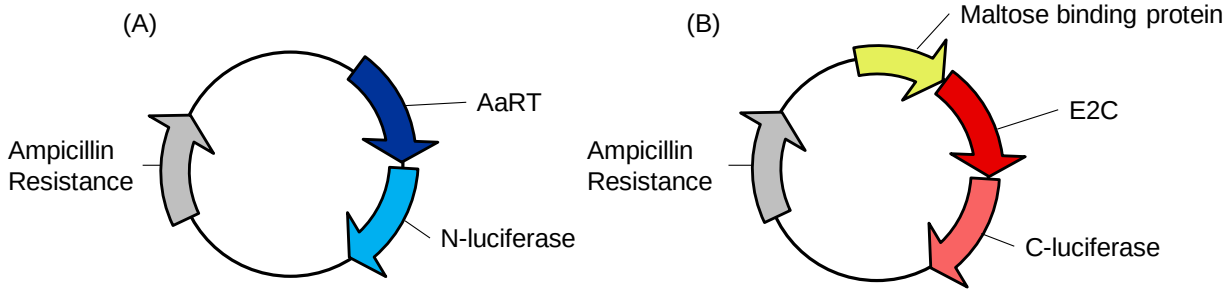


Figure 1.17. Expression plasmids (A) pDNC AaRT-NFluc and (B) pDNC MBP-E2C CFluc for the zinc finger-split luciferase fusion proteins. Constructs were provided by the Ghosh lab. The AaRT zinc finger is fused to the N-terminal luciferase while the E2C zinc finger is fused to the C-terminal luciferase. A maltose binding protein (MBP) is fused upstream of E2C to aid in the expression and purification of the protein. The MBP did not interfere with zinc finger binding and had a negligible effect on luciferase activity. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

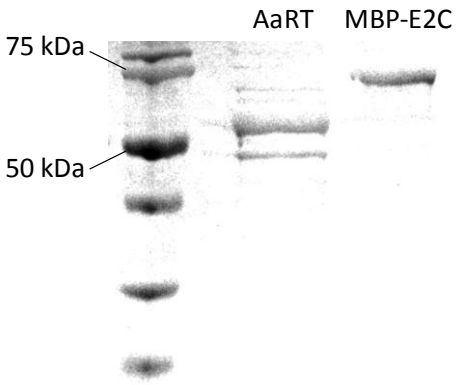


Figure 1.18. SDS-PAGE gel of the zinc finger-split luciferase proteins AaRT-Nluc and MBP-E2C-Cluc. Protein bands stained with Coomassie stain. The expected masses for AaRT (68 kDa) and MBP-E2C (81 kDa) are observed. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

A TEVp recognition site was inserted between the MBP and E2C proteins to enable cleavage of the fusion construct. Tobacco etch virus protease (TEVp) was added to the MBP-E2C protein to remove the MBP fusion. The TEVp-treated protein was combined with AaRT in a luciferase assay to determine if the MBP fusion protein inhibited zinc-finger binding or luciferase reconstitution (Figure 1.19). A DNA duplex containing the AaRT and E2C zinc-finger recognition sites was added to the proteins. The TEVp-treated and untreated proteins showed similar levels of activation. Addition of the MBP fusion protein was not negatively affecting the zinc-finger binding or split-luciferase reconstitution. All further experiments used non-treated MBP-E2C protein.

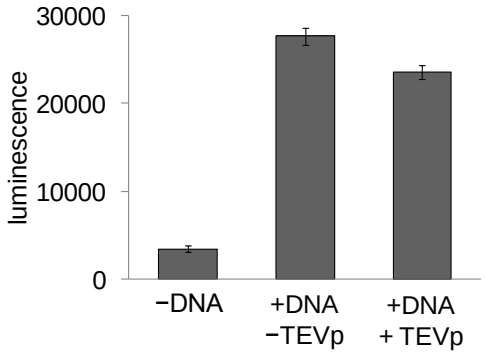


Figure 1.19. TEVp-treated E2C protein was assayed for activity. A DNA duplex containing both AaRT and E2C zinc-finger recognition sites (DNA) was added with the treated (+TEVp) and non-treated (−TEVp) proteins. No negative effect was caused by the MBP fusion to E2C.

In order to demonstrate the triggering of protein function by the DNA logic gate, an AND gate (Figure 1.20A) was selected as an initial example. This gate only produces an output when both inputs are present. The zinc finger AND gate follows the simplest design of the three gates: it consists of a single strand of DNA (G_{AND}) and is activated by two inputs (A and B). The two zinc-finger binding sites are only completed after the hybridization of both inputs to the gate strand. After binding of the zinc-finger proteins to the DNA scaffold, the split-luciferase halves are able to establish a functional luciferase enzyme. Therefore, activation of the AND gate only occurs in the presence of both inputs, as observed when A and B were added to the G_{AND} strand (Figure 1.20B). Only minimal baseline luminescence activity was detected in the presence of just A or just B; however, a 5-fold increase in luciferase activity was observed when both inputs were present. Owing to the stringent requirements for an AND gate, it is often used for the "carry" function in larger devices, like the half-adder.[69]

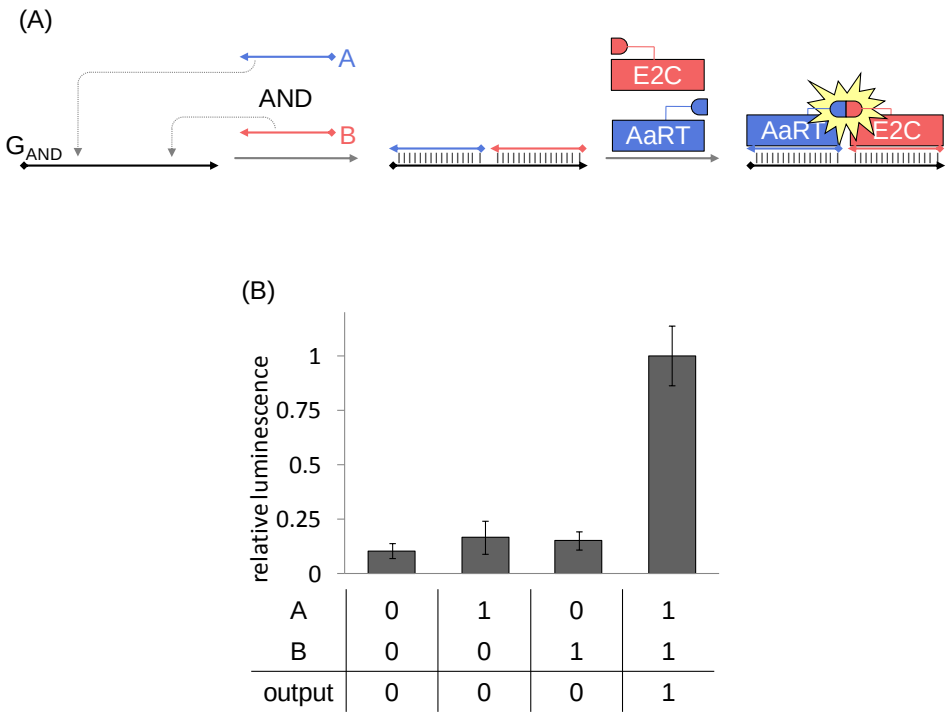
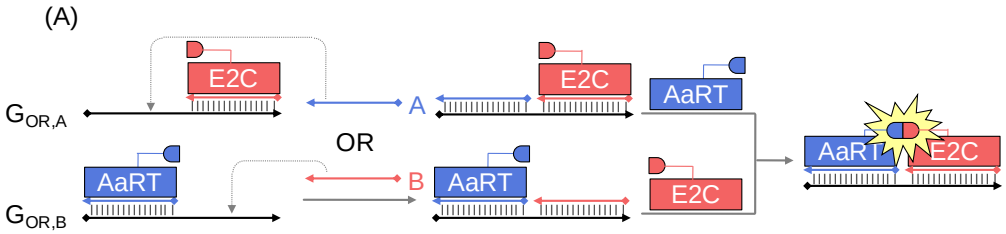


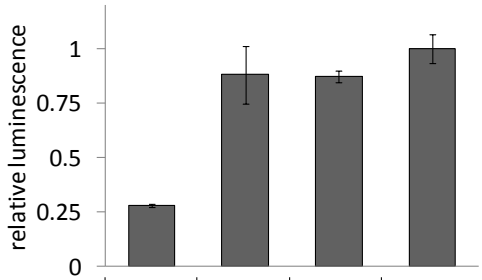
Figure 1.20. (A) Schematic of the AND gate based on DNA–protein interactions. Split-luciferase halves are epresented by solid half ovals, the zinc-finger proteins by rectangles, and DNA is shown as solid lines. (B) Luminescence readouts are shown for the AND gate after the inputs A and B were added in different combinations to the G_{AND} strand. The AND gate truth table is displayed below the chart. Three independent experiments were averaged and the error bars represent the standard deviation. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

In contrast to the AND gate, an OR gate (Figure 1.21A) yields an output in the presence of either of the two inputs. The corresponding DNA-based circuitry is generated from two preformed duplexes: one consisting of an incomplete AaRT binding site ($G_{OR,A}$) and the other containing the incomplete E2C binding site ($G_{OR,B}$). When input A or B is added and hybridized to $G_{OR,A}$ or $G_{OR,B}$, the missing DNA binding site for the zinc-finger is formed. The second zinc finger is then able to bind to the DNA,

thereby allowing the formation of an active luciferase enzyme. Thus, only a single input is necessary to activate this DNA-based OR gate (Figure 1.21B). As expected, when either A or B was added, a luminescence signal was produced. In the absence of any input strand, only basal levels of luciferase activity were detected. Slightly higher background activity was observed for the OR gate in comparison to the other gates. This effect may result from the doubled zinc-finger concentration, which is required owing to the use of two preformed duplexes, $G_{OR,A}$ and $G_{OR,B}$. However, a clear and significant difference in luminescence was detected in the presence of the inputs. OR gates are commonly found in electrical devices, including simple multiplexers, which can function as electronic rotary switches.[70]



(B)



A	0	1	0	1
B	0	0	1	1
output	0	1	1	1

Figure 1.21. (A) Schematic of the OR gate, based on DNA–protein interactions. Split-luciferase halves are represented by half ovals, the zinc-finger proteins by rectangles, and DNA is shown as solid lines. (B) Luminescence readouts are shown for the OR gate after the two inputs A and B were added in different combinations to the G_{OR}duplexes. The OR gate truth table is displayed below the chart. Three independent experiments were averaged and the error bars represent the standard deviation. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

As a third, essential logic gate operation, the NOR gate (Figure 1.22A) functions as an inverse (or negated) OR gate. Thus, the presence of either input signal will lead to no output signal. Translated into DNA circuitry interfaced with protein outputs, this occurs when either input A or B binds to a toe-hold region of the NOR gate, thereby displacing the A or B strand. With the two zinc fingers far apart, the split-luciferase halves are separated and do not produce a luminescence output. A minimum five-fold dynamic range can be observed with the addition of either input, thus distinguishing the high and low Boolean outputs (Figure 1.22B). A NOR gate is especially useful since it is often considered a universal gate; any other logic gate can be created through a combination of multiple NOR gates.[71] For example, a XOR gate can be created from five NOR gates through serial and parallel connections.

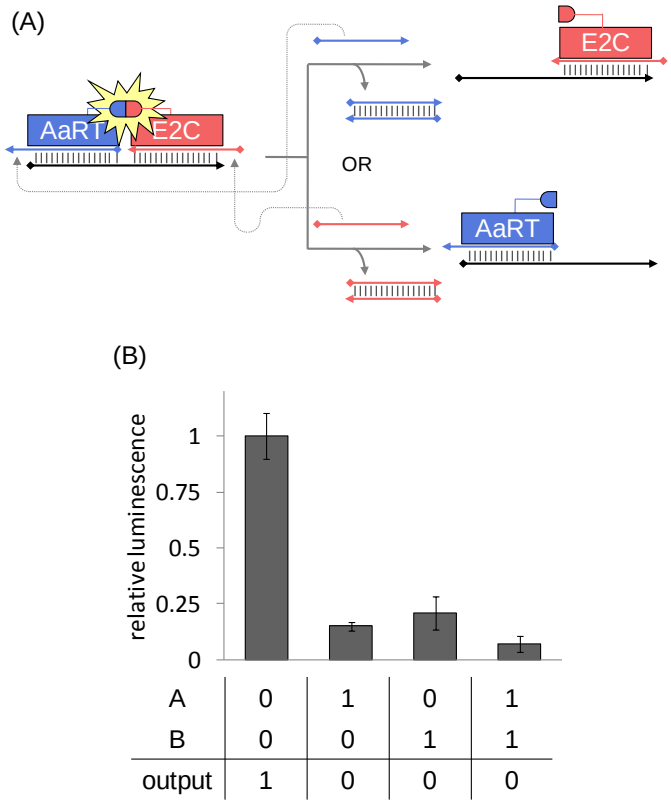


Figure 1.22. (A) A scheme of the NOR gate based on DNA–protein interactions. Split-luciferase halves are represented by half ovals, the zinc-finger proteins by rectangles, and DNA is shown as solid lines. (B) Luminescence readouts are shown for the NOR gate after the two inputs A and B were added in different combinations to the G_{NOR} duplex. The NOR gate truth table is displayed below the chart. Three independent experiments were averaged and the error bars represent the standard deviation. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

Connecting multiple logic gates in a larger network[2] enables the construction of more complex devices, like half-adders[69] and multiplexers.[70] The final Boolean computation in any of these networks could be readily performed by the zinc-finger-interfaced gates presented here. A sub-network was constructed by connecting an AND gate and a NOR gate in series (Figure 1.23A). When both inputs B and C are present,

the AND gate will release an output, which acts as an input for the NOR gate. Either the AND gate output or the input A will trigger the NOR gate and eliminate a zinc-finger binding site (Figure 1.23B). Luminescence measurements confirmed the expected activity of the network (Figure 1.23C). Consistent with the truth table for the circuit, high luminescence values are only observed in the absence of any input or in the presence of only B or only C.

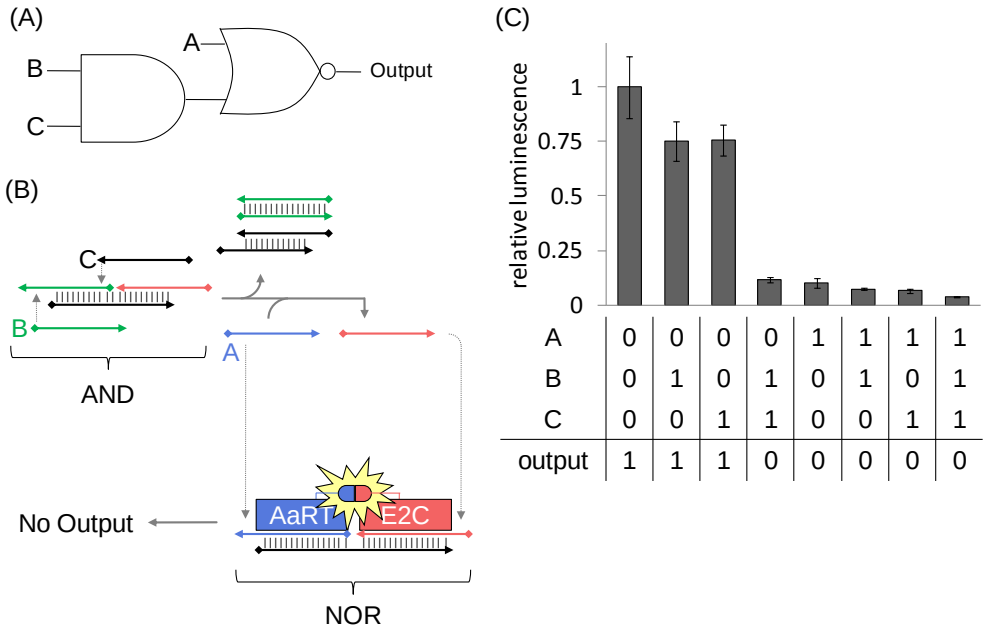


Figure 1.23. (A) A scheme of the corresponding DNA computation circuit. The ssDNA output from the AND gate becomes an input for the NOR gate. The final output from the NOR gate is emitted as luminescence. (B) Luminescence data for the AND–NOR circuit. Three independent experiments were averaged and the error bars represent the standard deviation. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

The encoding of multiple logic gates by simple DNA frameworks combined with zinc-finger proteins demonstrates the versatility of protein-interfaced DNA components. Obviously, a constraint is the strict sequence specificity required for zinc-finger proteins to bind. However, this limitation can be overcome by simply employing translator gates to convert any oligonucleotide sequence into the sequence of DNA inputs like A or B.[18] As a proof of principle, translator gates were designed for the microRNAs miR-21 and miR-122, which are implicated in several diseases (Figure 1.24A).[72,73,74] NOR gate input toe-holds are sequestered inside DNA duplexes, and in the absence of miRNAs, no decrease in luminescence output is observed. The addition of the miRNAs to the translator gates releases a NOR gate input with an exposed toe-hold binding site. The inputs are then able to interact with the NOR gate as shown in Figure 1.24B, thereby leading to a reduction in the luminescence output. Signal was reduced only about 2-fold after addition of the miRNA inputs. A decrease in the signal-to-background ratio may have been caused by inefficient displacement of the translator gate output. Since the displacement of output from the translator gate is reversible, the process may not be completely efficient and the NOR gate may remain active even in the presence of miRNA input. If the miRNA translator gates are used to detect or interact with biologically relevant nucleic acids, the operating conditions need to be further optimized.

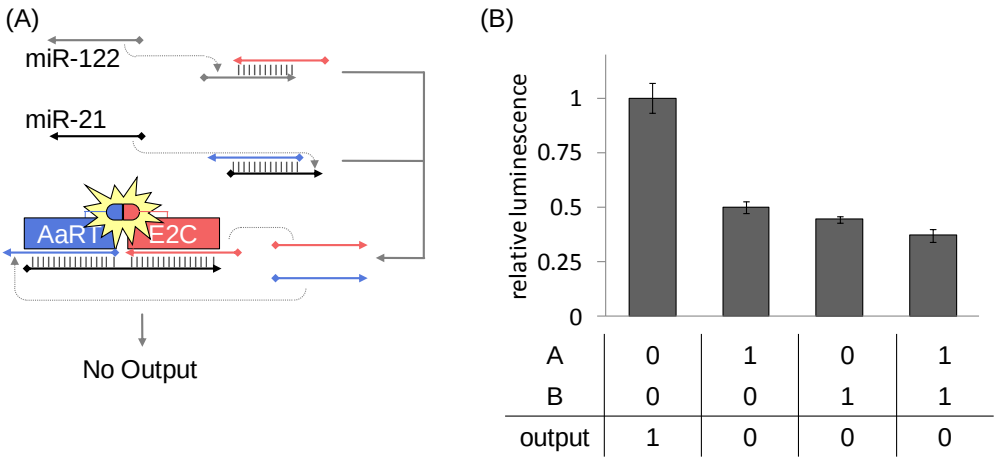


Figure 1.24. Translator gates for microRNA inputs to activate the NOR gate. (A) Schematic of the translator gate and NOR gate mechanism. Translator gates release NOR gate inputs upon addition of miR-21 or miR-122. (B) Luminescence data for the translator and NOR gate circuit. Three independent experiments were averaged and error bars represent standard deviations. Adapted with permission from Prokup *et al. Angew. Chem. Int. Ed. Engl.* 2014, *53* (48), 13192-13195.

The OR, AND, and NOR logic gates presented herein thus represent modular biological devices that can be used in the final layer of DNA computation circuits[4,23] in order to interface them with protein outputs. DNA logic gates typically yield single-stranded DNA output.[6,75,10] Luminescence outputs have not been incorporated into logic gate designs before and may show advantages for reducing background signal,[76] as well as for expanding the range of applications for DNA computational devices. The activation of an enzyme in response to a DNA computation event has the potential to provide a facile solution to the often absent signal amplification in DNA devices.[77,78] Moreover, the split-luciferase could be replaced by any number of split-proteins, which greatly expands the output possibilities to more than luminescence,

including direct activation of a protein of interest and the triggering of a wide range of biological processes by either activation (AND and OR gates) or deactivation (NOR gate) of a split-protein, such as a protease (TEV),[68] recombinase (Cre),[79] hydrolase (β-galactosidase),[67] or fluorophore (GFP).[66] A simple DNA circuit output could thus lead to significant modification and perturbation of the biological system it is employed in.

1.2.2 Experimental

Zinc finger protein expression and purification.

The plasmids pDNC AaRT-NFluc and pDNC MBP-E2C-CFluc were transformed into BL21 (DE3) *E. Coli cells.* Ampicillin (10 µL of a 100 mg/mL stock, 100 µg/mL) was added to Luria-Broth agar (LB agar, 10 mL). The transformed cells were plated on the agar and incubated overnight at 37 °C. Ampicillin (5 µL of a 100 mg/mL stock, 100 µg/mL) was added to LB broth (5 mL). A single colony from the plate was grown in the LB broth overnight at 37 °C. To express the protein, LB broth (25 mL, supplemented with 0.1 mM $ZnCl_2$) was inoculated with 2 mL of overnight culture and grown at 37 °C. When the culture reached an OD_{600} of 0.6, expression was induced with isopropyl β-D-1-thiogalactopyranosidase (250 µL of a 100 mM stock, 1 mM) and shaken at room temperature. After expression overnight, the cultures were centrifuged (9000 x *g*, 10 minutes, 4 °C), broth was decanted, and the cell pellet was stored overnight at −80 °C. Protease inhibitor (4 µL, Sigma) and lysozyme (40 µL of a 1 g/mL stock, 0.1 g/mL) were added to lysis buffer (4 mL, 40 mM tris-HCl pH 8, 1 M NaCl, 2 mM $MgCl_2$, 0.2 mM $ZnCl_2$, 0.2 mM TCEP, 20% glycerol, and 2 mL of elution buffer – see below). The cell pellets were resuspended in the lysis buffer. After lysis (1 hour, 4 °C) the cells were

sonicated (on ice, 40% duty cycle, output level 2 for 10 seconds, 3 for 10 seconds, 4 for 60 seconds) and centrifuged (20500 x *g*, 20 minutes, 4 °C). Ni-NTA resin (75 µL, Qiagen) was added to the supernatant and the suspension was rocked for 1 hour at 4 °C. The resin was loaded onto spin filters (0.4 µm, PVDF, Thermo) and washed with lysis (3 x 250 µL) and wash (10 mL lysis + 435 µL elution buffer, 2 x 250 µL) buffers. The protein was eluted with elution buffer (250 µL, 20 mM tris-HCl pH 8, 500 mM NaCl, 0.1 mM $ZnCl_2$, 250 mM imidazole, 1 mL 1 M HCl, 10% glycerol). Protein size and purity was determined using a Coomassie stained 12% SDS-PAGE gel (60V for 15 minutes then 150V for 60 minutes).

Testing the effect of the MBP fusion on zinc-finger binding and luciferase activity

MBP-E2C protein (100 µL) was added to TEVp (2 µL) and incubated at room temperature for 2 hours. The proteins (5 µL each of eluted protein from above) were added to a white walled and clear bottom 96-well plate in 50 µL total volume buffered with PBS buffer (137 mM NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7). DNA duplex containing both AaRT and E2C recognition sites was annealed according to the procedure outlined in “Preparation of the Logic Gate” was added at 1 µM. Each condition was tested in triplicate. To each well was added 50 µL of luminescence assay reagents (Bright-glo, Promega). Luminescence was measured on a BioTek Synergy 4 plate reader.

Preparation of the Logic Gate

Logic gates were assembled and quantified according to protocols developed by Winfree.[18] Noncaged strands G_Q, G_T, A_0, and B_0 were purchased from Integrated DNA Technologies (IDT, Table 1.2), and strand G_F was purchased from Alpha DNA. A 100 μM stock of all oligonucleotides was made by adding the appropriate amount of nuclease-free deionized water. The logic gates were assembled in a PCR tube (Table 1.3). To a PCR tube, 30 μL of G_{NOR}, G_A, and G_B strands were added to TEA/Mg^{2+} buffer (10 μL of a 10X stock, 0.4 M tris-acetate, 10 mM ethylenediaminetetraacetic acid, and 125 mM magnesium acetate). The DNA solution was slowly annealed in a PCR thermocycler (BioRad T100 Thermal Cycler) by heating to 95 °C and cooling to 12 °C with intermediate temperatures of 85, 75, 65, 55, 45, 35, 25, and 15 °C held for 1 min. After annealing, DNA loading buffer (25 μL, 50% glycerol, 50% water, bromophenol blue) was added. The sample was loaded onto a 16% native-gel (wells of the gel were removed by manual scraping with a metal spatula to create one large flat-bottomed well). The gel was run at 100 V for 45 min in native-PAGE running buffer (192 mM glycine, 25 mM tris base). Afterwards, the gel was removed, placed on a TLC plate (covered with clear plastic), and illuminated briefly with a handheld UV light. For caged DNA, only a small edge of the gel was illuminated with UV light to determine where the bands were with UV shadow. The illuminated area of the gel was not excised. The top-most band was excised, cut into smaller pieces, and added to a dialysis membrane tube (6-8 kDa MWCO, Fisher). The dialysis tubing was added to a gel apparatus, set up with another native-gel in a cold room to electroelute (100 V, ~12 hours). Next day, the liquid inside the dialysis tube was collected and concentrated by centrifugation in a 30

kDa MWCO spin-filter (Millipore). The concentrated AND gate (typically between 1-10 μM) was then stored at −20 °C and used as-is for further experiments. DNA was quantified by measuring the absorbance at 260 nm and using Beer's Law to calculate the concentration from the determined extinction coefficient. The coefficients for the individual strands can be determined through nearest-neighbor models and online from IDT.com.

Logic gate operations

All gate activation experiments used a white walled and clear bottom 96-well plates and PBS buffer (137 mM NaCl, 27 mM KCl, 100 mM Na2HPO4, 1.8 mM KH2PO4). Total volume of logic gate experiments was 50 μL before addition of 50 μL of luminescence assay reagents (Bright-glo, Promega). Inputs were added in all possible combinations according to the truth tables and information below. Oligonucleotide stock solutions were 100 μM. All experiments were performed in triplicate.

AND gate: AaRT (5 μL of a 2 μM stock, 200 nM), MBP-E2C (1 μL of a 10 μM stock, 200 nM), AND gate (1 μL of a 10 μM stock, 200 nM), and A and B (2 μL of a 10 μM stock, 400 nM) were added to PBS buffer (5 μL of a 10X stock, 1.37 M NaCl, 270 mM KCl, 1M Na_2HPO_4, 18 mM KH_2PO_4, pH 7) and water (34 μL). After incubation (room temperature, 1 hour), Bright-Glo Luciferase Assay System buffer was (50 μL) was added and luminescence was measured by a Tecan Infinite M1000 Pro plate reader (1000 ms integration time).

OR gate: AaRT (10 μL of a 2 μM stock, 400 nM), MBP-E2C (2 μL of a 10 μM stock, 400 nM), $G_{OR,A}$ and $G_{OR,B}$ gate duplexes (1 μL of a 10 μM stock, 200 nM), and A

and B (2 µL of a 10 µM stock, 400 nM) were added to PBS buffer (5 µL of a 10X stock, 1.37 M NaCl, 270 mM KCl, 1M Na_2HPO_4, 18 mM KH_2PO_4, pH 7) and water (27 µL). After incubation (room temperature, 1 hour), Bright-Glo Luciferase Assay System buffer was (50 µL) was added and luminescence was measured by a Tecan Infinite M1000 Pro plate reader (1000 ms integration time).

NOR gate: AaRT (5 µL of a 2 µM stock, 200 nM), MBP-E2C (1 µL of a 10 µM stock, 200 nM), NOR gate (1 µL of a 10 µM stock, 200 nM), and A_{comp} and B_{comp} (2 µL of a 10 µM stock, 400 nM) were added to PBS buffer (5 µL of a 10X stock, 1.37 M NaCl, 270 mM KCl, 1M Na_2HPO_4, 18 mM KH_2PO_4, pH 7) and water (34 µL). After incubation (room temperature, 1 hour), Bright-Glo Luciferase Assay System buffer was (50 µL) was added and luminescence was measured by a Tecan Infinite M1000 Pro plate reader (1000 ms integration time).

miR-translator gates: AaRT (0.625 µL of a 2 µM stock, 25 nM), MBP-E2C (0.125 µL of a 10 µM stock, 25 nM), NORmiR-translator gate (0.5 µL of a 10 µM stock, 100 nM), miR-21 and miR-122 (4 µL of a 10 µM stock, 800 nM), and miR-21 and miR-122 translator gates (1.9 µL of a 10.6 µM stock, 400 nM) were added to PBS buffer (5 µL of a 10X stock, 1.37 M NaCl, 270 mM KCl, 1M Na_2HPO_4, 18 mM KH_2PO_4, pH 7) and water (36 µL). After incubation (37 °C, 2 hours), Bright-Glo Luciferase Assay System buffer was (50 µL) was added and luminescence was measured by a Tecan Infinite M1000 Pro plate reader.

AND-NOR gate cascade: AaRT (6.2 µL of a 805 nM stock, 100 nM), MBP-E2C (0.5 µL of a 10 µM stock, 100 nM), NOR gate (0.5 µL of a 9.98 µM stock, 100 nM), AND gate (4.6 µL of a 4.35 µM stock, 400 nM), A (5 µL of a 4 µM stock, 400 nM), and B and

C (10 µL of a 4 µM stock, 800 nM) were added to PBS buffer (5 µL of a 10X stock, 1.37 M NaCl, 270 mM KCl, 1M Na_2HPO_4, 18 mM KH_2PO_4, pH 7) and water (8.2 µL). Bright-Glo Luciferase Assay System buffer (50 µL) was added and luminescence was measured by a Tecan Infinite M1000 Pro plate reader.

Table 1.2. DNA sequences for the construction and operation of the DNA logic gates.

AND gate	
DNA oligomer	**Sequence (5' → 3')**
A	TAGTAGGGAAAAGCCCGGTACCGA
B	TGCGTAGGGGCCGGAGCCGCAGTGG
G_{AND}	TCGGTACCGGGCTTTTCCCTACATCCACTGCGGCTCCGGCCCCTACGCA
OR gate	
DNA oligomer	**Sequence (5' → 3')**
A	TAGTAGGGAAAAGCCCGGTACCGA
B	TGCGTAGGGGCCGGAGCCGCAGTGG
G_{AND}	TCGGTACCGGGCTTTTCCCTACATCCACTGCGGCTCCGGCCCCTACGCA
NOR gate	
DNA oligomer	**Sequence (5' → 3')**
G_A	TAGTAGGGAAAAGCCCGGTACCGA
G_B	TGCGTAGGGGCCGGAGCCGCAGTGG
A	TCGGTACCGGGCTTTTCCCTACTA
B	CCACTGCGGCTCCGGCCCCTACGCA
G_{NOR}	CCGGGCTTTTCCCTACATCCACTGCGGCTCCGGCCCC
miR input NOR gate	
DNA/RNA oligomer	**Sequence (5' → 3')**
miR-21	UAGCUUAUCAGACUGAUGUUGA

Table 1.2 (continued)

miR-122	UGGAGUGUGACAAUGGUGUUUG
$G_{21\text{-miR}}$	TCAACATCAGTCTGATAAGCTA
$G_{21\text{-input}}$	ATCAGACTGATGTTGACCGGGCTTTTCCCTACTA
$G_{122\text{-miR}}$	CAAACACCATTGTCACACTCCA
$G_{122\text{-input}}$	CCACTGCGGCTCCGGCCCCTGGAGTGTGACAATGG
$G_{NOR,\ miR}$	CCGGGCTTTTCCCTACTACCACTGCGGCTCCGGCCCC
$G_{NOR,\ A}$	TAGTAGGGAAAAGCCCGGTCAACA
$G_{NOR,\ B}$	ACTCCAGGGGCCGGAGCCGCAGTGG
AND-NOR gate cascade	
DNA oligomer	**Sequence (5' → 3')**
B	TAGCTTATCAGACTGATGTTGA
C	CCGGCCCCTACGCATCAACA
$G_{AND,\ input}$	CCACTGCGGCTCCGGCCCCTACGCA
$G_{AND,\ top}$	TCAACATCAGTCTGATAAGCTA
$G_{AND,\ bottom}$	ATCAGACTGATGTTGATGCGTAGGGGCCGG

Table 1.3. A list of gates used and their corresponding strand components.

Gate strands	
Gate	**DNA oligomers**
AND	G_{AND}
OR	G_{AND} + A, G_{AND} + B
NOR	G_{NOR} + G_A + G_B
$AND_{AND\text{-}NOR}$	$G_{AND,\ bottom}$ + $G_{AND,\ top}$ + $G_{AND,\ input}$
miR-21 translator	$G_{21\text{-miR}}$ + $G_{21\text{-input}}$
miR-122 translator	$G_{122\text{-miR}}$ + $G_{122\text{-input}}$
$NOR_{miR\text{-}translator}$	$G_{NOR,\ miR}$ + $G_{NOR,\ A}$ + $G_{NOR,\ B}$

2.0 SIGNAL AMPLIFICATION FOR DNA COMPUTATIONAL SYSTEMS

The field of DNA computation encompasses a wide variety of DNA-based devices, which attempt to utilize the strict control inherent to electronic circuits for biological problems. A variety of nucleic acid tools exist with unique functions,[49,80,81] for example, logic gate operation,[6,18,7] memory simulation,[4] game simulation,[5,23] protein translation activation,[82] and edge detection.[83] In order to operate, the devices require an exchange of DNA strands between logic gates, which can become inefficient in large circuits, leading to a dampening in signal.[18,20,84] Signal amplification is a major limitation in many DNA computational devices. The problem arises from the design of many devices, in which a single input strand cannot release more than a single output strand. Additionally, a high concentration of input is required to maintain reasonable reaction rates, especially in complex multi-layered circuits.[18] However, cellular nucleic acids, such as microRNAs (miRNA), usually exist at low concentrations. Methods to overcome these signal amplification limitations have been developed,[80] such as hairpin-mediated quadratic amplification (HQEA),[85] seesaw gates, catalyzed hairpin assembly (CHA),[86,77] fuel-catalyst cycle,[20] and hybridization chain reaction (HCR).[19] These amplification cycles are typically driven by either a hairpin or fuel strand added in excess, to recycle or reuse the initiating strand. Sequence domains are built into the hairpin to prevent stem opening in the absence of initiator or analyte. For example, the hairpins in HCR

cannot open and form long concatemeric structures until the initiator strand is added. Similarly, the complementary toe-hold sequence of a fuel strand is designed to be available only after the initiating strand has interacted with the amplification cycle components. For instance, the fuel-catalyst cycle fuel strand cannot interact with the main duplex structure until the catalyst strand has interacted with the gate. Amplification cycles can react quickly to a small concentration of analyte, and the cycle will not be prematurely activated in the absence of an initiating strand.

Amplification systems offer a fast and direct route to reliably detect specific sequences of short nucleic acids. MiRNAs are 22 nucleotide long single stranded RNA molecules that have been shown to be important regulators of gene expression.[87] Biogenesis of miRNAs begins with the transcription of primary miRNAs (pri-miRNAs), typically by RNA polymerase II. The pri-miRNA transcript contains the mature miRNA within semi-complementary stem-loop structures. Once transcribed, the pri-miRNA is cleaved by an RNase III protein, Drosha, into a smaller hairpin structure known as precursor miRNA (pre-miRNA). Export of the pre-miRNA from the nucleus is facilitated by Exportin-5. In the cytoplasm, the pre-miRNA is bound by Dicer and cleaved into a 22 nt double-stranded RNA. The mature miRNA strand is bound by Argonaute, while the other strand is degraded. Dysregulation of any step of the miRNA biogenesis pathway can lead to over- or under-expression of miRNAs, which can cause cancers and other diseases. Circulating miRNAs are a type of extracellular miRNAs that exist at very low concentrations. Outside the cell, microvesicles or RNA-binding proteins, such as Argonaute, are thought to keep the circulating miRNAs protected against degradation by RNases. Previous studies have implicated circulating miRNAs as potential cancer

biomarkers. These RNA species are difficult to detect because of their small size and low concentrations of approximately 8,000 to 13,000 copies / μL plasma.[88] Methods such as qRT-PCR[89] or bead-based assays[90] are used to detect and quantify circulating miRNAs, however, these can be labor intensive or difficult to reproduce.[91] Therefore, the potential of DNA amplification systems for analyzing these and other cellular nucleic acids was investigated. This chapter discusses the design of various amplification methods and demonstrates the capability of these methods to amplify signal produced by catalytic amounts of single stranded nucleic acids.

Amplification cycles can be divided into two groups: enzymatic and non-enzymatic. Enzymes are intrinsically catalytic and can be powerful additions to any amplification system. However, additional factors such as stability, cofactor availability, temperature, and pH may limit the applicability of some enzymatic circuits. Interestingly, the non-enzymatic CHA cycle has been used as a transducer for the enzymatic loop-mediated isothermal amplification (LAMP) method.[92] LAMP is a very sensitive method, capable of enzymatically producing more than 10^9 DNA copies from less than 10 templates in a matter of hours (Figure 2.1). LAMP begins with the binding of the R_2R_1* primer to the DNA template and extension of the primer by Bst polymerase. Binding and extension of primer R_3 displaces the first extended DNA strand (from primers R_2R_1*) and forms a hairpin loop at the 3' end. Two subsequent primer extensions with F_1*F_2 and F_2 creates a hairpin on the 5' end. This double hairpin product can be further recycled multiple times to create various length LAMP products. Although the LAMP products are of differing lengths, all contain the same loop domain, which interacts with an acceptor hairpin to expose the CHA input. Binding of the input to hairpin 1 opens a

toe-hold for hairpin 2. The input is displaced from hairpin 1 and recycled after binding of hairpin 2. Finally, a reporter gate is activated by the binding of the hairpin 1 and 2 complex, producing a fluorescence output. A CHA cycle was added as the final readout due to ability to directly measure a fluorescent readout. Previous analysis steps for LAMP required much more laborious methods to indirectly measure amplification. Additionally, the use of CHA as a detection method greatly reduced false positive signals generated by similar amplicon sequences. Hairpin sequences were changed to target multiple LAMP products, enabling multiplexed detection of different inputs. Lastly, the LAMP-CHA combination was applied to the detection of two separate DNA templates by using an AND logic gate. After the LAMP cycle amplified the target DNA sequence, the two CHA cycles produced single stranded outputs capable of interacting with the AND gate. The design was similar to Winfree's AND gate,[18] and enabled a fluorescent output by separating quencher and fluorophore moieties. Connecting the CHA cycle in parallel to the LAMP cycle allowed for greater target specificity, multiplexed detection, and logic operations.

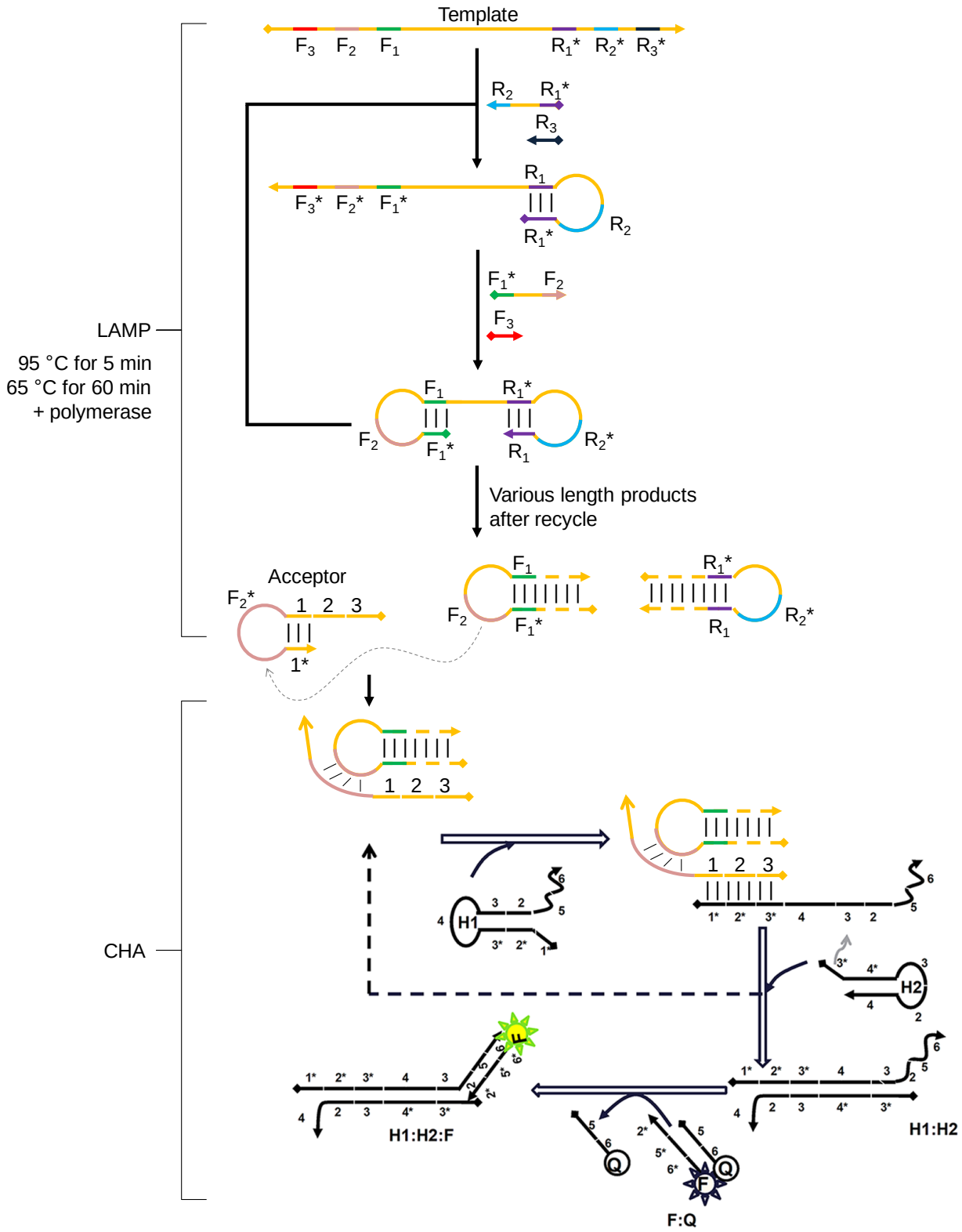

Template
F3 F2 F1
R1* R2* R3*
R2 R1*
R3
F3* F2* F1*
R1
R1*
R2
F1* F2
F3
F1
R1*
F2
F1*
R1
R2*
Various length products
after recycle
LAMP
95 °C for 5 min
65 °C for 60 min
+ polymerase
Acceptor
F2*
1 2 3
1*
F1
F2
F1*
R1*
R1
R2*
1 2 3
CHA
H1
H2
H1:H2:F
H1:H2
F:Q
Q
F

Figure 2.1. Schematic of the LAMP-CHA mechanism. Template DNA and primers were annealed by heating to 95 °C and cooling to 65 °C. Four extension steps using the primers R_2R_1*, R_3, F_1*F_2, and F_3 created an initial LAMP product with two hairpin loops. This double-hairpin loop structure was then recycled to produce multiple hairpins of various lengths (symbolized by dashed DNA line). This LAMP product then interacted with an acceptor hairpin to form the catalyst for CHA. After CHA, the reporter gate produced a fluorescent signal. Numbers indicate DNA sequence domains, with complementary sequences indicated by the same number and an asterisk. Adapted with permission from Li, B.; Chen, X.; Ellington, A. D., Adapting enzyme-free DNA circuits to the detection of loop-mediated isothermal amplification reactions. *Anal. Chem.* 2012, *84* (19), 8371-7. Copyright 2012 American Chemical Society.

Similarly, the exclusively nucleic acid-based HCR has been used in conjunction with enzymatic branched rolling circle amplification (BRCA).[93] This combination (HCR-BRCA) provided a 20-fold improvement over BRCA alone, and enabled the first chemiluminescent readout of DNA methylation. Here, a set of DNA oligomers were hybridized to form a DNA methylation site (Figure 2.2). Addition of the Dam methylase enzyme enabled the duplex to be methylated, and later cut by the restriction endonuclease DpnI. The cleaved DNA circularized to form the primer for BRCA by annealing the primers to complementary strands coupled to magnetic nanoparticles (MNP). Extension by a polymerase created a long ssDNA molecule containing multiple repeat sequences, which acted as initiator strands for HCR. The concatemeric HCR product was composed of hairpins modified with biotin. When a fusion of streptavidin and horse radish peroxidase (HRP) was added to the MNPs, a chemiluminescent signal could be produced. The merging of the enzymatic BRCA and non-enzymatic HCR enabled greater sensitivity and chemiluminescence detection, which would otherwise not be possible. The use of MNPs aided in improving the sensitivity of detection by

enabling separation of bound from unbound streptavidin-HRP, which resulted in accurate detection of the initial methylation event.

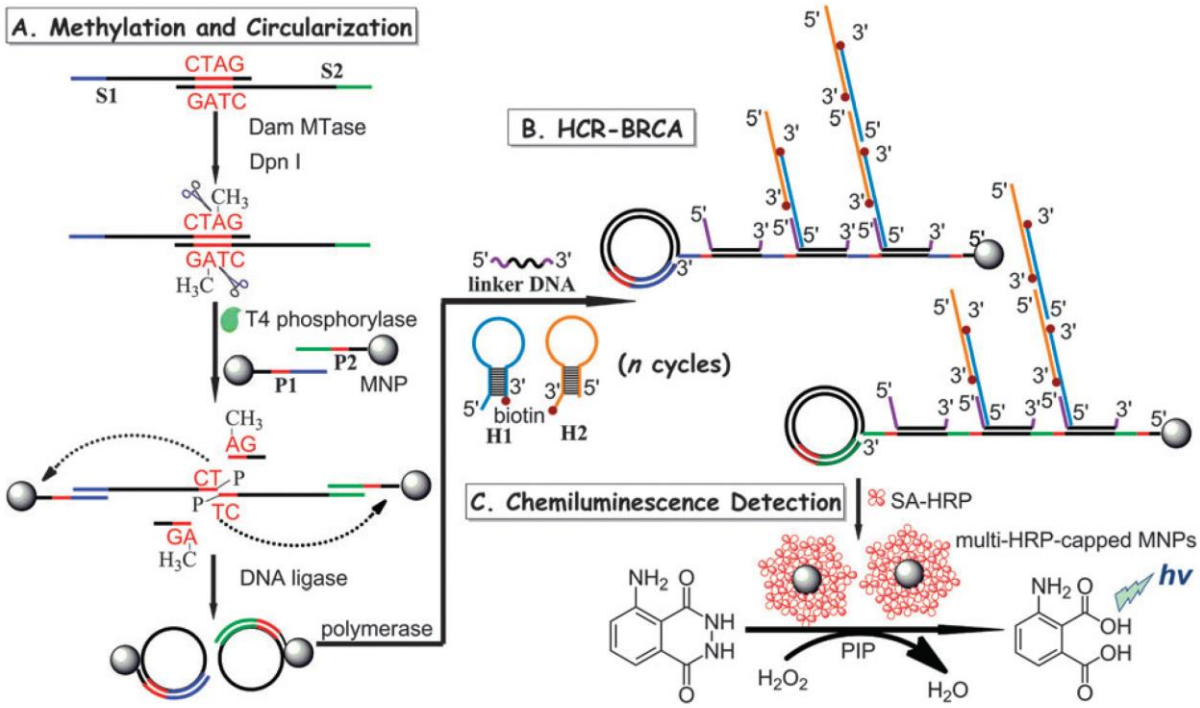


Figure 2.2. The mechanism of BRCA-HCR used to detect methylation via chemiluminescence. A DNA target is methylated by Dam methylase (MTase) and cut with DpnI. Magnetic nanoparticle coupled DNA strands (MNPs) circularize the digested DNA target after hybridization. Extension with a polymerase creates multiple copies of the same DNA domain on a single DNA strand. Addition of a linker DNA binds to the newly polymerized strand and acts as an initiator for HCR amplification. A 3'-biotin modification on the hairpins enables binding to streptavidin coupled horse radish peroxidase (SA-HRP) beads, where chemiluminescent detection is performed. Reproduced from Ref. 93 with permission from The Royal Society of Chemistry.

The amplification methods discussed herein are modular and can be applied to almost any existing nucleic acid circuit. Although the addition of amplification cycles can greatly increase signal, the catalytic nature of the cycles requires extra consideration to ensure that the generation of background signal is minimized. Even a small amount of prematurely activated gate or contaminating DNA strands could cause an uncontrolled cascade. This background can be minimized by proper duplex purification and hairpin annealing to ensure secondary structures are maintained.[19,20] In general, the variety of nucleic acid-based amplifiers allows for flexibility in selecting the proper amplification cycle for a specific application. These amplification cycles are garnering attention as possible diagnostic devices[80] and have recently been used in complex cellular environments.[29]

2.1 HAIRPIN-MEDIATED QUADRATIC ENZYMATIC AMPLIFICATION

2.1.1 Results and Discussion

Hairpin-mediated quadratic enzymatic amplification (HQEA) was introduced by Duan *et al.* to detect and quantify miRNAs from various cell lines.[85] Quadratic amplification refers to the non-linear increase in amplification generated by HQEA. Development of HQEA was motivated by limitations in current miRNA detection methods, such as Northern blots, microarrays, and RT-PCR. A molecular beacon-like hairpin is designed to open upon binding to a specific miRNA sequence (Figure 2.3)**.** Opening of the beacon causes a fluorophore and quencher to separate, allowing fluorescence

emission. Additionally, the opening enables a small primer strand to bind the beacon. Bst polymerase will extend the primer while removing the miRNA from the beacon. After extension, an Nb.BbvCI nicking enzyme recognition site is formed. Nicking the beacon duplex removes a portion of the beacon containing a phosphorothioate backbone. The phosphorothioate linkages are important to ensure that the beacon is not degraded prematurely by lamda exonuclease, which would cause a significant rise in background signal. Lamda exonuclease has a much higher affinity for 5'-phosphorylated ends of dsDNA than 5'-non-phosphorylated.[94] Therefore, after the phosphorothioate bonds are removed by the nicking enzyme, lambda nuclease can degrade the beacon. With the miRNA and extended primer strand recycling, the system is able to achieve quadratic signal amplification.

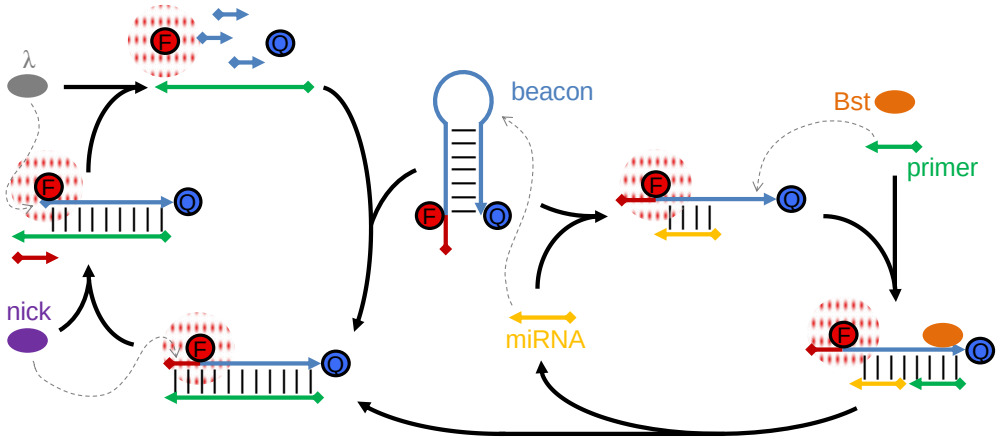


Figure 2.3. The HQEA mechanism. The miRNA input binds to the beacon loop, opening the hairpin and enabling fluorescence emission. Next, a primer strand binds the opened beacon and Bst polymerase extends the primer along the beacon template while removing the miRNA strand. After extension, a recognition site for the Nb.BbvCI is created and the phosphorothioate backbone is removed. Lambda exonuclease degrades the beacon at the open 5'-phosphorylated end of the dsDNA, freeing the extended primer to interact with another beacon. Oligonucleotides are represented by blue (molecular beacon), green (Bst primer strand), and yellow (miRNA) lines. The red line signifies the nicking enzyme recognition site. Enzymes are represented by orange (Bst polymerase), purple (Nb.BbvCI nicking endonuclease), and gray (lambda exonuclease) ovals.

To ensure sufficient amplification was produced, Duan *et al.* tested the HQEA system in the presence of miR-21. Fluorescence intensity rose significantly in the presence of the miRNA, beacon, and enzymes. Next, the effect of temperature was investigated. While quadratic amplification was observed at 25, 37, and 50 °C, maximum activity occurred at 37 °C, with a limit of detection of 10 fM. To further improve the limit of detection by decreasing the residual lambda exonuclease activity toward unbound beacons, the system was incubated at 4 °C.[95] A concentration of 4 aM was detectable at 4 °C, however, the reaction took 50 h. In addition to very low detection limits, HQEA was specific to certain miRNA target sequences. In case of a miR-122 probe, even a single mismatched base caused a complete loss in amplification. HQEA was evaluated in more complex biological environments by incubating the necessary components in MCF-7 and PC3 cell lysates. For cell lysates, the limit of detection was reported to be at the single cell level. Differential expression of miR-21 and miR-122 was distinguished in crude extracts of breast cancer tissue. Overall, HQEA represents a very robust enzymatic-based amplification system that is

capable of specific detection of very low concentrations of RNA even in complex biological environments.

HQEA was selected as a potential enzymatic amplification system for the detection of low concentration miRNAs. Since miRNA inputs were detected in previously published results, the same sequences were used in the following experiments (Table 2.1). Addition of reagents followed the protocol outlined by Duan *et al*.[96] This order of addition was important to minimize background accumulation of signal. Different input concentrations were tested (Figure 2.4) and fluorescence was monitored over 2 h (data not shown), however, no change in signal was observed. High background signal was observed in the absence of input. Effectively, no amplification had occurred. Two buffers were unsuccessfully tested, NEB buffer 2 and TE/Mg^{2+} buffer. According to the published protocol,[96] NEB buffer 2 should have enabled proper amplification.

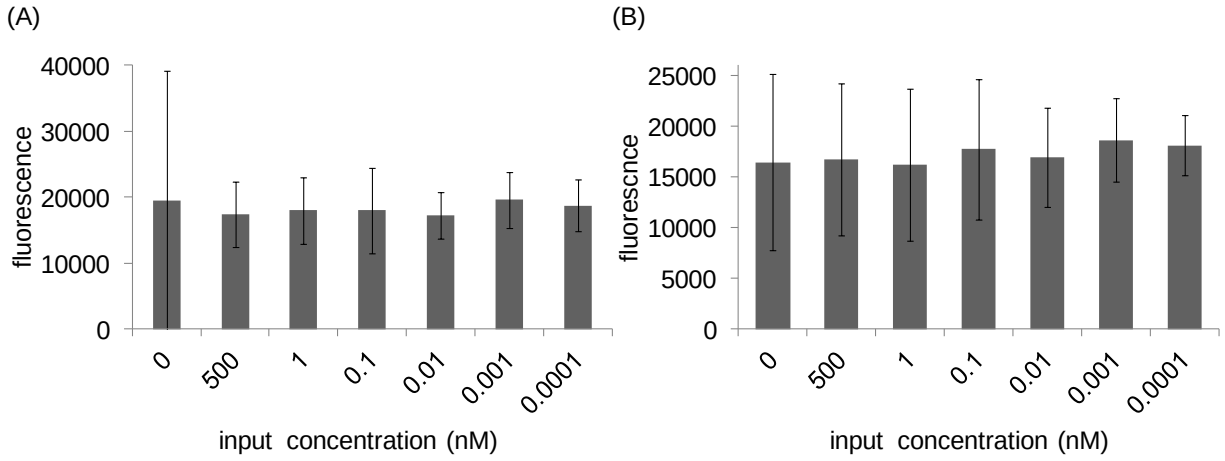


Figure 2.4. Initial test of the HQEA system with two different buffers. The miRNA input concentration was varied from 500 to 0.0001 nM and tested in (A) NEB buffer 2 and (B) TE/Mg^{2+} buffer. No change in signal was observed for any input concentration in either buffer. High background signal was observed at 0 nM input. An average of three experiments is shown and error bars represent standard deviations. Signal was measured after a 2 hour incubation.

Due to the high fluorescence output values for all input concentrations, the HQEA system is likely prematurely and non-specifically triggered. Possible reasons include degradation of the molecular beacon by the lambda exonuclease, misfolding of the beacon, or unfolding of the beacon. If the phosphorothioate backbone modifications were not sufficient for inhibiting the recognition of lambda exonuclease, then the beacon would be degraded and the fluorophore and quencher moieties would be separated before the addition of miRNA. Only two phosphorothioate linkages were added in order to follow the original protocol. However, another study found that three or more successive linkages were able to stop lambda exonuclease activity.[97] The HQEA system could be optimized by increasing the number of phosphorothioate backbone modifications within the beacon to fully inhibit the lambda exonuclease activity. Misfolded or unfolded molecular beacons would enable the catalytic cycle to progress in the absence of miRNA, by binding to the primer strand. After binding, the primer strand would be extended by Bst polymerase, which can continue through the cycle and activate the remaining beacons. Beacons were formed by slow-annealing through heating to 65 °C and cooling to 12 °C. No further purification was pursued according to the published protocol.[96] Strategic mismatches or clamp domains could be inserted into the design of the beacons to promote proper folding and stability of the molecular

beacon stem region. Ultimately, the HQEA system was not working correctly and did not show any increase in fluorescence output from the addition of any input concentration.

2.1.2 Experimental

DNA strands were ordered from Integrated DNA Technologies (IDT, Table 2.1) and enzymes were from New England Biolabs (NEB). The 10X NEBuffer 2.0 (500 mM NaCl, 100 mM Tris-HCl, 100 mM $MgCl_2$, and 10 mM dithiothreitol) and TE/Mg^{2+} buffer (10 mM Tris, 1 mM EDTA, 12.5 mM $MgCl_2$, pH 8) were used.

Table 2.1. DNA strands for HQEA amplification.[85] The fluorophores fluorescein (FAM) and dabcyl (DAB) are abbreviated.

Strand Name	**Sequence (5′ → 3′)**
Hairpin probe	AAGCTGAGGT-(FAM)CTTGGACAGAAACCCAGCAGACAATGTAGCTTGTCCAAGA-(DAB)
Primer	TCTTGGAC
Input	TAGCTTATCAGACTGATGTTGA

HQEA amplification

Water and NEB buffer 2 were added to hairpin beacons (100 μM stock concentration). Beacons (10 μL of a 100 μM stock added to 81 μL water and 9 μL 10X NEBuffer 2.0) were annealed by heating to 65 °C and cooling to 12 °C over 15 minutes. The hairpin was used as is. To test the amplification ability of HQEA, the hairpin components were added in the following order: DEPC-treated water (to bring the total volume to 50 uL),

NEBuffer 2.0 (5 µL of a 10X stock), dNTPs 16 µL of a 5 mM stock), beacon (10 µL of a 10 µM stock), RNase inhibitor (2 µL of a 20 Units/mL stock) (LifeTechnologies), BSA (2 µL of a 10 µg/mL) (NEB), Bst DNA polymerase (0.16 µL of a 800 Units/mL stock) (NEB), Nb.BbvCI nicking enzyme (0.9 µL of a 100 Units/mL stock) (NEB), Lambda exonuclease (0.25 µL of a 500 Units/mL stock) (NEB), primer (8 µL of a 100 µM stock, 4 µM), and a DNA oligo mimicking the miRNA sequence (0.0001-500 nM). The reactions were added to a 96-well black-walled clear bottom plate and fluorescence was measured (excitation / emission 490 / 515 nm) in a Tecan M1000 plate reader after a 2 hour incubation. Each condition was tested in triplicate.

2.2 TRIPLE SEESAW GATE

2.2.1 Results and Discussion

The seesaw gate was developed by Winfree, which used an excess of fuel strands to recycle an input.[15] Development of the seesaw gate was motivated by the need for a simple gate motif that can be easily scaled to create larger and more complex circuits. Additionally, the use of threshold gates enabled digital abstraction, by converting an analog signal (varying input concentration), into an ON or OFF digital signal (input concentration above or below a defined value). A threshold gate set a minimum concentration of input required for gate activation. The threshold gate will outcompete the gate duplex for input due to a longer input toe-hold and no possibility of a reverse reaction (Figure 2.5A). Therefore, the input will be completely consumed by the

threshold gate until its concentration exceeds the threshold gate. Then, the remaining amount of input will be left to activate the seesaw gate. To exemplify the usefulness of a uniform and scalable gate structure, Winfree created large networks containing 130 strands of DNA.[2] The circuit contained the seesaw and threshold gates, as well as a final reporter gate to quantify gate activity. Logic gates were derived from a simple arrangement of two seesaw gates and one threshold gate alternately connected in a serial circuit. By simply tuning the threshold concentration, the seesaw gates could be transformed into either AND or OR logic gates. Utilizing this modular logic gate framework, multi-gate circuits were constructed to investigate the effect of circuit size on signal delay and switching time. Circuit delay increased linearly with extra layers added to the circuit, while the switching time, which is the time necessary for signal to rise from 20 to 80%, was not affected by the number of circuit layers. With an understanding of optimal network conditions, a larger seesaw circuit was constructed to compute a four-bit square root value. A single gate consists of a preformed gate duplex, an input, and a fuel strand (Figure 2.5B). The input binds to the gate duplex, and the output is released through a toe-hold mediated strand displacement reaction. A fuel strand can then bind the gate duplex, liberating the input strand. Since the fuel strand releases the input, only a catalytic amount of input strand is necessary to fully activate a seesaw gate.

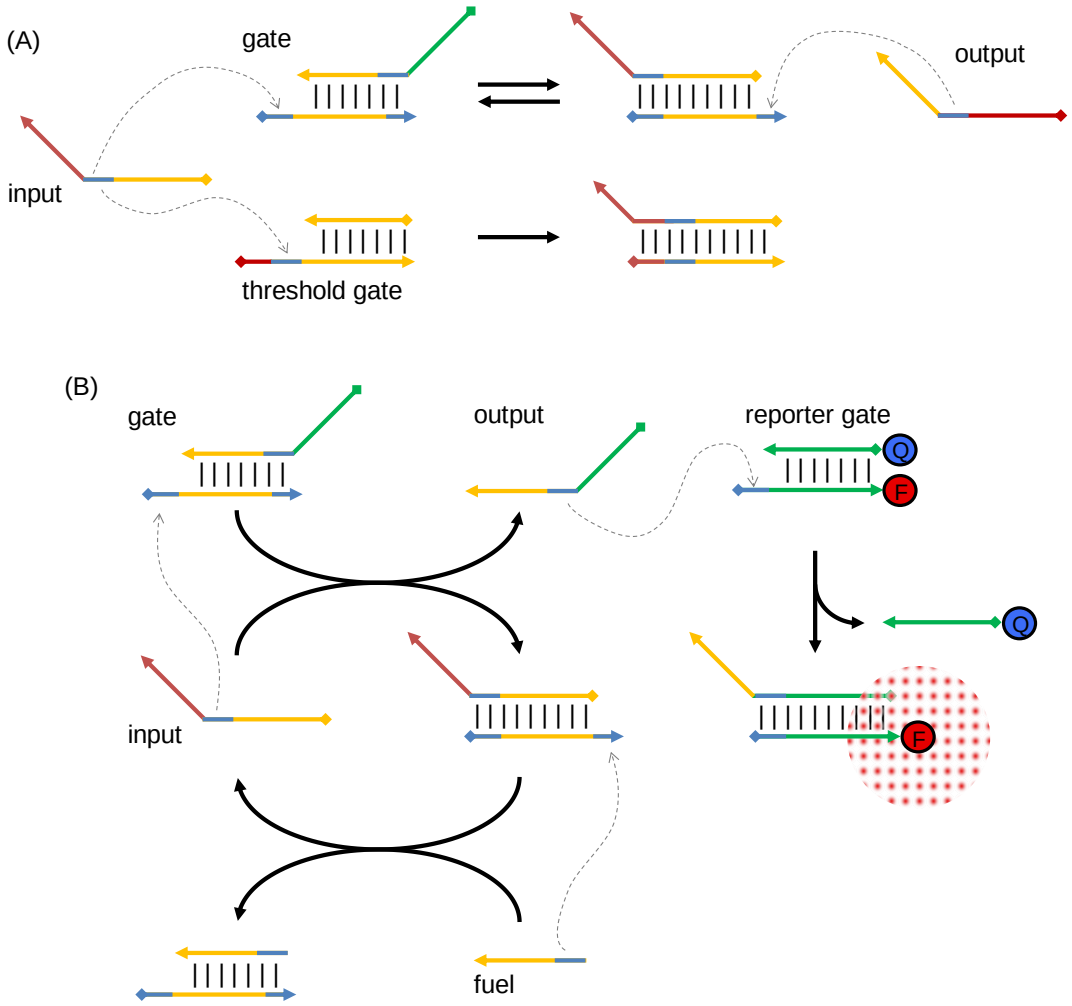


Figure 2.5. The seesaw gate motif. (A) A threshold gate sequesters input up to a certain concentration, due to a long toe-hold region and irreversible interaction. Threshold gates are used at concentrations below the seesaw gate concentration, in order to define a minimal amount of input necessary to trigger the gate. (B) An input initiates the seesaw by binding to the gate duplex. The released output strand can directly interact with the reporter gate. Other than the reporter gate interaction, all seesaw gate strand displacements are reversible. The excess fuel strands drive the seesaw gate to completion by freeing the input strand. Colored lines indicate complementary DNA domains. The 5' → 3' directionality is indicated by the respective diamond and arrowhead. F = fluorophore Q = quencher

An impressive example of seesaw gates, termed the “DNA neural network”, was also developed by Winfree.[4] Using only DNA strands, a four neuron Hopfield associative memory was modeled and tested. The same seesaw, threshold, and reporter gates were used. However, instead of using these three components to build logic gates, the different gates were made into linear threshold gates. Each linear threshold gate diagram displayed all inputs, outputs, threshold levels, and input weights. The input weights specified the amount an input is multiplied. For example, an input of four meant the input produced four times as much output. An initial demonstration consisted of a three-input four-output linear threshold gate. This circuit accepted a three input value, converted it into an analog value, and then decided if the value was less than or greater to the values 1, 3, 5, and 7. For instance, the input 101 would be converted to the analog value of 5. This value is greater than or equal to 1, 3, and 5 but less than 7. Therefore, three unique fluorescent reporters corresponding to 1, 3, and 5 would emit signal, whereas the fourth fluorophore would remain quenched. Similarly, linear threshold gates were designed to create a three-bit XOR logic gate, which accepted three inputs and yields an output only if a single input is present. The gate was tested with all possible combinations of inputs, and fluorescent signals were only generated in the presence of a single input. After successful utilization of linear threshold gates for the three-input XOR gate, a network was designed for the four-neuron Hopfield associative memory. A recurrent circuit was created, which cycled outputs back into the circuit as inputs. Four questions and answers were pre-programmed into the DNA, such that a yes was represented by a 1 and no was represented by a 0. A 1 and 0 were distinguished by the network depending upon the

specific DNA sequence added to the solution. For example, if the first three questions were correctly answered “yes” and the fourth “no,” then the correct answer would correspond to 1110, which was emitted by the corresponding three fluorophores. However, if only a portion of the information was provided, such as the first two “yes” questions, then the network would try to “read your mind” by guessing an answer that started with 11. This type of guessing is comparable to the associative memory that humans use and the neural network is the first example of nucleic acids making decisions based on environmental changes.

The catalytic potential of the seesaw gate was investigated for detection of low oligonucleotide concentrations through signal amplification. The same oligonucleotide sequences and concentrations reported by Winfree were initially tested (Table 2.2). A single seesaw gate was able to successfully amplify a sub-stoichiometric input concentration of 25 nM (Figure 2.6). However, only about a 2-fold change was observed between signal and background. The low signal to background ratio was likely due to a high reporter gate concentration (150 nM) and the result of prematurely released output strands or incorrectly formed gate structures. This was later optimized by lowering the reporter gate concentration to 75 nM (Figure 2.7).

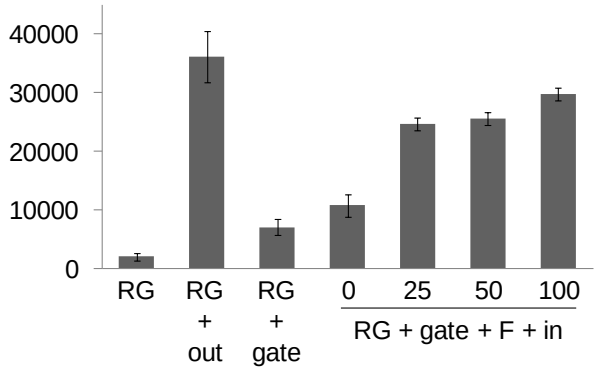


Figure 2.6. An initial test of seesaw gate amplification. The seesaw gate (100 nM) was initially tested with 0, 25, 50, and 100 nM input. A similar fluorescent output generated by the reporter gate (150 nM) is observed for all input concentrations. Abbreviations: reporter gate (RG), output strand (out), seesaw gate (gate), fuel strand (F), and input strand (in). An average of three independent experiments is shown and error bars represent standard deviations. Fluorescence was measured after 30 minutes.

To try to improve signal amplification generated by the seesaw gate, the reporter gate concentration was investigated (Figure 2.7). Reporter gate concentrations lower than the seesaw gate concentration (100 nM) yielded the best signal to background ratio (Figure 2.7), although this was not an improvement compared to the initial seesaw gate test (Figure 2.6). About a 2-fold difference in signal was still observed as before. This equates to an 2-fold signal amplification for the 25 nM input added to the 100 nM reporter gate (where amplification = (normalized signal generated by catalyst – background signal)*([reporter gate]/[input])).[20] However, as reporter gate concentration increased, the signal to noise ratio slowly decreased.

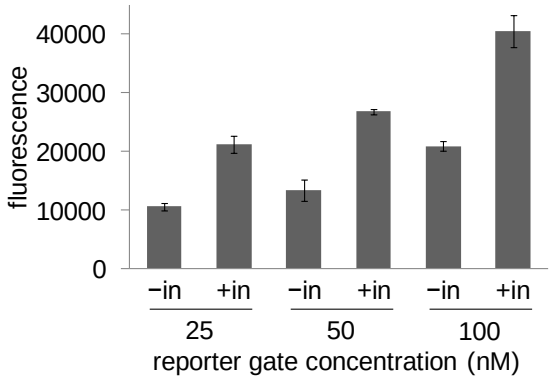


Figure 2.7. Optimization of the reporter gate concentration. Input was added at 25 nM to the seesaw gate (100 nM) and the reporter gate (25, 50, and 100 nM). An average of three independent experiments is shown and error bars represent standard deviations. Fluorescence was measured after 30 minutes.

To determine the limit of detection of the single seesaw gate, the concentration of input was varied from 0 to 100 nM (Figure 2.8). Only concentrations of 10 nM or higher were detectable. This corresponds to 1/10 of the seesaw gate concentration, or a 6.7-fold amplification, which is similar but still lower than what Winfree reported.[2]

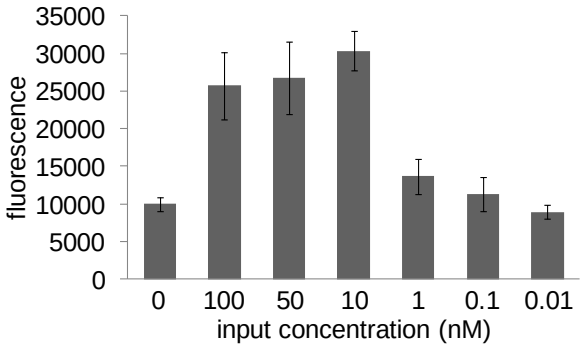


Figure 2.8. Determination of the detection limit for a single seesaw gate (10 nM). Using 100 nM seesaw gate and 75 nM reporter gate, only inputs at ≥0.1x the seesaw gate concentration could be detected. An average of three experiments is shown and error bars represent standard deviation. Fluorescence measured after 3 hours.

As discussed earlier, multiple gates have been layered before to create the "neural network"[4] and 4-bit square root circuit,[2] but not specifically for the purpose of amplifying the signal from a low concentration analyte. Thus, a double seesaw gate was constructed that uses the output of the first seesaw gate as the input. The goal was to

build upon the amplification of the first cycle by further amplifying the output. Since 10 nM input was detectable with a 100 nM seesaw gate concentration, adding another seesaw gate was predicted to enable detection of at least an additional 10-fold lower input concentration. Multiple concentrations of input were tested for the double seesaw gate, and at only 0.5 nM input, the signal was significantly higher than background (Figure 2.9). This represents an approximate 20-fold improvement in amplification compared to a single seesaw gate. However, the double gate was still not able to detect low picomolar concentrations, which is necessary to be considered competitive among other enzyme-free nucleic acid devices.[77] Compared to a single seesaw gate, the background noticeably increased. This increase is likely due to the serial coupling of two amplification systems, which would also increase sensitivity to any signal leakage. A threshold gate could be added between or after the two seesaw gates to counteract a small amount of output leakage. However, the overall goal of using seesaw gates is to detect extremely low input concentrations. Therefore, the threshold gate might decrease signal at the cost of lowering background.

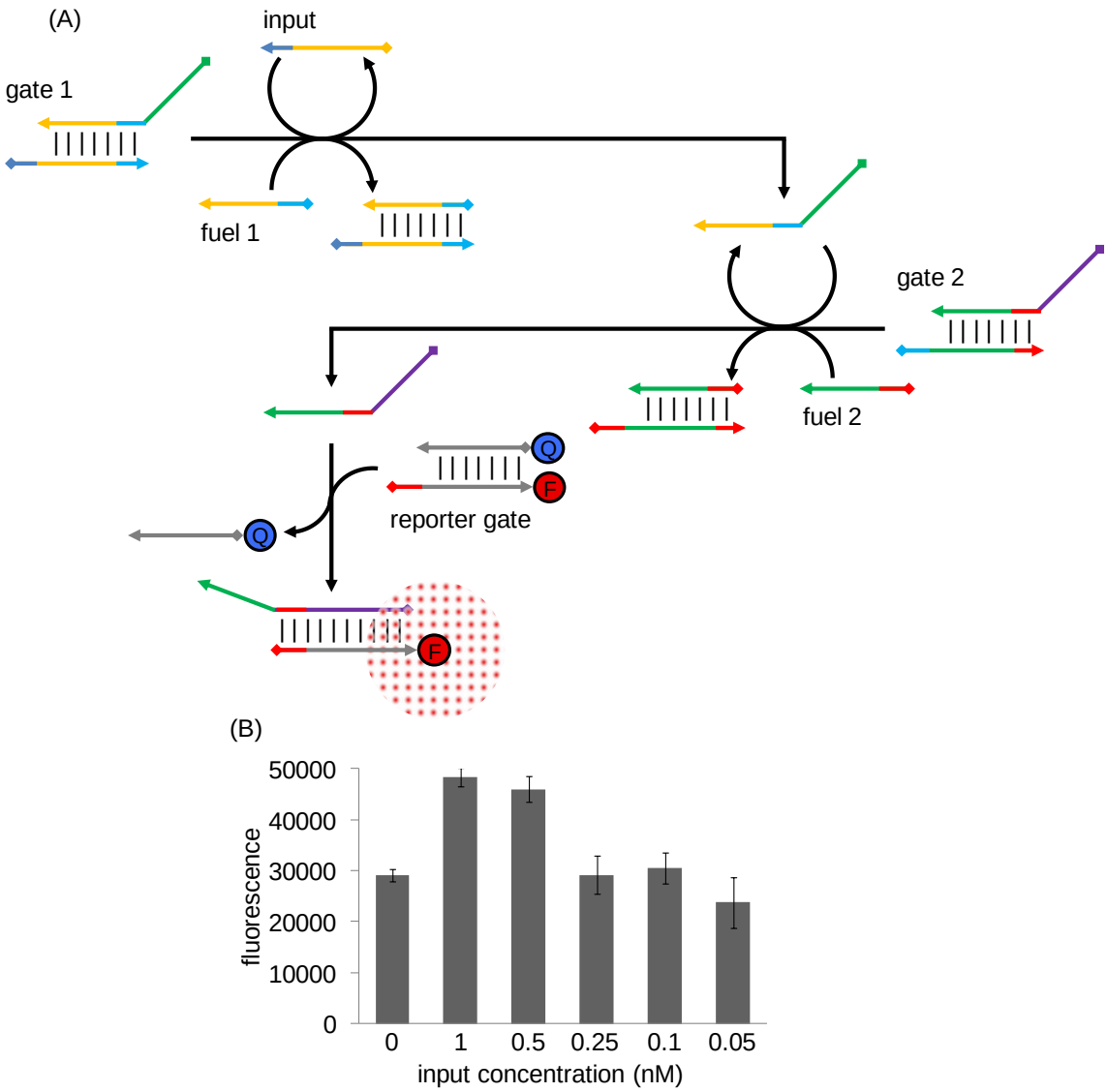


Figure 2.9. A double seesaw gate was designed to utilize the output of the first gate as an input for the second gate. (A) The input is recycled by an excess of fuel, which releases an output from the first gate. This output acts as an input to the second gate and a final output strand is released. The output from the second gate displaces a quencher strand in the reporter gate, enabling fluorescence emission. (B) The tandem seesaw gate system was successfully triggered with 0.5 nM input added to 100 nM first gate and 50 nM second gate. An average of three experiments is shown and error bars represent standard deviation. Fluorescence measured after 1.75 hours.

To further improve the detection limit of the seesaw gate system, a third gate was introduced (Figure 2.10). All three gates are serially connected such that the output of the previous gate acts as the input for the next gate (Table 2.2). The final output produced by the third gate interacts with the reporter gate and removes the quencher strand, leading to a fluorescent signal.

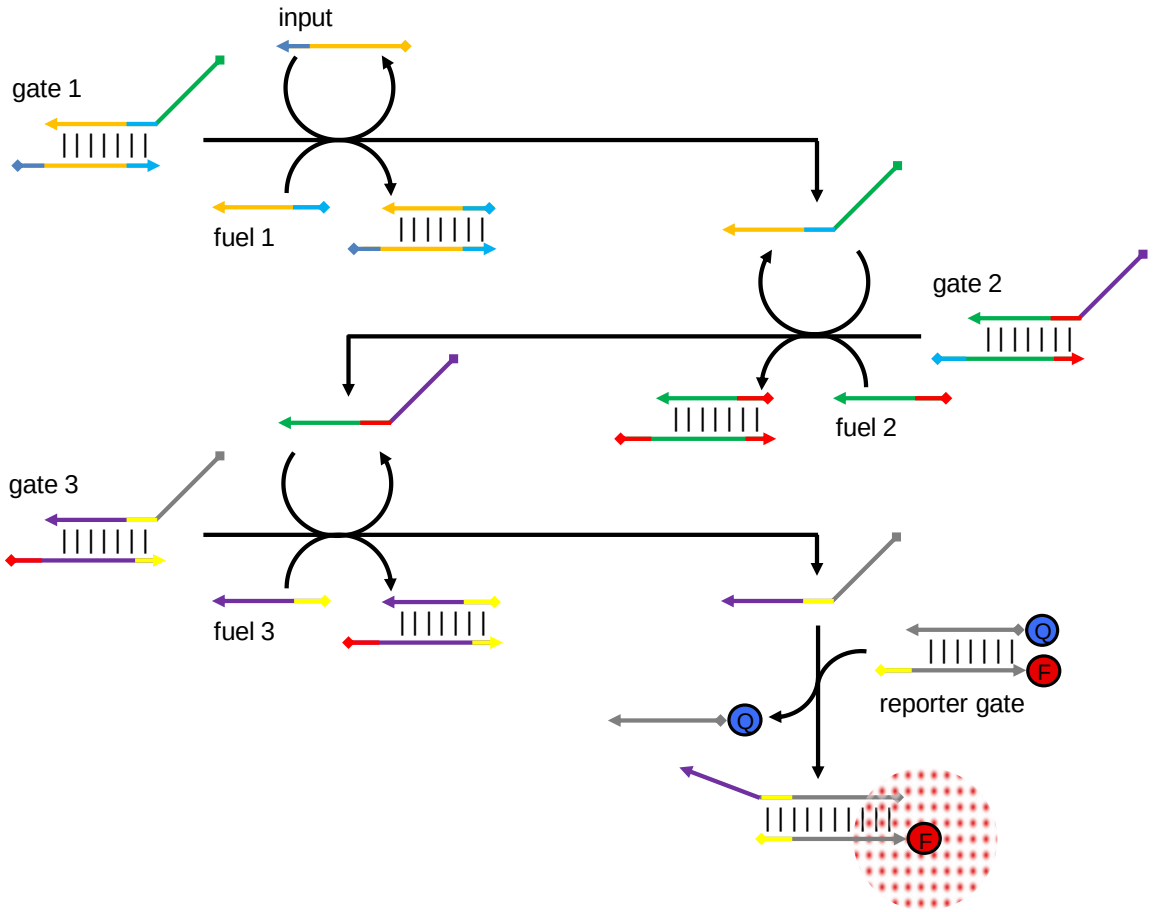


Figure 2.10. Schematic of the triple seesaw gate. Three seesaw gates are serially connected such that the output of the previous gate acts as the input for the next gate. Ultimately, a low concentration input will enable a fluorescence output. Directionality of the DNA strands from 5' to 3' is shown with a diamond and arrowhead, respectively. DNA is represented by lines, and colored areas indicate complementary domains. F = fluorophore, Q = quencher

Multiple concentrations of the first gate were tested (Figure 2.11). Addition of 10 nM input to the first, second (100 nM), and third (200 nM) gates produced minimal amplification. Lower concentrations of the third gate tended to produce less background signal, although high levels of background were observed with each concentration. Since each additional layer of a circuit can add to overall processing time, concentrations of the third gate were added in excess (100 and 200 nM) of the second gate (50 nM) were tested in an effort to increase the rate of detection. However, no increase in signal production was observed and a higher concentration of the first gate tended to produce higher background.

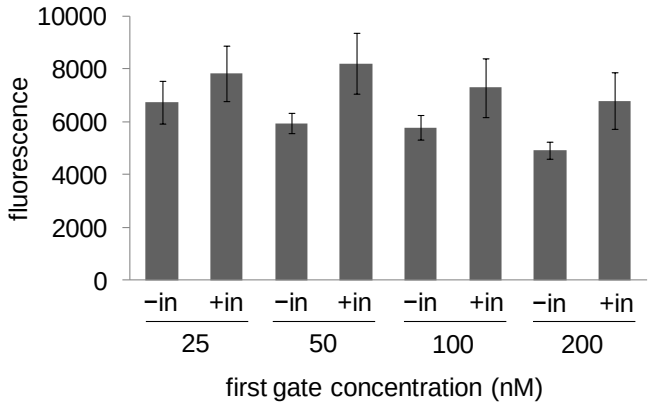


Figure 2.11. Triple seesaw gate concentration optimization. The first gate concentration was tested at 25, 50, 100, and 200 nM with (+in) or without (−in) 25 pM input. The second gate (50 nM), third gate (100 nM), second fuel strand (100 nM), and third fuel strand (200 nM) concentrations were held constant. Minimal amplification was produced by the low input concentration for all conditions. Data is shown as an average of three experiments and error bars represent standard deviations. Fluorescence measured after 3 hours.

To further improve the detection with the triple seesaw gate system, the gates were purified by incubating the seesaw gate duplex with fuel strand before separation on a PAGE gel.[2] Any misformed gate structures would be triggered by the added fuel strand, and the PAGE separation would allow the retention of only properly formed gates. As expected, the fluorescence levels of the purified gates were noticeably lower than the unpurified gates (Figure 2.12A). However, neither purified nor unpurified gates were able to detect low input concentrations. The limit of detection of the purified gates was also tested with input concentrations as low as 0.05 nM (Figure 2.12B). Although the system was able to detect such a low concentration of input, the results were not reproducible. Not surprisingly, the signal that was produced by low inputs was just barely noticeable above background levels.

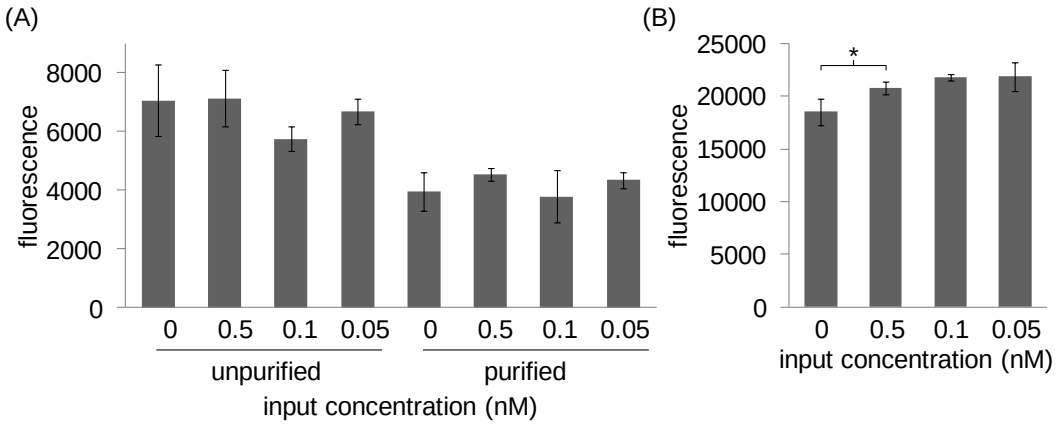


Figure 2.12. (A) Purified and unpurified gates were tested with multiple input concentrations. Fluorescence was measured after a 3 hour incubation at room temperature. (B) Purified triple seesaw gates were tested with lower input concentrations. An input concentration of 0.05 nM or higher was detectable. Asterisk (*) indicates significant results, with a p-value of 0.047. Data is shown as an average of three experiments and error bars represent standard deviations.

A delayed addition strategy was attempted to decrease background signal. The input was allowed to interact with the first gate for an hour before the second gate was added. Subsequent gate additions were delayed after hour incubations. Thus, any leaked signal generation would not be able to trigger the reporter gate until all gates have been activated. But, only an input at 1 nM or higher could be detected through staged gate addition (Figure 2.13). Unexpectedly, high input concentration (100 nM) was almost indistinguishable from background levels. The delayed addition strategy did not decrease background signal. The time between each addition may need to be further optimized. Longer interval times would ensure gates to be fully triggered before the next gate is added. However, if this time is too long, background in the absence of

input may fully trigger the gate as well. Therefore, a full range of time intervals should be investigated for further optimization.

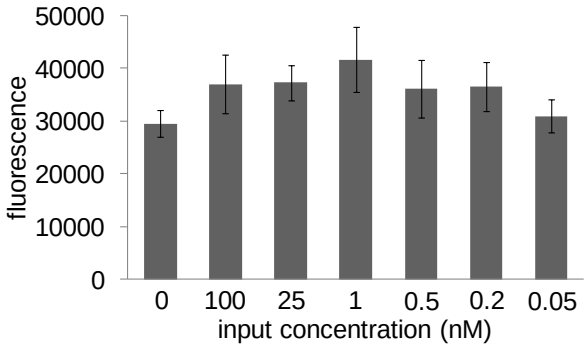


Figure 2.13. Delayed addition strategy for the triple seesaw system. Gates were added after subsequent 1 hour incubations to allow full release of output strands. No difference in signal was observed for any input lower than 1 nM. Data is shown as an average of three experiments and error bars represent standard deviation. Fluorescence measured 1 hour after addition of the reporter gate.

Overall, the triple seesaw gate system was not able to reproducibly detect low input concentrations. Addition of a second seesaw gate decreased the limit of detection from 10 nM to 0.5 nM. However, lower detection limits have been achieved by other DNA-based amplification systems.[77] Addition of a third gate was attempted to further decrease this limit of detection was not successful. Signal generated by input was barely noticeable above background levels. Additional optimization would be required to successfully and reproducibly use the triple seesaw gate system. Extra purification techniques, such as HPLC and PAGE purification of DNA strands may be able to remove some impurities from commercial DNA synthesis. Enzymatically synthesized strands have also been previously reported to decrease the rate of background signal

generation.[77] Sequences can be further optimized by using more G/C base pairs near the ends of DNA duplexes to strengthen the binding of DNA strands. Also, strategically placed mismatches have been shown to greatly reduce background signal generation.[9] Threshold gates can also be added to potentially decrease the amount of background produced in the seesaw gate cascades.[4,2] A threshold gate added before the reporter gate, third seesaw gate, or second seesaw gate may sequester any prematurely leaked signal. The concentration of each threshold gate would need to be carefully optimized. High threshold gate concentrations would sequester too much input, leading to very long processing times, whereas low threshold gate concentrations would not maximally reduce background signal accumulation.

2.2.2 Experimental

All DNA strands were purchased from Integrated DNA Technologies (IDT). 1X TE/Mg^{2+} buffer (10 mM Tris, 1 mM EDTA, 12.5 mM $MgCl_2$, pH 8) was used for all gate reactions and annealing, and diluted from a 10X stock.

Purification of gates

After receiving DNA from IDT, a 100 μM stock was made by adding the appropriate amount of nuclease-free deionized water. Gate strand (45 μL of a 100 μM stock) and output strand (45 μL of a 100 μM stock) were added to TE/Mg^{2+} buffer (10 μL of a 10X stock, 100 mM Tris, 10 mM EDTA, 125 mM $MgCl_2$, pH 8). For reporter gate, the strands Reporter Gate F and Reporter Gate Q were used. These DNA strands were slow annealed in a PCR thermocycler (BioRad T100 Thermal Cycler) by heating to 95 °C and

cooling to 12 °C with intermediate steps at 85, 75, 65, 55, 45, 35, 25, and 15 °C for 1 min each. After annealing, DNA loading buffer (25 µL, 50% glycerol, 50% water) was added. If the gate was purified further, fuel strand (1 µL of a 100 µM stock) was added to the annealed gate strands and, after 5 min, the sample was loaded onto a 16% native-gel (wells of the gel were removed by manual scraping with a metal spatula to create one large flat-bottomed well). The gel was run at 100 V for 45 min in native-PAGE running buffer (1X buffer, 192 mM glycine, 25 mM tris base). Afterwards, the gel was removed, placed on a TLC plate, and illuminated briefly with a handheld UV light. The top-most band was excised, cut into smaller pieces, and added to a 1.7 mL microcentrifuge tube with 1X TE/Mg^{2+} buffer (1 mL). The mixture was left overnight at room temperature. The seesaw gate solution was transferred into a fresh tube, frozen, and used as-is for further experiments. The DNA concentration was determined by measuring the absorbance on a NanoDrop ND-1000 instrument (260 nm, also check A_{260}/A_{280} for purity). This was then converted to concentration using Beer's Law and the determined extinction coefficient for the duplex (IDT Biophysics or other nearest-neighbor equations).

Table 2.2. DNA sequences used in the triple seesaw gate amplification system. Toe-hold regions are underlined. F = tetramethylrhodamine, Q = black hole quencher 2

Strand name	Sequence (5' → 3')
Seesaw gate 3	TGAGATGAAGTTTGGTGGTGAGATG
Seesaw output 3	CATAACACAATCACATCTCACCACCAAACTTCA
Seesaw fuel 3	CAACATATCAATTCATCTCACCACCAAACTTCA
Seesaw gate 2	TGGCTAGGACCAGGTTTTGTTTTGAGA
Seesaw output 2	CACCACCAAACTTCATCTCAAAACAAAACCTGGTCCT
Seesaw fuel 2	TCTCAAAACAAAACCTGGTCCT
Seesaw gate 1	TGACAGTTCATCTGGTACAGTCTGGCT
Seesaw output 1	AAACAAAACCTGGTCCTAGCCAGACTGTACCAGATGAAC
Seesaw fuel 1	AGCCAGACTGTACCAGATGAAC
Input	GACTGTACCAGATGAACTGTCA
Reporter gate F	F-CATAACACAATCACA
Reporter gate Q	TGAGATGTGATTGTGTTATG-Q

Gate activity assay

The gates were tested for activity by combining the purified seesaw gate, purified reporter gate, fuel strand, and input strands with 20 µL 10X TE/Mg^{2+} buffer according to the concentrations below. Water was added to a total volume of 200 µL. From the 200 µL solutions, 50 µL were added to three separate wells of a black-walled clear bottom 96-well plate and allowed to react (room temperature) for the specified amount of time. Fluorescence was measured on a Tecan M1000 plate reader (em/ex 545/585 nm) at the time points shown below.

Initial single seesaw test: To a PCR tube, reporter gate (1.8 µL of a 16 µM stock, 150 nM), seesaw gate 1 (4.5 µL of a 4.4 µM stock, 100 nM), fuel strand (4 µL of a 10 µM stock, 200 nM), input strand (0, 25, 50, 100 nM) were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). When necessary, output strand (3 µL of a 10 µM stock, 150 nM) was added as a positive control. Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Gates reacted for 30 minutes and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Reporter gate optimization: To a PCR tube, reporter gate (25, 50, 100 nM), seesaw gate 1 (4.5 µL of a 4.4 µM stock, 100 nM), fuel strand (4 µL of a 10 µM stock, 200 nM), and input strand (25, 50, 100 nM), were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Gates reacted for 30 minutes and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Limit of detection for single seesaw gate: To a PCR tube, reporter gate (0.9 µL of a 16 µM stock, 5 nM), seesaw gate 1 (4.5 µM of a 4.4 µM stock, 100 nM), fuel strand (4 µL of a 10 µM stock, 200 nM), and input strand (0.01, 0.1, 1, 10, 50, 100 nM), were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Gates reacted for 3 hours and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Double seesaw gate system initial test: To a PCR tube, reporter gate (0.9 µL of a 16 µM stock, 150 nM), seesaw gate 1 (4.5 µL of a 4.4 µM stock, 100 nM), fuel strand 1 (4 µL of a 10 µM stock, 200 nM), seesaw gate 2 (1.2 µL of a 16 µM stock, 100 nM), fuel

strand 2 (4 μL of a 10 μM stock, 200 nM), and input strand (0.05, 0.1 0.5 nM) were added to TE/Mg^{2+} buffer (20 μL of a 10X stock). Water was added to adjust the volume to 200 μL total. From the 200 μL solutions, 50 μL were added to three separate wells. Gates reacted for 1.7 hours and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Triple seesaw gate optimization: To a PCR tube, reporter gate (0.9 μL of a 16 μM stock, 75 nM), seesaw gate 3 (6.2 μL of a 3.2 μM stock, 100 nM), fuel strand 3 (4 μL of a 10 μM stock, 200 nM), seesaw gate 2 (0.6 μL of a 16 μM stock, 50 nM), fuel strand 2 (2 μL of a 10 μM stock, 100 nM), seesaw gate 1 (25, 50, 100, 200 nM), fuel 1 (2x seesaw gate 1 concentration, 50, 100, 200, 400 nM), and input strand (1 μL of a 50 nM stock, 25 pM), were added to TE/Mg^{2+} buffer (20 μL of a 10X stock). Water was added to adjust the volume to 200 μL total. From the 200 μL solutions, 50 μL were added to three separate wells. Gates reacted for 3 hours and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Purified gates: To a PCR tube, reporter gate (0.9 μL of a 16 μM stock, 75 nM), seesaw gate 3 (6.2 μL of a 3.2 μM stock, 100 nM), fuel strand 3 (4 μL of a 10 μM stock, 200 nM), seesaw gate 2 (0.6 μL of a 16 μM stock, 50 nM), fuel strand 2 (2 μL of a 10 μM stock, 100 nM), seesaw gate 1 (0.3 μL of a 16.9 μM stock, 25 nM), fuel 1 (2 μL of a 10 μM stock, 100 nM), and input strand (0.5 – 0.5 nM), were added to TE/Mg^{2+} buffer (20 μL of a 10X stock). Water was added to adjust the volume to 200 μL total. From the 200 μL solutions, 50 μL were added to three separate wells. Gates reacted for 3 hours and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

Delayed addition of gates: To a PCR tube, reporter gate (1.6 µL of a 9.3 µM stock, 75 nM) was added to seesaw gate 3 (3.9 µL of a 5.1 µM stock, 100 nM), fuel strand 3 (4 µL of a 10 µM stock, 200 nM), seesaw gate 2 (2 µL of a 4.9 µM stock, 50 nM), fuel strand 2 (2 µL of a 10 µM stock, 100 nM), seesaw gate 1 (1 µL of a 4.8 µM stock, 25 nM), fuel 1 (2 µL of a 10 µM stock, 100 nM), and input strand (0.5-100 nM), were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Gates reacted for 1 hour and fluorescence was measured (excitation / emission, 545 nm / 585 nm).

2.3 AND GATE CONNECTED TO AMPLIFYING REPORTER GATE

2.3.1 Results and Discussion

In an effort to combine logic gate operation with signal amplification, an AND gate serially connected to an amplifying reporter gate was designed. An AND gate was created to recognize two inputs with the sequences of miRNA-21 and -122. The design was similar to Winfree's gate, such that the first input, miR-21, would remove the first strand of the gate through a toe-hold mediated strand exchange, and reveal a toe-hold for the second input, miR-122 (Figure 2.14). The second branch migration reaction involving miR-122 would remove the output strand. The gate functioned as an AND gate, because an output (the removed DNA strand) only occurred in the presence of both inputs (miR-21 and miR-122). Amplification of the signal begins with the output

strand binding to the reporter gate and releasing the quencher strand to enable the emission of fluorescence. Essentially, the reporter gate is a seesaw gate modified with a 5'-fluorophore and 3'-quencher. Output is released and recycled after the fuel strand binds to a newly uncovered second toe-hold of the reporter gate. The output can continue to bind additional reporter gates, as long as the fuel is still present.

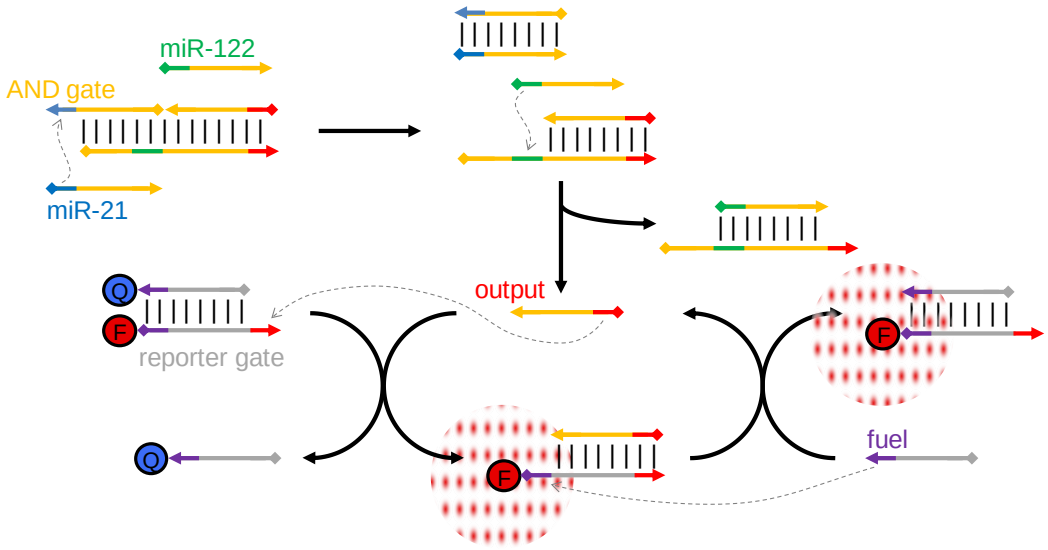


Figure 2.14. Reaction scheme for the serially connected AND and amplifying reporter gate. The miRNA inputs miR-21 and miR-122 interact with the AND gate to yield a single output strand, which acts as a catalyst for the seesaw-based reporter gate. An excess of the fuel strand drives signal amplification. Directionality of the DNA strands from 5' to 3' is shown with a diamond and arrowhead, respectively. Yellow lines represent strands that are part of the AND gate, and gray lines represent reporter gate strands. F = fluorophore, Q = quencher

To determine if the AND gate was capable of detecting nucleic acid inputs, the AND gate (200 nM) was tested with high concentrations of miR-21 and miR-122 DNA

mimics (800 nM each). At these high input concentrations, the AND gate should be fully activated and the efficiency of the reporter gate amplification will be exclusively investigated. In accordance with the AND gate truth table, a high signal was only observed with the addition of both inputs (Figure 2.15). The signal for the activated gate was only about 2-fold greater than the non-activated gate. A non-amplifying reporter gate used in the triple seesaw gate system yielded an approximately 3-fold change in signal for when connected downstream of a single seesaw gate. Therefore, the amplifying reporter gate could be increasing background signal. A small leakage would greatly amplify signal, leading to a larger increase in background. Although threshold gates can be added to seesaw gates to prevent a rise in background signal, the addition of one after a stoichiometric AND gate utilizing low input concentrations would be detrimental to the overall detection limit. A small input concentration would only be able to produce an equal amount of output, which would then be consumed by the threshold gate before being amplified by the reporter gate. Therefore, other optimization strategies, such as modifying the sequences and concentrations should be attempted. Strategic mismatches[9] placed within the reporter gate could potentially decrease background through a trial-and-error approach. Although high input concentration was added to fully activate the AND gate, the assumption of full activation may be false, and output strand is not actually being released. Due to the miRNA input sequences, further optimization of the AND gate is limited to the toe-hold region of the output strand, unless two translator gates are added to convert the miRNA sequences into new input sequences. The toe-hold region could be shortened to increase the rate of dissociation of the output strand from the AND gate. Also, the toe-hold sequence could be changed

to incorporate more A/T base pairs, which would weaken the binding and promote dissociation from the AND gate.

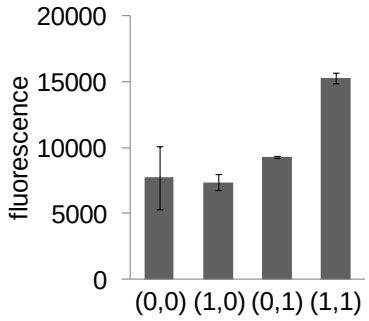


Figure 2.15. Initial test of the AND gate (200 nM) and amplification reporter gate (100 nM) with high input concentrations. A high signal is generated only in the presence of 800 nM miR-21 and 400 nM miR-122 inputs. Input addition is given in Boolean format, such that the first and second numbers indicate miR-21 and miR-122, respectively. Data is shown as an average of three experiments and error bars represent standard deviation. Fluorescence was measured after 3 hours.

The effect of reporter gate concentration was investigated (Figure 2.16). Lower concentrations barely produced any signal, and the highest concentration still only had about a 2-fold difference between activated and non-activated gate. Clearly, the fuel strand was causing a noticeable increase in background. A solution to reduce background would be changing the sequence of the reporter gate at the 5' and 3' ends to include more G and C bases. With stronger binding, the duplex should be more stable and become less susceptible to interactions with the fuel strand, until the quencher strand is removed by the AND gate output. The toe-hold for the fuel strand could be shortened as well. For example, a 4 nucleotide toe-hold for the fuel strand

would reduce binding energy and the rate of binding to the reporter gate, potentially decreasing background in the absence of input.

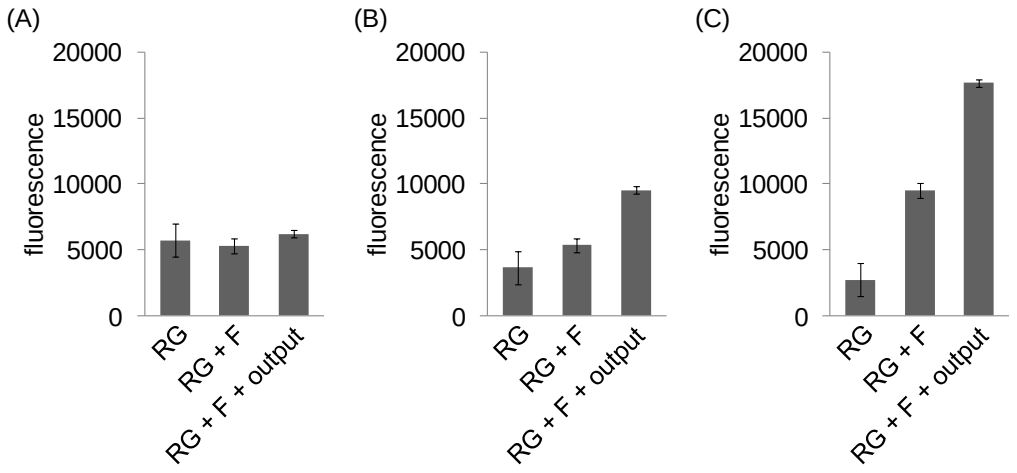


Figure 2.16. Optimization of the amplifying reporter gate concentration. The reporter gate concentration was tested at (A) 25 nM, (B) 50 nM, and (C) 100 nM. Fuel strand (F) was added at 10x the concentration of the reporter gate (250, 500, and 1000 nM) and output strand (output) was added at 100 nM. The best signal to background ratio was obtained for 100 nM RG. Data is shown as an average of three experiments and error bars represent standard deviations. Fluorescence measured after 3 hours.

The AND gate (100 nM) and amplifying reporter gate were again tested by adding either input miR-21 or miR-122 in excess (800 nM each, Figure 2.17A). Due to the poor performance of the AND gate and amplifying reporter gate in previous experiments, fuel strand was added in large excess (5 μM) in order to drive the amplification reaction forward. Additionally, the fluorescence was measured after an overnight incubation to allow for full gate activation. As expected, the gate only released output in the presence of both inputs, as shown by an increase in fluorescence.

However, when a sub-stoichiometric amount of input was added (10 nM) compared to the AND gate concentration (100 nM), the reporter gate did not produce a fluorescent signal (Figure 2.17B). The lack of signal production in the presence of low input concentrations could be due to minimal output strand release from the AND gate or poor amplification of signal by the reporter gate.

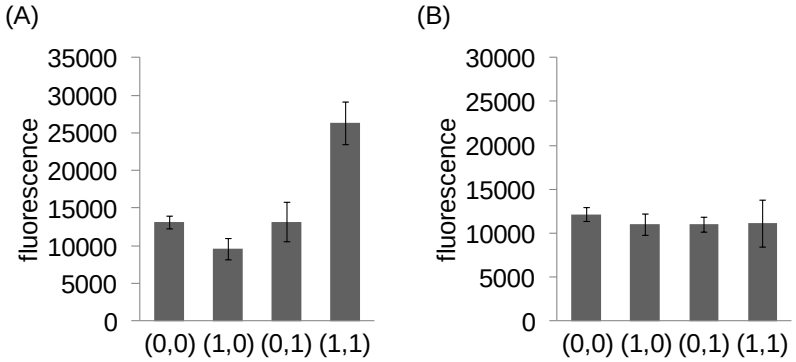


Figure 2.17. The AND gate (100 nM) and reporter gate (100 nM) were tested with (A) high (400 nM) and (B) low (10 nM) concentrations of inputs. A signal was only produced when a high input concentration was added, implying that amplification was not occurring. Data is shown as an average of three experiments and error bars represent standard deviations. Fluorescence measured after 15 hours (overnight).

Due to the poor response of the AND gate connected to an amplifying reporter gate, any further investigation was suspended and no extensive optimization was conducted. As shown in Figure 2.17, no amplification was occurring and the system was essentially functioning as a regular AND gate. Even in the presence of high input concentrations, the difference between activated gate and non-activated gate was about 2-fold. Compared to many other established amplification systems,[2, 77] a 2-fold change

is quite low. Different temperatures could also be investigated. While lower temperatures could be useful in reducing background, the rate of signal generation would decrease due to slower diffusion of DNA strands between gates. Additionally, the AND gate and amplification reporter gate depend upon the natural dissociation of the six nucleotide toe-hold of their respective output strands from the gate duplex. Lowering the temperature would only stabilize this binding, likely leading to less overall signal. While higher temperatures may encourage more interaction between the fuel and reporter gate, the toe-hold exchange reactions will be more complete, leading to more strand recycling. Higher temperatures have proven helpful in other toe-hold exchange reactions.[52]

2.3.2 Experimental

All DNA strands were purchased from Integrated DNA Technologies (IDT). The buffer 1X TE/Mg^{2+} (10 mM Tris, 1 mM EDTA, 12.5 mM $MgCl_2$, pH 8) was used for all gate reactions and annealing, and diluted from a 10X stock.

Purification of gates

Gates were assembled by mixing the corresponding Gate and Output strands in Table 2.3. Further purification was carried out according to the directions listed in Section 2.2 (seesaw gate purification).

Gate activation fluorescence assays

Gates were tested for activation by incubating the purified AND gate, purified reporter gate, miR-21 DNA mimic, miR-122 DNA mimic, and 20 µL 10X TE/Mg^{2+} buffer according to the concentrations below. Water was added to a total volume of 200 µL for each condition. From the 200 µL solutions, 50 µL from each sample was transferred into three separate wells of a 96-well black-walled clear bottom plate. Fluorescence was measured in a Tecan M1000 plate reader (excitation / emission 545 / 585 nm). Fluorescence was measured over 12 hours on a Tecan M1000 plate reader (em/ex 545/585 nm).

Initial gate test: To a PCR tube, AND gate (45.4 µL of a 1.4 µM stock, 200 nM), reporter gate (20 µL of a 0.99 µM stock, 100 nM), miR-21 DNA mimic (input 1, 16 µL of a 10 µM stock, 800 nM), miR-122 mimic (input 2, 16 µL of a 10 µM stock, 800 nM), and fuel strand (4 µL of a 10 µM stock, 200 nM) were added to TE/Mg^{2+} buffer (20 µL of a 10X stock) and water (78.6 µL). From the 200 µL solutions, 50 µL were added to three separate wells. Fluorescence was measured after 3 hours (excitation / emission, 545 nm / 585 nm).

Optimization of the reporter gate concentration: To a PCR tube, reporter gate (25, 50, 100 nM), miR-21 DNA mimic (input 1, 16 µL of a 10 µM stock, 800 nM), miR-122 mimic (input 2, 16 µL of a 10 µM stock, 800 nM), and fuel strand (2x reporter gate concentration, 50, 100, 200 nM) were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Fluorescence measured after 3 hours (excitation / emission, 545 nm / 585 nm).

High and low input concentrations: To a PCR tube, AND gate (10.8 µL of a 1.8 µM stock, 100 nM), reporter gate (0.4 µL of a 45 µM stock, 100 nM), miR-21 DNA mimic (input 1, 16 µL of a 10 µM stock, 800 nM) or miR-21 DNA mimic (0.2 µL of a 10 µM stock, 10 nM), miR-122 mimic (input 2, 16 µL of a 10 µM stock, 800 nM) or miR-122 mimic (0.2 µL of a 10 µM stock, 10 nM), and fuel strand (100 µL of a 10 µM stock, 5 µM) were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Fluorescence measured after 15 hours (excitation / emission, 545 nm / 585 nm).

Table 2.3. Sequences of DNA strands used in the serially connected AND gate and amplifying reporter gate. Toe-hold regions are underlined. F = tetramethylrhodamine, Q = BlackHole Quencher 2

Strand Name	Sequence (5' → 3')
AND gate miR 21	TGTTTGTCAACATCAGTCTGATAAGCTA
Output	CATGCATGGAGTGTGACAATGG
AND gate base	ATCAGACTGATGTTGACAAACACCATTGTCACACTCCATGCATG
Reporter gate Q	TGGAGTGTGACAATGGTCAGTA-Q
Reporter gate F	F-TACTGACCATTGTCACACTCCATGCATG
miRNA-21 DNA mimic	TGGAGTGTGACAATGGTCAGTA
miRNA-122 DNA mimic	TGGAGTGTGACAATGGTGTTTG

2.4 DUAL LAYER CATALYTIC HAIRPIN ASSEMBLY

2.4.1 Results and Discussion

The catalytic hairpin assembly (CHA) architecture was originally designed by Pierce,[86] and utilized the DNA hairpin motif as a programmable kinetic trap. Fundamentally, the work detailed a method to create an amplification circuit from ssDNA sequence and two meta-stable hairpins. Inspired by Pierce, Ellington simplified the design into the current CHA framework.[98] Unnecessary DNA domains were stripped away, and the lengths of toe-holds, hairpin loops, and hairpin stems were optimized for proper hairpin folding. Compared to Pierce's design, the new CHA boasted a 200-fold decrease in the rate of uncatalyzed background signal generation, while maintaining similar catalyzed reaction rates. Similar to Winfree's translator gate design,[18] Ellington used a molecular beacon-like hairpin to translate the siRNA sequence into a general catalyst sequence. Activation of the CHA cycle could only occur in the presence of siRNA. Correspondingly, an adenosine-triphosphate (ATP) aptamer hairpin was designed to release a single stranded region upon binding to ATP to interact with the orginal CHA design. When ATP was added to the DNA components, the catalytic cycle was initiated, resulting in small molecule detection. The translator and aptamer hairpins also demonstrated the modularity of the CHA design. Thus, re-designing and optimizing the entire CHA framework is not necessary for each new application.

The CHA mechanism begins with the hybridization of an oligonucleotide to the first hairpin, HP1 (Figure 2.18). Opening HP1 enables the second hairpin, HP2, to bind, which removes the catalyst strand. A reporter gate composed of fluorophore- and

quencher-modified strands binds to the HP1:HP2 duplex. With the quencher and fluorophore moieties separated, the fluorescence signal is emitted.

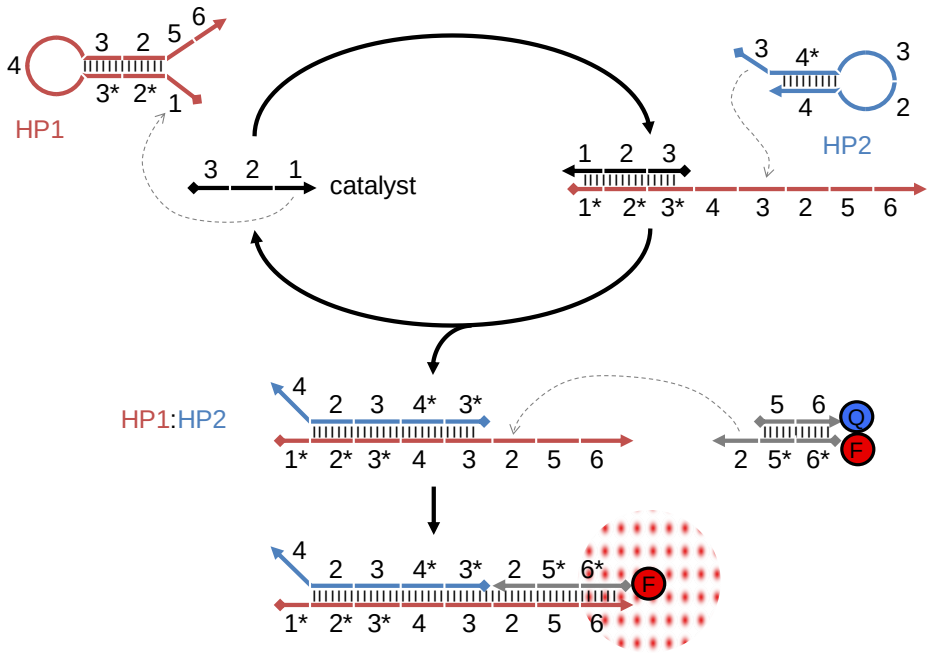


Figure 2.18. Schematic of the catalytic hairpin assembly. A catalyst strand interacts and opens a hairpin (HP1), exposing complementary domains to a second hairpin (HP2). The catalyst strand is removed after HP2 binds HP1. Fluorescence is generated from a reporter gate that is triggered by the HP1:HP2 complex. Directionality of the DNA strands from 5' to 3' is shown with a diamond and arrowhead, respectively. Numbers indicate unique DNA domains and complementary domains are distinguished by an asterisk (*). F = fluorophore and Q = quencher

Ellington expanded upon the idea of using a transducer to convert an input into a new sequence by utilizing a separate orthogonal CHA cycle. Two catalytic hairpin assemblies were layered to create the dual catalytic hairpin assembly (CHA-CHA).[77] The first layer operates the same as the CHA, except the reporter gate is omitted

(Figure 2.19). A portion of the HP1:HP2 complex that would have interacted with the reporter gate in a single-layered CHA, instead initiates the second layer CHA. Thus, two serial CHA cycles are initiated by a single catalyst strand. Ellington used the CHA-CHA system to achieve 7000-fold amplification. When a two-layer CHA-CHA system (a total of four CHAs) was used, up to 600,000-fold amplification was recorded. However, the two-layer CHA-CHA cascade required enzymatically-synthesized DNA strands, a longer total processing time, and suffered from a more rapidly growing background signal.[77] Overall, the successful cascading of catalytic cycles demonstrated the non-linear increase in signal generation by two relatively simple catalytic cascades. In addition to the large signal amplification advances, the sources of background signal generation were also investigated. Ellington discovered that signal leakage, or signal generated in the absence of a catalyst, was due to two types of leakage: asymptotic leakage and initial leakage. Asymptotic leakage is caused by the interaction between two perfectly formed structures to spontaneously trigger background amplification. Since asymptotic leakage is an intrinsic property of the circuit design, methods to minimize this leakage include careful selection of DNA sequences, such as using 50% G/C content, and may require trial and error. Impurities in the sequence of DNA or secondary structure of DNA duplexes cause initial leakage. This type of leakage is immediately present after all components are added together. Typically, secondary structures are used to temporarily block the hybridization of a single stranded region of DNA until a later timepoint. Therefore, if the duplex does not properly form, the single stranded DNA is available to prematurely bind a complementary strand. Thus, the background signal can be generated in the absence of a catalyst strand. To control for initial leakage, Ellington

proposed purification strategies involving elution from polyacrylamide gels, biotin-assisted strand purification, and enzymaticaly synthesized ssDNA. Ultimately, the enzymatically synthesized strands had the greatest effect on decreasing the background signal generation due to the creation of high-quality DNA strands. Truncations and deletions due to commercial DNA synthesis are thought to be contributing factors to signal leakage.[99]

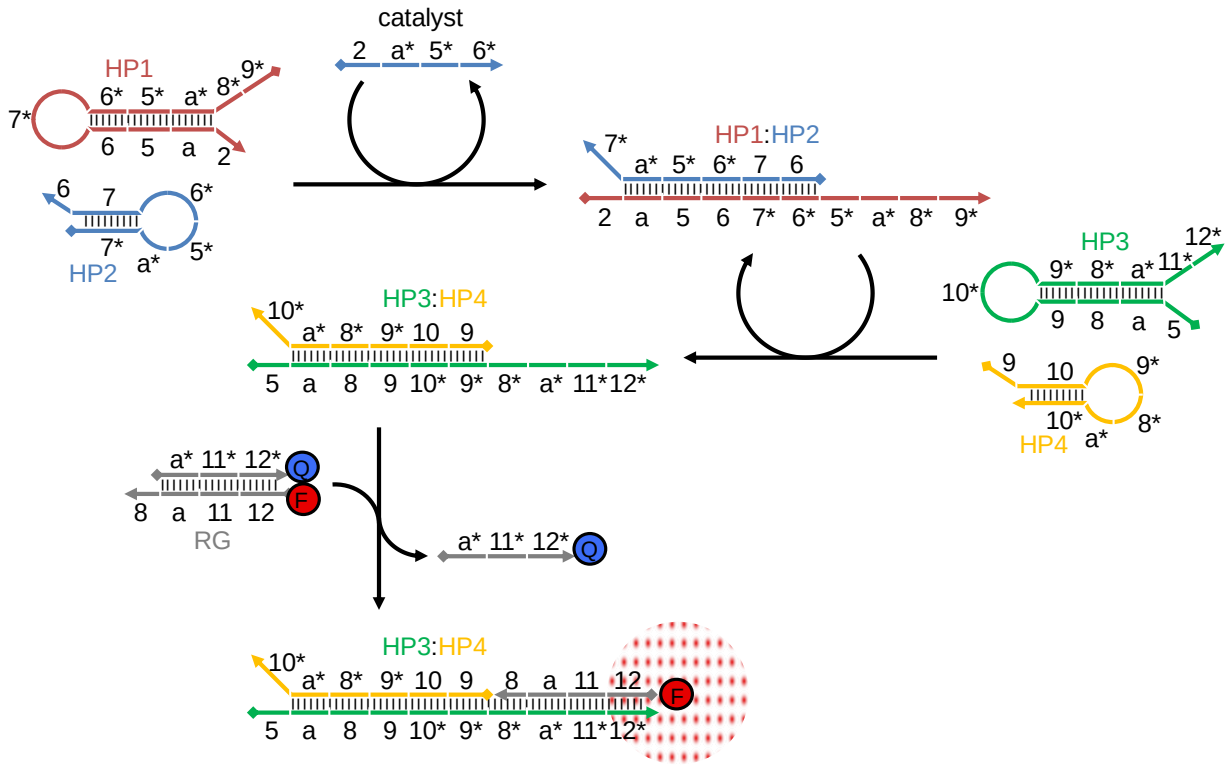


Figure 2.19. Schematic of the two-layer catalytic hairpin assembly. A catalyst strand is recycled after interacting and opening stem-loops of the HP1 and HP2 hairpins. A ssDNA region of the HP1:HP2 duplex can act as an initiator to open the HP3 and HP4 hairpins. Opening of these hairpins also recycles the HP1:HP2 duplex. A reporter gate is finally activated by HP3:HP4, and fluorescence can be emitted. Directionality of the DNA strands from 5' to 3' is shown with a diamond and arrowhead, respectively. Numbers indicate unique DNA domains and complementary domains are distinguished by an asterisk (*). F = fluorophore and Q = quencher

Initially, CHA was tested with multiple concentrations of input (Figure 2.20). The sequences were the same as those used in Ellington's design.[77] Only a concentration of 0.5 nM could be successfully detected above background signal, and signal generated by input was barely recognizable compared to background signal. The measured fluorescence for the addition of 0.05 nM is higher than expected. Fluorescence may be artificially high for this data point due to errors in handling. Ellington reported a detection limit of 5 pM a similar CHA design, which is two orders of magnitude lower than the limit of detection shown in Figure 2.20.[98] The CHA design was clearly detecting input, but either the signal production was not optimal or the background was too high. Reports from Ellington showed that essentially no background interaction was observed between the two hairpins even after 15 hours. To increase signal, the hairpin concentrations could be optimized. Previously, higher hairpin concentrations appeared to somewhat increase signal. But, this may correspondingly increase background.

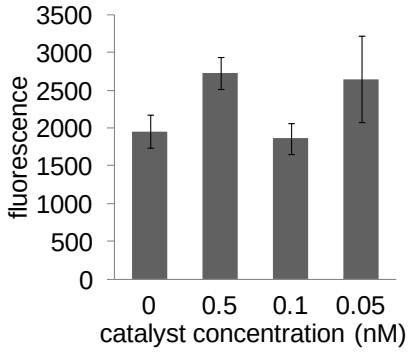


Figure 2.20. An initial test of the CHA system was performed with sub-nanomolar concentrations of input. Only 0.5 nM was detectable with CHA. Signal was collected 3 hours after addition of input. Data is shown as an average of three experiments and error bars represent standard deviation. Fluorescence was measured after 2.33 hours.

To try and further improve upon the limit of detection of the CHA cycle, a second CHA cycle was introduced. Hairpins were purified by elution from a PAGE gel. The CHA-CHA system was composed of two serially connected CHA cycles and used the same sequences designed by Ellington. However, a test of the CHA-CHA system showed only an input concentration in excess of 1 nM was able to be amplified (Figure 2.21). This limit of detection is higher than a single CHA cycle (0.5 nM) and raw fluorescence values show an almost ten-fold increase in background. As reported by Ellington, the background signal for the CHA-CHA cycle increased greatly unless enzymatically synthesized strands are used.[77] Thus, the serial connection of a second CHA cycle likely increased background with minimal effect on signal production. Impurities caused by DNA synthesis have also been reported elsewhere,[99] and could potentially trigger a DNA-based system in the absence of input. Base deletions or strand truncations would prevent proper hairpin folding leading to increased background

or inhibit toe-hold binding and causing decreased signal production. Purification via PAGE gel should have removed any of these impurities. This purification would not separate any incorrect strand sequences. Incorrect sequences would lead to improper hybridization of duplexes and prevent binding of toe-holds and other domains. The importance of synthesizing enzymatically pure strands may lie in preventing truncations, base deletions, and incorrectly added bases.

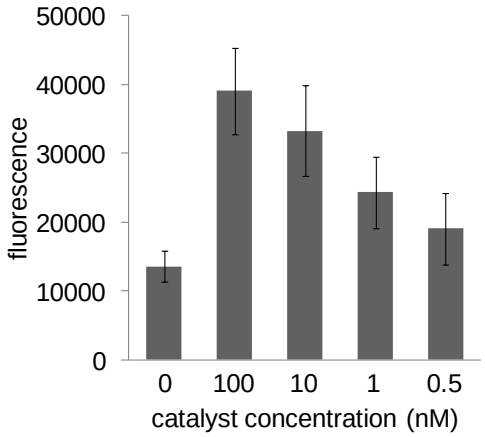


Figure 2.21. The CHA-CHA system was tested with multiple input concentrations. Successful amplification was achieved with greater than 1 nM input concentrations. An average of three experiments is shown with errors bars representing standard deviation. Fluorescence measured after 15 hours.

Despite the reported amplification power of the tandem CHA-CHA system, amplification was never observed. Furthermore, the CHA-CHA system should exhibit a lower detection limit than a single CHA cycle. The limit of detection for the CHA-CHA cycle was 1 nM, whereas the single CHA cycle the detection limit was 0.5 nM. Although unexpected, this result may be due to a rapid increase in background signal. As

discussed earlier, this is likely due to leakiness arising from misfolded hairpins or impure DNA originating from impurities in DNA synthesis. Any prematurely generated background would only be further amplified with the downstream CHA cycle. Due to a combination of added processing time for the dual CHA-CHA cycle and the potential for truncated, synthetic DNA oligonucleotides, the rise in background signal may cause the limit of detection for the single CHA cycle to be lower than the CHA-CHA cycle. A threshold gate, which is commonly used with seesaw gates,[15] could be added after the first and second CHA cycle to reduce accumulation of background signal. The threshold gates could sequester small amounts of output generated in the absence of catalyst. Processing time may be increased, however, the addition of threshold gates may circumvent the need for enzymatically synthesized DNA strands. Overall, the CHA and related CHA-CHA cycle did not perform as reported and the limit of detection was not even within the low picomolar range.

2.4.2 Experimental

All DNA strands were purchased from Integrated DNA Technologies (IDT). DNA stocks were made by adding nuclease-free water to a final concentration of 100 μM. These stocks were stored at −20 °C. The buffer TE/Mg^{2+} (10 mM Tris, 1 mM EDTA, 12.5 mM $MgCl_2$, pH 8) was used for all gate reactions and annealing, and diluted from a 10X stock.

Purification of hairpins and reporter gate

The reporter gate was assembled by mixing the corresponding components in Table 2.4. Further purification was carried out according to the directions listed in Section 2.2 (seesaw gate purification). Individual hairpin strands were annealed and not further purified for the initial CHA test. Hairpins used in the CHA-CHA systems were purified (see seesaw gate purification protocol) by individually annealing hairpins (30 μL of 100 μM hairpin added to 10 μL TE/Mg^{2+} buffer and 60 μL water, annealing procedure explained in seesaw gate purification protocol), adding the sets of hairpins together (hairpins 1 and 2 together, hairpins 3 and 4 together), and incubating overnight at room temperature. If there were any misformed hairpins, these should have hybridized overnight. Instead of excising the top band, the bottom two bands were excised and eluted into separate tubes. Purified hairpins were approximately 4 μM.

Catalyzed hairpin assembly fluorescence assays

CHA was tested for activity by incubating the hairpins, reporter gate duplex, catalyst strand, and TE/Mg^{2+} buffer (20 μL a 10X stock) according to the concentrations below. Water was added to a total volume of 200 μL for each condition. From the 200 μL solutions, 50 μL from each sample was transferred into three separate wells of a 96-well black-walled clear bottom plate. Fluorescence was measured in a Tecan M1000 plate reader (excitation / emission 545 / 585 nm).

Catalytic hairpin assembly test: To a PCR tube, Hairpin 1 (4 μL of a 10 μM stock, 200 nM), hairpin 2 (2 μL of a 10 μM stock, 100 nM), catalyst strand (0.05-0.5 nM), and reporter gate (1.8 μL of a 16 μM stock, 150 nM) were added to TE/Mg^{2+} buffer (20 μL of

a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Fluorescece measured after 2.33 hours.

Serially connected CHA-CHA: To a PCR tube, purified hairpin 1 (10.9 µL of a 3.7 µM stock, 200 nM), purified hairpin 2 (4.7 µL of a 4.3 µM stock, 100 nM), purified hairpin 3 (6.1 µL of a 6.6 µM stock, 200 nM), purified hairpin 4 (4.8 µL of a 4.1 µM stock, 100 nM), catalyst strand (0.5-100 nM), and reporter gate (5.4 µL of a 5.6 µM stock, 150 nM) were added to TE/Mg^{2+} buffer (20 µL of a 10X stock). Water was added to adjust the volume to 200 µL total. From the 200 µL solutions, 50 µL were added to three separate wells. Fluorescence measured after 15 hours.

Table 2.4. Sequences of DNA strands used in the serially CHA and CHA-CHA systems. F = tetramethylrhodamine, Q = BlackHole Quencher 2

Strand name	Sequence (5' → 3')
Hairpin 1	GCTACTACGTGTTTTAGCGAGCGGTAGTAGCGAAAGTGCACACG CTCGCTAAAACAC
Hairpin 2	GTTTCACCTGTGCACTTTCGCTACTACCGCTCGCTAAAACACGTA GTAGCGAAAGTGCACATCTCAGAAATGGC
Hairpin 3	GCTATATCCTCCACGACATCTCAGAAATGGCCGTGGAGGATATAG CGCCATTTC
Hairpin 4	CATAACACAATCACATCTCAGAAATGGCGCTATATCCTCCACG GCCATTTCTGAGATGTGCACTTTC
Reporter gate F	F-CATAACACAATCACA
Reporter gate Q	TGAGATGTGATTGTGTTATG-Q
Catalyst strand	GTAGTAGCGAAAGTGCACAGGTGAAAC

2.5 OPTICALLY CONTROLLED SIGNAL AMPLIFICATION FOR DNA COMPUTATION

2.5.1 Results and Discussion

As discussed above, several options are available to amplify a low DNA output signal: for example, the use of branched DNA to accumulate labeled DNA strands on an output strand,[100] a kinetically trapped metastable DNA fuel,[101] the hybridization chain reaction (HCR),[19,102] and an entropy-driven fuel–catalyst cycle.[20,84] However, the aforementioned amplification methods are limited in their ability to be externally controlled, particularly in a temporal and spatial fashion. Therefore, light-regulated variants of HCR and fuel–catalyst cycle were developed for DNA computation.[28] This design relies on introducing nucleobase-caging groups into DNA strands[41,39,40] to sterically block DNA/DNA hybridization until irradiation with UV light induces caging group removal and DNA duplex formation.[103,32] Thus, a simple chemical modification to an existing structure enables the DNA amplification devices to be either turned ON or OFF through application of photochemical triggers in a spatial and temporal manner.

HCR allows for the detection of small concentrations of nucleic acids by generating an amplified signal through the opening of metastable hairpins to form a long nicked duplex, even in complex biological environments.[29] Three components are required for the reaction: two hairpins and an initiator strand. In the absence of the initiator, the hairpins will not cross react, as there are no complementary sequences exposed. However, when the initiator is added, it will hybridize to the toehold of hairpin 1 (H1) and expose a toehold for hairpin 2 (H2). After H2 binds to H1, a new toehold will

be revealed, allowing the concatemer to continue growing. Overall, the presence of the input signal (i.e., the initiator strand) is amplified through the production of high molecular weight duplexes. To obtain optical control over HCR, the initiator strand was blocked with nucleobase caging groups to prevent binding to H1. Activation of the initiator strand is achieved by irradiation with UV light to photochemically remove the caging groups. Consequently, UV irradiation can act as a switch to turn on HCR (Figure 2.22A). This design utilizes the known hairpin and initiator sequences published by the Pierce group,[19] with modified toehold regions to increase the number of thymidines. These hairpin sequences have been designed to prevent premature signal generation in the absence of initiator. The extra thymidine residues also allowed for increased flexibility in selecting nucleobases to cage.

Before optical control could be investigated, proper function of HCR needed to be ensured. Oligonucleotides were combined in solution, incubated overnight, and analyzed by gel electrophoresis. As expected, HCR only occurred when both hairpins and initiator strand were present (Figure 2.22B). To photochemically control HCR, four caging groups were added to the initiator, creating strand N_4. Four caging groups were selected based on previous experiments with caged oligomers, which demonstrated photochemical control of DNA/DNA hybridization with a nucleobase caging group every 4–5 bases.[32] In the absence of UV light, no HCR occurred and no higher molecular weight products were formed. Removal of the caging groups through UV exposure restored activity to the initiator, which was evident by the formation of the same HCR products as were produced by the noncaged initiator. Thus, photochemical control of HCR was achieved through a synthetic modification of a single DNA component.

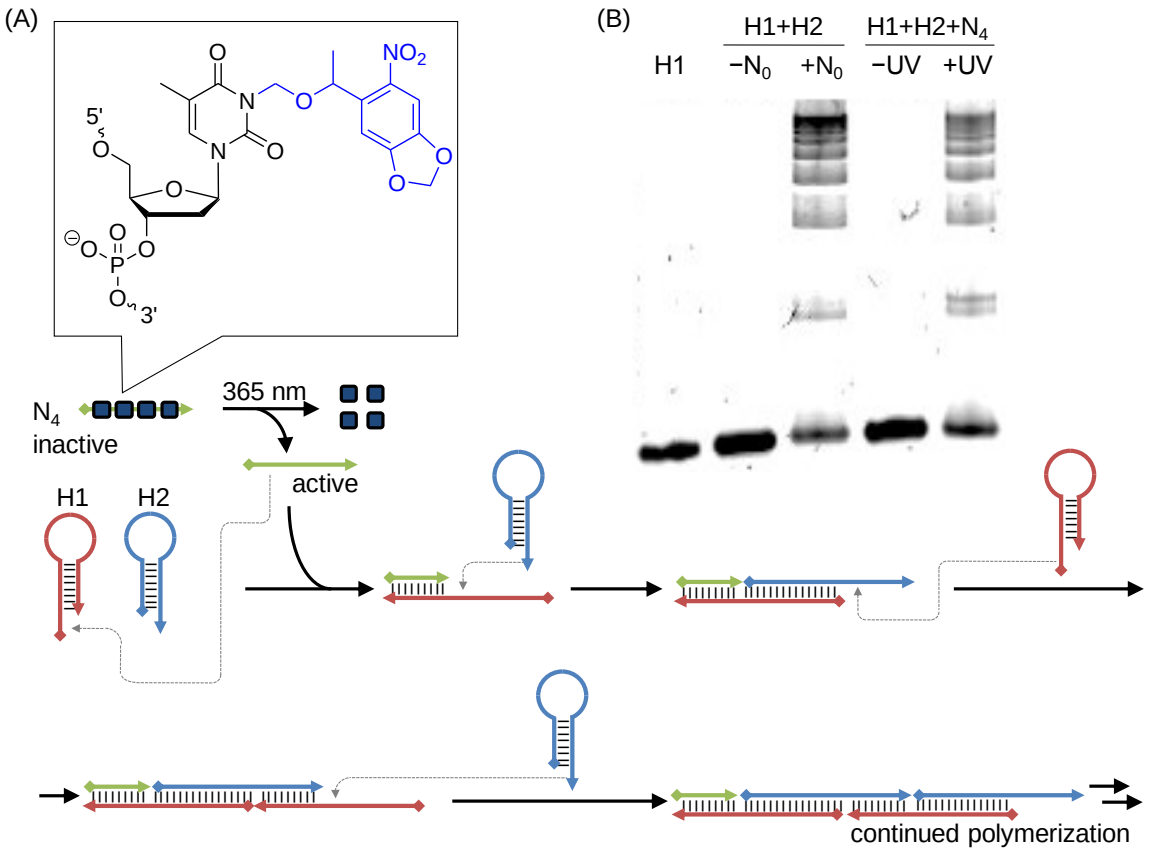


Figure 2.22. (A) Schematic of the hybridization chain reaction with the caged initiator strand N_4. Oligonucleotides are shown as colored lines, and caging groups are represented by blue boxes. Two hairpins (H1 and H2) are metastable until light-triggering of the initiator strand causes the formation of higher molecular weight product strands. (B) PAGE analysis of HCR reactions with noncaged (N_0) and caged initiator (N_4) strands. No background is observed in the absence of UV light. Irradiation of N_4 with UV light triggers the formation of higher molecular weight products similar to the HCR products produced by N_0. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

Previously, HCR has been photochemically controlled with a cleavable backbone linker incorporated into a third hairpin structure.[104] When irradiated, the photolyzed

hairpin would create initiator strands for HCR. However, applying a third hairpin as a blocked initiator strand caused a steady increase in background signal over the course of the reaction. After only 45 min, the background signal was nearly 20% of the highest obtained signal. This is in contrast to the photocaged HCR system described herein, which showed no observable signs of product formation before decaging, even after an overnight reaction (Figure 2.22B and Figure 2.23) and thus show no background leakiness of the system. Additionally, full HCR activation is observed after only 90 s of irradiation (Figure 2.23) compared to 20 min in the case of the previous design.[104]

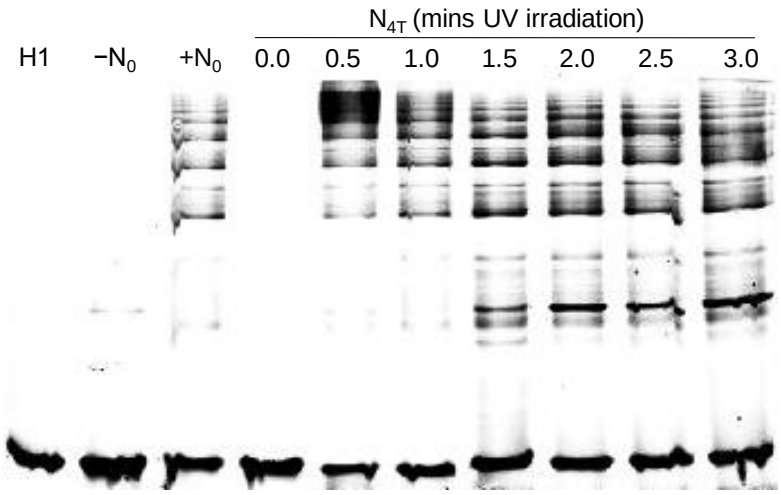
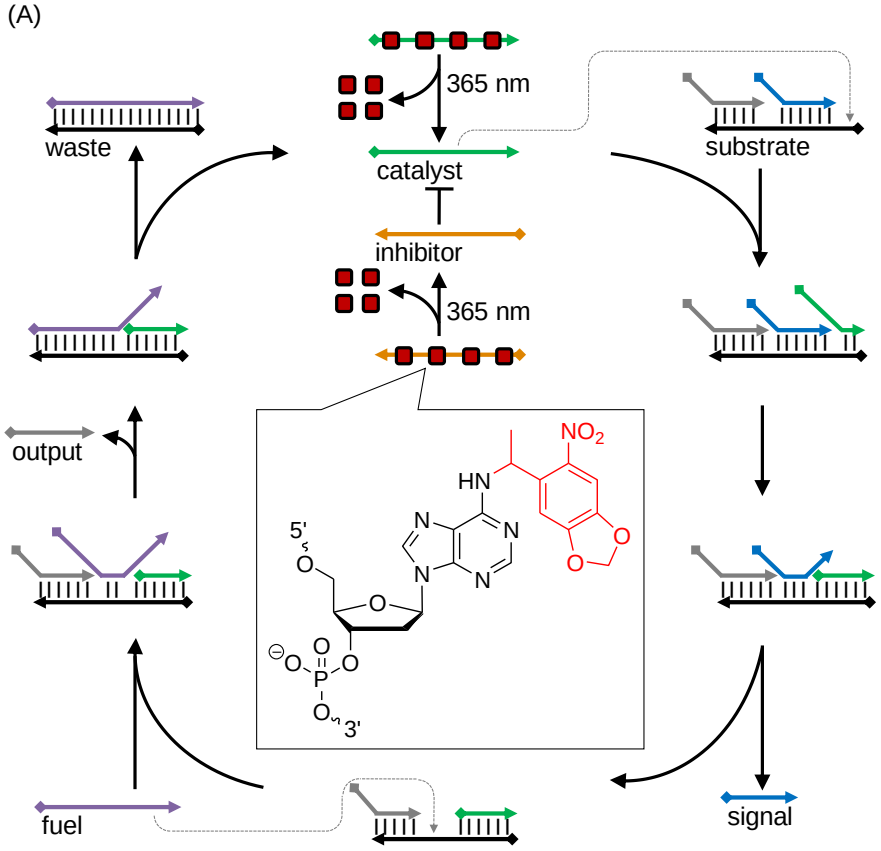


Figure 2.23. Native-PAGE gel (16%) stained with SYBR Gold nucleic acid stain. The caged initiator strand N_4 was irradiated with UV light for different durations. Activation of HCR is observed after only 0.5 minutes of irradiation although 1.5 minutes yields a more complete activation of the HCR. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

A second DNA-based device to achieve signal amplification involves a fuel–catalyst cycle.[20] The cycle begins with binding of the catalyst strand to the substrate complex (duplex containing the substrate, signal, and output strands). After a toehold mediated strand exchange, the catalyst displaces the signal strand revealing a toehold for the fuel strand. Binding of the fuel strand completely removes the output strand and catalyst, which creates a waste duplex. Displacement of the catalyst strand allows the cycle to continue, and the output strand is then able to interact with the reporter gate, releasing the quencher strand from the fluorophore strand. Thus, an increase in fluorescence corresponds to an active cycle. The fuel–catalyst cycle amplification arises from the release of a surplus of signal and output strands from a limited supply of catalyst. If the catalyst strand is caged (as in C_4), the caging groups will prevent activation of the cycle by blocking hybridization of the catalyst strand to the substrate complex. Removal of the caging groups with 365 nm light will create free catalyst strands thereby initiating the cycle (Figure 2.24A). To turn the amplification cycle OFF, an inhibitor strand I_4 was conceived, which is completely complementary to the catalyst thereby blocking its function. When the inhibitor strand is caged, hybridization to the catalyst is prevented and the cycle operates normally. After irradiation, the inhibitor will bind to the catalyst prohibiting continuation of the cycle. In order to generate a fluorescent output signal, a reporter gate can be added (Figure 2.24B and Figure 2.25). The gate will interact with the output strand, releasing the fluorophore strand. The free fluorophore is then able to emit a fluorescent signal. Activation or deactivation of the cycle through photochemical means enables reliable regulation of amplification by precisely optically switching the cycle from either OFF → ON or ON → OFF.

(A)


365 nm
waste
catalyst
substrate
inhibitor
365 nm
output
NO2
HN
5'
N
N
N
N
O
O
O
O
O
P
O
O
3'
fuel
signal

(B)

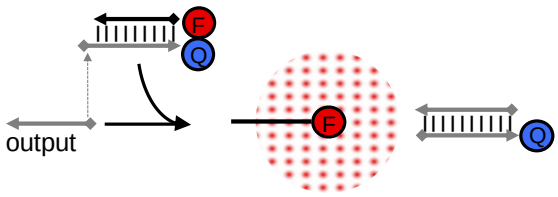

F
Q
output
F
Q

Figure 2.24. Schematic of the (A) fuel–catalyst cycle with caged inhibitor and caged catalyst strands and (B) fluorescence reporter gate. Colored lines represent DNA oligomers and red boxes indicate caging groups. In the absence of caging groups, adding a catalyst strand to the substrate complex and fuel strands will release the signal and output strands while creating a waste duplex. The output strand is able to interact with the reporter gate, releasing a fluorophore strand. F = TAMRA fluorophore, Q = BlackHole quencher-2. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

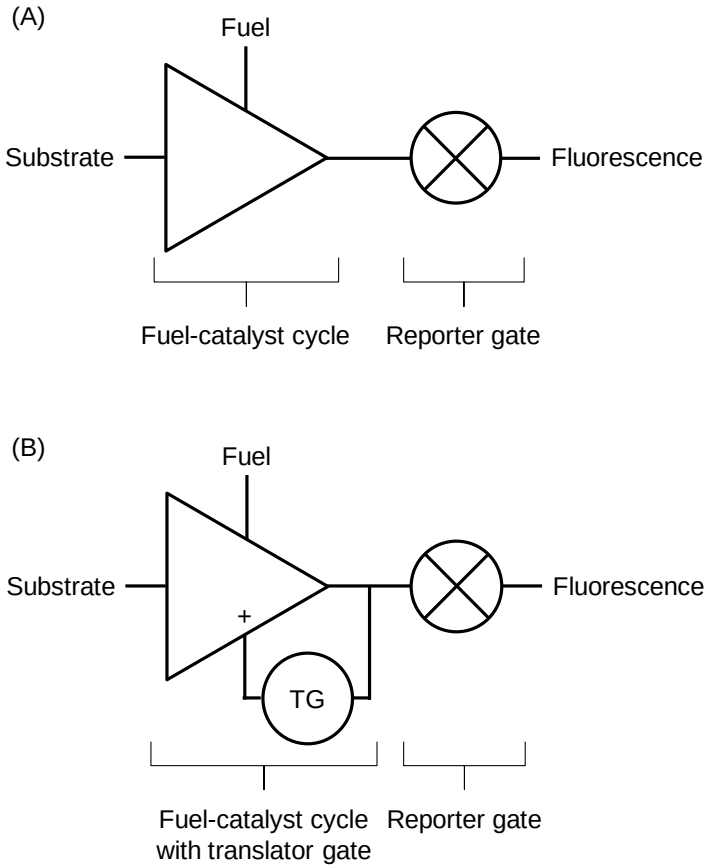


Figure 2.25. Schematics using modified electrical circuit symbols of the (A) fuel-catalyst cycle and (B) fuel-catalyst cycle with translator gate. The triangle represents the fuel-catalyst cycle, the cross-hatched circle represents the reporter gate, and the TG-labeled circle represents the translator gate. A plus sign indicates that the translator gate produces extra catalyst strand and operates as a positive feedback loop. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

Oligonucleotide strands and preformed gate structures were mixed in solution and assessed by gel electrophoresis (Figure 2.26A and Figure 2.27) or fluorescence emission. After replacing C_0 with the caged C_4, no signal was produced, effectively turning the amplification cycle OFF (Figure 2.26B). Removal of the caging groups with UV light restored catalyst activity, generating a signal. Conversely, introduction of I_4 to the substrate complex, fuel strand, and C_0 did not affect normal operation of the cycle, and strand exchange cascades continued to produce signal. However, irradiating the caged inhibitor strand prevented amplification by sequestration of the catalyst strand, which switched the cycle OFF. Thus, the activity of the fuel–catalyst cycle could be photochemically controlled through the use of the caged inhibitor or caged catalyst strand. A fluorescence reporter gate was also used to measure the activity of the fuel–catalyst cycle, and confirmed the switching behavior of the caged oligonucleotides in the presence of UV irradiation (Figure 2.26B and C). More than 5-fold change in fluorescence signal was observed upon light-activation of the caged catalyst or caged inhibitor strand. This corresponds to efficient photoswitching of the amplification cycle, since a similar change was observed for the addition of noncaged catalyst and inhibitor strands.

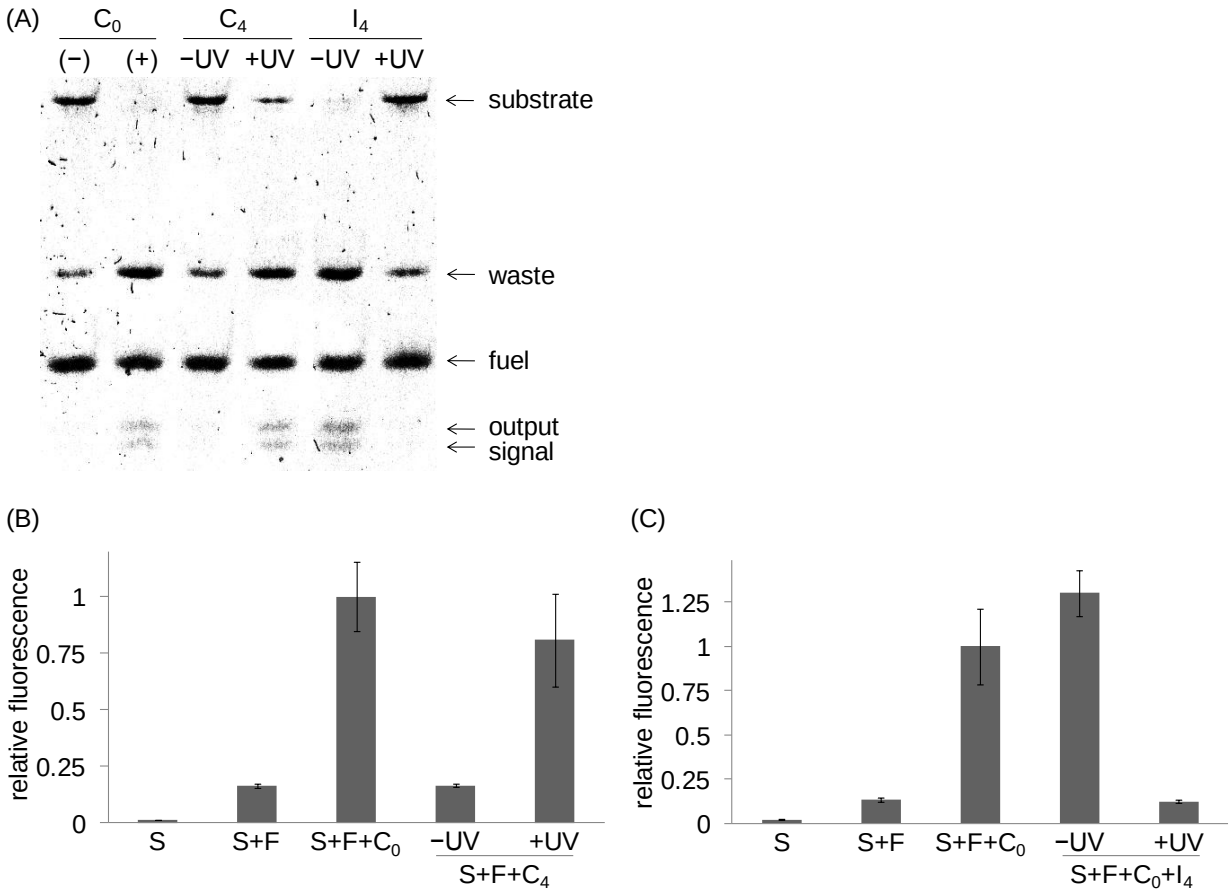


Figure 2.26. Optical OFF → ON and ON → OFF switching of the fuel–catalyst cycle by using caged catalyst and caged inhibitor strands. (A) Native-PAGE gel (16%) stained with SYBR Gold nucleic acid stain. Photochemical control is accomplished with the addition of caged catalyst (C_0) or caged inhibitor (I_4). Irradiation with UV light turns the fuel–catalyst cycle (without reporter gate) ON or OFF using the caged catalyst or caged inhibitor, respectively. Fluorescence quantification for the fuel–catalyst cycle after the addition of (B) caged catalyst (C_4) or (C) caged inhibitor (I_4) using a reporter gate. Samples labeled +UV were irradiated with 365 nm light for 10 min before addition. Additional single letter abbreviations are used for substrate (S) and fuel (F). Excitation and emission wavelengths were 545 and 585 nm, respectively. Error bars represent standard deviations from three independent experiments. Fluorescence was measured after 3 hours. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

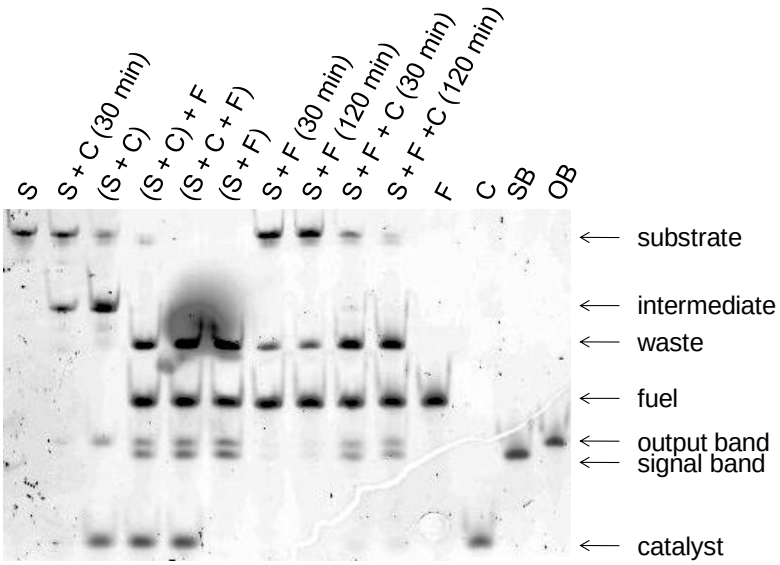


Figure 2.27. Native-PAGE gel (16%) stained with SYBR Gold nucleic acid stain. Non-caged strands were tested to ensure the fuel-catalyst cycle functioned properly. Parentheses indicate the components that were pre-annealed for the given amount of time. Abbreviations for strands and complexes were used: S (substrate complex), C (catalyst), F (fuel), SB (signal strand), and OB (output strand). Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

When fluorescence was measured over time for the fuel–catalyst cycle in the presence of the caged catalyst (Figure 2.28A), a steep increase in the fluorescence signal was observed only after irradiation with UV light. Addition of the caged inhibitor to the cycle did not affect normal output, whereas the decaged inhibitor strand blocked all signal production (Figure 2.28B).

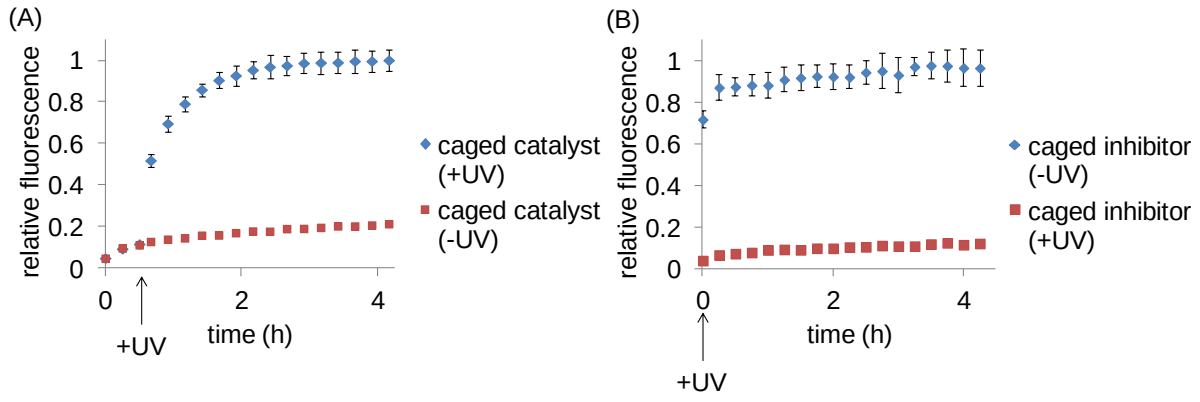


Figure 2.28. Time-course fluorescent measurements for the fuel-catalyst cycle with (A) caged catalyst and (B) caged inhibitor. (A) After 0.5 h, the caged catalyst was irradiated with UV light. (B) Before starting fluorescence recording, the caged inhibitor was irradiated and then added to the fuel-catalyst cycle. Measurements are normalized to the highest value. Error bars represent standard deviations of three independent experiments. Fluorescence measured over 4 hours. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

With successful optical control of a solution-based amplification cycle, the system was next transitioned to semisolid media. Amplification of a signal inside a semisolid can expand the applications of DNA computation systems beyond solution-based devices, since it greatly constrains diffusion thereby enabling spatial control. Additionally, a solid structure creates a modular unit that could facilitate the physical separation of components in a DNA cascade. Much like electric components, embedded DNA computation devices can act as stand-alone elements of a larger circuit. To demonstrate spatial control of the fuel–catalyst cycle using the caged catalyst C_4, oligonucleotide components were embedded into low-melt agarose. Spatially restricted illumination of the gel with a fiber optics probe (Figure 2.29A) or through a

mask (Figure 2.29B) enabled the amplification cycle to be activated in specific and independent regions. Although masks allowed for customizable shapes, the edges were not as well-defined as those produced by an LED fiber optic light source. To create a gradient effect, the gel was irradiated for different time intervals (Figure 2.29C). The gradient demonstrated how signal intensity could be tuned by varying the applied UV irradiation. Variable light intensities will create diverse populations of signal intensity, adding depth to the recognition of an OFF → ON transition. The ability to create signals in any desired pattern using optical regulation could allow for better control in investigating or modeling biological events.

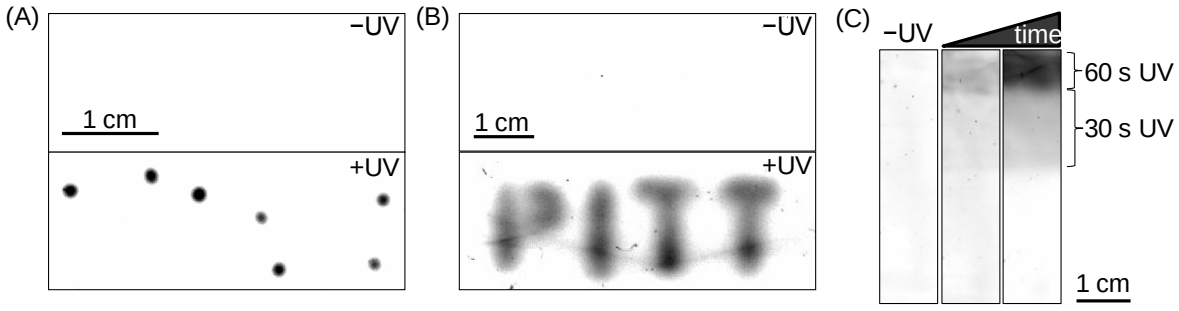


Figure 2.29. Spatial control of the fuel–catalyst cycle by using low-melt agarose gels (1–2%) embedded with the substrate complex, fuel strand, and caged catalyst C_4. (A) A fiber optics probe was used to irradiate (1 min, 365 nm) spatially independent areas in the shape of the big dipper constellation. (B) Illumination through a mask was used to pattern “PITT”. (C) Multiple irradiation time intervals were used to create a gradient effect, tuning activation of the catalyst cycle. All fluorescent imaging was performed using green LED excitation, and emitted light was detected at 580–630 nm. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

In addition to patterning, the system can be engineered to allow for signal propagation from a defined start point through a gel. A translator gate was added to the fuel–catalyst system to produce a second catalyst strand from the signal strand (Figure 2.30).

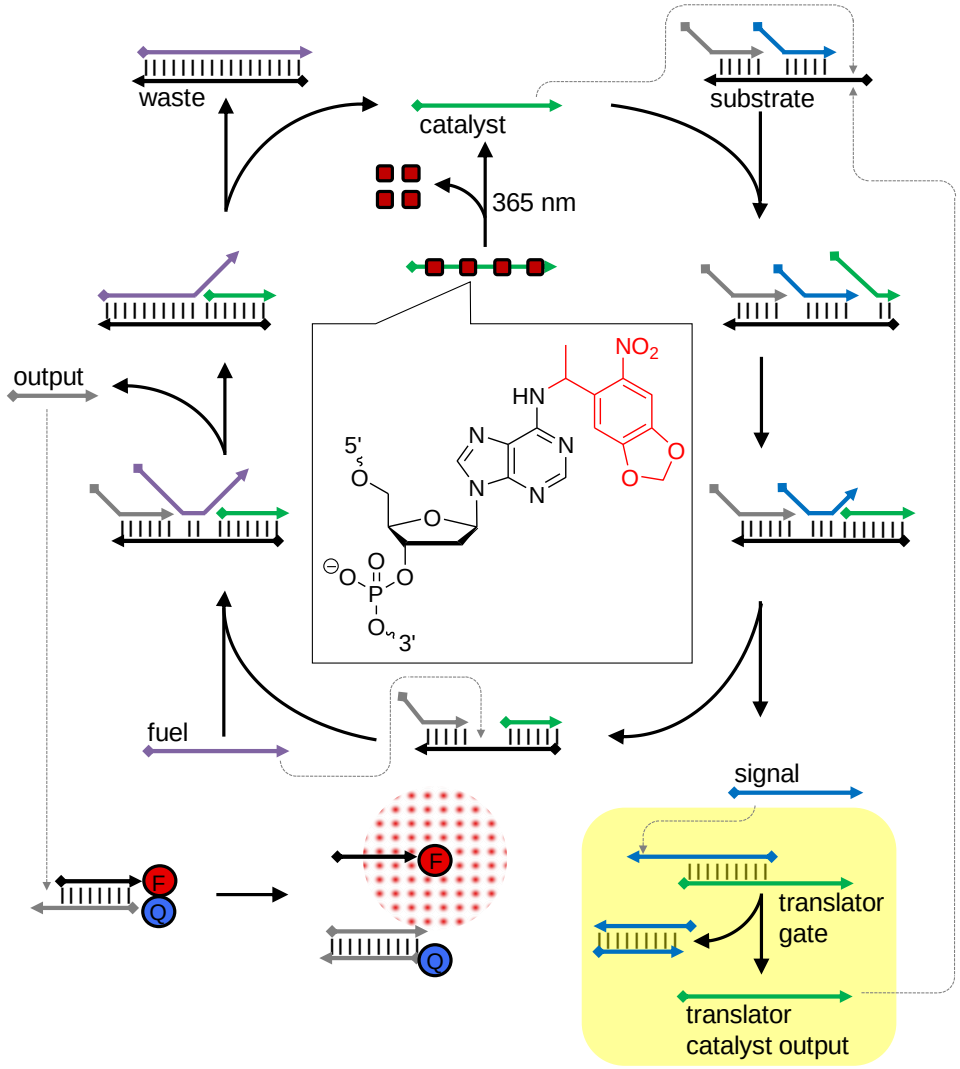


Figure 2.30. Schematic of the fuel-catalyst cycle modified with a translator gate (highlighted in yellow). The translator gate produces an extra catalyst strand from the signal strand, which enables the cycle to act autocatalytically. Colored lines represent DNA oligomers and red boxes indicate caging groups. F = TAMRA fluorophore, Q = BlackHole-2 quencher. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

When the cycle is triggered through light exposure, the signal propagates from the point of irradiation. Unlike the cycle without the translator gate, which expands minimally after an hour through simple diffusion of the decaged catalyst (Figure 2.31), the signal strand can activate new catalyst cycles in neighboring regions by interacting with the new translator gate, thereby counteracting dilution of the catalyst and instead facilitating signal propagation. In regions where the signal strand is not available yet, the translator gate will remain hybridized as a duplex and will not be able to release additional catalyst. When the translator gate is added but the circuit is not irradiated, the signal only increases minimally. Thus, localized UV irradiation generates a signal that continues to travel in the shape of the gel. Movement of the signal along a specific, predefined path is analogous to a chemical wire. Similar to electrical circuits, this may allow for the communication of spatially separated DNA computation devices with each other.

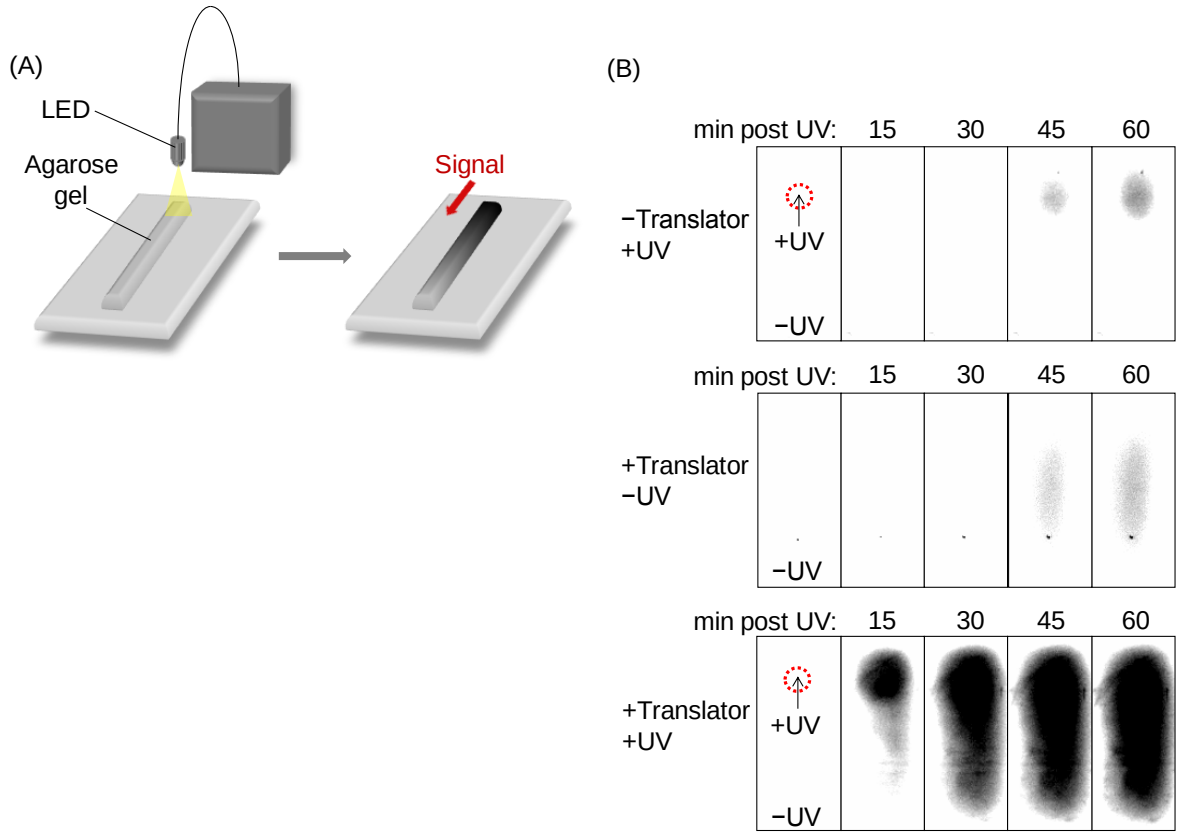


Figure 2.31. (A) Schematic of the spatial control setup to demonstrate signal propagation through an agarose gel. (B) Three 1.0% agarose gels with or without added translator gates were irradiated with 365 nm light in a single spot. The presence of the translator gate led to a more intense and rapidly expanding signal compared to the cycle without translator gate. Minimal background is observed in the gel containing the translator gate in the absence of UV exposure. The dotted red circle indicates the area of UV irradiation. The addition of translator gate also enables the signal to propagate in a single direction. Numbers above the images represent time post irradiation in minutes. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

Activation of spatially defined areas at multiple time intervals demonstrated spatiotemporal activation of the fuel–catalyst cycle (Figure 2.32). Three separate signals

appear relatively quickly after irradiation with a fiber optics probe. Temporal control is advantageous in order to tune the cycle to external events. A signal output is no longer restricted to a single location or pattern and can be changed manually to desired specifications with minimal irradiation. Thus, spatial and temporal control of the gel-based fuel–catalyst cycle offers enhanced flexibility in controlling signal amplification.

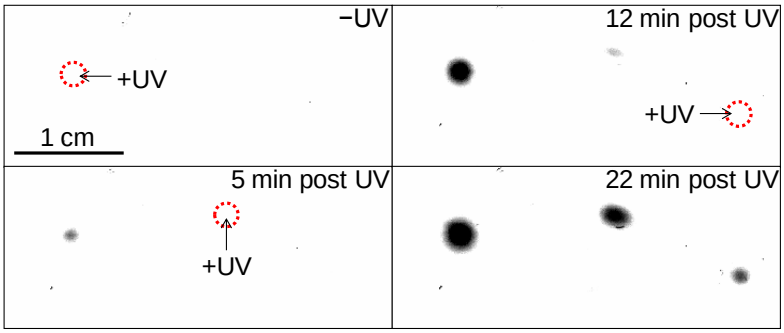


Figure 2.32. Spatio-temporal control of the fuel-catalyst cycle in a semi-solid. A 1% low-melt agarose gel containing the substrate complex, fuel strand, and C_4 was irradiated with 365 nm fiber optics probe in a spatial and temporal manner. Three spots appear after light activation at different times. Red circles indicate the area of UV irradiation. Adapted with permission from Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS Synth. Biol.* 2015 ASAP. Copyright 2015 American Chemical Society.

In conclusion, modification of oligonucleotides with nucleobase-caging groups enabled optical control over HCR and a fuel–catalyst cycle, DNA devices that allow for a DNA signal amplification. Crucial DNA strands were modified with photocleavable caging groups to optically control the individual reaction circuits. For HCR, a caged initiator strand was synthesized. Upon decaging, the initiator strand was able to interact

with the hairpins, causing amplification through DNA strand polymerization. Only a low concentration of initiator is necessary to start the HCR, which can be achieved through minimal light exposure. A fuel–catalyst cycle was also successfully optically switched from OFF → ON or ON → OFF by using either a caged catalyst or a caged inhibitor, respectively. To prevent DNA/DNA hybridization in the absence of illumination, and thus to control the cycle with light, four evenly spaced caging groups were added to the DNA strands. Quantification of the output was made possible with a reporter gate generating a fluorescent signal. Conducting light-activation in a semisolid containing the DNA circuits led to remarkable spatial and temporal control. Localized illumination of the gel embedded with the DNA circuits enabled triggering of signal amplification in customizable patterns as well as tunable gradients. Furthermore, the fuel–catalyst cycle was modified with an additional translator gate and converted into a light-triggered autocatalytic cycle that enabled directional signal propagation in a preshaped agarose gel. In addition to spatial activation, the amplification cycle was also temporally activated. For each cycle, light acts as a dependable switch for triggering computational events as it is tunable and noninvasive. Caged oligonucleotides represent a modular framework that can be easily fitted to existing DNA-based architectures.

The methodology development reported here may find application in more complex DNA computation circuits that contain an output amplification. Temporal control enables precise sequencing of gate and subnetwork functions and, in the case of temporal control of an amplification cycle, allows for upstream circuit completion before output amplification, thereby potentially reducing the background signal by preventing premature activation. In addition, temporal control of DNA circuits enables

modification of any system (e.g., drug treatment of cells) that is interfaced with a DNA computation network before circuit activation. Light is an excellent external control element that can be used as a switch with very high temporal and spatial resolution without the need for other physical or chemical alterations (e.g., injections). The application of light-activated DNA circuits and amplification devices is especially advantageous in systems that do not allow for later addition of oligonucleotide triggers, for example, in semisolid media as shown here or in enclosed biological environments, such as organisms.

2.5.2 Experimental

Oligonucleotide Synthesis

DNA synthesis was performed by James Hemphill (Deiters Lab) using an Applied Biosystems (Foster City, CA) model 394 automated DNA/RNA synthesizer and standard β-cyanoethyl phosphoramidite chemistry. The caged oligonucleotides were synthesized on a 40 nmol scale, with solid-phase supports obtained from Glen Research (Sterling, VA). Reagents for automated DNA synthesis were also obtained from Glen Research. Standard synthesis cycles provided by Applied Biosystems were used for all normal bases with 25 s coupling times. The coupling time was increased to 2 min for incorporation of caged thymidine modified phosphoramidite. The NPOM-caged thymidine phosphoramidite[33] was resuspended in anhydrous acetonitrile to a concentration of 0.1 M. Unmodified oligonucleotides were purchased from Integrated DNA Technologies (IDT). Fluorophore and quencher labeled oligos were purchased from AlphaDNA.

HCR with Caged Initiator

Hairpins H1 (0.3 μL of a 100 μM stock) or H2 (0.9 μL of a 100 μM stock) were added to 8.7 μL water and 1 μL HCR buffer (50 mM Na_2HPO_4, 0.5 M NaCl, pH 6.8) and annealed (95 to 4 °C, 3 μM) independently (see seesaw gate purification for a more detailed protocol). Hairpins H1 (3 μL of 3 μM, 1 μM) and H2 (3 μL of 3 μM, 1 μM) were combined with initiator strand (3 μL of 1 μM, 0.33 μM) and allowed to react (overnight, room temperature). Reactions were analyzed by native-PAGE (16%, 100 V, 1 h), stained (SYBR Gold nucleic acid stain, LifeTechnologies), and imaged (General Electric Typhoon FLA 7000, SYBR fluorescence setting).

Assembly of Substrate Complex, Reporter Gate, and Translator Gate

The signal (30 μL of a 100 μM stock, 30 μM), output (30 μL of a 100 μM stock, 30 μM), and substrate (30 μL of a 100 μM stock, 30 μM) oligonucleotide strands (Table 2.5) were combined with of TE/Mg^{2+} buffer (10 μL of 10X stock, 100 mM Tris, 10 mM EDTA, 125 mM $MgCl_2$, pH 8) and annealed slowly (95 to 4 °C, see seesaw gate purification for more detailed protocol). The solution was then run on a 16% native-PAGE gel (100 V, 1 h). Bands were illuminated on a TLC plate with a hand-held UV light and excised with a razor blade (about 1000 mm^3). The substrate complex was eluted (using 1 mL of TE/ Mg^{2+} buffer) from the gel on a shaker (room temperature, overnight) and quantified using Beer's law and the calculated extinction coefficient (using nearest-neighbor models and IDT biophysics website).[105] The reporter gate (RG) was made using the same procedure, but with the R_F and R_Q oligonucleotides; similarly, the translator gate

used the oligonucleotides T_{base} and T_{output}. Final concentrations from the elution were about 10 μM substrate complex, reporter gate, and translator gate. The solutions were separated from the gel slices after overnight elution by pipette into a new tube, stored at −20 °C, and used as-is.

Gel Analysis of Triggering the Fuel–Catalyst Cycle with Caged Oligonucleotides

Substrate complex (S, 0.4 μL of 10 μM, 200 nM), fuel strand (F, 0.4 μL of 10 μM, 200 nM), and catalyst (C, 0.4 μL of a 1 μM stock, 20 nM) were added to TE/Mg^{2+} buffer (2 μL of a 10X stock) and water (16.8 μL). When appropriate, the caged catalyst strand (C_4, 0.4 μL of a 1 μM stock, 20 nM), inhibitor (I, 0.4 μL of ab10 μM stock, 100 nM) or caged inhibitor strand (I_4, 2 μL of a 1 μM stock, 100 nM) was added and decaged (365 nm light, 10 min) or kept in the dark. After the fuel–catalyst cycle was completed (room temperature, overnight), the reactions were analyzed by 16% native-PAGE (100 V, 1 h), stained (SYBR Gold nucleic acid stain), and imaged (General Electric Typhoon FLA 7000, SYBR fluorescence setting, SYBR fluorescence setting).

Fluorescence Quantification

Substrate complex (S, 25 μL of a 6.4 μM stock, 800 nM), reporter gate (RG, 8.1 μL of a 4.9 μM stock, 200 nM), and fuel strand (F, 17 μL of a 10 μM stock, 850 nM) were added to TE/Mg^{2+} buffer (20 μL of a 10X stock) and water (129.9 μL). Where necessary, the catalyst (C, 16 μL of a 1 μM stock, 80 nM), caged catalyst (C_4, 40 μL of a 1 μM stock, 200 nM), inhibitor (I, 4.8 μL of a 10 μM stock, 240 nM), or caged inhibitor strand (I_4, 48 μL of a 1 μM stock, 240 nM) were added to TE/Mg^{2+} buffer (20 μL of a 10X stock).

Water was added to a total volume of 200 µL for each condition. If irradiated with UV light, the caged oligo was transferred into a plastic cuvette and decaged (365 nm light, 10 min, transilluminator, cooled with an ice pack). To three wells of a black-walled and clear bottom well plate, 50 µL of the reaction was added. The fuel–catalyst cycle was allowed to react (room temperature, 3 h) and quantified by a plate reader (excitation / emission 545 / 585 nm, Tecan Infinite M1000 Pro). Fluorescence measurements were recorded every 15 min (room temperature) in a black walled 96-well plate. The caged inhibitor was irradiated (365 nm, 10 min) before measuring fluorescence, whereas the caged catalyst was irradiated after recording fluorescence for 30 min.

Fuel–Catalyst Spatial Control

Spatial control was performed in low-melt agarose gel (1–2%) prepared with TE/Mg^{2+} buffer and solidified on a glass microscope slide. Before solidification, the substrate complex (S, 6.8 µL of a 14.7 µM stock, 200 nM), fuel strand (F, 10 µL of a 10 µM stock, 200 nM), reporter gate (RG, 8.8 µL of a 11.4 µM stock, 50 nM), caged catalyst (C_4, 25 µL of a 1 µM stock, 50 nM), and if necessary, translator gate (TG, 2.1 µL of a 9.5 µM stock, 20 nM) were added to TE/Mg^{2+} buffer (50 µL of a 10X stock) and water (397.3 µL). This solution was added to a 2X concentrated low-melt agarose solution (500 µL of a 2-4% stock) so that the final gel percentage was 1–2% in a 1 mL total volume solution. Agarose solution was also allowed to cool to almost room temperature before mixing to prevent denaturating the preformed DNA/DNA duplexes. The gel was irradiated with UV light (0.5–1 min) through a patterned foil mask on a transilluminator (UVP high performance UV transilluminator) or with a UV fiber optics

probe (1–5 min). The gel was then imaged for fluorescence (green excitation LEDs, 605/50 filter, ChemiDoc MP Imaging System).

Concentrations for experiments shown in Figure 2.29 and Figure 2.32: substrate complex (S, 6.8 μL of a 14.7 μM stock, 200 nM), fuel strand (F, 10 μL of a 10 μM stock, 200 nM.), reporter gate (RG, 8.8 μL of a 11.4 μM stock, 50 nM), and caged catalyst (C_4, 25 μL of a 1 μM stock, 50 nM) were added to TE/Mg^{2+} buffer (50 μL of a 10X stock) and water (399.4 μL). Concentrations for experiments shown in Figure 2.31: substrate complex (S, 6.8 μL of a 14.7 μM stock, 200 nM), fuel strand (F, 10 μL of a 10 μM stock, 200 nM), reporter gate (RG, 35.2 μL of a 11.4 μM stock, 50 nM), caged catalyst (C_4, 25 μL of a 1 μM stock, 50 nM), and (TG, 1 μL of a 9.5 μM stock, 20 nM) were added to TE/Mg^{2+} buffer (50 μL of a 10X stock) and water (372 μL).

Table 2.5. Sequences of oligonucleotides used in HCR and the fuel-catalyst cycle. Underlined sequences indicate toe-hold regions, while caged nucleotides are bolded and indicated by an asterisk (*). F = tetramethylrhodamine fluorophore, Q = black hole quencher 2. Sequences for the fuel-catalyst system were obtained from Winfree,[20] and sequences for the HCR were adapted from Pierce.[19]

Strand Name	Abbreviation	Sequence (5’ → 3’)
Hairpin 1	H1	TATATACACGCCGAATCCTAGACTCAAAGTAGTC TAGGATTCGGCGTG
Hairpin 2	H2	AGTCTAGGATTCGGCGTGGGTTAACACGCCGAA TCCTAGACTTATATA
Initiator	N	AGTCTAGGATTCGGCGTGTATATA
Caged initiator	N_4	AGTCT*AGGATT*CGGCGT*GTATAT*A

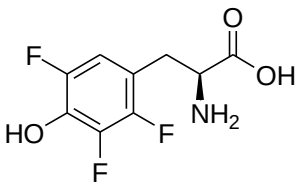
Table 2.5 (continued)

Signal	S_{signal}	CCACATACATCATATTCCCTCATTCAATACCCTACG
Output	S_{output}	CTTTCCTACACCTACGTCTCCAACTAACTTACGG
Substrate base	S_{base}	TGGAGACGTAGGGTATTGAATGAGGGCCGTAAGTTAGTTGGAGACGTAGG
Fuel	F	CCTACGTCTCCAACTAACTTACGGCCCTCATTCAATACCCTACG
Catalyst	C	CATTCAATACCCTACGTCTCCA
Caged catalyst	C_4	CA*TTCA*ATA*CCCTA*CGTCTCCA
Inhibitor	I	TGGAGACGTAGGGTATTGAATG
Caged inhibitor	I_4	TGGA*GACGTA*GGGTA*TTGAA*TG
Reporter fluorophore	R_F	F-CTTTCCTACACCTACG
Reporter quencher	R_Q	TGGAGACGTAGGTGTAGGAAAG-Q
Translator output	T_{output}	CATTCAATACCCTACGTCTCCACCACATACATCATATTCCC
Translator base	T_{base}	GAATGAGGGAATATGATGTATGTGGTGGAGA

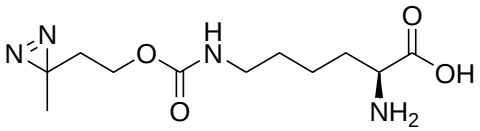
3.0 UNNATURAL AMINO ACID MUTAGENESIS

Proteins are naturally comprised of the twenty canonical amino acids. Thus, the available functional groups are limited and ubiquitous among all proteins. Expanding the genetic code to incorporate unnatural amino acids (UAAs) allows for further investigation into the specific roles and functions of proteins, such as mimicking post translational modifications[106] or through protein labeling.[107] There have been many UAAs incorporated,[108] including **1** and **2** (Figure 3.1). Previously, 2,3,5-trifluorotyrosine, **1**, was incorporated into ribonucleotide reductase.[109] The fluorinated amino acid was used as a spectroscopic handle in EPR experiments to study tyrosyl radicals. Similarly, **1** was incorporated into myoglobin, which was used as a model oxidase protein.[110] Oxidase activity increased due to incorporation of fluorotyrosines at position 33 in the protein. The increase in activity was attributed to the lower pKa of **1** (pKa 6.4) compared to tyrosine (pKa 9.9). In Section 3.1, amino acid **1** was enzymatically synthesized and used to study abasic site bypass in the DNA polymerase KlenTaq. The diazirine lysine **2** has previsouly been incorporated into glutathione S-transferase (GST) to investigate protein-protein crosslinking.[111] A structurally similar diazirine lysine amino acid was synthesized and incorporated into HdeA, an acid-protection chaperone protein, to identify protein-binding partners of HdeA through protein-protein crosslinking under

acidic stress conditions.[112] In Section 3.2, **2** was incorporated into an RNA-binding protein Hfq and used for UV-induced protein-RNA crosslinking.



1



2

Figure 3.1. UAAs discussed in this chapter, which were incorporated into proteins by unnatural amino acid mutagenesis.

Incorporation of UAAs into proteins requires specialized methods, which have a wide array of advantages, disadvantages, and applications. Native chemical ligation (NCL) is a method of creating proteins from smaller synthetic peptides.[113] NCL works by reacting a C-terminal α-thioester of one peptide with a N-terminal cysteine residue on a second peptide. After an S → N acyl rearrangement, an amide bond is formed, creating a native protein backbone. To incorporate a UAA into the final protein, the UAA must be included in one of the smaller synthetic peptides. Major disadvantages of NCL include limits on overall protein size and the need to ensure proper folding of proteins for correct function. To form larger proteins from NCL, multiple NCL reactions may be required. An extension of NCL is expressed protein ligation (EPL), which uses a combination of molecular cloning and NCL to fuse a peptide to a recombinant protein.[114] The first step of EPL involves cloning the gene of interest to the N-terminus of an intein. This intein naturally forms a thioester intermediate through an N → S acyl shift. Addition of

thiophenol cleaves the recombinant protein from the intein to form an α-thioester. The synthetic peptide can then be ligated to the recombinant protein through NCL. Another incorporation method utilizes auxotrophic strains of bacteria.[115] If the structure of the UAA is comparable to a canonical amino acid, the endogenous cellular machinery may be able to integrate the UAA naturally. Biosynthetic methods have been developed, which microinject or transfect tRNAs chemically acylated with the UAA into cells.[116] However, without continuous addition of the chemically aminoacylated suppressor tRNAs, the overall amount of unnatural protein is limited by the amount of tRNA.[117] The methods described herein utilize a TAG (amber) stop codon and corresponding suppressor tRNA/aminoacyl tRNA synthetase (aaRS) pair (Figure 3.2).[118] The TAG stop codon is recognized by a $tRNA_{CUA}$, which has been genetically aminoacylated by a mutant aaRS with the UAA. Therefore, the use of an amber stop codon allows for site specific incorporation of UAAs into a protein.

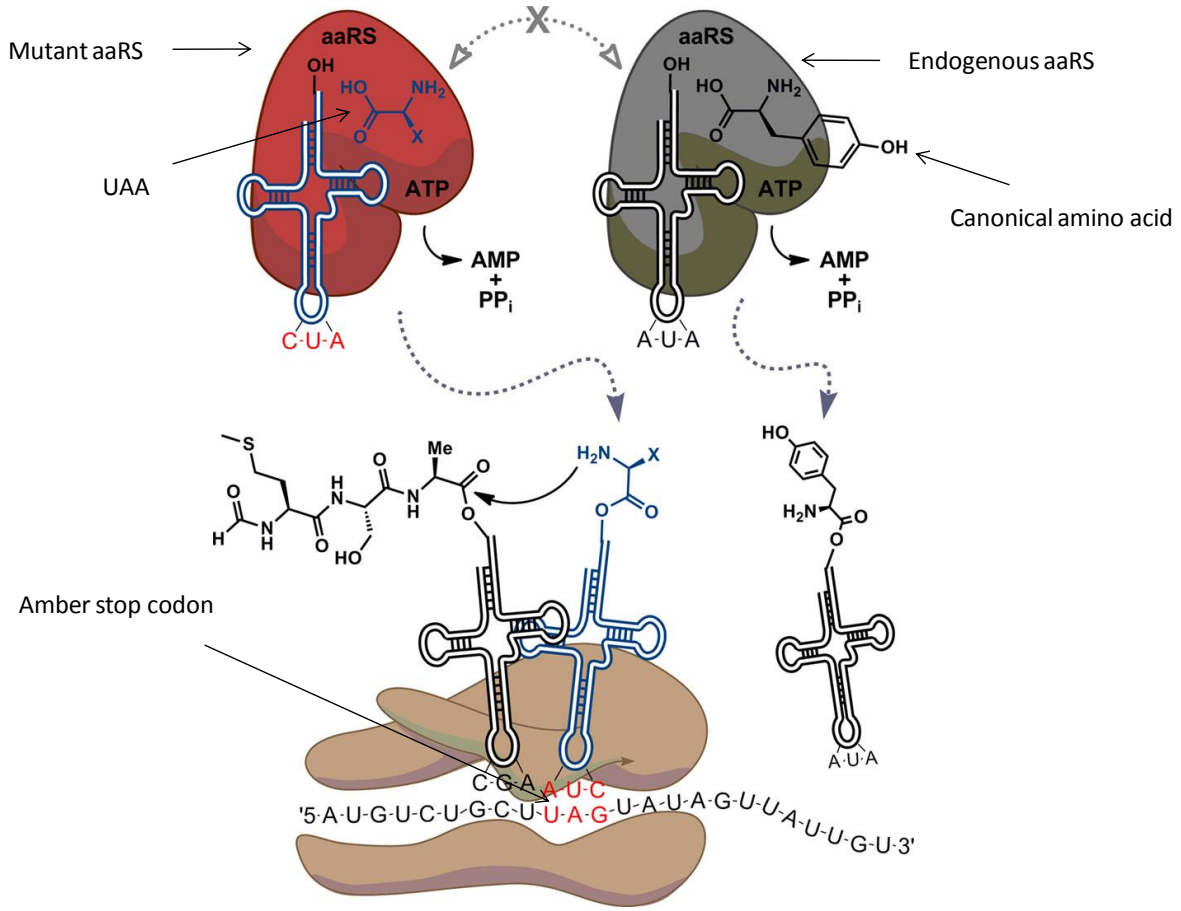


Figure 3.2. An endogenous synthetase (right) aminoacylates a $tRNA_{AUA}$ with a canonical amino acid while the mutant synthetase (left) aminoacylates a $tRNA_{CUA}$ with the UAA. After the ribosome assembles around the mRNA, the aminoacylated tRNA recognizes and binds to the complementary codon. The amino acid is then added to the C-terminus of the nascent peptide chain. Importantly, the peptide is not truncated at the amber stop codon (UAG) in the presence of aminoacylated $tRNA_{CUA}$.[119] This research was originally published in The Journal of Biological Chemistry. Young, T.; Schultz, P., Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. Journal of Biological Chemistry 2010, 285:11039-11044. © the American Society for Biochemistry and Molecular Biology.

Bioorthogonality is an important issue in the incorporation of unnatural amino acids. To be bioorthogonal, an aaRS/tRNA pair must not interfere or disrupt natural cellular processes, including no aminoacylation of $tRNA_{CUA}$ with canonical amino acids

and no recognition of endogenous tRNAs by the mutant aaRS.[120] Additionally, the suppressor tRNA must only recognize the TAG amber stop codon and must be only aminoacylated with the UAA. Therefore, the aminoacylation must be performed by the complementary mutant aaRS, which has been evolved to accept a specific UAA. If bioorthogonality is not properly achieved and the suppressor tRNA is aminoacylated with another amino acid, then a dysfunctional mixture of full length protein populations will be created.[121] To prevent misaminoacylation, a tRNA/aaRS pair which naturally accepts a non-canonical amino acid is used. One well studied example is the pyrolysyl tRNA/aaRS pair from archaebacterium *Methanosarcina barkeri*.[108] This pair naturally incorporates an amino acid that is not present in bacterial, yeast, or mammalian cells, pyrrolysine, making it well suited as a starting point in developing completely bioorthogonal machinery. To obtain orthogonal tRNA/aaRS pairs, a double sieve selection is performed.[122,123] The first round involves a negative selection, which removes endogenously aminoacylated tRNAs through the suppression of stop codons in a toxic barnase gene. A positive selection is then performed to select for the tRNAs that are aminoacylated by the cognate aaRS by suppressing stop codons in a beneficial chloramphenicol acetyl transferase gene.

Construction of the pEVOL system was motivated by the desire to increase unnatural amino acid incorporation efficiency and to standardize the unnatural amino acid mutagenesis systems. Two aaRS^{tyr} copies were added to the pEVOL plasmid, one controlled by a *glnS* constitutive promoter and another under the inducible *pBAD* promoter.[124] The *glnS* promoter enabled basal level aaRS expression until induction of gene expression. After induction of gene expression, a larger supply of aaRS was

maintained by the *pBAD* promoter. Although previous results suggested an increase in incorporation efficiency with polycistronic $tRNA_{CUA}$,[125] no effect was observed with extra copies.[124] Additional mutations were made to the $aaRS^{tyr}$ and $tRNA_{CUA}$ to boost incorporation efficiency.[126,127] In a direct comparison to other unnatural amino acid mutagenesis systems (pBK[128] and pSUP[129]), the optimized pEVOL system was able to increase incorporation by 250%. Additionally, growth rates of DH10B and BL21 *E. coli* strains were not severely affected by the pEVOL system. Many amino acids were incorporated into GFP using the pEVOL system containing the appropriate aaRS. Overall, the pEVOL plasmid offered increased unnatural amino acid incorporation compared to other similar systems, and was easily modified to accept other amino acids.

3.1 MODULATING THE PKA OF A TYROSINE IN KLENTAQ DNA POLYMERASE BY INCORPORATION OF AN UNNATURAL AMINO ACID

3.1.1 Results and Discussion

Material was used with permission from ChemBioChem.[130] Protein expression and purification were performed by Alex Prokup in the Deiters lab. Protein mass spectrometry and nucleotide incorporation assays were performed by Nina Blatter in the Marx lab.

DNA is continuously damaged by endogenous and exogenous agents. Under physiological conditions the most frequent type of DNA damage is the abasic site, which

results from spontaneous hydrolysis of the bond that connects the sugar to the nucleobase (Figure 3.3A).[131] It has been estimated that approximately 10000 abasic sites are formed in a human cell each day.[132, 133] These lesions can stall DNA synthesis, as they present a major challenge for replicative DNA polymerases.[134, 135] Furthermore, they are noninstructive, as genetic information is lost by cleavage of the nucleobase.[132, 136] However, members of DNA polymerase families A and B were observed to preferentially insert dATP opposite an abasic site, a phenomenon termed the "A-rule".[132, 136, 137]

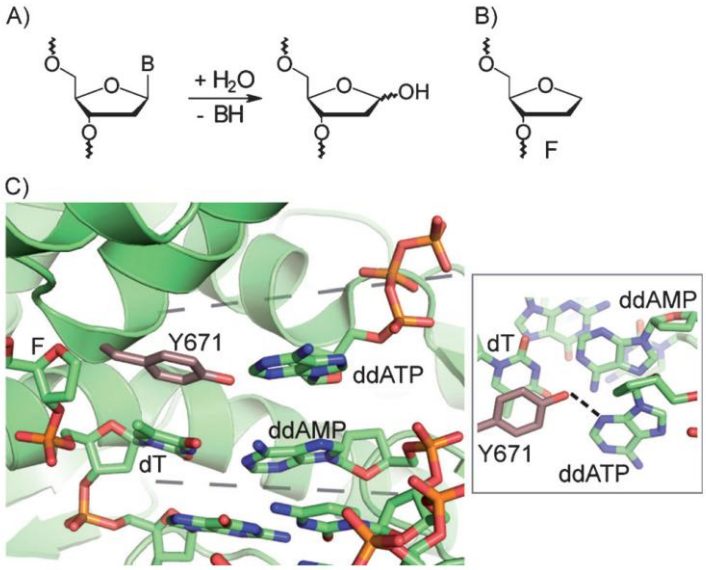


Figure 3.3. (A) Hydrolysis of the glycosidic bond results in formation of an abasic site (B: base). (B) Chemical structure of the abasic-site analogue F. (C) Crystal structure of the active site of wild-type *KlenTaq* with ddATP opposite the abasic site F (PDB ID: 3LWL).[138] The penultimate base pair, the incoming ddATP, and Y671 are shown. Detail shows hydrogen bond between Y671 and N3 of the incoming ddATP (dashed line). Reprinted with permission from Blatter, N. *et al. ChemBioChem* 2014, *15* (12), 1735-7. Reprinted with permission from Modulating the pKa of a tyrosine in KlenTaq DNA polymerase that is crucial for abasic site bypass by in vivo incorporation of a non-canonical amino acid. Blatter, N.; Prokup, A.; Deiters, A.; Marx, A. *ChemBioChem* 15 (12). Copyright © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

In vitro studies of *KlenTaq* DNA polymerase (family A polymerase) showed that this enzyme also follows the A-rule in the presence of the stabilized tetrahydrofuran abasic-site analogue F (Figure 3.3B).[138-139] Crystal structures of this enzyme in complex with a template containing an abasic-site analogue F showed an "amino acid templating" mechanism, facilitated by a tyrosine residue at position 671 (Y671).[138-139] The tyrosine residue mimics the shape and size of a six-membered pyrimidine nucleobase in the template strand, thus directing purine incorporation opposite the abasic site, by providing an optimal geometric fit for the active site (Figure 3.3C). Furthermore, structural and functional data indicate a hydrogen bond interaction between the incoming nucleotide at N3 of adenine and the hydroxy group of Y671 (Figure 3.3C).[138]

The importance of the hydrogen bond between Y671 and the incoming nucleotide was demonstrated by mutational analysis as well as by substituting the N3 of adenine with a nonpolar CH-group.[138] Herein, we studied this hydrogen bond by modulating the pK_a value of the Y671 hydroxy group. Fluorinated tyrosine analogues

have already been employed as valuable tools for the investigation of acid–base catalyzed reactions[140] and of biologically generated tyrosine radicals used for catalysis in proteins.[109, 141] Furthermore, fluorotyrosine analogues provide an opportunity to probe hydrogen bonding interaction networks in proteins.[140, 142] The fluorine atom combines isosteric properties comparable to those of hydrogen with increased electronegativity, thereby resulting in lower pK_a values for the fluorinated analogues.[140, 109, 141] Thus, the correct choice of a fluorotyrosine analogue in combination with an appropriate reaction pH allows the protonation state of the tyrosine hydroxy group to be modulated. Site-specific incorporation of fluorotyrosines into proteins offers the possibility to elucidate and validate hydrogen bonding interactions in proteins, for example, in the polymerase active site where this hydroxy group is involved.

Reported herein is the site-specific incorporation of **1** into *KlenTaq* DNA polymerase at position 671 and its application to validate the impact of a specific hydrogen bond in abasic site bypass. **1** was employed because its pK_a value (pK_a=6.4) is much lower than that of natural tyrosine (pK_a=9.9).[141b]

The amino acid, **1**, was enzymatically synthesized from pyruvate, ammonia, and 2,3,6-trifluorophenol, by using the enzyme tyrosine phenol-lyase (Figure 3.1).[143] Mass spectrometric analysis confirmed the identity of **1**. The synthesized amino acid was incorporated into *KlenTaq* Y671TAG (generated through site-directed mutagenesis) by non-canonical amino acid mutagenesis, by using a mutated *M. jannaschii* tyrosyl aminoacyl tRNA synthetase and its cognate aaRS.[109] An SDS-PAGE analysis demonstrated that protein expression was only observed in the presence of **1** (Figure 3.4). Additionally, the expressed protein was digested with trypsin, and fluorinated

tyrosine incorporation was confirmed by MALDI-MS/MS analysis. This *KlenTaq* mutant harboring the fluorinated tyrosine analogue at position 671 was termed "**1**-*KlenTaq*".

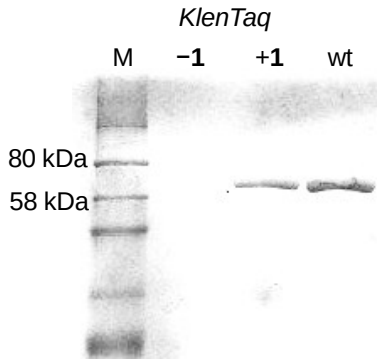


Figure 3.4. Incorporation of 1 into *KlenTaq* DNA polymerase. SDS-PAGE analysis showing the incorporation of 1 into *KlenTaq*. M: Marker; wt: wild-type *KlenTaq*. These experiments were conducted by Alex Prokup in the Deiters lab. Adapted with permission from Modulating the pKa of a tyrosine in KlenTaq DNA polymerase that is crucial for abasic site bypass by in vivo incorporation of a non-canonical amino acid. Adapted with permission from Blatter, N.; Prokup, A.; Deiters, A.; Marx, A. *ChemBioChem* 15 (12).

First, the *KlenTaq* mutant was tested to determine if it follows the A-rule in single primer-extension experiments with a 5′-^{32}P-radiolabeled DNA primer (23 nt) annealed to a template containing an abasic-site analogue, F (Figure 3.5A). Single nucleotide incorporation opposite F was conducted with all four nucleotides. All reactions were performed at pH 7.5 and with an incubation time of 120 min. Subsequent analysis by denaturing PAGE showed preferential incorporation of dAMP for both wild-type *KlenTaq* and **1**-*KlenTaq*, with incorporation of dGMP also detected in both cases (Figure 3.5B). Thus, preferential incorporation of purines was shown to be retained in

the **1**-*KlenTaq* mutant (primer elongation with pyrimidine substrates was not observed under these conditions). Next, we compared the activity of the enzymes when incorporating dAMP opposite a natural dT in the template strand or opposite the abasic-site analogue, F. The polymerases were diluted in a stepwise manner. This assay showed slightly reduced activity for **1**-*KlenTaq* (compared to wild-type) when using a natural template, but there was more pronounced activity loss when dAMP incorporation opposite F was measured. As dAMP incorporation opposite the lesion was reduced substantially for both enzymes, the reaction time was adjusted to 30 min as opposed to 5 s for the natural template. To further compare the difference in the incorporation efficiency opposite the abasic-site analogue, we performed further single-nucleotide incorporation studies over time, with an excess of enzyme (relative to primer/template) to negate potential different binding efficiencies of the enzymes. As observed previously, there was lower incorporation efficiency for **1**-*KlenTaq*, as almost full conversion was achieved after 5 min for the wild-type *KlenTaq*, but 30 min were required for the **1** mutant (Figure 3.5C). We propose that the tyrosine hydroxy group in wild-type *KlenTaq* (pK_a = 9.9) is protonated at pH 7.5, thus allowing formation of a hydrogen bond between the hydroxy group and N3 of the incoming adenine. However, reaction conditions presumably favor the deprotonated state of the hydroxy group of fluorotyrosine in **1**-*KlenTaq* because of the lower pK_a (6.4) of the non-canonical amino acid. Thus, stabilization of the incoming nucleotide in the active site decreases when the hydrogen-bond donor is altered. The possibility of the presence of a counter ion in the active site cannot be ruled out, and this might further hamper interactions between the tyrosine hydroxy group and N3 of the incoming dATP.

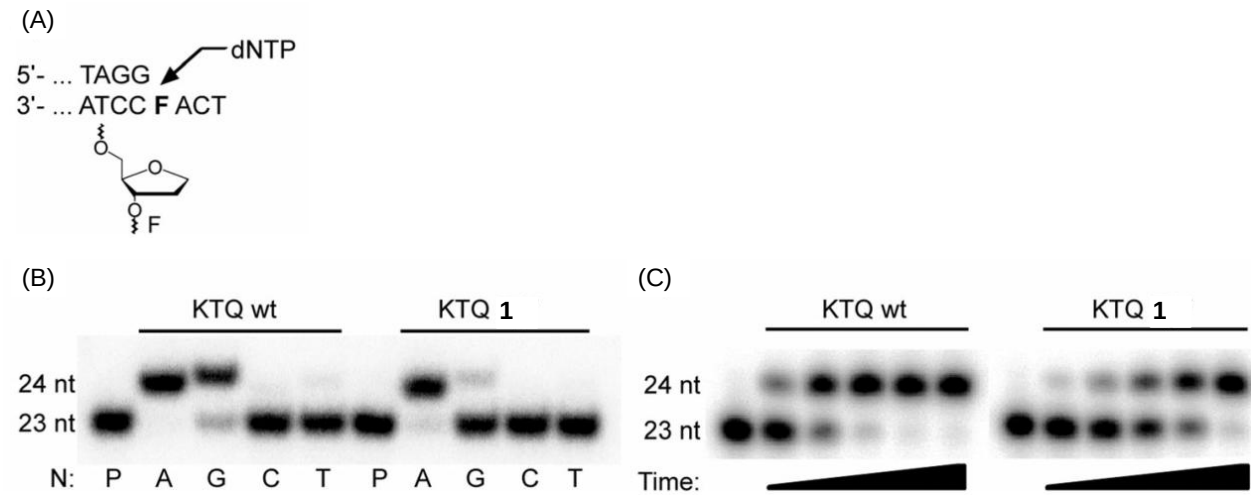


Figure 3.5. Nucleotide incorporation opposite abasic site F. (A) Partial sequence of primer/template. (B) Single nucleotide incorporation opposite F in 120 min by wild-type *KlenTaq* (KTQ wt) and 1-*KlenTaq* (KTQ 1). N: respective dNTP; P: primer. (C) Time course experiment of dATP incorporation opposite abasic site F: 30 s, 2 min, 5 min, 10 min, or 30 min reaction time; left lane: primer alone. These experiments were conducted by Nina Blatter in the Marx lab. Adapted with permission from Modulating the pKa of a tyrosine in KlenTaq DNA polymerase that is crucial for abasic site bypass by in vivo incorporation of a non-canonical amino acid. Blatter, N.; Prokup, A.; Deiters, A.; Marx, A. *ChemBioChem* 15 (12). Copyright © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

In summary, **1** was site-specifically incorporated into *KlenTaq* DNA polymerase by in vivo incorporation of a non-canonical amino acid.[109] This confirmed the importance of hydrogen bonding in abasic-site bypass, especially the impact of the hydrogen bond between the N3 of adenine and the hydroxy group of Y671. Previous site-directed mutagenesis studies suggested the importance of a potential hydrogen bond between N3 of adenine and the hydroxy group of Y671, as a Y671F mutant showed reduced

ability to bypass the abasic site, and reduced incorporation efficiency was observed for the incorporation of 3-deaza-dATP.[138] Our results corroborate our previous findings and confirm that the hydrogen bond formed between the tyrosine hydroxy group and N3 of the adenine plays an important role in stabilizing the active site. Furthermore, they demonstrate the utility of fluorinated tyrosine analogues in studies investigating hydrogen bond formation at protein active sites.

3.1.2 Experimental

Oligonucleotides were purchased from Metabion (Martinsried, Germany) or Biomers (Ulm, Germany). Wild-type *KlenTaq* was expressed and purified as described previously.[144] To express **1**-*KlenTaq*, the plasmids pEVOL-3Fy-E3 (1 µL of a 100 ng/µL stock) and pGDR11-Y671TAG (1 µL of a 100 ng/µL stock) were added to 20 µL *Escherichia coli* BL21-Gold (DE3) cells (Agilent Technologies). After incubating (ice, 15 min), the tubes were submerged in a 42 °C water bath for 45 seconds and then immediately placed on ice for 2 minutes. To the tube was added SOC (180 µL) and the tube was incubated (37 °C, 1 hour). The SOC media and cells were transferred to a Luria broth (LB) agar plate (10 mL) containing ampicillin (10 µL of a 100 mg/mL stock, 100 µg/mL) and chloramphenicol (10 µL of a 25 mg/mL stock, 25 µg/mL), and incubated (37 °C, overnight).

Protein expression and purification

A colony was picked and grown (37 °C, overnight) in 2×YT medium (5 mL) with ampicillin (5 µL of a 100 mg/mL stock, 100 µg/mL) and chloramphenicol (5 µL of a 25 mg/mL stock, 25 µg/mL). Expression of **1**-*KlenTaq* was carried out in 2×YT medium (100 mL) with ampicillin (10 µL of a 100 mg/mL stock, 100 µg/mL) and chloramphenicol (5 µL of a 25 mg/mL stock, 25 µg/mL), and **1** (1 mM), by inoculation with the overnight culture. At OD_{600}=0.6, the cultures were induced with L-arabinose (100 µL of a 20% stock, 0.2 %) and isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mL of a 100 mM stock, 1 mM). Cells were further incubated to express protein (6 h, 37 °C), then centrifuged in a Thermo IEC Multi RF model 120 centrifuge (9000 x *g*, 4 °C, 10 min), and stored (−80 °C). The cell pellet was then suspended in 16 mL lysis buffer (Tris·HCl (500 mM, pH 9.2), NaCl (300 mM), $MgCl_2$ (2.5 mM), Triton X100 (0.1 %)), lysozyme (160 µL of a 1 mg/mL stock, 0.1 mg/mL), and protease inhibitor cocktail (Sigma, 16 µL of stock solution, 0.1 %). After incubation (1 h, 4 °C), the cells were lysed by sonication (Branson Sonifier S-450A, 40 % output, power level: 2–15 s, 3–15 s, 4–45 s), heated (85 °C, 30 min), and centrifuged (20500 x *g*, 4 °C, 15 min). The supernatant was removed and Ni-NTA resin (200 µL) was added. After incubation to bind protein (1 h, 4 °C) the mixture was centrifuged (800 x *g*, 4 °C, 5 min). The resin was removed by pipetting and placed into a filter spin-column (0.45 µm, PVDF, Fisher), and centrifuged (5800 x *g*, 4 °C, 0.5 min) to eliminate residual supernatant. The resin was then washed with 250 µL of lysis buffer, wash buffer (4:1 ratio of lysis/elution buffer), and eluted in elution buffer (Tris·HCl (100 mM, pH 8), $MgCl_2$ (5 mM), Tween 20 (0.2 %), imidazole (200 mM)). The buffer was exchanged by dialysis (Tris·HCl (50 mM, pH 9.2), $(NH_4)_2SO_4$

(16 mM), $MgCl_2$ (2.5 mM), Tween 20 (0.1%), and glycerol (50%)), and the protein was stored at −80 °C. The purified protein was loaded onto a 12 % SDS-PAGE gel, electrophoresed (60 V/15 min, 150 V/45 min) and stained with Brilliant Blue G protein stain (Sigma). Concentration was determined by Coomassie blue staining comparison to BSA protein standards (NEB, 62.5, 125, and 250 ng/uL, load 10 µL of each on the gel with 10 µL Klentaq).

Single nucleotide incorporation experiments.

For radioactive labeling, primer (400 nM) was incubated in the presence of γ-^{32}P-ATP (400 nCi/µL, Hartmann Analytic), T4 polynucleotide kinase (0.4 Units/µL, Fermentas) in supplied 1× reaction buffer A for 60 min at 37 °C (50 µL scale). The reaction was terminated (95 °C, 2 min), and the labeled primer was purified by gel filtration (Sephadex G25). Unlabeled primer (20 µL) was added to obtain a final stock concentration of 3 µM. Incorporation opposite F was tested with all four nucleotides. Reaction mixtures (20 µL) contained radioactively labeled primer (100 nM, 5′-d(CGTTGG TCCTGA AGGAGG ATAGG)-3′), F-containing template (130 nM, 5′-d(AAATCA FCCTAT CCTCCT TCAGGA CCAACG TAC)-3′), the respective dNTP (100 µM), and the respective *KlenTaq* DNA polymerase (25 nM), in Tris·HCl (20 mM, pH 7.5), NaCl (50 mM) and $MgCl_2$ (2 mM). Reaction mixtures were incubated at 37 °C and terminated at 120 min by addition of stop solution (45 µL, formamide (80 %, *v*/*v*), EDTA (20 mM), Bromophenol Blue (0.25 %, *w*/*v*), xylenecyanol (0.25 %, *w*/*v*)). After denaturation (95 °C, 5 min), reaction mixtures were separated on a 12 % denaturing PAGE gel. Visualization was performed by phosphorimaging.

Assay for incorporation opposite F against time: experiments were performed as described above with the respective *KlenTaq* DNA polymerase (1 μM). Incubation times are given in the figure legend.

3.2 PROTEIN CROSSLINKING TO RNA USING A DIAZIRINE LYSINE

3.2.1 Results and Discussion

mHfq was first discovered in the 1960s from *E. Coli* as a host factor for the bacteriophage Qβ. The Hfq protein binds with many small RNAs (sRNAs) to regulate bacterial gene expression at the post-transcriptional level by promoting the binding of the sRNA to complementary 5' regions of messenger RNA (mRNA) to suppress[145] or increase translation.[146] Additionally, the binding of Hfq can prevent RNase degradation of the sRNA, [147] or allow for degradation of both the sRNA and the mRNA target.[148] The Hfq protein adopts a ring-like quaternary structure, comprised of six monomer subunits. RNA will bind preferentially to either the distal or proximal face. Typically, adenosine-rich RNAs will bind the distal face, while uridine-rich RNAs bind the proximal face.[149]

The Hfq protein has been co-crystallized with a polyuridinylate[150] and polyriboadenylate RNA.[149] Similar to other reports[151] of Hfq binding, the poly(A) RNA was bound to the distal face, whereas poly(U) RNA was bound to the proximal face. Selective binding of RNA molecules to the distal face of Hfq was generalized to sequences that follow the ARN and ARNN' rules (A is an adenosine, R is a purine, and N and N' represent any nucleotide). The lack of discrimination for the third position is

due to the rotation of the nucleotide away from the distal face. Dissociation constants (K_d) measured by Link *et al.* provided additional insight into the specificity of RNA binding to Hfq.[149] As the tract length of ARN repeats increased, the binding also increased. For poly(A) RNAs, the K_d decreased greatly from A_3 (1770 nM) to A_6 (160 nM) and A_{16} (1.4 nM). After a certain length, the binding is no longer improved, as exemplified by A_{27} (1.7 nM).

Understanding the binding patterns of Hfq could aid in the discovery of new RNA binding partners and an understanding how Hfq regulates gene expression. The main methods to identify these RNA include co-immunoprecipitation, the use of transcriptomics and proteomics, and systematic evolution of ligands by exponential enrichment (SELEX).[152] Crosslinking immunoprecipitation (CLIP) enables the identification of RNA-protein interactions by crosslinking nucleotides and amino acid residues after irradiation at 254 nm. The RNA can be reversed transcribed and identified by high-throughput sequencing. A major disadvantage of CLIP, especially for the identification of Hfq binding partners, is the low crosslinking efficiency of purines, which may limit the detection of mRNA targets and other distal face binding RNAs. However, due to the success of crosslinking in the microRNA field, the use of an improved crosslinking system may prove extremely useful. Previously, CLIP has been used to identify RNA binding partners of Hfq, and more specifically identify the facial binding preferences of each RNA.[153] The study used UV irradiation to non-specifically crosslink all RNAs bound to Hfq, which were later identified by reverse transcription to cDNA and sequencing. In total, Tree *et al.* identified fifty-five sRNAs that bind to Hfq in the *E. coli* genome in the CLIP experiment.[153] The investigation of mRNA binding was

limited, but confirmed the ARN binding motif associated with Hfq. Experimentally, Tree *et al.* showed a preference of Hfq for AGN motifs,[153] which are commonly associated with Shine-Dalgarno ribosome binding site sequences. This similarity is understandable since Hfq acts as a mediator for the binding of sRNAs and mRNAs. Binding of the sRNA close to the Shine-Dalgarno sequence was suggested to play an important role in translation inhibition, by blocking the binding of ribosomes to the mRNA. Identification of the facial binding preference of each RNA was based on previous information on binding motifs[149,154] and RNA sequence deletions.[153] A method to directly identify facial binding of each RNA would be possible by specific covalent crosslinking RNA to one face of Hfq.

The diazirine **2** was previously genetically encoded and validated in the photocrosslinking of superoxide glutathione S-transferase dimers.[111] In order to specifically crosslink distal and proximal-binding RNAs, **2** was incorporated into the Hfq monomer at several positions. A two-plasmid system was used, one containing the Hfq gene and $tRNA_{CUA}$ sequence, and another containing the a pyrrolysyl aaRS containing the mutations D203N, Y271C, L274V, C313V, and M315Y It is hypothesized that the site-specific incorporation of **2** into Hfq will enable the determination of facial binding preferences of Hfq-binding RNAs after a CLIP-like experiment under *in vivo* conditions (Figure 3.6).

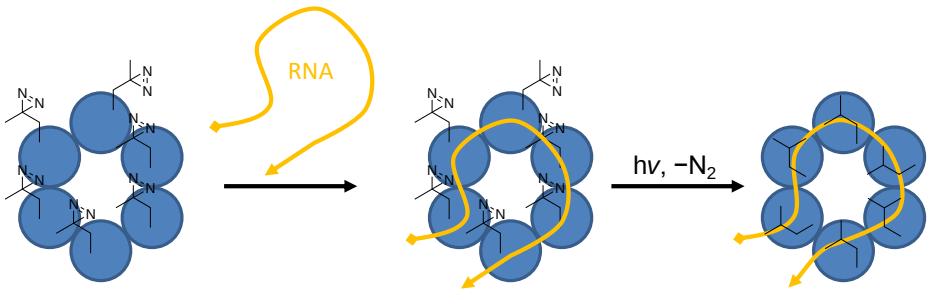


Figure 3.6. Schematic of diazirine-modified Hfq and crosslinking to RNA. An RNA strand (yellow line) binds to hexameric diazirine-Hfq (blue circles), and crosslinking is initiated by irradiation with UV light.

Expression of Hfq in the presence of **2** will incorporate the unnatural amino acid into the protein at the specific residue containing the TAG mutation. Multiple residues were selected based upon their facial position within the Hfq structure. For distal crosslinking, the residues N28, I30, K31, Q33, N48, and V63 were selected, whereas residues Q8, Q41, and H57 were selected for proximal crosslinking. Proteins were obtained in good yield after purification by nickel affinity chromatography and treatment with RNase A (Figure 3.7).

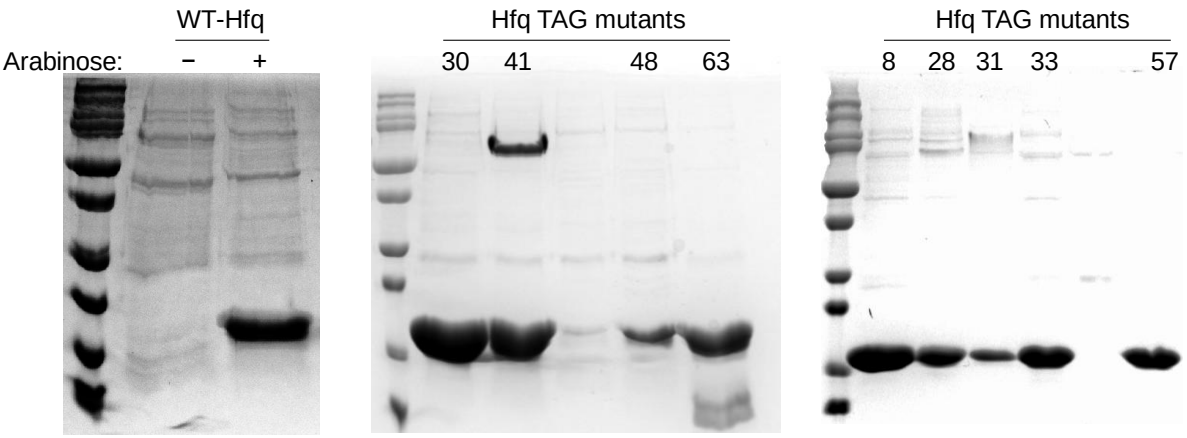


Figure 3.7. Expression of wild-type and diaizirine modified Hfq proteins. Residues Q8, N28, I30, K31, Q33, Q41, N48, H57, and V63 of Hfq were mutated to a TAG stop codon. The proteins were expressed in the presence of 1 mM 2 (except wild-type), purified by nickel affinity chromatography, and treated with RNase A. Yields ranged from 2.5 – 12.5 mg/L.

Optimal irradiation time was determined by irradiating a single mutant, 48, in the presence of a 16-mer poly(A) RNA for increasing time intervals (Figure 3.8). The RNA was radiolabeled using T4 polynucleotide kinase (T4 PNK) and [γ-^{32}P]ATP. Samples were analyzed by native- and SDS-PAGE to determine the binding and crosslinking efficiencies, respectively. Longer irradiation times enabled greater crosslinking between Hfq and the RNA. Therefore, 40 minutes was used as the irradiation time for all further crosslinking investigations. A large black band was observed at the bottom of the gel, corresponding to free or non-crosslinked radiolabeled RNA.

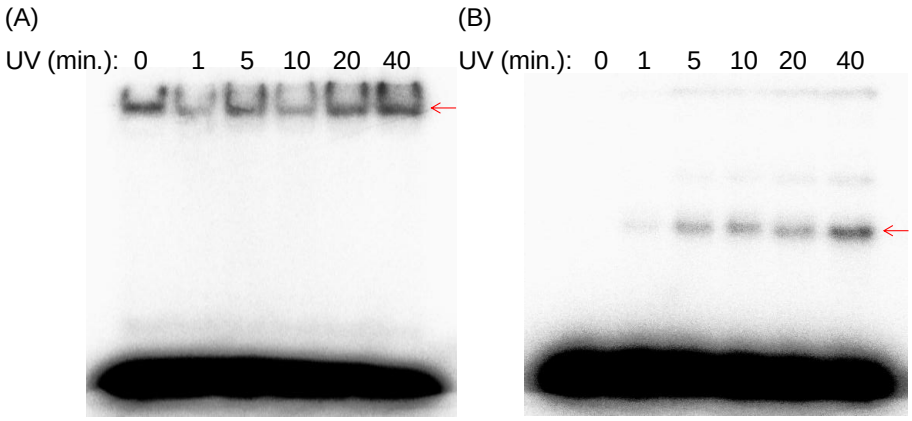


Figure 3.8. Optimization of UV irradiation time for crosslinking of 2-Hfq mutant 48 (1 μM) and RNA-A (25 nM). (A) Native- and (B) SDS-PAGE analysis of photocrosslinking depending on UV exposure time. (A) Binding was essentially unaffected by irradiation as observed on the native-PAGE. The red arrow indicates RNA-bound protein. (B) An SDS-PAGE analysis shows the intensity of the higher molecular weight band increased as irradiation time was increased. A red arrow indicates the crosslinked RNA and protein. RNA was radiolabeled with [γ-^{32}P]ATP by T4 PNK. Unbound or non-crosslinked RNA can be seen as a large black band at the bottom of the gels.

Wild-type Hfq and radiolabeled 16-mer poly(A) and poly(U) RNAs were irradiated with UV light for 40 minutes to determine that crosslinking did not non-specifically occur in the absence of the diazirine group, and binding was not affected as shown by the native-PAGE gel (Figure 3.9).

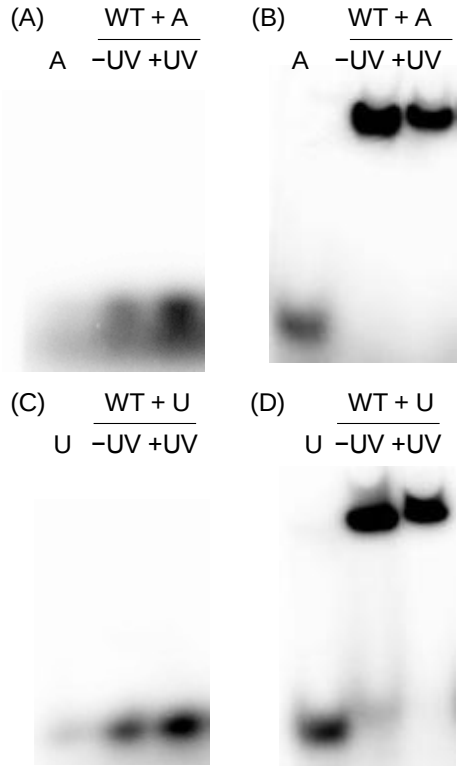


Figure 3.9. UV irradiation of WT-Hfq and poly(A) and poly(U) RNAs. WT-Hfq (WT-Hfq, 1 μM) was incubated with RNA (25 nM) at room temperature for 10 minutes before UV irradiation (40 min). Wild-type Hfq was analyzed by SDS-PAGE (left) and native-PAGE (right). The 16-mer (A) poly(A) and (B) poly(U) RNAs were abbreviated as "A" and "U," respectively. No RNA-protein crosslinking was observed after UV irradiation, but normal RNA and protein binding was observed. RNA was radiolabeled with [γ-^{32}P]ATP by T4 PNK.

With successful incorporation of the diazirine moiety into the Hfq monomer, the RNA can be crosslinked (Figure 3.10). After binding, the RNA will be within close proximity of the diazirine functionality. Irradiation with UV light will form a carbene, which can react and covalently bind with nearby CH, NH, or OH bonds. Since the diazirine is located on only one face of the Hfq hexamer, an RNA binding the other face should not be crosslinked. Initial crosslinking tests showed that the Hfq mutants with most promising crosslinking efficiency were the mutants 8, 48, and 57. These mutants were selected for further crosslinking experiments (Figure 3.11). A radiolabeled 16-mer poly(A) or poly(U) RNA was added to the protein and irradiated with UV light. Samples were analyzed by SDS-PAGE to remove non-covalently bound RNA from protein. Therefore, a gel shift indicated potentially crosslinked RNA and protein monomers. As expected, when the diazirine moiety was incorporated on the distal face of Hfq (mutant 48), crosslinking was only observed with poly(A) RNA (Table 3.1). Similarly, when positioned on the proximal face of Hfq (mutants 8 and 57), the crosslinking was only observed with poly(U) RNA. The incorporation of diazirine lysine was well tolerated, as binding to either face was not disrupted. However, K_d values

are needed to quantitatively determine if addition of diazirine lysine to Hfq disrupted binding (see further discussion below).

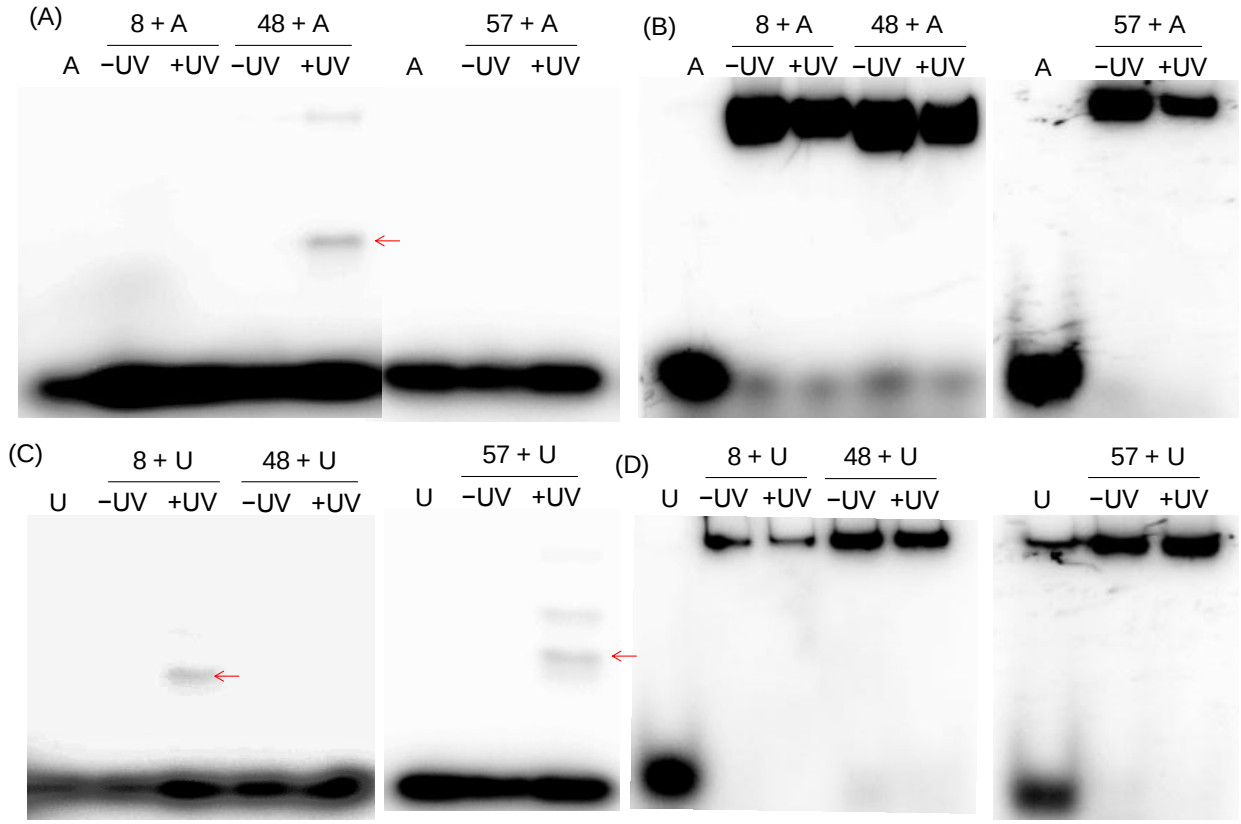


Figure 3.10. Crosslinking between diazirine-modified Hfq and RNAs. (A) SDS-PAGE and (B) native-PAGE analysis for the crosslinking between diazirine-modified proteins and RNA-A. (C) Native-PAGE and (D) SDS-PAGE analysis for the crosslinking between diazirine-modified proteins and RNA-U. Hfq (1 μM) was incubated with RNA (25 nM) at room temperature for 10 minutes before UV irradiation (40 min). Expected crosslinking was observed for mutants 8, 48, and 57. RNA was radiolabeled with [γ-^{32}P]ATP by T4 PNK.

Table 3.1. Summary of the protein expression, RNA binding, and RNA crosslinking experiments. Expected RNA facial preference based on the RNA sequence and overall conclusion for wild-type Hfq and diazirine Hfq mutants 8, 48, and 57 is included.

Mutant	RNA	Expression	Binding	Observed Crosslinking	Expected Facial Pref.	Conclusion
Wild-type	poly(A)	Moderate yield	Good	None	Distal	No crosslinking as expected
Wild-type	poly(U)		Good	None	Proximal	No crosslinking as expected
Q8	poly(A)	High yield	Good	None	Distal	No crosslinking as expected
Q8	poly(U)		Good	Proximal	Proximal	Crosslinking as expected
N48	poly(A)	Moderate yield	Good	Distal	Distal	Crosslinking as expected
N48	poly(U)		Good	None	Proximal	No crosslinking as expected
H57	poly(A)	High yield	Good	None	Distal	No crosslinking as expected
H57	poly(U)		Good	Proximal	Proximal	Crosslinking as expected

To ensure that the natural binding affinity of Hfq to RNA was not interrupted by the incorporation of **2**, the K_d was measured (Figure 3.11). Due to the importance of surface residues, a small increase may be expected after modification of a residue. None of the mutated residues were lysines, and incorporation of an unnatural lysine analogue at these positions would be expected to disrupt natural binding. A significant increase in K_d of diaizirine-proteins would indicate the binding of protein and RNA is being interrupted. Multiple concentrations of protein were added to radiolabeled RNA

and analyzed by native-PAGE. The measured K_d value for WT-Hfq and RNA-A was 6.25 nM, which is similar to the literature reported value.[149] Unmodified wild-type Hfq functions as expected, tightly binding poly(A) RNA on the distal face. Further measurements are needed to determine the effect of **2** on RNA binding.

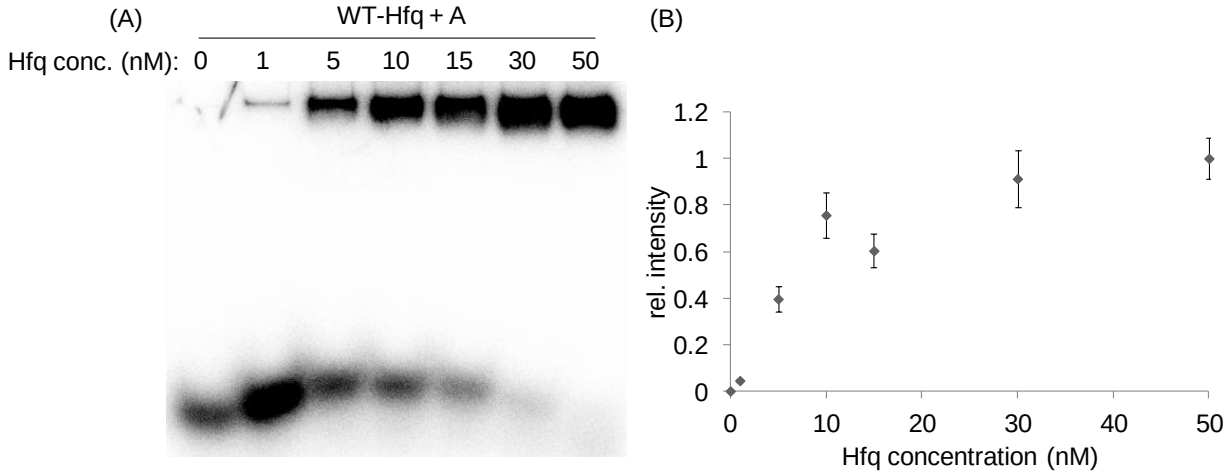


Figure 3.11. Dissociation constant measurement for WT-Hfq and RNA-A. (A) A representative native-PAGE analysis containing WT-Hfq and RNA-A, abbreviated as "A." Hfq was added at multiple concentrations (0-50 nM) to 1 nM [γ-^{32}P] ATP-radiolabeled RNA at room temperature for 10 minutes. (B) The RNA-protein bands from the three native-PAGE gels were analyzed for band intensity and plotted against Hfq concentration. The K_d value was determined to be 6.25 nM, which is similar to the value reported in literature.

To try and improve upon the low crosslinking efficiency shown above, the protein was expressed and irradiated in bacterial cells before purification (Figure 3.12). Previously, the crosslinking of GST monomers was much more efficient in *E. coli* than in a test tube.[111] Immunoblot analysis revealed the presence of higher molecular weight

bands for Hfq mutant 57 when crosslinked *in vivo*, suggesting that protein and RNA are crosslinked. However, further experiments are necessary to determine if RNA, and not just protein, was crosslinked. Unexpectedly, mutant 48 did not appear to crosslink with any RNAs. The absence of crosslinking may be due to low efficiency of crosslinking *in vivo* or poor expression of the protein.

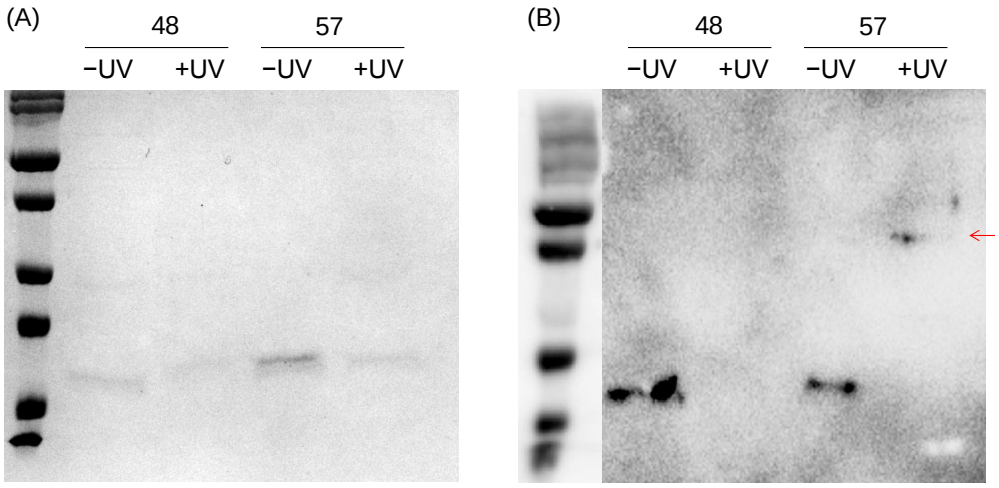


Figure 3.12. A Western blot image of Hfq proteins detected in the absence or presence of UV irradiation of bacterial cells expressing Hfq diazirine mutants 48 and 57. Mutants were expressed overnight at 37 °C in the presence of 1 mM of 2, irradiated with UV light, then purified without RNase A treatment. The same protein samples were detected by (A) Coomassie blue staining and (B) immunodetection. A mouse anti-His primary antibody with secondary goat anti-mouse conjugated horse radish peroxidase enzyme secondary antibody was used for the Western blot. Only faint Hfq monomer protein bands were observed after Coomassie staining, whereas more intense Hfq monomer and crosslinked protein bands were apparent after immunodetection. A higher molecular weight band was detected on the Western blot, potentially signifying successful crosslinking between Hfq and RNA.

Incorporation of diazirine amino acid **2** into Hfq was successful. The diazirine mutants 8, 48, and 57 enabled facial-specific crosslinking of RNA molecules. Crosslinking was specific and only occurred in the presence of the diazirine moiety. As expected, the incorporation of diazirine to the distal or proximal face enabled poly(A) or poly(U) crosslinking, respectively. Dissociation constants are needed to determine to what extent RNA binding was disrupted from the incorporation of **2** to the two faces of Hfq. RNA was successfully crosslinked using mutants 48 and 57. However, more work is necessary to determine if the *in vivo* crosslinking is between protein and RNA or just Hfq monomers. The identity of higher molecular weight bands observed on the Western blot could be determined by staining with SYBR Gold or labeling with T4 polynucleotide kinase to determine if nucleic acids are present.

3.2.2 Experimental

Site-directed mutagenesis

In a PCR tube, plasmid pBAD-Hfq-pylT (1 µL of a 5 ng/µL stock) was added to forward primer (0.25 µL of a 100 µM stock, Table 3.2), reverse primer (0.25 µL of a 100 µM stock, Table 3.2), dNTPs (1 µL of a 5 µM stock), Phusion buffer (5 µL of a 5X stock, NEB), deionized water (16.5 µL), and Phusion polymerase (1 µL of a 2,000 units/mL stock, NEB). The PCR tube was incubated in a Bio-Rad T100 Thermal Cycler (95 °C for 5 min, then 12 cycles of: 95 °C for 1 min, 60 °C for 1 min, and 72 °C for 6 min, then a final incubation at 72 °C for 30 min). DpnI (2.5 µL of a 20000 Units/mL stock, NEB) and 10X CutSmart buffer (2.7 µL of a 10X stock, NEB) were added to the PCR tube and

incubated (37 °C, 1 hour). The PCR tube was then cooled on ice and transformed into Top10 competent cells according to the protocol in Section 3.1.

Table 3.2. DNA primer strands used for the site-directed mutagenesis of Hfq to create TAG mutants. Underlined nucleotides indicate mutations.

Strand	Sequence (5' → 3')
Hfq Q8TAG forward primer	ATCTTTATAGGATCCGTTCCTGAACGCACTGCGTCGGGAACGTGTTCCAGTTTC
Hfq Q8TAG reverse primer	ACGGATCCTATAAAGATTGCCCCTTAGCCATGGTTAATTCCTCCTGTTAG
Hfq N48TAG forward primer	GTTGAAATAGACGGTCAGCCAGATGGTTTACAAGCACGCGATTTCTACTG
Hfq N48TAG reverse primer	TGACCGTCTATTTCAACAGGATCACGAACTGATCAAAAGACTCGATTTGC
Hfq H57TAG forward primer	TTACAAGTAGGCGATTTCTACTGTTGTCCCGTCTCGCCCGGTTTCTC
Hfq H57TAG reverse primer	AAATCGCCTACTTGTAAACCATCTGGCTGACCGTGTTTTTCAACAGGATCACGAAC

Expression and purification of Hfq protein

The protocol outlined in Section 3.13.2 can be followed for the expression of WT-Hfq with a few changes: Top10 competent cells were transformed with the plasmid pBAD-Hfq-pylT. For expression, tetracycline (25 µL of a 10 mg/mL stock, 10 µg/mL) was added to LB media (25 mL).

For **2**-Hfq, EV11 (0.5 µL of a 100 ng/µL stock, 50 ng, pBKpylRS D203N Y271C L274V C313V M315Y) and the TAG mutated pBAD-Hfq-pylT plasmid (0.5 µL of a 100 ng/µL stock, 50 ng, the pBAD-Hfq-pylT plasmid contained either the Q8TAG, N48TAG,

or H57TAG mutation) were transformed into Top10 competent cells. Ampicillin (10 µL of a 100 mg/mL stock, 100 µg/mL) and tetracycline (10 µL of 10 mg/mL stock, 10 µg/mL) were added to LB agar (10 mL). After cooling, the transformed compentent cells were grown on the LB agar (overnight, 37 °C). For expression of the proteins, ampicillin (25 µL of a 100 mg/mL stock, 100 µg/mL), tetracycline (25 µL of a 10 mg/mL stock, 10 µg/mL), and **2** (250 µL of a 100 mM stock, 1 mM, except for wild-type Hfq), were added to LB media (25 mL) and inoculated with 2 mL of the overnight culture. At OD_{600} > 0.6, the cultures were induced by adding arabinose (25 µL of a 20% stock, 0.2%) and expressed for 6 hours (37 °C). The cells were pelleted in a Thermo IEC Multi RF model 120 centrifuge (9000 x *g*, 4 °C, 10 min), the supernatant was discarded, and the pellet was stored (−80 °C). After resuspension in lysis buffer (5 mL), RNase A (5 µL of a 10 mg/mL stock, Amresco) was added to the 5 mL suspension of cells. The Ni-NTA purification protocol in Section 3.1 was used.

Binding and crosslinking of radiolabeled RNA

RNA was radiolabeled using T4 polynucleotide kinase (PNK) and [γ-^{32}P]ATP. To a PCR tube, RNA (0.5 µL of a 100 µM stock, 1 µM), PNK (2 µL of a 500 Units/mL stock, 20 Units), 10X PNK buffer (5 µL), and [γ-^{32}P]ATP (15 µL of a 3.33 µM stock, 1 µM) were added to DEPC-treated water (27.5 µL). The PCR tube was incubated in a thermocycler (Bio-Rad T100, 45 min, 30 °C) and protein was denatured (10 min, 60 °C). The reaction mixture was transferred to a 7 kDa MWCO Zeba spin-column (Thermo Scientific) (pre-equilibrated with DEPC-treated water) and centrifuged (1500 x *g*, 1 min). The filtrate contained labeled RNA at 1 µM without unreacted [γ-^{32}P]ATP.

To a PCR tube, Hfq (0.6 µL of a 208 µM stock, 1 µM), radiolabeled RNA (0.5 µL of a 1 µM stock, 25 nM), and Hfq buffer (2 µL of a 10 X stock, 500 mM tris pH 8, 100 mM $MgCl_2$, 1 M NaCl) were added to a PCR tube with DEPC-treated water (16.9 µL). The protein was incubated at room temperature for 10 minutes. The sample was transferred to a 1 mL plastic cuvette, which was then capped. An ice pack was added on top of the cuvette. The cuvette was placed on the transilluminator and irradiated for 40 minutes. After irradiation, the sample was transferred from the cuvette into a clean PCR tube. The 20 µL samples were then split into two 10 µL volumes. SDS loading buffer (200 mM tris, 20% glycerol, 10 mM beta-mercaptoethanol, 10% sodium dodecylsulfate, bromophenol blue for color) was added to one sample and boiled (10 min, 95 °C) before being analyzed by 12% SDS-PAGE (60 V / 15 min, 150 V / 45 min). The other 10 µL sample was added to native loading buffer (3 µL, 50% glycerol in water, bromophenol blue added for color) and analyzed by 10% native-PAGE (100 V, 45 min). The gels were then wrapped completely in a single layer of plastic wrap and placed in a phosphorimager cassette (overnight). The phosphorimager screen was then imaged for phosphorescence using the Typhoon FLA 7000 imager and software.

K_d measurements

The process to measure K_d values followed the same protocol as “Binding and crosslinking of radiolabeled RNA,” but without any UV irradiation steps. The final concentration of the RNA was changed to 1 nM (0.2 µL of a 100 nM stock). Gels were imaged and bands were integrated using the Typhoon FLA 7000 software.

In vivo crosslinking

The *in vivo* crosslinking protocol follows the same protocol as the “Expression and purification of Hfq,” up until the purification (expression volumes were 5 mL). After the expression cultures were centrifuged, the supernatant was removed, the cell pellet is resuspended in 10X PBS buffer buffer (5 mL), and centrifuged again (9000 x *g*, 10 min, 4 °C). This PBS washing was repeated twice more. The cultures were then resuspended again in PBS and split into two equal volumes (2.5 mL each). One of the 2.5 mL cultures was irradiated (40 min) on a UV transilluminator inside a covered petri dish. An ice pack was placed on top of the petri dish. After irradiation, the cultures were again purified as described above, except no RNase A was added.

After the eluted protein samples were collected, they were analyzed by 12% SDS-PAGE as described above. The gel was then immediately soaked in ~50 mL transfer buffer (100 mL 10X transfer buffer (250 mM Tris, 1.92 M glycine), 200 mL methanol, 0.025% SDS). Proteins were transferred to a nitrocellulose membrane (soaked for 5 min in pure methanol, then 5 min in transfer buffer) using a Western blot cassette (from black side up: 2 pieces filter paper, gel, nitrocellulose membrane, 2 pieces of filter paper). The transfer was started (cooled on all sides by ice, cassette in transfer buffer, 80 V, 1 h). The blot was then removed and blocked with milk (50 mL TBST, 5% milk, 1 hour, 4 °C, cooled with an ice pack). Some of the milk (15 mL) was poured into a 50 mL conical tube and mixed with a mouse anti-His antibody (15 µL, 1:1000 dilution, Santa Cruz). The primary antibody was incubated at 4 °C overnight. Next day, the blot was washed 5x with TBST buffer (50 mL, 0.1% Tween20, 1X TBS buffer). Then the secondary goat anti-mouse antibody conjugated to horse radish

peroxidase was added (0.5 µL into 15 mL TBST buffer, 1:30000 dilution, 1 hour, 4 °C, cooled with an ice pack). The membrane was washed with TBST buffer (5x 50 mL). Then reagents A and B (Visi-glo substrates) were mixed (2 mL each) and added to the membrane (allow it to shake on membrane for 2 min). The blot was transferred to a clear plastic notebook cover (for easier handling) and imaged on a ChemiDoc imager.

4.0 CONCLUSION

Chapter 1.0 describes methods to improve the capabilities of DNA computation devices. The field of DNA computation attempts to create DNA-based structures that mimic the operation of electrical devices. One of the fundamental building blocks of computational circuits is the logic gate, which utilizes two or more inputs to predictably produce an output. Thus, logic gates are a practical starting point for the creation of DNA-based circuitry. DNA is used as a uniform building material, similar to the use of transistors in electrical devices. Generally, DNA logic gates are composed of a duplex structure that binds inputs and releases outputs. The operation of DNA logic gates is controlled by a process known as toe-hold mediated strand exchange, which uses a short extension of a DNA duplex, known as a toe-hold, to kinetically benefit strand displacement.[14] Although DNA logic gates are not as complex as electrical devices, the main advantage to DNA-based devices is the ability to directly interact with biological systems. DNA logic gates can be used to interface electrical circuits with biological systems through the installation of photocleavable caging groups onto the DNA framework. In the absence of UV irradiation, the caging groups remain intact, and DNA:DNA hybridization is not possible. After irradiation, the caging groups are removed and normal DNA binding is possible. Thus, with the added caging groups, light produced from electrical circuits could theoretically be used to initiate a DNA computation event. When applied to

a DNA-based AND gate, the caging groups facilitated unprecendented temporal and spatial control.[6] Furthermore, this demonstrated a merging of electrical and biological systems through the creation of step-like responses and spatially activated logic gate activation. To better interface DNA circuitry with biological systems, logic gates directly activating protein function were also designed.[52] These gates used DNA inputs to produce protein function as an output. Direct activation of protein output was mediated by zinc-finger proteins, which can naturally bind DNA structures and modified as a fusion protein. Split-protein components were fused to the zinc-finger proteins, such that reconstitution of the protein structure would enable protein function as an output. In total, three different logic gates were designed and successfully used in a mini-circuit and the detection of two micro-RNAs.

Chapter 2.0 describes signal amplification systems for DNA-based circuits. For electrical circuitry, signal amplification is important to increase the voltage or current through a circuit. Similarly, DNA signal amplification cycles increase the production of DNA strands in DNA-based circuits and can be applied to larger DNA-based networks and logic gates. In larger networks, multiple logic gates can be connected to create complex networks. However, due to inefficient or incomplete gate activation, a process known as signal dampening can occur, which can cause a decrease in output signal. For stoichiometric logic gates, where one input creates one output, a low input concentration can only produce a low output concentration. Therefore, a signal amplification system can be applied to increase the generated signal to practical concentrations. Photocleavable caging groups were added to crucial DNA strands in two different amplification cycles, the hybridization chain reaction[19] and fuel-catalyst

cycle.[20] The caging groups further increased control of amplification by enabling switching of the cycles ON or OFF with UV light.[28] When the components were embedded into a semi-solid agarose gel, the amplification could be temporally and spatially controlled. The fuel-catalyst cycle was further modified and converted into an autocatalytic cycle, which enabled propagation of signal along a predefined path. Electrical circuits require spatially activated components and the use of light as a switch to activate DNA amplification systems further demonstrates how the interface of electrical and biological circuits is further merged through the addition of photocleavable caging groups.

Chapter 3.0 discusses unnatural amino acids and their application to the investigation of DNA polymerases in abasic site bypass[130] and protein:RNA crosslinking. Many unnatural amino acids have been synthesized and incorporated into proteins using a method known as unnatural amino acid mutagenesis. Using this method, an unnatural amino acid can be genetically encoded and site-specifically incorporated into a protein using an evolved amino acyl tRNA synthetase, its cognate $tRNA_{CUA}$, and a gene containing a TAG mutation. The $tRNA_{CUA}$ is specifically aminoacylated with the unnatural amino acid using the evolved synthetase and adds the unnatural amino acid to a growing peptide chain through the suppression of the TAG stop codon. A fluorotyrosine amino acid was enzymatically synthesized and used to investigate the role of a phenolic hydrogen bond in a DNA repair mechanism known as abasic site bypass. The addition of fluorines to the tyrosine structure increased the acidity of the phenolic hydrogen, leading to a decrease in DNA polymerase activity. A lysine analog containing a diazirine moiety was incorporated into the RNA-binding

protein hfq. The diazirine enabled covalent crosslinking to RNA bound on either the proximal or distal face of hfq after irradiation with UV light.

BIBLIOGRAPHY

1. Adleman, L. M., Molecular computation of solutions to combinatorial problems. *Science* **1994,** *266* (5187), 1021-4.
2. Qian, L.; Winfree, E., Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **2011,** *332* (6034), 1196-201.
3. Orbach, R.; Wange, F.; Lioubashevski, O.; Levine, R. D.; Remacle, F.; Willner, I. A full-adder based on reconfigurable DNA-hairpin inputs and DNAzyme computing modules **2014**, p. 3381-3387.
4. Qian, L.; Winfree, E.; Bruck, J., Neural network computation with DNA strand displacement cascades. *Nature* **2011,** *475* (7356), 368-72.
5. Stojanovic, M. N.; Stefanovic, D., A deoxyribozyme-based molecular automaton. *Nat Biotechnol* **2003,** *21* (9), 1069-74.
6. Prokup, A.; Hemphill, J.; Deiters, A., DNA computation: a photochemically controlled AND gate. *J Am Chem Soc* **2012,** *134* (8), 3810-5.
7. Hemphill, J.; Deiters, A., DNA Computation in Mammalian Cells: microRNA Logic Operations. *J Am Chem Soc* **2013**.
8. Church, G. M.; Gao, Y.; Kosuri, S., Next-generation digital information storage in DNA. *Science* **2012,** *337* (6102), 1628.
9. Jiang, Y. S.; Bhadra, S.; Li, B.; Ellington, A. D., Mismatches Improve the Performance of Strand-Displacement Nucleic Acid Circuits. *Angew Chem Int Ed Engl* **2014**.
10. Yoshida, W.; Yokobayashi, Y., Photonic Boolean logic gates based on DNA aptamers. *Chem Commun (Camb)* **2007,** (2), 195-7.
11. Lederman, H.; Macdonald, J.; Stefanovic, D.; Stojanovic, M. N., Deoxyribozyme-based three-input logic gates and construction of a molecular full adder. *Biochemistry* **2006,** *45* (4), 1194-9.
12. Zhu, J.; Zhang, L.; Li, T.; Dong, S.; Wang, E., Enzyme-free unlabeled DNA logic circuits based on toehold-mediated strand displacement and split G-quadruplex enhanced fluorescence. *Advanced materials* **2013,** *25* (17), 2440-4.
13. Lake, A.; Shang, S.; Kolpashchikov, D. M., Molecular logic gates connected through DNA four-way junctions. *Angew Chem Int Ed Engl* **2010,** *49* (26), 4459-62.
14. Zhang, D. Y.; Winfree, E., Control of DNA strand displacement kinetics using toehold exchange. *J Am Chem Soc* **2009,** *131* (47), 17303-14.
15. Qian, L.; Winfree, E., A simple DNA gate motif for synthesizing large-scale circuits. *Journal of the Royal Society, Interface / the Royal Society* **2011,** *8* (62), 1281-97.

16. Srinivas, N.; Ouldridge, T. E.; Sulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P.; Winfree, E., On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res* **2013,** *41* (22), 10641-58.
17. Lakin, M. R.; Youssef, S.; Polo, F.; Emmott, S.; Phillips, A., Visual DSD: a design and analysis tool for DNA strand displacement systems. *Bioinformatics* **2011,** *27* (22), 3211-3.
18. Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E., Enzyme-free nucleic acid logic circuits. *Science* **2006,** *314* (5805), 1585-8.
19. Dirks, R. M.; Pierce, N. A., Triggered amplification by hybridization chain reaction. *Proc Natl Acad Sci U S A* **2004,** *101* (43), 15275-8.
20. Zhang, D. Y.; Turberfield, A. J.; Yurke, B.; Winfree, E., Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **2007,** *318* (5853), 1121-5.
21. Saghatelian, A.; Volcker, N. H.; Guckian, K. M.; Lin, V. S.; Ghadiri, M. R., DNA-based photonic logic gates: AND, NAND, and INHIBIT. *J Am Chem Soc* **2003,** *125* (2), 346-7.
22. Frezza, B. M.; Cockroft, S. L.; Ghadiri, M. R., Modular multi-level circuits from immobilized DNA-based logic gates. *J Am Chem Soc* **2007,** *129* (48), 14875-9.
23. Macdonald, J.; Li, Y.; Sutovic, M.; Lederman, H.; Pendri, K.; Lu, W.; Andrews, B. L.; Stefanovic, D.; Stojanovic, M. N., Medium scale integration of molecular logic gates in an automaton. *Nano Lett* **2006,** *6* (11), 2598-603.
24. Riggsbee, C. W.; Deiters, A., Recent advances in the photochemical control of protein function. *Trends Biotechnol* **2010,** *28* (9), 468-75.
25. Deiters, A., Principles and applications of the photochemical control of cellular processes. *Chembiochem* **2010,** *11* (1), 47-53.
26. Deiters, A., Light activation as a method of regulating and studying gene expression. *Curr Opin Chem Biol* **2009,** *13* (5-6), 678-86.
27. Young, D. D.; Deiters, A., Photochemical control of biological processes. *Org Biomol Chem* **2007,** *5* (7), 999-1005.
28. Prokup, A.; Hemphill, J.; Liu, Q.; Deiters, A., Optically Controlled Signal Amplification for DNA Computation. *ACS synthetic biology* **2015**.
29. Choi, H. M.; Chang, J. Y.; Trinh, l. A.; Padilla, J. E.; Fraser, S. E.; Pierce, N. A., Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat Biotechnol* **2010,** *28* (11), 1208-12.
30. Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E., An autonomous molecular computer for logical control of gene expression. *Nature* **2004,** *429* (6990), 423-9.
31. Dunlop, M. J., Engineering microbes for tolerance to next-generation biofuels. *Biotechnology for biofuels* **2011,** *4*, 32.
32. Young, D. D.; Edwards, W. F.; Lusic, H.; Lively, M. O.; Deiters, A., Light-triggered polymerase chain reaction. *Chem Commun (Camb)* **2008,** (4), 462-4.
33. Lusic, H.; Young, D. D.; Lively, M. O.; Deiters, A., Photochemical DNA activation. *Org Lett* **2007,** *9* (10), 1903-6.
34. Young, D. D.; Govan, J. M.; Lively, M. O.; Deiters, A., Photochemical regulation of restriction endonuclease activity. *Chembiochem* **2009,** *10* (10), 1612-6.
35. Heckel, A.; Mayer, G., Light regulation of aptamer activity: an anti-thrombin aptamer with caged thymidine nucleobases. *J Am Chem Soc* **2005,** *127* (3), 822-3.

36. Young, D. D.; Garner, R. A.; Yoder, J. A.; Deiters, A., Light-activation of gene function in mammalian cells via ribozymes. *Chem Commun (Camb)* **2009,** (5), 568-70.
37. Richards, J. L.; Seward, G. K.; Wang, Y. H.; Dmochowski, I. J., Turning the 10-23 DNAzyme on and off with light. *Chembiochem* **2010,** *11* (3), 320-4.
38. Höbartner, C.; Silverman, S. K., Modulation of RNA tertiary folding by incorporation of caged nucleotides. *Angew Chem Int Ed Engl* **2005,** *44* (44), 7305-9.
39. Young, D. D.; Lusic, H.; Lively, M. O.; Yoder, J. A.; Deiters, A., Gene silencing in mammalian cells with light-activated antisense agents. *Chembiochem* **2008,** *9* (18), 2937-40.
40. Young, D. D.; Lively, M. O.; Deiters, A., Activation and deactivation of DNAzyme and antisense function with light for the photochemical regulation of gene expression in mammalian cells. *J Am Chem Soc* **2010,** *132* (17), 6183-93.
41. Deiters, A.; Garner, R. A.; Lusic, H.; Govan, J. M.; Dush, M.; Nascone-Yoder, N. M.; Yoder, J. A., Photocaged morpholino oligomers for the light-regulation of gene function in zebrafish and Xenopus embryos. *J Am Chem Soc* **2010,** *132* (44), 15644-50.
42. Tang, X.; Swaminathan, J.; Gewirtz, A. M.; Dmochowski, I. J., Regulating gene expression in human leukemia cells using light-activated oligodeoxynucleotides. *Nucleic Acids Res* **2008,** *36* (2), 559-69.
43. Shestopalov, I. A.; Sinha, S.; Chen, J. K., Light-controlled gene silencing in zebrafish embryos. *Nat Chem Biol* **2007,** *3* (10), 650-1.
44. Shah, S.; Jain, P. K.; Kala, A.; Karunakaran, D.; Friedman, S. H., Light-activated RNA interference using double-stranded siRNA precursors modified using a remarkable regiospecificity of diazo-based photolabile groups. *Nucleic Acids Res* **2009,** *37* (13), 4508-17.
45. Blidner, R. A.; Svoboda, K. R.; Hammer, R. P.; Monroe, W. T., Photoinduced RNA interference using DMNPE-caged 2'-deoxy-2'-fluoro substituted nucleic acids in vitro and in vivo. *Mol Biosyst* **2008,** *4* (5), 431-40.
46. Mikat, V.; Heckel, A., Light-dependent RNA interference with nucleobase-caged siRNAs. *RNA* **2007,** *13* (12), 2341-7.
47. Cunningham, C. W.; Mukhopadhyay, A.; Lushington, G. H.; Blagg, B. S.; Prisinzano, T. E.; Krise, J. P., Uptake, distribution and diffusivity of reactive fluorophores in cells: implications toward target identification. *Mol Pharm* **2010,** *7* (4), 1301-10.
48. Young, D. D.; Lusic, H.; Lively, M. O.; Deiters, A., Restriction enzyme-free mutagenesis via the light regulation of DNA polymerization. *Nucleic Acids Res* **2009,** *37* (8), e58.
49. Chen, X.; Ellington, A. D., Shaping up nucleic acid computation. *Curr Opin Biotechnol* **2010,** *21* (4), 392-400.
50. Janssen, B. M.; van Rosmalen, M.; van Beek, L.; Merkx, M., Antibody activation using DNA-based logic gates. *Angew Chem Int Ed Engl* **2015,** *54* (8), 2530-3.
51. You, M.; Zhu, G.; Chen, T.; Donovan, M. J.; Tan, W., Programmable and multiparameter DNA-based logic platform for cancer recognition and targeted therapy. *J Am Chem Soc* **2015,** *137* (2), 667-74.
52. Prokup, A.; Deiters, A., Interfacing synthetic DNA logic operations with protein outputs. *Angew Chem Int Ed Engl* **2014,** *53* (48), 13192-5.
53. Gabriel, R.; Lombardo, A.; Arens, A.; Miller, J. C.; Genovese, P.; Kaeppel, C.; Nowrouzi, A.; Bartholomae, C. C.; Wang, J.; Friedman, G.; Holmes, M. C.; Gregory, P.

D.; Glimm, H.; Schmidt, M.; Naldini, L.; von Kalle, C., An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **2011,** *29* (9), 816-23.
54. Choo, Y.; Sánchez-García, I.; Klug, A., In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **1994,** *372* (6507), 642-5.
55. Minczuk, M.; Papworth, M. A.; Kolasinska, P.; Murphy, M. P.; Klug, A., Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. *Proc Natl Acad Sci U S A* **2006,** *103* (52), 19689-94.
56. Reynolds, L.; Ullman, C.; Moore, M.; Isalan, M.; West, M. J.; Clapham, P.; Klug, A.; Choo, Y., Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *Proc Natl Acad Sci U S A* **2003,** *100* (4), 1615-20.
57. Miller, J.; McLachlan, A. D.; Klug, A., Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **1985,** *4* (6), 1609-14.
58. Beerli, R. R.; Schopfer, U.; Dreier, B.; Barbas, C. F., Chemically regulated zinc finger transcription factors. *J Biol Chem* **2000,** *275* (42), 32617-27.
59. Segal, D. J.; Crotty, J. W.; Bhakta, M. S.; Barbas, C. F.; Horton, N. C., Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol* **2006,** *363* (2), 405-21.
60. Rebar, E. J.; Pabo, C. O., Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **1994,** *263* (5147), 671-3.
61. Jester, B. W.; Cox, K. J.; Gaj, A.; Shomin, C. D.; Porter, J. R.; Ghosh, I., A coiled-coil enabled split-luciferase three-hybrid system: applied toward profiling inhibitors of protein kinases. *J Am Chem Soc* **2010,** *132* (33), 11727-35.
62. Hashimoto, T.; Adams, K. W.; Fan, Z.; McLean, P. J.; Hyman, B. T., Characterization of oligomer formation of amyloid-beta peptide using a split-luciferase complementation assay. *J Biol Chem* **2011,** *286* (31), 27081-91.
63. Furman, J. L.; Badran, A. H.; Ajulo, O.; Porter, J. R.; Stains, C. I.; Segal, D. J.; Ghosh, I., Toward a general approach for RNA-templated hierarchical assembly of split-proteins. *J Am Chem Soc* **2010,** *132* (33), 11692-701.
64. Ghosh, I.; Stains, C. I.; Ooi, A. T.; Segal, D. J., Direct detection of double-stranded DNA: Molecular methods and applications for DNA diagnostics. *Mol Biosyst* **2006,** *2* (11), 551-60.
65. Shekhawat, S. S.; Porter, J. R.; Sriprasad, A.; Ghosh, I., An autoinhibited coiled-coil design strategy for split-protein protease sensors. *J Am Chem Soc* **2009,** *131* (42), 15284-90.
66. Stains, C. I.; Porter, J. R.; Ooi, A. T.; Segal, D. J.; Ghosh, I., DNA sequence-enabled reassembly of the green fluorescent protein. *J Am Chem Soc* **2005,** *127* (31), 10782-3.
67. Broome, A. M.; Bhavsar, N.; Ramamurthy, G.; Newton, G.; Basilion, J. P., Expanding the utility of beta-galactosidase complementation: piece by piece. *Mol Pharm* **2010,** *7* (1), 60-74.
68. Wehr, M. C.; Laage, R.; Bolz, U.; Fischer, T. M.; Grünewald, S.; Scheek, S.; Bach, A.; Nave, K. A.; Rossner, M. J., Monitoring regulated protein-protein interactions using split TEV. *Nat Methods* **2006,** *3* (12), 985-93.

69. Stojanović, M. N.; Stefanović, D., Deoxyribozyme-based half-adder. *J Am Chem Soc* **2003,** *125* (22), 6673-6.
70. Credi, A., Molecules that make decisions. *Angew Chem Int Ed Engl* **2007,** *46* (29), 5472-5.
71. Zhou, J.; Arugula, M. A.; Halámek, J.; Pita, M.; Katz, E., Enzyme-based NAND and NOR logic gates with modular design. *J Phys Chem B* **2009,** *113* (49), 16065-70.
72. Medina, P. P.; Nolde, M.; Slack, F. J., OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **2010,** *467* (7311), 86-90.
73. Coulouarn, C.; Factor, V. M.; Andersen, J. B.; Durkin, M. E.; Thorgeirsson, S. S., Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* **2009,** *28* (40), 3526-36.
74. Jopling, C., Liver-specific microRNA-122: Biogenesis and function. *RNA Biol* **2012,** *9* (2), 137-42.
75. Gardner, L.; Deiters, A., Light-controlled synthetic gene circuits. *Curr Opin Chem Biol* **2012,** *16* (3-4), 292-9.
76. Roda, A.; Pasini, P.; Mirasoli, M.; Michelini, E.; Guardigli, M., Biotechnological applications of bioluminescence and chemiluminescence. *Trends Biotechnol* **2004,** *22* (6), 295-303.
77. Chen, X.; Briggs, N.; McLain, J. R.; Ellington, A. D., Stacking nonenzymatic circuits for high signal gain. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5386-91.
78. Zhang, Y.; Tang, Z.; Wang, J.; Wu, H.; Maham, A.; Lin, Y., Hairpin DNA switch for ultrasensitive spectrophotometric detection of DNA hybridization based on gold nanoparticles and enzyme signal amplification. *Anal Chem* **2010,** *82* (15), 6440-6.
79. Jullien, N.; Sampieri, F.; Enjalbert, A.; Herman, J. P., Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res* **2003,** *31* (21), e131.
80. Jung, C.; Ellington, A. D., Diagnostic applications of nucleic acid circuits. *Acc Chem Res* **2014,** *47* (6), 1825-35.
81. Krishnan, Y.; Simmel, F. C., Nucleic acid based molecular devices. *Angew Chem Int Ed Engl* **2011,** *50* (14), 3124-56.
82. Anderson, J. C.; Voigt, C. A.; Arkin, A. P., Environmental signal integration by a modular AND gate. *Mol Syst Biol* **2007,** *3*, 133.
83. Tabor, J. J.; Salis, H. M.; Simpson, Z. B.; Chevalier, A. A.; Levskaya, A.; Marcotte, E. M.; Voigt, C. A.; Ellington, A. D., A synthetic genetic edge detection program. *Cell* **2009,** *137* (7), 1272-81.
84. Zhang, D. Y.; Winfree, E., Dynamic allosteric control of noncovalent DNA catalysis reactions. *J Am Chem Soc* **2008,** *130* (42), 13921-6.
85. Duan, R.; Zuo, X.; Wang, S.; Quan, X.; Chen, D.; Chen, Z.; Jiang, L.; Fan, C.; Xia, F., Lab in a tube: ultrasensitive detection of microRNAs at the single-cell level and in breast cancer patients using quadratic isothermal amplification. *J Am Chem Soc* **2013,** *135* (12), 4604-7.
86. Yin, P.; Choi, H. M.; Calvert, C. R.; Pierce, N. A., Programming biomolecular self-assembly pathways. *Nature* **2008,** *451* (7176), 318-22.
87. He, L.; Hannon, G. J., MicroRNAs: small RNAs with a big role in gene regulation. *Nature reviews. Genetics* **2004,** *5* (7), 522-31.

88. Mitchell, P. S.; Parkin, R. K.; Kroh, E. M.; Fritz, B. R.; Wyman, S. K.; Pogosova-Agadjanyan, E. L.; Peterson, A.; Noteboom, J.; O'Briant, K. C.; Allen, A.; Lin, D. W.; Urban, N.; Drescher, C. W.; Knudsen, B. S.; Stirewalt, D. L.; Gentleman, R.; Vessella, R. L.; Nelson, P. S.; Martin, D. B.; Tewari, M., Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **2008,** *105* (30), 10513-8.
89. Kroh, E. M.; Parkin, R. K.; Mitchell, P. S.; Tewari, M., Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* **2010,** *50* (4), 298-301.
90. Chen, J.; Lozach, J.; Garcia, E. W.; Barnes, B.; Luo, S.; Mikoulitch, I.; Zhou, L.; Schroth, G.; Fan, J. B., Highly sensitive and specific microRNA expression profiling using BeadArray technology. *Nucleic Acids Res* **2008,** *36* (14), e87.
91. Zampetaki, A.; Mayr, M., Analytical challenges and technical limitations in assessing circulating miRNAs. *Thrombosis and haemostasis* **2012,** *108* (4), 592-8.
92. Li, B.; Chen, X.; Ellington, A. D., Adapting enzyme-free DNA circuits to the detection of loop-mediated isothermal amplification reactions. *Anal Chem* **2012,** *84* (19), 8371-7.
93. Bi, S.; Zhao, T.; Luo, B.; Zhu, J. J., Hybridization chain reaction-based branched rolling circle amplification for chemiluminescence detection of DNA methylation. *Chem Commun (Camb)* **2013,** *49* (61), 6906-8.
94. Mitsis, P. G.; Kwagh, J. G., Characterization of the interaction of lambda exonuclease with the ends of DNA. *Nucleic Acids Res* **1999,** *27* (15), 3057-63.
95. Zuo, X.; Xia, F.; Xiao, Y.; Plaxco, K. W., Sensitive and selective amplified fluorescence DNA detection based on exonuclease III-aided target recycling. *J Am Chem Soc* **2010,** *132* (6), 1816-8.
96. Duan, R.; Zuo, X.; Wang, S.; Quan, X.; Chen, D.; Chen, Z.; Jiang, L.; Fan, C.; Xia, F., Quadratic isothermal amplification for the detection of microRNA. *Nat Protoc* **2014,** *9* (3), 597-607.
97. Liu, X. P.; Liu, J. H., The terminal 5' phosphate and proximate phosphorothioate promote ligation-independent cloning. *Protein Sci* **2010,** *19* (5), 967-73.
98. Li, B.; Ellington, A. D.; Chen, X., Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. *Nucleic Acids Res* **2011,** *39* (16), e110.
99. Zhang, D. Y.; Winfree, E., Robustness and modularity properties of a non-covalent DNA catalytic reaction. *Nucleic Acids Res* **2010,** *38* (12), 4182-97.
100. Collins, M. L.; Irvine, B.; Tyner, D.; Fine, E.; Zayati, C.; Chang, C.; Horn, T.; Ahle, D.; Detmer, J.; Shen, L. P.; Kolberg, J.; Bushnell, S.; Urdea, M. S.; Ho, D. D., A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res* **1997,** *25* (15), 2979-84.
101. Seelig, G.; Yurke, B.; Winfree, E., Catalyzed relaxation of a metastable DNA fuel. *J Am Chem Soc* **2006,** *128* (37), 12211-20.
102. Choi, H. M.; Beck, V. A.; Pierce, N. A., Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS nano* **2014,** *8* (5), 4284-94.
103. Govan, J. M.; Lively, M. O.; Deiters, A., Photochemical control of DNA decoy function enables precise regulation of nuclear factor kappaB activity. *J Am Chem Soc* **2011,** *133* (33), 13176-82.

104. Huang, F.; You, M.; Han, D.; Xiong, X.; Liang, H.; Tan, W., DNA branch migration reactions through photocontrollable toehold formation. *J Am Chem Soc* **2013,** *135* (21), 7967-73.
105. Cantor, C. R.; Warshaw, M. M.; Shapiro, H., Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* **1970,** *9* (9), 1059-77.
106. Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W., Traceless and site-specific ubiquitination of recombinant proteins. *J Am Chem Soc* **2011,** *133* (28), 10708-11.
107. Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; VanDelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A., Click strategies for single-molecule protein fluorescence. *J Am Chem Soc* **2012,** *134* (11), 5187-95.
108. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010,** *79*, 413-44.
109. Minnihan, E. C.; Young, D. D.; Schultz, P. G.; Stubbe, J., Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J Am Chem Soc* **2011,** *133* (40), 15942-5.
110. Yu, Y.; Lv, X.; Li, J.; Zhou, Q.; Cui, C.; Hosseinzadeh, P.; Mukherjee, A.; Nilges, M. J.; Wang, J.; Lu, Y., Defining the role of tyrosine and rational tuning of oxidase activity by genetic incorporation of unnatural tyrosine analogs. *J Am Chem Soc* **2015,** *137* (14), 4594-7.
111. Chou, C.; Uprety, R.; Davis, L.; Chin, J.; Deiters, A., Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chemical Science* **2011,** *2* (3), 480-483.
112. Zhang, M.; Lin, S. X.; Song, X. W.; Liu, J.; Fu, Y.; Ge, X.; Fu, X. M.; Chang, Z. Y.; Chen, P. R., A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nature Chemical Biology* **2011,** *7* (10), 671-677.
113. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.
114. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **1998,** *95* (12), 6705-10.
115. Strable, E.; Prasuhn, D. E.; Udit, A. K.; Brown, S.; Link, A. J.; Ngo, J. T.; Lander, G.; Quispe, J.; Potter, C. S.; Carragher, B.; Tirrell, D. A.; Finn, M. G., Unnatural amino acid incorporation into virus-like particles. *Bioconjug Chem* **2008,** *19* (4), 866-75.
116. Roesser, J. R.; Xu, C.; Payne, R. C.; Surratt, C. K.; Hecht, S. M., Preparation of misacylated aminoacyl-tRNA(Phe)'s useful as probes of the ribosomal acceptor site. *Biochemistry* **1989,** *28* (12), 5185-95.
117. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* **2007,** *4* (3), 239-44.
118. Gautier, A.; Deiters, A.; Chin, J. W., Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc* **2011,** *133* (7), 2124-7.
119. Young, T.; Schultz, P., Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry* **2010,** *285* (15), 11039-11044.
120. Xie, J.; Schultz, P. G., An expanding genetic code. *Methods* **2005,** *36* (3), 227-38.
121. Dougherty, D. A., Unnatural amino acids as probes of protein structure and function. *Curr Opin Chem Biol* **2000,** *4* (6), 645-52.

122. Wilkins, B. J.; Marionni, S.; Young, D. D.; Liu, J.; Wang, Y.; Di Salvo, M. L.; Deiters, A.; Cropp, T. A., Site-specific incorporation of fluorotyrosines into proteins in Escherichia coli by photochemical disguise. *Biochemistry* **2010,** *49* (8), 1557-9.
123. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.
124. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G., An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **2010,** *395* (2), 361-74.
125. Chen, S.; Schultz, P. G.; Brock, A., An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in Saccharomyces cerevisiae. *J Mol Biol* **2007,** *371* (1), 112-22.
126. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S., Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nature structural biology* **2003,** *10* (6), 425-32.
127. Guo, J.; Melancon, C. E., 3rd; Lee, H. S.; Groff, D.; Schultz, P. G., Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew Chem Int Ed Engl* **2009,** *48* (48), 9148-51.
128. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Proc Natl Acad Sci U S A* **2003,** *100* (1), 56-61.
129. Ryu, Y.; Schultz, P. G., Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat Methods* **2006,** *3* (4), 263-5.
130. Blatter, N.; Prokup, A.; Deiters, A.; Marx, A., Modulating the pKa of a tyrosine in KlenTaq DNA polymerase that is crucial for abasic site bypass by in vivo incorporation of a non-canonical amino acid. *Chembiochem* **2014,** *15* (12), 1735-7.
131. Lindahl, T., Instability and decay of the primary structure of DNA. *Nature* **1993,** *362* (6422), 709-15.
132. Loeb, L. A.; Preston, B. D., Mutagenesis by apurinic/apyrimidinic sites. *Annual review of genetics* **1986,** *20*, 201-30.
133. Lindahl, T.; Nyberg, B., Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **1972,** *11* (19), 3610-8.
134. Hubscher, U.; Maga, G.; Spadari, S., Eukaryotic DNA polymerases. *Annu Rev Biochem* **2002,** *71*, 133-63.
135. Goodman, M. F., Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **2002,** *71*, 17-50.
136. Lawrence, C. W.; Borden, A.; Banerjee, S. K.; LeClerc, J. E., Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res* **1990,** *18* (8), 2153-7.
137. (a) Sagher, D.; Strauss, B., Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: uniqueness of adenine nucleotides. *Biochemistry* **1983,** *22* (19), 4518-26; (b) Schaaper, R. M.; Kunkel, T. A.; Loeb, L. A., Infidelity of DNA synthesis associated with bypass of apurinic sites. *Proc Natl Acad Sci U S A* **1983,** *80* (2), 487-91; (c) Goodman, M. F.; Cai, H.; Bloom, L. B.; Eritja, R., Nucleotide insertion and primer extension at abasic template sites in different sequence contexts. *Annals of the New York Academy of Sciences* **1994,** *726*, 132-42; discussion 142-3; (d) Shibutani, S.; Takeshita, M.; Grollman, A. P., Translesional

synthesis on DNA templates containing a single abasic site. A mechanistic study of the "A rule". *J Biol Chem* **1997,** *272* (21), 13916-22; (e) Avkin, S.; Adar, S.; Blander, G.; Livneh, Z., Quantitative measurement of translesion replication in human cells: evidence for bypass of abasic sites by a replicative DNA polymerase. *Proc Natl Acad Sci U S A* **2002,** *99* (6), 3764-9; (f) Pages, V.; Johnson, R. E.; Prakash, L.; Prakash, S., Mutational specificity and genetic control of replicative bypass of an abasic site in yeast. *Proc Natl Acad Sci U S A* **2008,** *105* (4), 1170-5.
138. Obeid, S.; Blatter, N.; Kranaster, R.; Schnur, A.; Diederichs, K.; Welte, W.; Marx, A., Replication through an abasic DNA lesion: structural basis for adenine selectivity. *EMBO J* **2010,** *29* (10), 1738-47.
139. Obeid, S.; Welte, W.; Diederichs, K.; Marx, A., Amino Acid Templating Mechanisms in Selection of Nucleotides Opposite Abasic Sites by a Family A DNA Polymerase. *Journal of Biological Chemistry* **2012,** *287* (17), 14099-14108.
140. Brooks, B.; Phillips, R. S.; Benisek, W. F., High-efficiency incorporation in vivo of tyrosine analogues with altered hydroxyl acidity in place of the catalytic tyrosine-14 of Delta 5-3-ketosteroid isomerase of Comamonas (Pseudomonas) testosteroni: effects of the modifications on isomerase kinetics. *Biochemistry* **1998,** *37* (27), 9738-42.
141. (a) Seyedsayamdost, M. R.; Yee, C. S.; Reece, S. Y.; Nocera, D. G.; Stubbe, J., pH Rate profiles of FnY356-R2s (n = 2, 3, 4) in Escherichia coli ribonucleotide reductase: evidence that Y356 is a redox-active amino acid along the radical propagation pathway. *J Am Chem Soc* **2006,** *128* (5), 1562-8; (b) Seyedsayamdost, M. R.; Reece, S. Y.; Nocera, D. G.; Stubbe, J., Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J Am Chem Soc* **2006,** *128* (5), 1569-79.
142. Thorson, J. S.; Chapman, E.; Murphy, E. C.; Schultz, P. G.; Judice, J. K., Linear Free-Energy Analysis of Hydrogen-Bonding in Proteins. *Journal of the American Chemical Society* **1995,** *117* (3), 1157-1158.
143. Nagasawa, T.; Utagawa, T.; Goto, J.; Kim, C. J.; Tani, Y.; Kumagai, H.; Yamada, H., Syntheses of L-tyrosine-related amino acids by tyrosine phenol-lyase of Citrobacter intermedius. *Eur J Biochem* **1981,** *117* (1), 33-40.
144. Holzberger, B.; Rubini, M.; Möller, H. M.; Marx, A., A highly active DNA polymerase with a fluorous core. *Angew Chem Int Ed Engl* **2010,** *49* (7), 1324-7.
145. Aiba, H., Mechanism of RNA silencing by Hfq-binding small RNAs. *Current opinion in microbiology* **2007,** *10* (2), 134-9.
146. Frohlich, K. S.; Vogel, J., Activation of gene expression by small RNA. *Current opinion in microbiology* **2009,** *12* (6), 674-82.
147. Folichon, M.; Arluison, V.; Pellegrini, O.; Huntzinger, E.; Regnier, P.; Hajnsdorf, E., The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **2003,** *31* (24), 7302-10.
148. Masse, E.; Escorcia, F. E.; Gottesman, S., Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes & development* **2003,** *17* (19), 2374-83.
149. Link, T. M.; Valentin-Hansen, P.; Brennan, R. G., Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A* **2009,** *106* (46), 19292-7.

150. Schumacher, M. A.; Pearson, R. F.; Moller, T.; Valentin-Hansen, P.; Brennan, R. G., Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* **2002,** *21* (13), 3546-56.
151. Brennan, R. G.; Link, T. M., Hfq structure, function and ligand binding. *Current opinion in microbiology* **2007,** *10* (2), 125-33.
152. Faner, M. A.; Feig, A. L., Identifying and characterizing Hfq-RNA interactions. *Methods* **2013,** *63* (2), 144-59.
153. Tree, J. J.; Granneman, S.; McAteer, S. P.; Tollervey, D.; Gally, D. L., Identification of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. *Molecular cell* **2014,** *55* (2), 199-213.
154. Panja, S.; Schu, D. J.; Woodson, S. A., Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res* **2013,** *41* (15), 7536-46.