2022

Characterization of a Malachite Green DNA Aptamer and Development of Malachite Green DNA Aptamer Sensor

Martin O'Steen
University of Central Florida

CHARACTERIZATION OF A MALACHITE GREEN DNA APTAMER AND DEVELPOMENT OF MALACHITE GREEN DNA APTAMER SENSOR

By

MARTIN O'STEEN
B.S. University of Central Florida, 2014
MS. University of Central Florida, 2020

A dissertation submitted in partial fulfillment of the requirements
For the Degree of Doctor of Philosophy
In the Department of Chemistry
In the College of Sciences
At the University of Central Florida
Orlando, Florida

Fall Term 2022

Major Professor: Dmitry M. Kolpashchikov

ABSTRACT

Aptamers are functional nucleic acid (FNA) species capable of specifically binding a target ligand with high affinity generated through a combinatorial chemistry technique known as SELEX, or systematic evolution of ligands through exponential enrichment. Often compared to antibodies, aptamers enjoy a distinct advantage by being the product of a synthetic technique which allows for a larger of prospective targets and the ability to be tailored in an application specific manner. The inherent modularity of their nucleic acid base and the wide range of perspective targets make aptamers particularly suitable for development of biosensors.

A subset of aptamers, Light Up Aptamers, are able to bind non-fluorescent ligands. This binding often institutes torsional restrictions of the cognate ligand, resulting in the generation of the fluorescence. These subset of aptamers make for enticing prospects for the development of label free biosensors that obviate the need for expensive covalent modifications common to other hybridization based approaches. One such example, a DNA malachite green aptamer, was developed into a split aptameric biosensor capable of detecting SARS-COVID19 sequences with a sensitivity in the lower nM range. This was then used in concert with another light up aptamer to form the first example of a solution based multiplex assay comprised of self-assembling aptamer sensors. Further work with this malachite green aptamer determined that the functional aspect of this aptamer can be attributed to 6 total nucleotides, making it one of the smallest aptamers known to date. Characterization of the sequence and structural requirements of this minimized aptamer provides a more comprehensive understanding and can allow for its application in a host of bioanalytic techniques.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Evan Cornett and Dr. Carlo Ledezma, without your patience and guidance I would not be here today. Marcella whose support kept me going even in the worst of times. My friends and family who are responsible for this as much as I am. Lastly to the people who have been in life and are no longer here, each of you contributed in your own way and I can't ever express enough gratitude. Additionally I would like to acknowledge that this work was supported by the National Science Foundation through the CCF: Software and Hardware Foundations under cooperative agreement SHF-1907824.

DEDICATION

I would like to dedicate this to my father. He always wanted me to have the education that he never could. He may no longer be with us but he is never forgotten.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid
XNA	Xeno Nucleic Acid
PNA	Peptide Nucleic Acid
SMG	Split Malachite Green Aptamer Sensor
SDA	Split Dapoxyl Aptamer Sensory
MG	Malachite Green
AO	Auramine O
Mini-MG	Mini-Malachite Green Aptamer
MGB	Malachite Green B Strand
MGA	Malachite Green A Strand
SMG-A	Split Malachite Green Aptamer Sensor Strand A
SMG-B	Split Malachite Green Aptamer Sensor Strand B
NC	Novel Coronavirus

CHAPTER 1: INTRODUCTION

Nucleic acids have occupied a central role in biology since their function in the hereditary transfer of information, for deoxyribonucleic acid (DNA), and the direction of protein synthesis, via ribonucleic acid (RNA), were established.[2, 3] Initially the roles of nucleic acids in the cellular context was passive, acting as information carriers and as blueprints for the production of proteins necessary for cellular function. It was not until the discovery of the ribosome, the macromolecular complex responsible for protein synthesis, that that the roles of nucleic acids were expanded. Initially, ribosomal RNA (rRNA) was relegated to acting as a structural element in ribosomes, facilitating the necessary positioning of ribosomal proteins to achieve optimal protein synthesis. It was not until a series of seminal experiments utilizing deproteinized ribosomes that the true scope of rRNA and its contribution to protein synthesis was established.[2] Far from being a passive element, rRNA was found to be the principal actor in protein synthesis via their function in catalyzing peptide bond formation. This discovery fundamental altered the role of nucleic acids in biology and presented the possibility that other, none rRNA, nucleic acids could perform similar catalysis among other functions.

The study of nucleic acids and their roles in biology was accelerated by the discovery of polymerases that could function at higher temperatures and the development of its associated amplification technique, polymerase chain reaction (PCR).[4] This technique provided a reliable method of amplifying nucleic acids from small initial amounts with high sequence fidelity and its widespread adoption rapidly advanced genomic science to scale not previously obtainable with standard synthetic techniques. Building upon this, two independent groups concurrently developed a combinatorial chemistry technique that could exploit the straightforward synthesis

and enzymatic compatibility of oligonucleotides to produce nucleic acids species that perform novel functions not seen in nature. This technique was later dubbed SELEX, Sequential Evolution of Ligands by Exponential Enrichment.[5, 6]

SELEX is a combinatorial chemical technique in which a large library of randomized oligonucleotides are challenged to perform a specific task. Through iterative cycles of partitioning functional and non-functional library members with subsequent amplification, the initial library's composition is changed becoming enriched in functional sequences. This procedure continues until no further changes in the library's composition is observed, at which point the library is sequenced to provide the identity of the members. In some aspects the term evolution is a misnomer, as the term evolution implies that the sequence space sampled by the library is altered through subsequent rounds by the influence of selective pressures. In reality, SELEX is a technique not for generation of novel sequences, but rather for the isolation of predefined sequences that are intrinsically capable of performing the desired function. This isolation requires sequence information to be preserved from the initial library state throughout the selection, as loss of even one to two bases worth of information can result in inactive species. As the selected oligonucleotide's function extends from its sequence and the subsequent 3D shape of its fold.[7-9] The product of this technique produces oligonucleotides that can perform a desired function, deemed functional nucleic acids (FNAs). These FNAs can be stratified into two groups based upon their activity. Oligonucleotides that perform a catalytic function are deemed deoxyribozymes, for DNA, and ribozymes, for RNA.[10] If the oligonucleotide's function is to bind a target, the resulting FNA is an aptamer. Subsequent discussion of SELEX technique will

focus on its application to the selection of aptamers and the influence of the selection process on the end product rather than catalytic nucleic acids.

The general SELEX process, shown in Figure 1, can be divided into three phases: library construction, partitioning and regeneration/sequencing. Holistically, SELEX technologies are considered matured after several decades of refinement and study[11-13]. Still selections are a technical challenge, and the majority of selections fail to produce a functional product after sequencing. Estimates of the success rate of SELEX can vary but have been reported to have an upper limit 30%, which is most likely an overestimation based on the reporting bias of only successful selections.[14-17] This failure rate is difficult to attribute to any singular cause, as many aspects of the selection process are understood but not precisely described. Much of this difficulty arises from the scare information concerning the library in its initial state and a lack of control over unintended interactions between library members.[18] These interactions can negatively impact the selection, leading to longer selections, low-affinity end products, or complete failure.[18, 19] Compounding this lack of control is that the proportion of functional sequences within any given library is incredibly small compared to the number of total members. This minority status remains even after significant enrichment, with one selection estimating that 25.8% of the final library was composed of functional sequences after 10 rounds of enrichment[20]. This scarcity means that any loss of functional members early in the selection process can deplete the library of potential hits, leaving non-functional sequences to dominant the pool. Even with these difficulties and limitations, SELEX has demonstrated to be a powerful combinatorial technique which warrants further discussion.

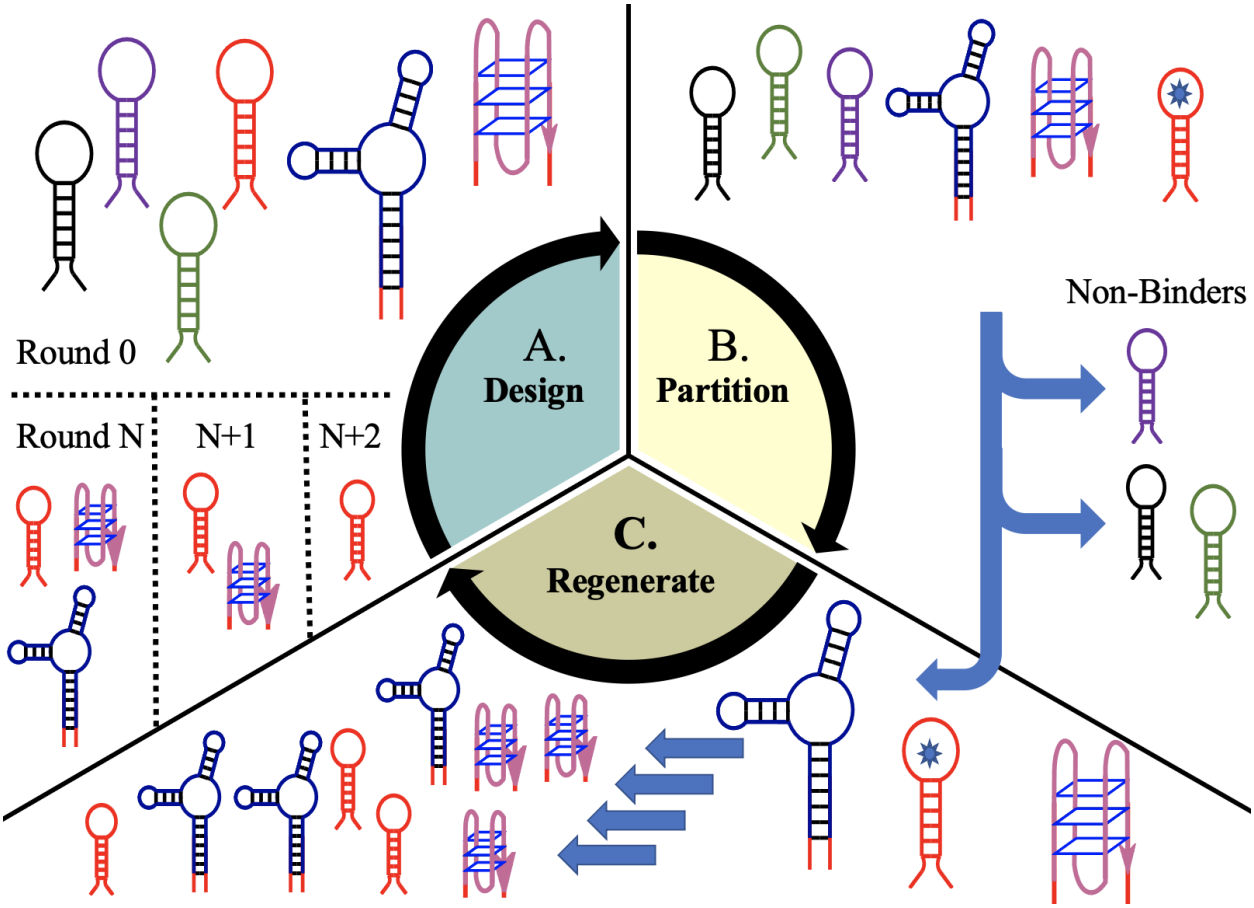


Figure 1. Illustration of SELEX Process. SELEX can be roughly divided into 3 distinct phase involving library design (A), partitioning (B) and regeneration of the library (C). The initial library consists of a series of randomized oligonucleotides which assume a range of secondary structures such hairpins or G-quadruplexes. Here the desired aptamer is shown, highlighted red, and is the minority of the library. Partitioning results the separation of non-binders and binders from the library, as the target is shown as a star contained within the red aptamer. This partition reduces the overall sequence diversity of the library and results in a significant loss of mass. Regeneration of the library, shown in C, involves amplification of the remaining species followed by processing to return the library to its initial single stranded state. This library is then utilized in as the input material for successive rounds. The diversity of the remaining library is reduced progressively (as shown in Round 0, N, N+1, etc. until further enrichment is not achieved, and the remaining oligonucleotides are sequenced.

The most relevant aspects of the SELEX procedure to the functioning of the resulting aptamer lie with the initial construction of the library and the conditions utilized during the partitioning phase[21, 22]. The choices made for these steps dictate the general approach of the selection, for example amplification strategies, and impose a series of constraints upon which the final aptamer will optimally function. Given the importance of these areas, more discussion of

these stages is needed. The other stages, such as regeneration and sequencing can also influence the resulting aptamer but to a lesser extent. As such discussion of the SELEX process will focus on the initial library construction and the partitioning step and how these parameters affect the resulting aptamer.

Library Construction:

The initial phase, library construction, revolves around three axes: polymer type, library size and the inclusion of modified nitrogenous bases. The type of polymer references the initial decision between using deoxyribose (for DNA selections), ribose (for RNA selections) or less commonly Xeno Nucleic Acids (XNA), which encompass all non-phosphodiester backbone nucleic acid analogs[23]. This initial selection is determined, in part, by the desired function of the aptamer. When in vivo applications are desired, the polymer of choice is typically RNA. Usage of RNA as the backbone polymer facilitates the usage of selected aptamers in in vivo applications more straightforward. As the small size, generally less than 100 nucleotides, and RNA nature allow expression of aptameric sequences in live cells through the use of standard molecular techniques[24].

The use of RNA aptamers for in vivo applications has proven fruitful, with a range of RNA aptamers specific for fluorescent analytes being utilized to track RNA processes in real time [25]. In contrast the use of DNA, with its comparatively higher stability and simpler amplification workflow, is the polymer of choice when in vitro functions are intended. DNA's greater stability than RNA is attributed to the absence of the 2' hydroxyl group on the deoxyribose sugar, as presence of the nucleophilic 2' OH on RNA accelerates the rate of

intramolecular transesterification reactions, particular in alkaline conditions[26]. This inherent instability is compounded by the ubiquity of RNases in the both the cellular and extracellular environment resulting in short half-lives, on the orders of seconds to minutes, which requires specialized handling, equipment and storage conditions. Even with these challenges, RNA dominated early selections. This dominance extended from the assumption that the diversity of RNA structures in any random library is superior to those of an analogous DNA library. This difference in diversity was attributed to the 2' OH group, an additional hydrogen bonding center, which would allow the formation of more complex secondary structures on a per-unit basis. The increased complexity, or diversity, of structures was thought to be essential to the selection higher quality aptamers as binding is a product of their specific shape.[27] While the increased structural diversity of RNA over DNA holds true for smaller libraries, this inherent advantage tends to dissipate as the overall size of the library increases. As larger libraries will intrinsically be more diverse, but at the sacrifice that any starting sample will fail to adequately represent all possible structures.

As the number of generated aptamers increased, no statistically significant difference between the Kds of RNA or DNA aptamers were found [22].While in theory the extra functionality found in RNA could lead to more interactions with its target, this relies on the binding between aptamer and target utilizing the 2' OH in its binding interface. This particular interaction is certainly dependent on individual the aptamer, as some RNA aptamers can be converted to a DNA analog and still retain binding activity even with the loss of 2' OH group[28]. More recently, the application of other nucleic acids with more exotic backbones such as PNA (peptide nucleic acids) [29], XNA (xeno nucleic acids) [30], or L-DNA[31] have been completed. These diverse

modifications are intended to increase in vivo stability, as the modified backbones are poor substrates for endogenous nucleases. However, the prohibitive cost and technical challenges in working with such heavily modified libraries have limited their wide-spread adoption, in particular as the requisite amplification step is difficult to achieve with the heavily modified oligonucleotides[32, 33]. Furthermore, the advantages of these modified backbones over standard DNA or RNA in terms of binding affinity remains an open question and may be limited to specific applications rather than a generalized advantage. After the library's polymer base has been established, the next relevant decision point rest with determining the size of library.

Library size is less impactful to the selection in comparison to polymer type or the use of modified nucleotides, which fundamentally alter aspects of the selection. Most SELEX libraries are constructed to contain a randomized region of bases flanked by two regions of defined sequence. These defined regions have a known sequence, deemed constant regions, which serve as primer binding sites for PCR amplification or as handles for immobilization during column affinity chromatography. Contained between these regions is a run of randomized nucleotides that can sample from any of the four nitrogenous bases, deemed the variable region which is where the prospective aptamer sequence is contained. The variable region for standard selections spans from 40-70 random nucleotides, producing a corresponding library which contains $1.2 * 10^{24}$ to $1.4 * 10^{40}$ distinct sequences[22, 23]. A schematic representation of a typical SELEX library is given in Figure 2. Early thought was that larger libraries, with more complex structure, would

produce higher affinity aptamers.

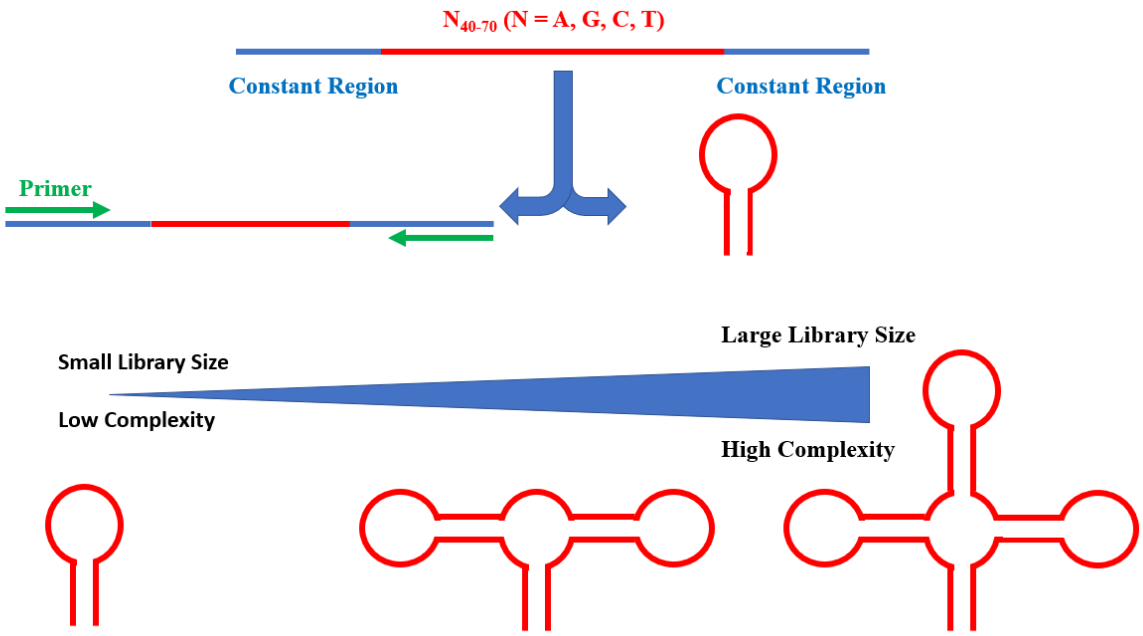


Figure 2. Abstract representation of a typical SELEX library and the relation of library size to complexity. Presented top is a diagram of a generalized SELEX library with the random region shown in red and the constant regions denoted by blue. The random region of the library is responsible for generation of the aptamer pool, represented with a hairpin structure, while the constant regions serve as primer binding sites or handles for immobilization. The bottom of the image relates random region size to complexity of possible structures.

This proposed correlation between library size and aptamer quality is not well supported via statistical analysis of over 300 reported aptamers and their associated $K_{DS,}$ as previously stated [22]. Library size might have more subtle impacts, as the size-affinity correlation is stronger when the target type, such as small molecules vs viruses, is considered. However, these correlations are variable and still relatively weak ($R < 0.5$) , which underscores the lack of fine detail in relating library size to expected outcome[22].

This lack of strong correlation probably extends from the observation that many of the interactions between the aptamer and its target, which define specificity and affinity, are

mediated by select bases of the entire sequence with other auxiliary bases providing structural support as opposed to direct interaction. An example of this disconnection in relation to the A1-Domain specific von Willebrand Factor (VMF) aptamer, ARC1172, is shown in Figure 3 [34]. Larger libraries tend to be more technically challenging to work with than smaller libraries, stemming from non-specific interactions between library members. The commonly utilized 40-70 nucleotide long random region is a fair compromise between complexity and simplicity. The next decision point lies on the inclusion of modified nucleotides, which seek to extend the chemical tool kit of the final aptamer.

The use of modified nucleotides in selections have received growing attention, as the synthetic routes to modified nucleotides are now well establish and many of the common modifications having been commercialized. The nature of these modifications can vary ranging

ARC1172:
5’GGCGTGCAGTGCCTTCGGCCGTGCGGTGCCTCCGTCACGCCT-3’

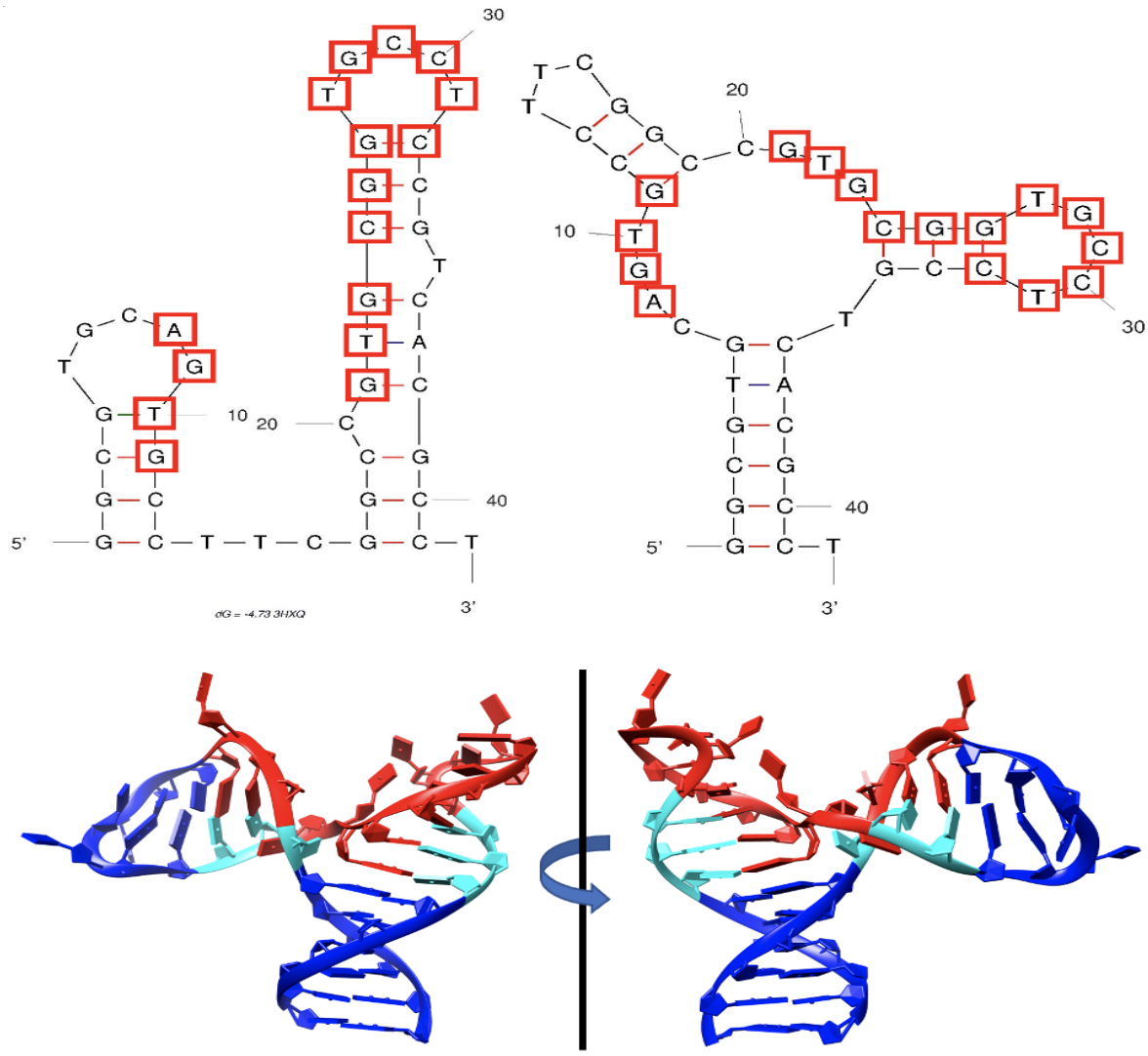


Figure 3. ARC1172 aptamers disconnection between sequence and structure. ARC1172 is an DNA aptamer with specificity toward Von William Factor A1 domain, a protein factor involved in thrombogenesis. The primary sequence, show top, is annotated to demonstrate which bases form contacts with VWF (red text) . Nucleotides that are underlined are immutable and function in ensuring proper positioning of the red nucleotides for interaction. The secondary structure prediction, shown middle, depicts many of these bases as being base-paired which would exclude them from interaction with the target and complicate development of the aptamer into a functional device as false leads. The actual structure of the aptamer, shown bottom in the absence of the protein, is color coded to demonstrate the connection between nucleotide and binding. Red residues are involved in direct contacts via a mixture of hydrogen bonding, salt bridges and van dar wall interactions. Cyan residues have no contact directly with VWF but provide structural stability to facilitate nucleotide positioning. Blue residues are not involved in either of these activities and are interchangeable provided their general secondary structure is maintained.

from simple transformations, such as addition of an alkyl group, to more complex additions such as multi-ring systems[35]. Depending on the extent of these modifications, these modified nucleotides can be incorporated during the amplification step, assuming polymerase compatibility, or added in a post amplification step using conjugative techniques[36]. The rational for including modified nucleotides stems from the need to overcome the narrow chemical space of nucleotides. Amino acids, in comparison, explore a much larger chemical space due to the variety of their side chain functionalities. Nucleotides possess a relative narrow array of chemical functionalities with little differentiation between one another. This narrow chemical space excludes certain interactions between the aptamer and target, reducing the potential binding strategies that aptamers can exploit. As such most modifications intend to provide additional strategies unavailable to standard nucleotides rather than enhancement of intrinsic capabilities. A common theme is the installation of hydrophobic moieties or the addition of acid-base functionalities that are functional in more neutral pH ranges, both areas in which natural nucleotides tend to be deficit [37]. Once the polymer type, size and modification status of the library are set, the next stage of selection is devising a partition system which efficiently separates the more numerous non-binders from the desired high affinity binders.

Partitioning:

The partition step relies on establishing the mechanism of segregating of functional and nonfunctional library members. The classical SELEX partitioning approach is a separation via affinity chromatography, in which the target is immobilized via covalent conjugation to a column material such as agarose beads. The library is then passed through the column, with functional

members binding to the immobilized target while non-binding library members are removed. These bound members are retained until denaturing conditions disrupt their interactions resulting in release from the column for collection and reamplification. The efficiency of this partitioning is a key determinant in the required number of selection rounds necessary to achieve a sufficiently enriched library, with high efficiency partitions producing shorter selection times[21]. Initially functional sequences represent a minority of the total sequence space. In smaller libraries this could amount to a few dozen distinct sequences out of potentially trillions. Given such small numbers, even high efficiency separations still take considerable time to remove competing sequences and achieve sufficient enrichment of the library for sequencing. Typical selections will encompass around 10 rounds of enrichment and the usage of larger libraries may extend the selection further into 15 rounds or more[21]. Even with a well-designed library and expertly crafted partitioning scheme, the feasibility of aptamer development relies on the interplay to the aptamers' natural characteristics and the prospective target's chemical nature.

Target characteristics influence on success or failure of selections is difficult to describe relying on general observations rather than controlled empirical study. The aptamer target space is exceptionally large and successful selections have been performed on targets ranging from ions to whole cells. This large, potentially infinite, range is a strength of aptamer selections, but decades of development have found that certain target characteristics prove to be challenging for aptamer binding. These factors are not random and correlate well with those that naturally discourage interaction with nucleic acids. Some of the strongest negative factors for successful selection include significant negative charge, hydrophobicity, and high degrees of conformational freedom.

The naturally anionic phosphodiester backbone of aptamers, in the absence of heavy modification, dissuades interactions with anionic targets. This can be attributed to the electrostatic repulsion felt by both parties, which introduces significant energetic penalties to binding. The inclusion of counter ions, namely Na^{+} and Mg^{2+}, are common constituents of the buffers used during selections which could alleviate some of this repulsion.

A second deficiency of natural nucleotides is the inability to exert significant hydrophobic interactions. While the bases themselves possess hydrophobic nature, due to their natural aromaticity. The extent this can be used for binding this is limited as the hydrophobic bases themselves contain hydrophilic edges and are punctuated by the highly hydrophilic backbone. This high hydrophilicity tends to result in the exclusion of hydrophobic regions for proteins for binding and makes small molecules with a predominately hydrophobic nature challenging targets. While the electrostatic repulsion of negative charged targets can be alleviated with the use of environmental factors, the hydrophobic effect is not so easily circumvented. The bias that this introduces are deleterious enough that some of the more commercially successful aptamer-bioanalytic ventures, namely the company Somalogic, entire technology is based around utilizing hydrophobically modified nucleotides to achieve aptamers with distinct kinetic profiles[38].

These other factors, negative charge and hydrophobicity, which dissuade aptamer development are a natural consequence of nucleic acid structure. In contrast, conformational freedom negative impact stems from aptamers exquisite specificity for their targets[39]. Aptamer binding relies on the highly flexible nature of nucleic acid backbones and the conformational freedom of the sugar bases to produce species that have high degrees of shape complementarity.

This degree of shape complementarity is so high that even subtle shifts in multi ring systems, such as those found in steroids[40], can result in significant impacts to binding. Interconversion of the target between different conformational states during selection can make existing binding pockets inefficient or lead to a complete loss in binding. In effect this conformational switching results in enrichment of a wider pool of potential aptamers, with portions of the pool being selective for one conformation over the other. This makes enrichment inconsistent between rounds slowing the selection and potentially leading to selection failure. These factors, singly or in combination, tend to dissuade aptamer development but do not prohibit it and examples of successful aptamer generation against traditionally difficult targets are found in the literature, such as the aforementioned steroid specific aptamers. In contrast, targets characteristics that would impart a natural affinity toward nucleic acids have proven to be reliable targets for aptamer generation.

Compounds that have natural affinity toward interacting with nucleic acids, such as possessing positive charge or aromatic rings, tend to increase the success rate of selections. Targets containing electropositive regions, such as those in proteins, or discrete positive charges, for small molecules, lend themselves to being excellent targets for aptamer development. This arises from the natural electrostatic attraction between the anionic aptamer and the target's opposing charge, providing additional energetic incentives for aptamer-target complex formation as the free energy of the resulting complex is lowered. The natural attraction in these regions is enticing enough that it tends to bias prospective aptamers to these discrete areas within larger targets, such as in full proteins. This can lead to a competition between library members, providing an avenue for selection of higher affinity aptamers. The natural incentive for aptamers

to bind naturally electropositive regions can tend to dissuade the development of aptamer specificity but can produce aptamers which bind with high affinity. In one study a series of aptamers were selected for platelet-derived growth factor (PDGF) at a variety of pHs, showing a positive correlation between the selected aptamers K_D and the net charge of PDGF[41].

Aptamers, utilizing natural nucleotides, contain aromatic bases which are intrinsically capable of engaging in pi stacking interactions, which are considered the major energetic driver of natural helix formation, or pi-cation interactions. Thus, small molecule targets that contain aromatic ring systems tend to interact favorably with aptamers, and these pi stacking and pi cation interactions can result intercalation of the target into the aptamer structure much as a natural base would stack within a helix. Here the relationship between shape complementarity and binding affinity is less apparent, as many of these aptamers' crystal structures tend to be relatively simple in comparison to protein binding aptamers. Rather selection in this capacity is focused around optimization of a binding pocket rather than larger changes in overt structure[42].

Taken together, both positive and negative target factors reinforce that the chemical nature of the aptamer is inexorably tied to their binding strategies. Given the natural limitations imposed by the use of natural nucleotides, the motivation for expanding the aptamer chemical toolkit via modified nucleotides is apparent. This large target space is an undisputed strength of aptamers and makes them interesting tools to bridge between classes of biomolecules or biomolecules and small molecules. This ability does come at a cost though, in that an aptamer's complex structure often requires the use of environmental conditions to achieve their proper folding. This requirement makes them intrinsically sensitivity to environmental conditions and could be considered a weakness of aptamer technologies.

The partitioning step often imposes a set of environmental constraints that influence aptamer function which encompasses such factors as the pH of the solution, the temperature of selection and overall ion content. Aptamers binding to their target ligands are dictated by the folded confirmation of the aptamer. The exact structures, or range of structures, achieved by aptamers is a function of both its sequence and the presence or absence of ions, particularly those which are cationic such as Na and Mg. The anionic nature of nucleic acids, as imparted by the phosphodiester backbone, represents a natural barrier to the folding of complex nucleic acid structures as chain repulsion disfavors the both the initiation, continuation and maintenance of hybridization. Cations shield the anionic backbone, facilitating proper folding and providing additional stabilization. Thus, an aptamer's final conformation can be stated as a product of both its sequence and environmental conditions, such as ion content or temperature. This interplay means consistent ionic strength during selection is needed to adequately enrich the pool. This consistency is also important due to the proclivity of aptamers to incorporate ions into their target binding, as utilization of salt bridges or the coordination of water molecules facilitated by ions are common facets of aptamer binding[43]. This salt dependence for nucleic acid structure is well established and some canonical secondary structures of nucleic acids exhibit strict ionic requirements to achieve their proper folded confirmation. One such structure is the planar arrangement of a tetrad of guanine residues referred to as G quadruplex, which relies on the presence of monovalent cations for stabilization and folding. The G quadruplex exhibits both a high thermodynamic and kinetic stability which naturally lends itself to stable aptamer-target interactions. The use of conditions that are conducive to G quadruplex folding are especially relevant given that guanine residues are overrepresented in aptamer sequences[44]. This would

seem to indicate that aptamers exploit the kinetic and thermodynamic stability of G quadruplexes, among others, to develop highly specific and stable interactions. Since these structures are a product of both sequence and conditions, and these conditions can be thought of as integral part of the aptamer similarly to its sequence. When both the conditions and sequence align, the resulting product is one that can bind its target with both high affinity and specificity similarly to their protein analogs antibodies.

The similar functions of aptamers and antibodies led to the proposal that many antibody based technologies would be viable for the application of aptamers. With distinct advantages and well-developed methodologies, the anticipated widespread adoption and supplantation of certain antibody-based technologies has failed to materialize. This hesitancy can be attributed to several factors that restrain the wide-spread adoption of aptamer technology. First there is a historical argument in that antibody techniques are well established and enjoy robust production and are supported by a well-established approach to validation and characterization. The exchange of antibody for aptamer technology is not straightforward and would require a significant investment. The second factor is the relative standardization in antibody based techniques in comparison to aptamer. The protein nature of antibodies, its associated production and validation allows for exchange of antibodies between various application with little lost in performance from a technical perspective. In comparison, aptamers lack a set of standardized conditions for selection and characterization. The wide range of possible initial selection conditions and the associated sensitivity to environmental conditions, such as salt concentration, for aptamer functioning makes universal application difficult. The next stage of aptamer technologies will have to rely on achieving a consensus among aptamer researchers to propose reasonable limits

on selection conditions and an agreed upon characterization standard. This standardization would result in a more cohesive approach for the field as a whole and facilitate adoption of different aptamers between technologies and research groups.

Post -selection Optimization:

Post-selection aptamer sequences encompass the full range of the original library selection, in that it comprises both the consensus regions utilized for primer binding and the variable region which contains the supposed aptamer core (apta-core) which mediates the binding to its target. Traditionally, sequences undergo optimization in which non-essential bases are trimmed via a series of progressive deletions in a process known as crawling deletion. These extra bases, in the most charitable cases, provide structural stability which facilitate the proper folding of the aptamer core and at worst can impede the aptamers folding or its interaction with the target[45]. The effects of these extraneous bases lie at how nucleic acids structures fold. The folding of complex nucleic structures proceeds similar to that of proteins, in that the resulting structure is a factor of both sequence and environmental conditions and proceeds through an initial collapse of the chain[46]. Here the first difference in protein vs nucleic acid folding occurs. The driving force for this initial collapse in proteins is attributed to the burying of hydrophobic residues, while for nucleic acids requires the condensation of counter ions to alleviate electrostatic repulsion to form intermediate secondary structures. This folding pathway is inherently rough with competing pathways which result from the inability of nucleic acids to achieve the lowest energy for all residues simultaneously, thus the competition between residues results in the formation of traps, where the local energy minima are stable and the transition

energies are too high to overcome. As the number of bases increases, these traps become proportional higher in number and deeper resulting in a higher proportion of misfolded aptamer. The improvement with removal of these extraneous bases can be thought as facilitate the folding pathway and the interaction between the aptamer and its ligand.

The aptitude of crawling deletion strategies in optimizing aptamer function is an extension from how aptamers interact with their ligands. The binding affinity and specificity are imparted by a select bases within the structure, which are the minority of the sequence[34]. These principal actors are supported by adjacent bases in a distance dependent manner. Nucleotides which are removed from the binding site, often by as little as 3 or 4 bases, contribute very little to the overall energetics of the binding interaction[34, 47]. The contribution of these bases is often more generic in nature, in that they provide structural stability to the active bases but are interchangeable.

After deletion of extraneous bases, a more detailed sequence interrogation by mutagenesis studies can be accomplished to determine which bases and their structural features are responsible for ligand binding in the minimized aptamer. However these studies can be time consuming and expensive, and are often neglected in the majority of aptamer publications. This lack of fine sequence data can hinder aptamer applications such as biosensor development. This process could be facilitated by sequence comparison to previous aptamers, similarly to the approach in determination of a protein's structure and function. However, data repositories of this kind are much less developed for aptamers than for proteins, both in terms of primary sequence and end structure of active aptamers.

A simple search of the RSCB repository can demonstrate the divide between protein crystal structures and aptamer crystal structures. This disparity is not for a lack of effort, but centers associated difficulties in producing high quality crystals of nucleic acids species. Proteins, by comparison, have a well-developed background that aids in determining necessary crystal growth conditions. These factors and the relatively disparate sequences that comprise aptamers, even for similar ligands, results in what can be considered the black box of binding. This represents a significant obstacle in the use and design of functional devices from aptamers. This difficulty is shown in Figure 3, which demonstrates the disconnect that occurs between the primary sequence, secondary structure and actual structure of aptamers.

ARC1172 is a DNA aptamer specific for A1 domain of von Willebrand factor (VWF), a protein factor involved in the thrombogenesis cascade[34]. This aptamer is relatively high affinity, Kd of 2 nM, and has been the subject of intense study for application as anti-coagulation agent. The primary sequence of this aptamer is shown in Figure 2 and is annotated to denote the nucleotides on a per function basis. Out of the 42 nucleotide sequence binding is mediated by 16 bases which form direct contacts with VWF, shown in red text in Figure 2. An additional 5 of these bases are structurally required to facilitate binding, as indicated by underlining, but do not for direct contacts with VWF. The sum of these nucleotides, 21 or 50%, are considered vital for binding with the rest being interchangeable as long as secondary structure for these areas is maintained. Initially, most aptamers are not crystalized instead relying on secondary structure predictions to give insight into which bases could form the basis of interaction. Secondary structural predictions, as done by Mfold, are seen below the sequence and have poor correlation with that actual crystal structure as the active bases are indicated by red squares. The presented

secondary structures are roughly equivalent in energy, and thus are both reasonable solutions to the folded state of the sequence. However, many of the key bases involved in binding would not be considered by general aptamer principles as they are involved in base pairing interactions. This disconnect occurs as most structure predictions have difficulty in determining long range interactions between bases within a structure and cannot incorporate the influence of the protein on the resulting structure.

The relatively small number of distinct aptamer crystal structures is a detriment but does provides the ability to perform comprehensive analysis with the few reported structures. The results of these aptamer binding surveys offer a unique snapshot of the binding interface between aptamer and ligand, the contribution of bases to this binding and general modes of interactions between the aptamer and its ligand.

The approaches aptamers use in binding their targets can be broadly divided into two strategies: shape complementarity between the surface of the target and aptamer, or encapsulation into the aptamer structure. Aptamers are particularly well suited to demonstrated high degrees of shape complementarity. This aptitude is derived from the flexible phosphodiester bond that comprises the backbone of nucleic acids, and the ability of the sugar base to assume several confirmations. This allows for a high degree of shape complementarity to develop between aptamer and target which is reflected by the area of interactions between aptamer and its epitope, which is similar to that of antibodies but exhibits a larger range, and mean higher degree of shape complementarity than those found in antibodies, 0.71 and 0.69 respectively [22]. This strategy is often utilized by aptamers when a significant size disparity is present between the target and aptamer. In theory the entire surface of the protein could function as the aptamer

binding site, however aptamers tend to favor particular surface features for binding regions. Particularly regions that are electropositive, such as those produced by high concentrations of basic amino acid residues which provide electrostatic driving force. However, repeated selections with highly positively charged surfaces can demonstrate that this attraction can lead to a relatively low specificity as the binding requires little contribution from the individual bases themselves and relies on general structural features of nucleic acid.

Besides exploiting electrostatic interactions, aptamers binding to large targets often encompass regions that possess a high degree of hydrogen bonding potential. This approach is intuitive for aptamers to utilize, as hydrogen bonding is a natural feature of nucleic acids and contributes significantly to the specificity of nucleic acid binding. Other strategies aptamers have utilized include the coordination of metal ions, another general feature of nucleic acids, and more rarely hydrophobic interactions. These same driving forces when applied to smaller targets often result in the target being incorporated into the aptamers structure, a strategy deemed encapsulation.

The distinction between shape complementarity and encapsulation-based binding is often blurred. Utilization of the target into the aptamer structure requires target characteristics that are complimentary to the natural structure forming principals of nucleic acids. In particular, encapsulated targets often possess the ability to hydrogen bond, allowing bridging to occur between even distant bases, or possess the ability to coordinate metal ions cohesively with the bases resulting in the formation of salt bridges. Of particular importance is the ability to engage in pi-stacking or pi-cation interactions, which often facilitate the positioning and incorporation of flat aromatic ligands into the aptamers structure. Incorporation into the aptamer's folded

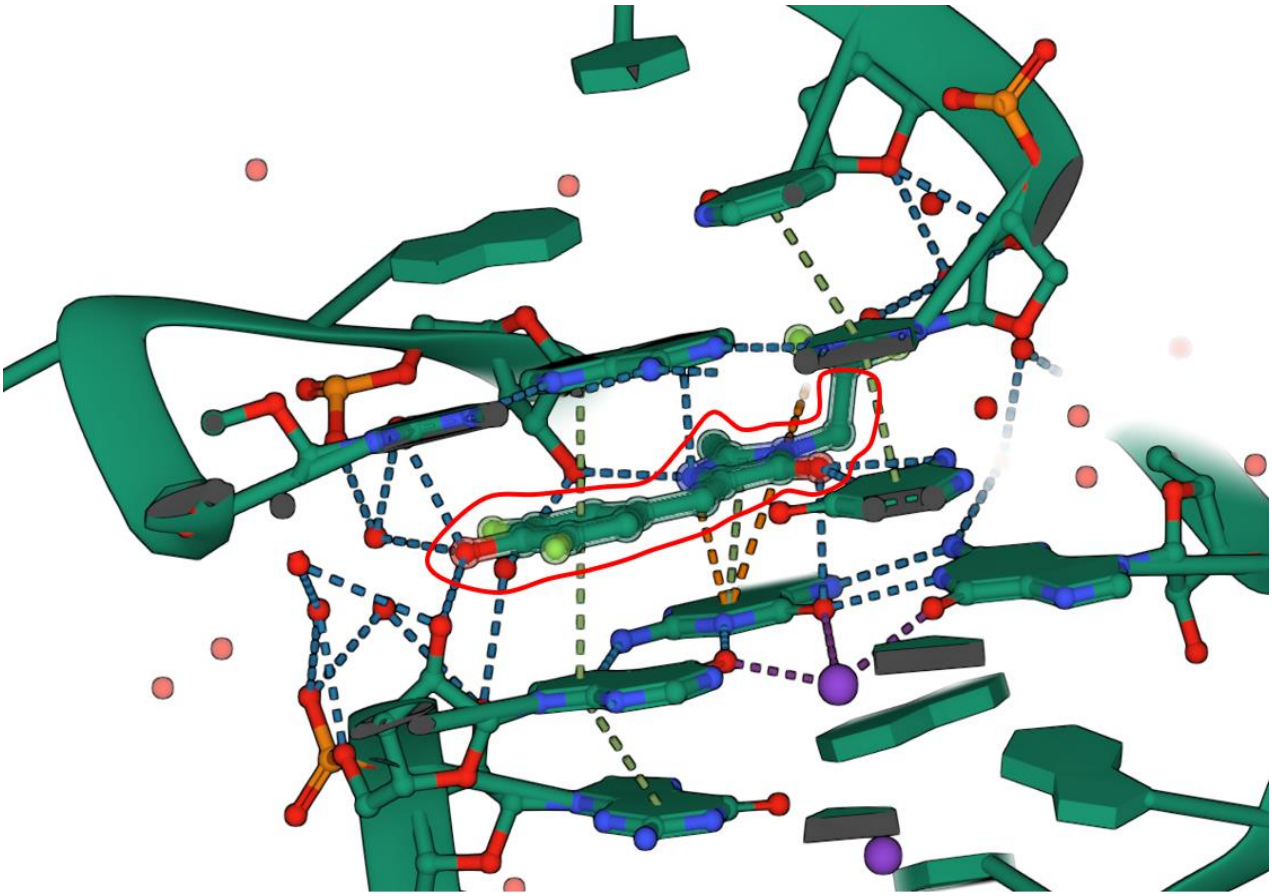


Figure 4. Immobilization of DFHBI by a Spinach Aptamer Derivative. Here the planar aromatic DFHBI is incorporated into the structure of the Spinach Aptamer. The intermolecular forces governing the incorporation are shown as colored dotted lines, while the structure of DFHBI is outlined in red. Pi stacking interactions are depicted by green lines, salt bridges by purple lines, pi-cation interactions are shown via orange, and hydrogen bonding is indicated by blue lines. Structural incorporation often relies on a complex network formed by these interactions with a variety of the nucleotide's structural features. The above image was derived from Fernandez-Millian [1] and processed using UCSD Chimera program.

structure also tends to disfavor target with high degrees of rotatable bonds, as a certain degree of structural consistency is necessary for the initial selection and stabilization of the resulting structure. The crystal structures of encapsulated ligands often show ligands which reside in plane with the bases themselves, facilitating pi stacking interactions as shown in Figure 4 which shows the interaction between a fluorescent RNA Aptamer, Spinach, and its target DFHBI [1]. This strategy is often utilized by aptamers that induce torsional restriction-based fluorescence of their target.

While the selection of aptamers and their characterization is with merit, aptamers are often selected to be potential drug candidates in the treatment of disease or bioanalytical tools for the detection of a variety of targets. Once selected and characterized, aptamers represent an attractive prospect for usage as pharmaceuticals. Aptamers occupy a unique niche that combines the benefit small molecule drugs, such as biodistribution and synthetic production, with some of the specificity associated with the use of larger antibodies. This synthetic production lends itself to a process that is inexpensive and with a production capability that is scalable. In particular the use of automated solid phase synthesis (SPS) allows production to occur continuously and with high degree of batch to batch fidelity which is an area that antibodies can present production shortfalls. Even so, the examples of aptamer based drugs that have made it to market, i.e. progressed through all phases of FDA trials, are very small and encompass only a few examples. An illustrative example of an aptamer based drug encompasses Macugen, pegaptanib sodium, which is utilized in the treatment of wet age macular degeneration [48]. This XNA aptamer is specific for VEGF which is a necessary transcription factor in the angiogenesis of blood vessels. In the context of macular degeneration, this angiogenesis results in the formation of scar tissue which results in progressive vision loss and effects a significant amount of patients over the age of 55. This aptamer consists of a heavily modified, both the backbone and base, RNA aptamer [48]. Direct injection into the eye results in antagonism of VEGF and reduction in angiogenesis thus preserving vision without correct the underlying damage. The heavy modification of aptamers is necessary for in vivo functioning. These modifications often focus on improving several characteristics that make application of the unmodified aptamer as drugs difficult.

The first hurdle for any nucleic acid-based drugs, like aptamers, lies in the low half life such candidate drugs exhibit when used in vivo. These reduced half lives, often measured in the range of seconds or hours for unmodified RNA and DNA respectively [49], is attributed to nucleic acid instability in biological systems. The extracellular and intracellular environment is hostile to foreign nucleic acid due to endogenous nucleases resulting in significant or total degradation. As such many modifications to the aptamer bases and backbone are done to increase the nuclease resistance. Often these modifications focus on the terminal ends of the sequence, alterations to the backbone, or 2' position modifications inhibiting the recognition by exonuclease and generating a non-suitable substrate for degradation respectively [49]. The intrinsic highly charged nature of nucleic acids makes traversing the cellular membrane difficult. Neutralization of these charges can improve permeability, the direct encapsulation into liposome or similar structures is more commonly applied. The translation of aptamer to therapeutic has proven difficult and even in the cases of successful aptamers, such as Macugen. The associated cost with these modifications has led to the supplantation of Macugen by more inexpensive antibody-based drug. Given the constraints of the aptamers as drug candidates, there use as analytical tools such as biosensor is more straightforward.

Aptamers as Bionsensors:

In contrast to their applications as drug candidates, the use of aptamers as bioanalytical tools and biosensors is more developed and robust. Aptamers fall into the category of nucleic acid based sensors. Utilizing nucleic acids as the basis of the sensor allows exploitation of the

natural tendency for nucleic acids to fold reliably and allows for a degree of control of the spatial arrangement of the resulting sensor architecture.

Less apparent benefits of the aptamer-based sensors include the ability to bridge the sensing gap that exist between nucleic acids and other biomolecules, particularly proteins. Most hybridization-based sensor technology are only capable of interacting with, and thus determining, changes in nucleic acid sequence. These changes, often labeled single nucleotide polymorphism or single nucleotide variations, are often causative of disease in their own right making them high value targets. Aptamers, as nucleic acids themselves, are intrinsically capable of interacting with other nucleic acids but also possess the ability to specifically interaction with a wide range of the other target types. This broad scope presents the ability to combine the robust methods of hybridization-based detections and extend them to small molecules and other biomolecules which would naturally be intractable to direct hybridization-based sensing methods [50] [39]. Their nucleic acid base also allows for a certain degree of modularity in design, as the structural principles for nucleic acids sensors are more developed. This modularity allows for aptamers to be combined with other detection methods and other aptamers using short sequences, called communicating domains, which tie the sensing and signaling functions together or provide the ability to interrogate several aspects of the target simultaneously [51].

Similarly to other nucleic acid based technologies, such as molecular beacons, simple aptamer sensors can be designed which utilize confirmational upon target binding to induce a change in signal [52]. Often these confirmation sensors utilize technology analogous to the established molecular beacon technology, in which binding results in a confirmational shift that alters the distance between a fluorophore and quencher pair. The confirmational shift results in a

spatial separation of the pair restoring fluorescence of the fluorophore which can be detected through spectroscopic methods. While this method has proven reliable and widely used, there is an economic cost associated with synthesis of the oligonucleotide modified with these pairs. This approach also requires the aptamer to undergo conformational shifts upon binding which is not a universal characteristic. More recent focus has been given on utilizing aptamers which are naturally able to induce fluorescence upon binding their target. This removes the necessity of costly synthesis, greatly reducing the cost per test and provide a low-cost avenue to develop various biosensors.

The modularity of nucleic acids and the possibility of utilizing aptamers with naturally fluorescent ligands raises the possibility of developing inexpensive multiplex detection schemes. Multiplex detection is routinely applied to other hybridization based techniques, such as multiplex PCR. The use of a multiplex format allows for gathering of more actionable data by interrogating several variables simultaneously. Similarly to other hybridization based technologies, the primer pairs utilized in multiplex PCR require individual synthesis of expensive covalently modified primers. Furthermore, altering the target sequence in question does not allow use of previously synthesized primer requiring. This application of aptamers to multiplex analysis is lacking in literature, and the examples that do exist tend to focus on relying on attachment of the aptamer to surfaces rather than solution-based approaches which would be preferable.

Here in, I describe my efforts to develop a multiplex assay which utilizes pairs of aptamers to perform analysis simultaneously of two distinct SARS-COVID19 sequences. Development of this technology would produce an aptamer toolkit which allows for easy change

in target specificity without relying on expensive covalent modifications of the oligonucleotides. This change will make this form of analysis feasible and establish the principals to guide development in this field.

References:

1. Fernandez-Millan, P.; Autour, A.; Ennifar, E.; Westhof, E.; Ryckelynck, M., Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *Rna* **2017,** *23* (12), 1788-1795.
2. Noller, H. F.; Hoffarth, V.; Zimniak, L., Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science* **1992,** *256* (5062), 1416-1419.
3. Crick, F. H., On protein synthesis. *Symp Soc Exp Biol* **1958,** *12*, 138-63.
4. Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N., Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **1985,** *230* (4732), 1350-1354.
5. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-10.
6. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990,** *346* (6287), 818-822.
7. CAROTHERS, J. M.; DAVIS, J. H.; CHOU, J. J.; SZOSTAK, J. W., Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *RNA* **2006,** *12* (4), 567-579.
8. Russo Krauss, I.; Merlino, A.; Randazzo, A.; Novellino, E.; Mazzarella, L.; Sica, F., High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Research* **2012,** *40* (16), 8119-8128.
9. Lin, C. H.; Patel, D. J., Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. *Chem Biol* **1997,** *4* (11), 817-32.
10. Liu, L. M. a. J., Catalytic Nucleic Acids: Biochemistry, Chemical Biology, Biosensors, and Nanotechnology. *iScience* **2020,** *23* (1).
11. Klug, S. J.; Famulok, M., All you wanted to know about SELEX. *Molecular Biology Reports* **1994,** *20* (2), 97-107.
12. Gopinath, S. C. B., Methods developed for SELEX. *Analytical and Bioanalytical Chemistry* **2007,** *387* (1), 171-182.
13. Qi, S.; Duan, N.; Khan, I. M.; Dong, X.; Zhang, Y.; Wu, S.; Wang, Z., Strategies to manipulate the performance of aptamers in SELEX, post-SELEX and microenvironment. *Biotechnology Advances* **2022,** *55*, 107902.

14. Chan, K. Y.; Kinghorn, A. B.; Hollenstein, M.; Tanner, J. A., Chemical Modifications for a Next Generation of Nucleic Acid Aptamers. *ChemBioChem* **2022,** *23* (15), e202200006.
15. Zhuo, Z.; Yu, Y.; Wang, M.; Li, J.; Zhang, Z.; Liu, J.; Wu, X.; Lu, A.; Zhang, G.; Zhang, B., Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *International Journal of Molecular Sciences* **2017,** *18* (10), 2142.
16. Famulok, M.; Mayer, G., Aptamers and SELEX in Chemistry & Biology. *Chemistry & Biology* **2014,** *21* (9), 1055-1058.
17. Wang, T.; Chen, C.; Larcher, L. M.; Barrero, R. A.; Veedu, R. N., Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances* **2019,** *37* (1), 28-50.
18. Kohlberger, M.; Gadermaier, G., SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry n/a* (n/a).
19. Tolle, F.; Wilke, J.; Wengel, J.; Mayer, G., By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLOS ONE* **2014,** *9* (12), e114693.
20. Schütze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Mörl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glökler, J., Probing the SELEX Process with Next-Generation Sequencing. *PLOS ONE* **2011,** *6* (12), e29604.
21. Saito, S., SELEX-based DNA Aptamer Selection: A Perspective from the Advancement of Separation Techniques. *Analytical Sciences* **2021,** *37* (1), 17-26.
22. McKeague, M.; McConnell, E. M.; Cruz-Toledo, J.; Bernard, E. D.; Pach, A.; Mastronardi, E.; Zhang, X.; Beking, M.; Francis, T.; Giamberardino, A.; Cabecinha, A.; Ruscito, A.; Aranda-Rodriguez, R.; Dumontier, M.; DeRosa, M. C., Analysis of In Vitro Aptamer Selection Parameters. *Journal of Molecular Evolution* **2015,** *81* (5), 150-161.
23. Vorobyeva, M. A.; Davydova, A. S.; Vorobjev, P. E.; Pyshnyi, D. V.; Venyaminova, A. G., Key Aspects of Nucleic Acid Library Design for in Vitro Selection. *International Journal of Molecular Sciences* **2018,** *19* (2), 470.
24. Bai, J.; Luo, Y.; Wang, X.; Li, S.; Luo, M.; Yin, M.; Zuo, Y.; Li, G.; Yao, J.; Yang, H.; Zhang, M.; Wei, W.; Wang, M.; Wang, R.; Fan, C.; Zhao, Y., A protein-independent fluorescent RNA aptamer reporter system for plant genetic engineering. *Nature Communications* **2020,** *11* (1), 3847.
25. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J., RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* **2014,** *9* (10), 2412-2420.
26. Semlow, D. R.; Silverman, S. K., Parallel Selections In Vitro Reveal a Preference for 2′–5′ RNA Ligation upon Deoxyribozyme-Mediated Opening of a 2′,3′-Cyclic Phosphate. *Journal of Molecular Evolution* **2005,** *61* (2), 207-215.
27. Schneider, B.; Morávek, Z.; Berman, H. M., RNA conformational classes. *Nucleic Acids Res* **2004,** *32* (5), 1666-77.
28. Amero, P.; Lokesh, G. L. R.; Chaudhari, R. R.; Cardenas-Zuniga, R.; Schubert, T.; Attia, Y. M.; Montalvo-Gonzalez, E.; Elsayed, A. M.; Ivan, C.; Wang, Z.; Cristini, V.; Franciscis, V. d.; Zhang, S.; Volk, D. E.; Mitra, R.; Rodriguez-Aguayo, C.; Sood, A.

K.; Lopez-Berestein, G., Conversion of RNA Aptamer into Modified DNA Aptamers Provides for Prolonged Stability and Enhanced Antitumor Activity. *Journal of the American Chemical Society* **2021,** *143* (20), 7655-7670.
29. Lee, E. J.; Lim, H. K.; Cho, Y. S.; Hah, S. S., Peptide nucleic acids are an additional class of aptamers. *RSC Advances* **2013,** *3* (17), 5828-5831.
30. Rangel, A. E.; Chen, Z.; Ayele, T. M.; Heemstra, J. M., In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Research* **2018,** *46* (16), 8057-8068.
31. Chen, J.; Chen, M.; Zhu, T. F., Directed evolution and selection of biostable l-DNA aptamers with a mirror-image DNA polymerase. *Nature Biotechnology* **2022**.
32. Fouz, M. F.; Appella, D. H., PNA Clamping in Nucleic Acid Amplification Protocols to Detect Single Nucleotide Mutations Related to Cancer. *Molecules* **2020,** *25* (4), 786.
33. Christensen, T. A.; Lee, K. Y.; Gottlieb, S. Z. P.; Carrier, M. B.; Leconte, A. M., Mutant polymerases capable of 2′ fluoro-modified nucleic acid synthesis and amplification with improved accuracy. *RSC Chemical Biology* **2022,** *3* (8), 1044-1051.
34. Huang, R. H.; Fremont, D. H.; Diener, J. L.; Schaub, R. G.; Sadler, J. E., A structural explanation for the antithrombotic activity of ARC1172, a DNA aptamer that binds von Willebrand factor domain A1. *Structure* **2009,** *17* (11), 1476-84.
35. Antipova, O. M.; Zavyalova, E. G.; Golovin, A. V.; Pavlova, G. V.; Kopylov, A. M.; Reshetnikov, R. V., Advances in the Application of Modified Nucleotides in SELEX Technology. *Biochemistry (Moscow)* **2018,** *83* (10), 1161-1172.
36. Tolle, F.; Brändle, G. M.; Matzner, D.; Mayer, G., A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angewandte Chemie International Edition* **2015,** *54* (37), 10971-10974.
37. Bowater, R. P.; Gates, A. J., Nucleotides: Structure and Properties. In *eLS*, pp 1-9.
38. Kraemer, S.; Vaught, J. D.; Bock, C.; Gold, L.; Katilius, E.; Keeney, T. R.; Kim, N.; Saccomano, N. A.; Wilcox, S. K.; Zichi, D.; Sanders, G. M., From SOMAmer-Based Biomarker Discovery to Diagnostic and Clinical Applications: A SOMAmer-Based, Streamlined Multiplex Proteomic Assay. *PLOS ONE* **2011,** *6* (10), e26332.
39. Yu, H.; Alkhamis, O.; Canoura, J.; Liu, Y.; Xiao, Y., Advances and Challenges in Small-Molecule DNA Aptamer Isolation, Characterization, and Sensor Development. *Angewandte Chemie International Edition* **2021,** *60* (31), 16800-16823.
40. Skouridou, V.; Schubert, T.; Bashammakh, A. S.; El-Shahawi, M. S.; Alyoubi, A. O.; O'Sullivan, C. K., Aptatope mapping of the binding site of a progesterone aptamer on the steroid ring structure. *Anal Biochem* **2017,** *531*, 8-11.
41. Ahmad, K. M.; Oh, S. S.; Kim, S.; McClellen, F. M.; Xiao, Y.; Soh, H. T., Probing the Limits of Aptamer Affinity with a Microfluidic SELEX Platform. *PLOS ONE* **2011,** *6* (11), e27051.
42. Ruscito, A.; DeRosa, M. C., Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Frontiers in Chemistry* **2016,** *4* (14).
43. Kuo, T. C.; Lee, P. C.; Tsai, C. W.; Chen, W. Y., Salt bridge exchange binding mechanism between streptavidin and its DNA aptamer--thermodynamics and spectroscopic evidences. *J Mol Recognit* **2013,** *26* (3), 149-59.

44. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S., Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* **2016,** *4*, 38.
45. Zheng, X.; Hu, B.; Gao, S. X.; Liu, D. J.; Sun, M. J.; Jiao, B. H.; Wang, L. H., A saxitoxin-binding aptamer with higher affinity and inhibitory activity optimized by rational site-directed mutagenesis and truncation. *Toxicon* **2015,** *101*, 41-47.
46. Thirumalai, D.; Hyeon, C., RNA and Protein Folding: Common Themes and Variations. *Biochemistry* **2005,** *44* (13), 4957-4970.
47. Ketterer, S.; Fuchs, D.; Weber, W.; Meier, M., Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. *Nucleic Acids Research* **2015,** *43* (19), 9564-9572.
48. Tobin, K. A., Macugen treatment for wet age-related macular degeneration. *Insight* **2006,** *31* (1), 11-4.
49. Effect of Chemical Modifications on Aptamer Stability in Serum. *Nucleic Acid Therapeutics* **2017,** *27* (6), 335-344.
50. Strehlitz, B.; Nikolaus, N.; Stoltenburg, R., Protein Detection with Aptamer Biosensors. *Sensors* **2008,** *8* (7), 4296-4307.
51. Stojanovic, M. N.; Kolpashchikov, D. M., Modular Aptameric Sensors. *Journal of the American Chemical Society* **2004,** *126* (30), 9266-9270.
52. Moutsiopoulou, A.; Broyles, D.; Dikici, E.; Daunert, S.; Deo, S. K., Molecular Aptamer Beacons and Their Applications in Sensing, Imaging, and Diagnostics. *Small* **2019,** *15* (35), 1902248.

CHAPTER 2: DEVELOPMENT OF A SPLIT MALACHITE GREEN DNA APTAMER SENSOR

Reproduced in part with permission from Kolpashchikov, D. M. and O'Steen, M.: A self-assembling split aptamer multiplex assay for SARS-COVID19 and miniaturization of a malachite green DNA-based aptamer. Sensors and Actuators Reports, 2022, 4, 100125.

Abstract

Multiplex assays often rely on expensive sensors incorporating covalently linked fluorescent dyes. Herein, we developed a self-assembling aptamer-based multiplex assay. This multiplex approach utilizes a previously established split aptamer sensor in conjugation with a novel split aptamer sensor derived from a malachite green (MG) DNA aptamer. This system was capable of simultaneous fluorescent detection of two SARS COVID-19-related sequences in one sample with individual sensors that possesses a limit of detection (LOD) in the low nM range. Structure-based optimization of the Split Malachite Green (SMG) sensor yielded a surprisingly small active aptamer construct, Mini-MG, capable of inducing fluorescence of malachite green in both a DNA hairpin and switch-sensor format.

Introduction

With the potential threat of new or resurgent diseases, the need for both economical and rapidly deployable diagnostics are paramount. Commonly, nucleic acid analysis includes the use of fluorescently labelled oligonucleotides, such as TaqMan [53] or molecular beacons probes [54]. While these detection methods can achieve low limits of detection[55], the synthesis of fluorescently modified oligonucleotides incurs a significant economic cost. Several economical

alternatives to these techniques have been proposed: including the use of functional nucleic acids such as deoxyribozymes [56], or aptamers[57-61] which often do not rely on expensive modifications for their function.

Aptamers are functional nucleic acid species generated through a series of iterative cycles of partitioning non-binding members from binding members of a randomized oligonucleotide library - a technique pioneered concurrently by the Szostak [62] and Gold [63] group in the early 90s and termed SELEX (Sequential Evolution by Exponential Enrichment). The selected aptamers are generally short, less than 100 nucleotides, thus they are suitable for automated solid phase synthesis, which decreases cost and increases production potential in comparison to antibodies. The relative ease in which aptamers can be selected has enabled the production of aptamers with a wide variety of target analytes including small molecules, toxins, [64] and whole cells[65]. A variety of detection schemes are possible with aptamers but one of the most widely utilized schemes involves covalent conjugation of a fluorophore and quencher pair to the 5' and 3' terminus of the aptamer [66], analogous to molecular beacon technology. The binding of the aptamer's target ligand involves a confirmational shift, increasing the distance between the fluorophore and quencher and subsequent increase in fluorescent output. A subset of aptamers are capable of inducing a detectable signal upon binding their cognate target often through restriction of torsional rotation [67-69]. These aptamers, designated as Light-Up Aptamers, are enticing for the development of label-free and inexpensive biosensors as unmodified oligonucleotides can be utilized. This avoids costly synthesis of specially modified oligos while still being compatible with optical-based detection methods.

Light up aptamers have been successfully selected utilizing both RNA and DNA. In particular, RNA-based Light-Up aptamers have found use in intracellular imaging such as the spinach [70], mango [25], and broccoli [71] aptamers. The use of RNA allows intracellular imaging through the use of expression vectors but presents both economic and workflow challenges as the synthesis of RNA is more costly than DNA while also being susceptible to degradation. Thus, the use of DNA based Light-Up aptamers presents an alternative that is suitable to long term storage and provides a simpler workflow while being more amenable to in vitro diagnostic approaches.

Conventional hybridization probes can demonstrate a lack of selectivity towards mismatches in its target sequence. This lack of selectivity, particular at ambient temperatures, is a by-product of the length oligonucleotide probe that are commonly utilized in these hybridization-based assays [72]. The splitting paradigm relies on destabilization of the aptamer-core by scission of the intact strand into two smaller halves. These split aptamers can be converted to sensors by appending complimentary oligonucleotide sequences specific for the analyte of interest, deemed analyte-binding arms, to the ends of the split aptamer [33]. In the absence of its target, the two halves are incapable of hybridizing, resulting in low signal. In the target's presence hybridization occurs between the analyte binding arms and analyte bringing the halves in close proximity, facilitating hybridization via a proximity effect with subsequent refolding of the aptamer core for signal generation. The advantage of this approach is that by utilizing shorter binding arms, selectivity for mismatches is increased as they are less tolerated by the shorter helices while affinity is maintained via the cooperative binding of the small fragments[72]. The inverse relationship between affinity and selectivity central to hybridization-

based technologies is circumvented by use of the splitting paradigm, as the affinity is imparted by cooperation between shorter fragments while the shortened individual strands maintain selectivity as each individual piece is sensitive to mismatches. Split aptamer sensors have been produced for a variety of targets, with a recent review highlighting some of the detection schemes and designs [73, 74]. While many single target interrogating aptamer sensors have been developed, the literature is much less developed for multiplex aptamer-based assays particularly for those that are solution-based which eliminates complex fabrication procedures. Addressing this gap could lead to substantial development of inexpensive multiplex approaches whose basis is anchored in a self-assembling sensors and is usable with the conceptually simpler solution-based format.

Multiplex assays allow the interrogations of one or more analytes of interest simultaneously and is commonly applied in PCR analysis of nucleic acid targets. Multiplexing reduces both cost and time necessary to determine if a sample contains an analyte of interest, while providing an avenue to generate more actionable data and unambiguous identification for potentially serious pathogens. While multiplex analysis is routinely accomplished via the use of expensive fluorescently modified primers, a gap in the literature exists for the application of split aptamer sensors in a multiplex format. With only a few examples of multiplex analysis utilizing aptamers in the literature and no examples of a split-aptamer based multiplex assays in solution. Our work here will lay the foundation for the development and optimization of other split aptamer based multiplex assays, describing some of the pitfalls and design approaches for this new direction in multiplex analysis.

Here in, we describe our efforts in the development of a self-assembling multiplex aptamer assay utilizing all-DNA split aptameric sensors: malachite green (MG) DNA aptamer [67] and Dapoxyl DNA aptamer [75]. This is the first demonstration of multiplex analysis utilizing self-assembling aptamers in solution. In this work we first conduct a screening of several available light up DNA aptamers and discuss some of the selection criteria that are essential to produce a functioning sensor. Next, we outline principles for the conversion and optimization of a new split aptamer sensor, using the MG DNA aptamer, to produce a sensor platform that is able to detect its target analyte in the nM range and retains much of the activity of the parent aptamer. This was accomplished by introducing a split in the parent MG DNA aptamer, followed by verification of its ability to enhance intrinsic MG fluorescence was retained after conversion to a split-aptamer sensor format, Split Malachite Green Sensor (SMG). Central to the split-sensor paradigm is a decoupling of the sensing and signaling portion of the aptamer sensor. This is accomplished via structural optimization, producing a platform which should allow for quick exchange of target specificity by altering the analyte binding arms while maintaining similar performance, an approach analogous to lateral flow-based assays. Following this paradigm, iterative structural optimization of the analyte binding arms, stem elements, and linkages between the sensor's sensing and signaling elements was performed to produce a sensor with detection limits in the low nM range. To demonstrate the ability to alter target specificity without time-consuming and costly optimization, a previously optimized split aptamer sensor, the split-dapoxyl aptamer (SDA) [32], was altered to detect new target analytes via exchange of its analyte binding arm with little loss in performance. This demonstrates the generality of the design with a successful decoupling of activity from the sensing function. Next, the individual aptamer sensors were then

used in combination to interrogate two SARS-COVID19 DNA analytes simultaneously. Further work focused on quantifying the impact on the limit of detection (LOD) for the individual sensors in conditions that differ from their native buffer conditions and can provide guidelines for the selection of future aptamer pairs for multiplex detection. Lastly, we demonstrate that the MG aptamer can be further reduced to an active miniaturized construct, Mini-MG, which retains the ability to enhance MG fluorescence in a switch-sensor format.

Results and Discussion:

For our model system we selected two fragments of the SARS-CoV-2 genome as pulled from the NCBI reference sequence (NC_045512.2). Here we chose a region of the coronavirus genome that contains single nucleotide variation, SNV, that results in a glutamate to glycine mutation in the spike protein [76]. This mutation is associated with higher infectivity and more severe diseases and is present in several variants of concern such as B1.617.2 (Delta) [76]. In this study, we used synthetic DNA sequences, titled NC-45 and NC-14, and corresponding mutant sequences containing single or double mismatches, NC-45A containing a guanine to adenine mutation and NC-14C containing a thymine to cytosine mutation. Derivatives of the NC-45 analyte consisting of the same mismatch (G>A) were designed that encompasses more of the reference genome (NC-45L or NC-45AL). The sequence diagrams for each of our target analytes are shown in Figure 5. The conformational state of the analyte can present obstacles for hybridization based sensors, as folded analytes are inherently more stable than linear analytes. This folded state also result in slower hybridization kinetics, particularly if both the analyte and sensing strands possess

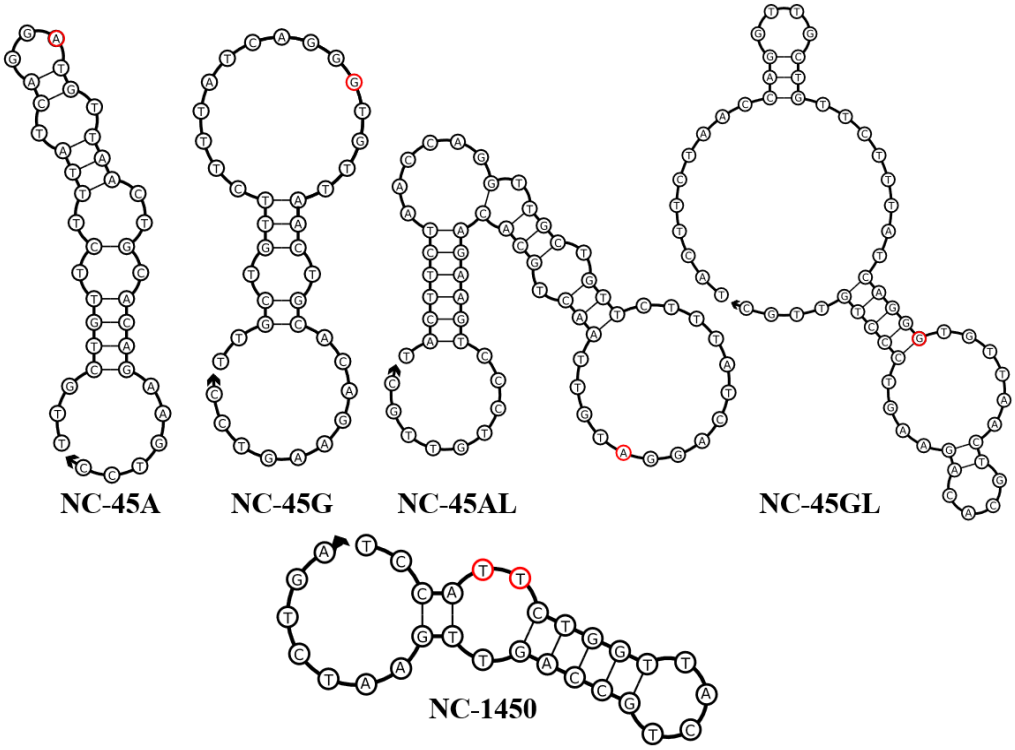


Figure 5: Sequence diagrams displaying the predicted secondary structures of all analytes used in this study. From left to right the order of display corresponds to NC45A, NC45G, NC45AL, NC45GL, NC1450. Single nucleotide variations (SNVs) are indicated by red arrows. The mutation of A > G is associated with a significant change in the folded character of the analytes, with the alternative structure generated from G being much less stable than the alternative A containing analyte, as such all those images are presented. NC-1450 mutations, indicated by red highlight, do not result in a significant structural change as the initial folded structure is quite stable due to the presence of the GC rich stem.

secondary structure as reciprocal melting is necessary to achieve full hybridization. Conventional hybridization approaches would utilize longer strands or attack regions of free bases to displace the intramolecularly hybridized strand through a toe-hold mediated displacement mechanism. This often increases the kinetics and completeness of hybridization but at the cost of selectivity. To circumvent this cost the split approach can utilize asymmetric arms, one long analyte binding arm to unwind the secondary structure and shorter arm that is more inherently selective. Singly neither of these arms should produce detectable signal. It is only when both arms bind that a detectable

signal is produced. This in combination with the short analyte binding arm's inherent intolerance of mismatches results in an overall complex that retains selectivity and sensitivity.

A literature survey of DNA light up aptamers produced several candidates that were relatively small, less than 100 nucleotides, and whose target ligand was commercially available dyes. These encompassed a variety of sequences and target ligands, and the predicted secondary structure of these sequences is shown in figure 6.

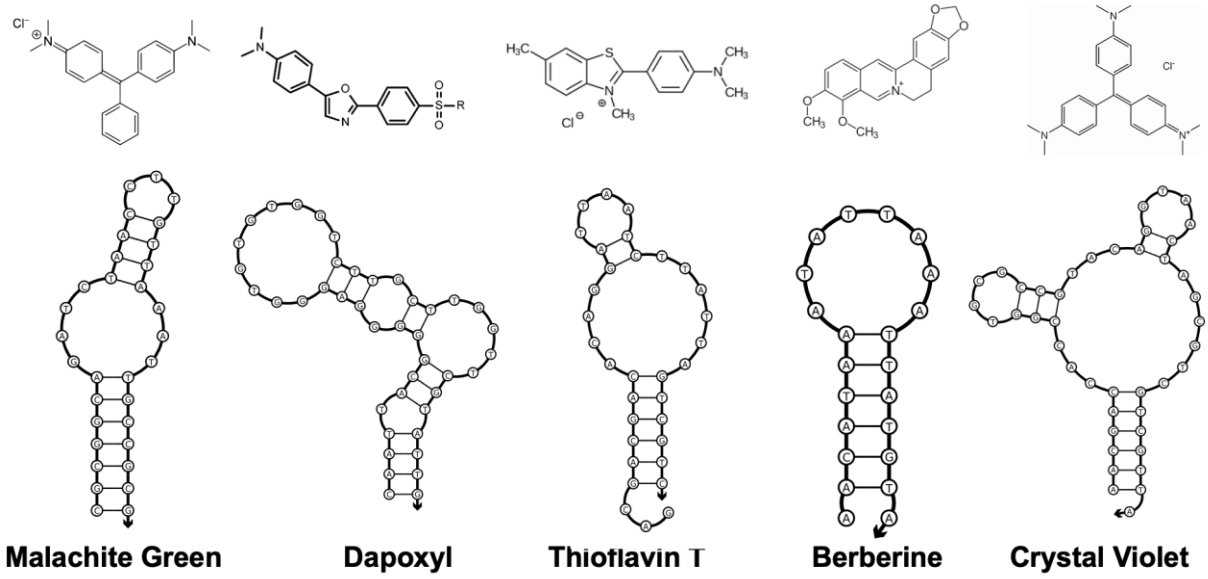


Figure 6. Selected Light Up Aptamers for initial screening. Here the secondary structure predictions, via NUPACK, of the primary sequences of our selected aptamers are shown with the structure of their light up dye target shown above. The malachite green [67], thioflavin T [77], crystal violet [78], dapoxyl aptamer [75] and berberine [69] aptamers were screened for the usefulness in developing light up DNA aptamer sensors.

Initial screening focused on finding light up aptamer-ligand pairs that possessed several distinct qualities. First the dye in question must exhibit low non-specific interaction with nucleic acids. Since the best aptamer targets tend to be compounds that have some baseline affinity for nucleic acids, a high degree of non-specific interaction would obscure fluorescence enhancement achieved from the aptamer's binding. For this initial screening, the dye concentrations utilized in

the original publication were used with a random DNA sequence of equivalent length to our target analyte (40 nucleotides). This screening resulted in exclusion of the crystal violet and berberine aptamer, as significant fluorescence enhancement was observed even in the absence of the aptamer sensors analyte. Crystal violet in particular has been shown to have significant non-specific interactions and thus was deemed unsuitable for sensor development. This trend was maintained for berberine, and to a less extent, thioflavin T. The dapoxyl aptamer is the base aptamer used for the development of the split dapoxyl aptamer sensor. The cognate ligand for this aptamer was DFHBI, which was commercially available but discontinued. However, the dapoxyl aptamer as a whole is relatively non-specific and can enhance the fluorescence of a variety of dyes particularly if they contain conjugated ring systems. This promiscuity was outlined by Connelly [79], which provided a suitable and inexpensive alternative in Auramine O. Also shown in figure 7, is a generalized graphical approach to the design of the split aptamer sensors. As depicted in the top of panel B, a functional split sensor should have low fluorescence in the absence of the target. This low signal is a function of low non-specific interactions and the inability of the aptamer to signal in the absence of the target analyte. Once analyte is added to the system, the proximity of the two halves facilitates their hybridization generating strong fluorescence. Ideally this fluorescence turn on would produce a signal far above the intrinsic fluorescence of the compound, which would correlate to a more sensitive assay. Fluorescence turns on in the absence of the aptamer is undesirable, as it negatively impacts the achievable differentiation between presence and absence of target. This scenario is shown in the bottom right of panel B. Another scenario is depicted on the left bottom of panel B. Here the dye has little non-specific interaction but exhibits a weak turn on signal. This scenario could be attributed to the characteristics of the dye molecule in question,

representing a poor quantum yield, or an aptamer which is inefficient in inducing fluorescence. Both of these scenarios would be undesirable much for the same reason as depicted in the bottom right of panel B. In that the achievable signal difference in the presence or absence of the analyte would be small resulting in less sensitivity.

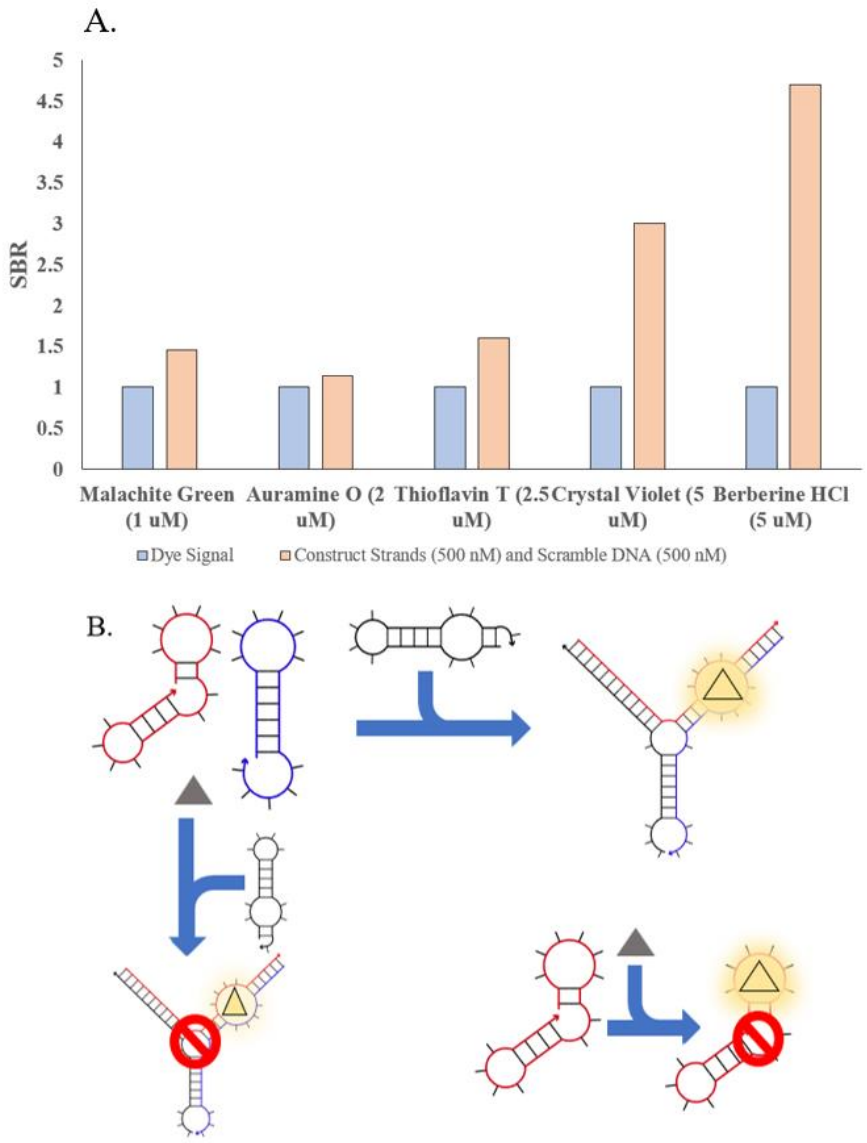


Figure 7: Initial Light Up Aptamer Screening Results and Graphical Depiction of the screening process. The potential dyes of each light up aptamer were incubated with 500 nM of 40 nucleotide long unrelated DNA in 100 uL at concentrations one-half of their reported K_d in the buffer in which their selection

occurred. The crystal violet and berberine aptamers were discarded as the fluorescence was highly elevated in the absence of the aptamer sequence via non-specific DNA interactions. The graphical depiction of this process is shown below. The ideal scenario, Panel B Top, shows a split aptamer sensor that is non-fluorescent in the absence of the target analyte. Fluorescence is only trigger when both halves of the aptamer, shown in red and blue, are brought into proximity by the target analyte, shown in black. The resulting complex should be highly fluorescent as indicated by the highlighted triangle while initially having low fluorescence show by the gray triangle. A scenario where the dye is weakly fluorescent is show in panel B bottom left, which would result in its exclusion. A scenario where the dye has strong non-specific interactions with DNA is depicted in panel B, right, would also result in disqualification.

Figure 8 contains the sequence diagram of the optimized MG aptamer as reported by Hyung et al. [67]. The MG aptamer is a relatively small aptamer of 32 nucleotides with a predicted secondary structure consisting of a centralized asymmetric bulge flanked by two stem elements. Alterations to an aptamer's structure, such as sequence minimization or splitting, is a laborious process that relies on combination of experience and intuition as most aptamers lack structural detail such as crystal structures or in depth nucleotide contributions to the binding pocket [80, 81]. This editing can be guided by utilizing general observations related to aptamer-target interactions. Two such relevant observations are that target-aptamer interaction are mostly mediated by free bases rather than those that are base-paired, such as the found in stem elements. Secondly, nucleotide contribution to binding tends to decrease as the nucleotide's distance from the purported aptamer core increases [82]. This allows for exchange or elimination of nucleotides with only small changes in aptamer affinity to its ligand [82]. A structural analysis of the folded MG aptamer via NUPACK [83] shows a trinucleotide loop, consisting of CTT, that is removed from the proposed core by 4 base-paired nucleotides. This decreases the likelihood that the loop is intrinsic to MG aptamer's function and represented the first attempted structural modification to achieve a split MG DNA aptamer. This trithymine loop was excised, severing the strands thus generating a split MG aptamer (SMG). The functioning of this split aptamer was verified by fluorescence spectroscopy with no significant impact on fluorescent enhancement of MG or its affinity toward

MG (data not shown). The split aptameric sensor paradigm relies on the inability of the two aptamer strands to reform an active binding pocket in the absence of target; instead relying on a proximity effect induced by binding to the target analyte to stabilize hybridization of the stem elements to reform the active pocket as shown via the diagram in Figure 1C[84, 85] Once split, analyte binding arms specific for our model analytes were appended onto the 5' and 3' ends of the new MG aptamer strand halves, SMG-#A and SMG-#B, as shown in Figure 8. The two strands are unable to stably hybridize in the absence of the target analyte as the melting temperature of the resulting stem elements are below ambient temperature, resulting in low fluorescence output. Once hybridized to the target analyte, the aptamer halves hybridize via a proximity-induced effect with an enhancement in the fluorescence of MG as the core is reformed. The process of developing the SMG sensor utilized an iterative approach. Thus, the convention for naming our sensor designs follow the format of SMG-#A/B where # refers to the iteration of the sensor, A indicates strong analyte binding arm, and B for the weak analyte binding arm. Within each iteration, designs are further classified (SMG-A/B-#.#L/T) to denote different pairs within the same generation. The usage of the letters L or T specifies the inclusion of a hexaethylene glycol or tri-thymidine linkers, respectively

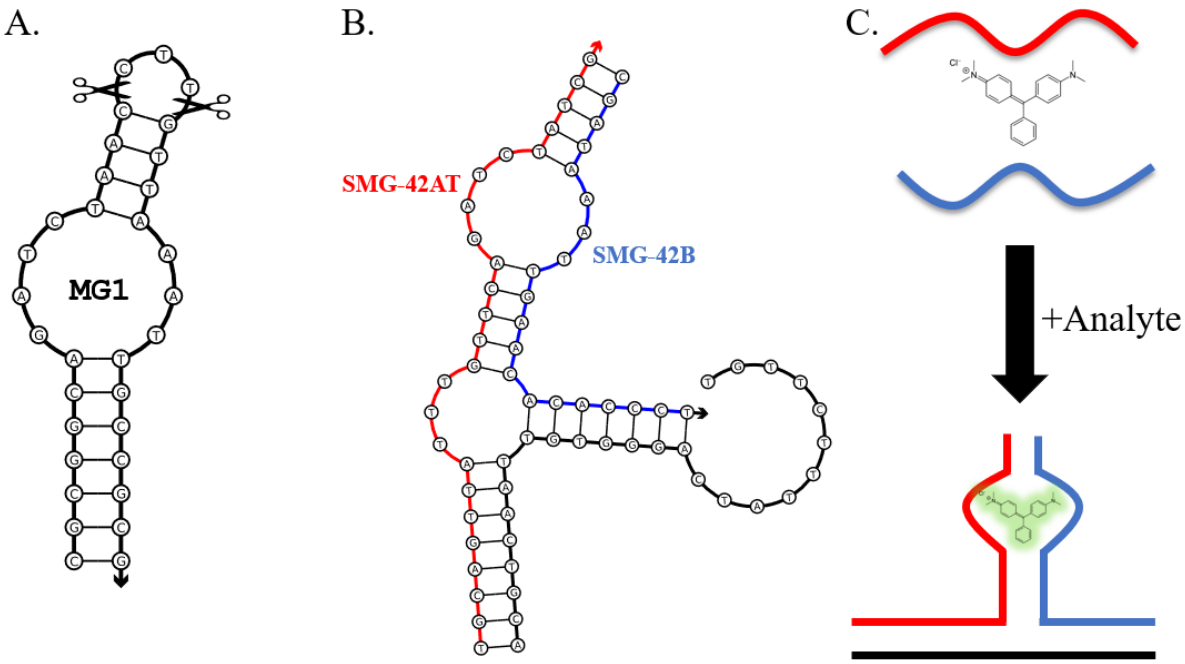


Figure 8: The parent aptamer structure for Malachite Green (A), the sequence diagram of the most active complex SMG4.2T (B) and graphical depiction of the split aptamer sensor approach (C). The reported aptamer sequence was folded via NUPACK and produced the resulting secondary structure prediction. The trinucleotide loop present on the upper stem was excised resulting in splitting of the aptamer into two halves. The result of our series of optimization is shown with the most active construct, SMG4.2T, which is shown in B. C depicts the approach to split aptameric sensing, in which the two halves of the aptamer are incapable of hybridization, and therefore signaling, prior to the hybridization with the target analyte. This induces a proximity hybridization event which restores the aptamers activity for detection.

Second Generation Split Malachite Green Sensor (SMG2)

Our initial work focused on establishing the lowest limit of detection (LOD) of the SMG sensor platform with no regards to the specificity. To this end, a sensor was developed utilizing the split aptamer sequence with long near symmetric analyte binding arms of 20 and 21 nucleotides. At these lengths, the binding between the aptamer sensor and its target analyte will be strong (–65.0 kcal/mol as predicted via NUPACK [83]) but will lack SNV selectivity particularly at ambient temperatures. The resulting complex between the sensor and its analyte is

shown in Figure 9 (top) with its accompanying calibration curve (bottom). With this design, an LOD was obtained in the low nM range (1.52 nM).

SARS-CoV-2 is a positive sense RNA virus and the direct application of the SMG2 sensor to interrogation of viral RNA is attractive [86]. This direct application could eliminate the need for a prior amplification and reverse transcription step by directly sampling purified viral cultures, though feasible the amount of culture that would be necessary is impractical. This is mentioned because an RNA analog of our target analyte sequence was procured and applied in the sensor assay. Unexpectedly the aptamer failed to produce any detectable signal above background, i.e. it achieved an SBR of 1 indicating no fluorescent enhancement. This result is shown in Figure 10. No further work with this RNA analyte was attempted but the significant difference in activity warrants a discussion, if one that is more conjecture that empirical in nature. This lack of activity could at first be attributed to the relative difference in stability between the RNA and DNA analyte, but at the lengths of the binding arms used in this sensor the influence of the folded structure is mote due to the large free energy associated with complex formation. Rather this effect may reflect a geometric sensitivity of the aptamer to adjacent RNA-analyte binding arm helix. RNA duplexes adopt a different confirmation than that of DNA, referred to as confirmation A and B helices respectively, which possess different geometric parameters such as bases per turn, pitch and groove dimensions. The formation of RNA-DNA hybrid helix results in a geometry that is hybrid between the two. The influence of this hybrid confirmation on the folding of the aptamer could result in complete ablation of its activity and could be restored by some of the structural modifications undertaken in further iterations. Further

study of this disparate activity could elucidate some of the influence of adjacent helices on the folding and function of aptamer sensors.

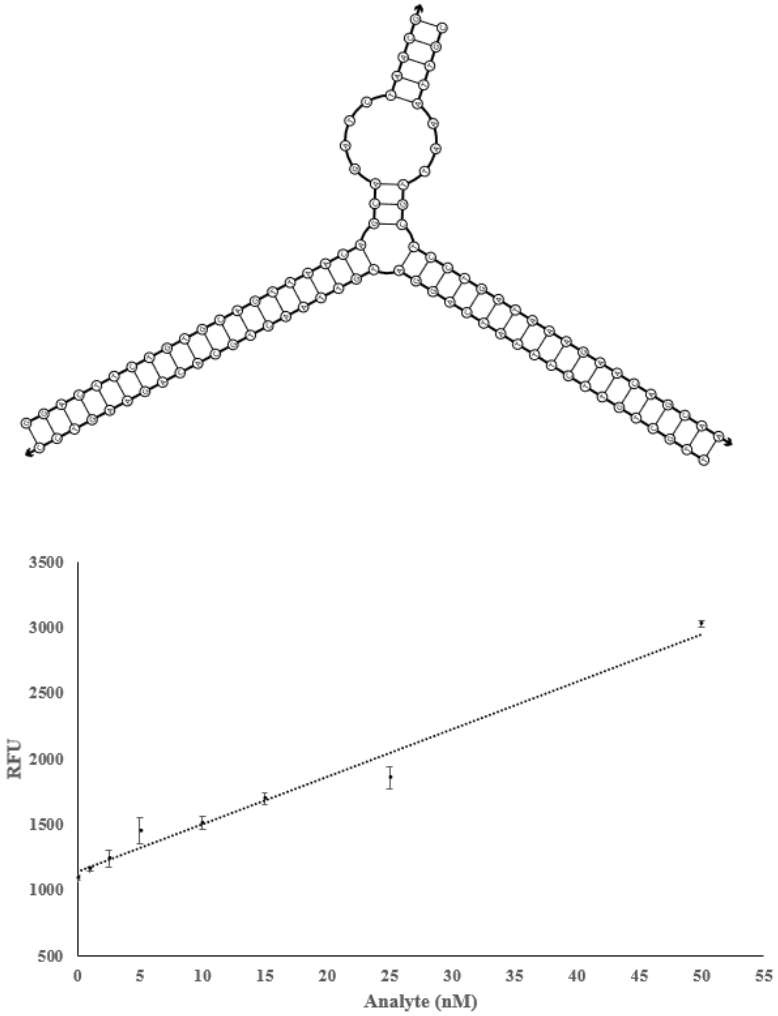


Figure 9 Sequence diagram and calibration curve for SMG2. SMG2 full sequence diagram is shown on top which depicts the long binding arms that cover the entire analyte of NC45A. This is expected to produce a highly sensitive assay that is non-selective against mutations in its target sequence. This sensor was utilized to detect NC45A within a concentration range of 1-50 nM. Here the sensor strands, at a concentration of 500 nM, were incubated with NC45A at the indicated concentrations for 105 minutes prior to the addition of MG to a final concentration of 1 uM and final volume of 100 uL. This was followed by an additional 15 minutes of incubation before the fluorescence was measured at 658 nm with an excitation of 634 nm. The LOD was determined by the finding the threshold signal, equal to the average background + 3 standard deviations, and was solved via Microsoft Excel. The above represents the average curve of three independent trials.

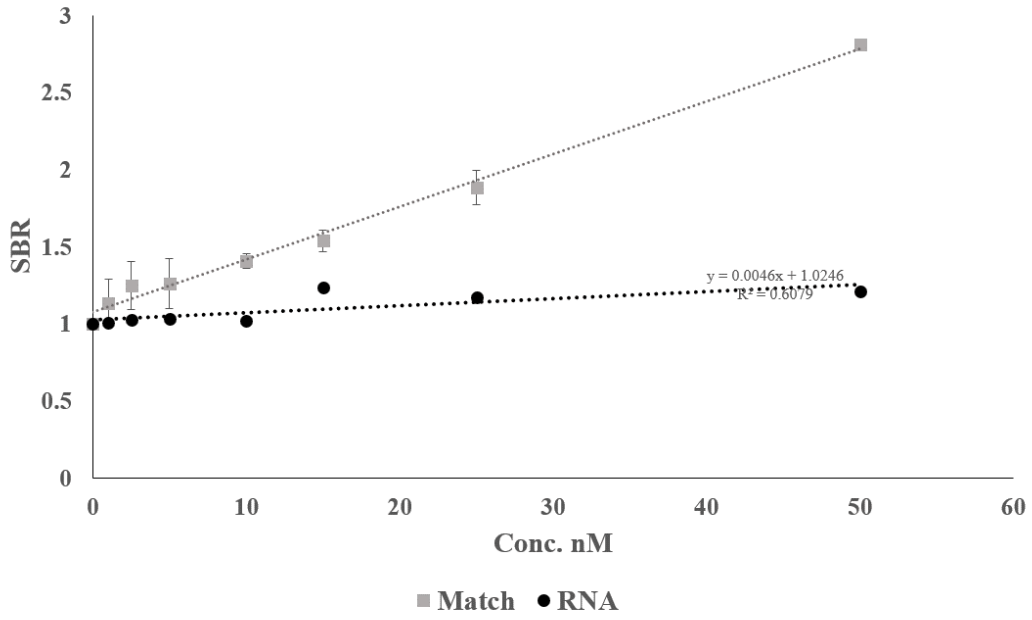


Figure 10: SMG2 is incapable of detected RNA analytes. Here SMG2 was unable to detect the RNA analog of NC45A. Here the sensor strands, at a concentration of 500 nM, were incubated with either NC45A or rNC45A at the indicated concentrations for 105 minutes prior to the addition of MG to a final concentration of 1 uM and final volume of 100 uL. This was followed by an additional 15 minutes of incubation before the fluorescence was measured at 658 nm with an excitation of 634 nm. The LOD was determined by the finding the threshold signal, equal to the average background + 3 standard deviations, and was solved via Microsoft Excel. The above represents the average curve of three independent trials.

Within this generation a second derivative was developed, SMG2B, which shared common binding arms but reduced the top stem by two base pairs. This reduction in stability of the top helix was anticipated to result in a performance reduction, which was observed with an achieved LOD of 7.9 nM. This reduction can be attributed to the relatively weak binding, which is depicted in Figure 11, but required lowering the temperature parameter of the NUPACK folding to 10C. Lower temperatures favor the formation of weak nucleic acid structures as transient base melting, which is compensated for by the presence of more base pairs, is decreased in frequency. At room temperature, the formation of a helix is not predicted to occur even in high salt concentrations (NaCl = 1M). These predictions are accurate for the formation of canonical

structures, such as the double helix, but fail to take into account any additional stabilizing factors. One such factor is the potential contribution of the folded aptamer core, which may help stabilize the upper helix and allow for signaling. The relatively small changes in structure, two base pairs, had significant impacts on structure and function of the split sensor.

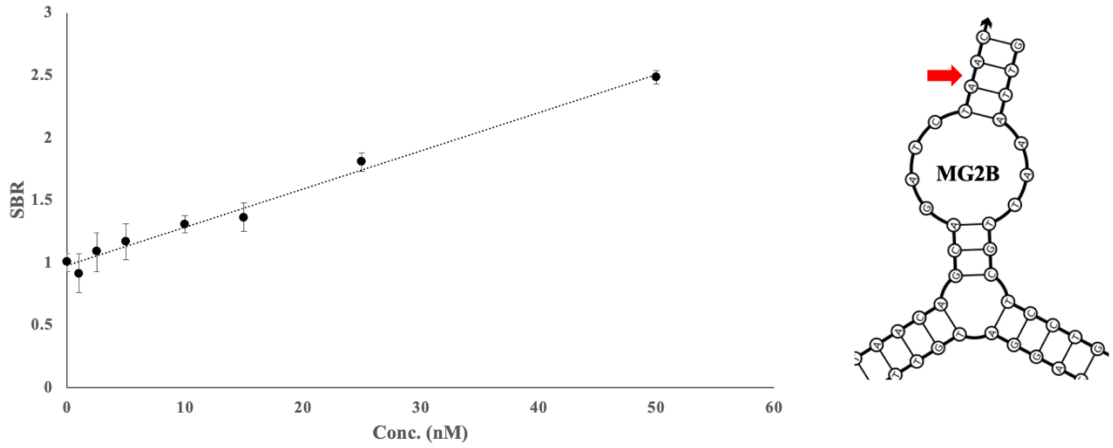


Figure 11. Calibration curve and partial sequence diagram of SMG2B. The calibration curve for SMG2B is shown left, which obtained an LOD of 7.2 nM. A partial sequence diagram is depicted on the right with an arrow indicating the position which was reduced by two base pairs in comparison to SMG2. Here the sensor strands, at a concentration of 500 nM, were incubated with NC45A at the indicated concentrations for 105 minutes prior to the addition of MG to a final concentration of 1 uM and final volume of 100 uL. This was followed by an additional 15 minutes of incubation before the fluorescence was measured at 658 nm with an excitation of 634 nm. The LOD was determined by the finding the threshold signal, equal to the average background + 3 standard deviations, and was solved via Microsoft Excel. The above represents the average curve of three independent trials.

Third Generation Split Malachite Green Sensors (SMG3)

From these initial designs, the next iteration of SMG sensor variants (SMG-3.#A/B) were designed to determine the influence of linkers between the sensing elements (analyte binding arms) and the reporter unit (the aptamer core) on sensor functioning. The analyte binding arms from SMG-2 were edited to produce two asymmetric arms consisting of 17 nucleotides (SMG-3A) and 8 nucleotides (SMG-3B). The longer arm, designated SMG-3A,

provides a stable and high affinity interaction with the target while the shorter arm, SMG-3B, provides selectivity by being intolerant of mismatches within its shortened helix. This intolerance is attributed to the short length, 8 nucleotides, which results in a melting temperature below ambient temperatures and requires stabilization from SMG-3.1A to facilitate binding to its target site. Thus, binding of SMG-3.1A to the target is required to facilitate the binding of SMG-3.1B which is sensitive to the single nucleotide variation found in the sequence, increasing the selectivity of the sensor with little reduction in affinity via the cooperative binding effect. However, addition of the analyte binding arms to the aptamer can result in alterations to the binding pocket as the geometry of the complex differs from that of the native aptamer, lowering performance. This tension between binding pocket formation and tight analyte binding can be circumvented by use of linkers, such as hexaethylene glycol or trithymine linkers [32], which separates the sensing element of the SMG sensor (the analyte binding arms) from the signaling core. Similar investigations focused on linker length optimization to improve affinity for its target or its performance when functionalized onto a solid support, such as gold nanoparticles [87, 88]. Presumably, these linkers provide the aptameric sensor flexibility to facilitate binding between the sensor and target while preserving its aptameric core's fold. Many of these alterations can be made "in the dark" as fine structural detail for aptamers is generally absent. This lack of detail does not rest with the absence of crystal structures but often extends to the sequence requirements or base contributions to aptamer binding, which is a core aspect of aptamer function. The neglect of these kind of investigations is attributed to a cost-benefit analysis, where the time and investment in elucidating these core principles fail to produce a compelling impetus. However, the lack of these details often make application of aptamer

sensors inherently difficult, as further work is stymied by pursuing unproductive avenues which makes application of aptamers to functional devices difficult. As such many of these modifications rely on intuition and a general sense of how the aptamer folds. Undeterred, derivatives of the split SMG-3 sensor were designed that incorporated trithymine or hexaethylene glycol linkers in the analyte binding arms of SMG3.1A, SMG3.1B or both (Figure 3). These variations were screened for their ability to induce MG fluorescence with emphasis being placed on determining which architecture achieves lower background signal and a high turn on fluorescence in the presence of target. Both of which result in the higher SBRs and should produce a more sensitive device. As shown in Figure 12, linker position and composition had effect on both the background and turn on fluorescence signal for the SMG sensor. The inclusion of hexaethylene glycol linkers did not provide a fluorescence increase but did result in reduction of the background fluorescence when compared to the split aptamer without the linkers. This effect was present when the linker was present on either or both analyte binding arms (compare columns A-D, with background RFU values of 1063, 1026, 725 and 858 respectively). The lowest background fluorescence was obtained when the hexaethylene glycol linker was placed on both analyte binding arm (SMG3.1AL and SMG3.1BL), reducing the background fluorescence by 32% from a RFU of 1063 (for SMG-3, column A) to 725 (For SMG-3.1AL/BL, column C) for the unmodified and modified sensor respectively. The use of a trithymidine linker on the strong analyte binding arm (SMG-3.1AT, where T signifies the trithymidine linker) produced a simultaneous increase in turn on signal and a reduced background fluorescence (compare columns A, B, and E). This reduction was similar to the value when utilizing ethylene glycol linkers on the weak analyte binding arm or both arms

(SMG-3.1AL and SMG3.1BL, column C), with RFUs of 725 and 736 for column C (dark bar) and column E (dark bar), respectively. In contrast, the inclusion of the trithymidine linker on the strong analyte binding arm) resulted in a significant enhancement of ON fluorescence when compared to the unmodified sensor (Columns A and E, light bars with values of 2777 and 3925 respectively). While the structural impact behind this improvement has not been studied in great depth, previous work with aptamers have shown similar preferences [84, 85]. The exact mechanism behind how these modifications influence aptamer function is not well understood and may require individualized study due to heterogenous folding patterns across aptamers. This heterogeneity extends in multiple directions, in which aptamers of disparate sequences may bind the same ligand but exhibit similar folds or possess similar sequences but fold into disparate structures by utilizing different aspects of the bases to achieve their binding. Qualitatively, the inclusion of linkers results in separation of the target analyte-analyte binding arm helix from the aptamer core. This increased space may allow for better accommodation of the folded core, thus improving the ON signal in presence of the analyte. The difference in activity between the use of hexaethylene glycol linkers and trithymidine linkers may be attributed to the qualitative difference in flexibility or the separation distance between the core and binding arm-target helix. The internal hexaethylene glycol linker is roughly equivalent to two base pairs and provides more flexibility than the naturally occurring phosphodiester backbone of the oligonucleotide and demonstrates a lack of charge. When the trithymidine are utilized as linkers, the linkage is longer and more rigid due to base stacking interactions between the thymines and there is maintenance of charge. A possibility exists that this increase in flexibility imparted by the linkers may lower background fluorescence by stabilizing the secondary structures of the non-complexed aptamer

sensor strands, reducing the proportion of strands folded into an active confirmation in the absence of target. When the aptamer-analyte complexed is formed the trithymidine linkers, being slightly longer than its hexaethylene glycol counterpart, may facilitate the folding of the aptamer into adopting of a more natural fold. This more natural fold would result in an increased activity of the aptamer core, reflected in higher fluorescence signal by the tri-thymidine modified analyte binding arm.

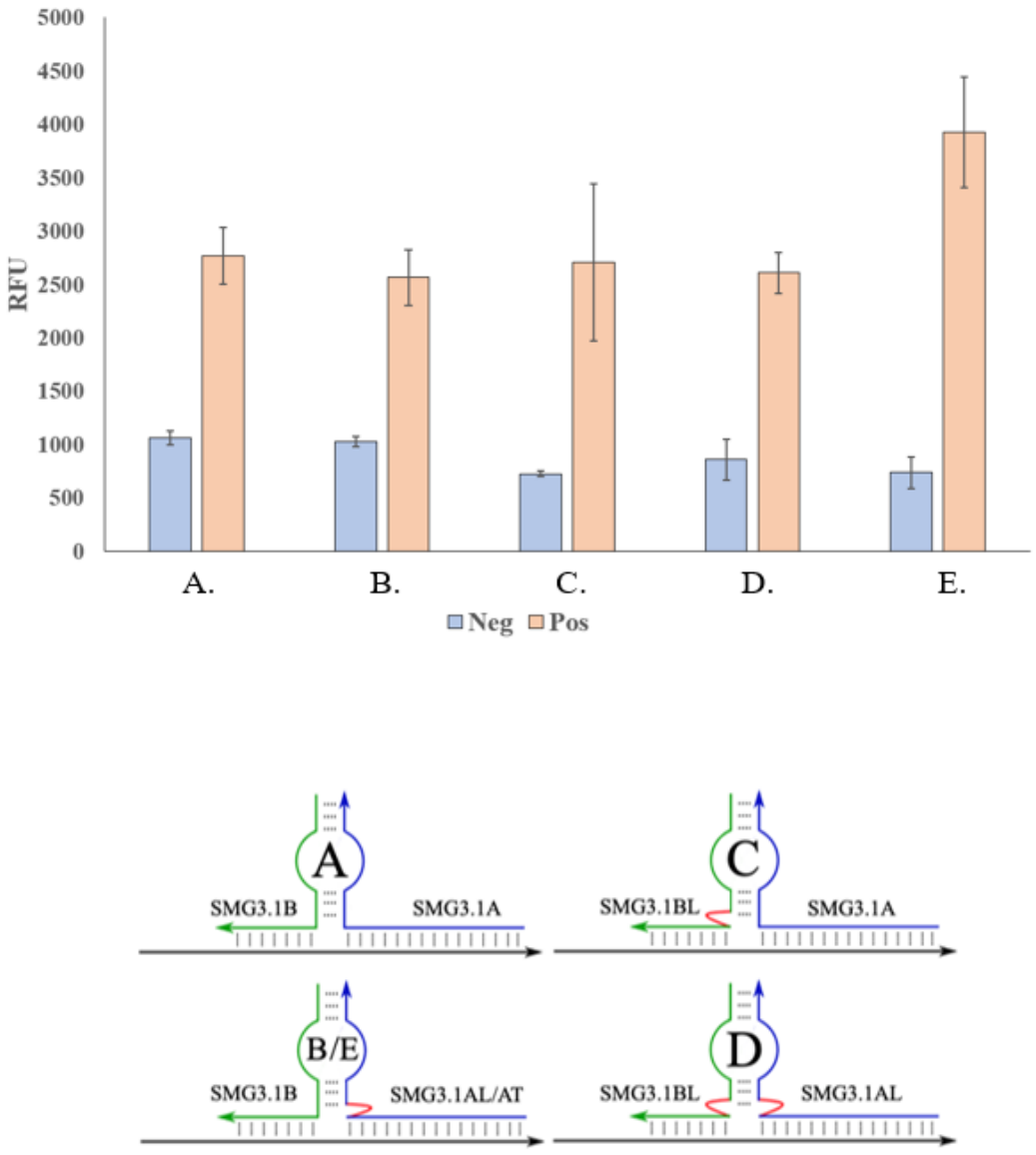


Figure 12. Linker Composition and Position Influence SMG-3 Performance. Top a schematic of SMG-3.1 complexed with its target analyte (NC-45) is shown. The strong analyte binding arm is depicted blue, with the SNV sensitive arm depicted green. Linker position is denoted via red line. Composition of the linker is shown by the use of L (hexaethylene glycol) and T (tri-thymidine). Bottom is MG fluorescence measured when 250 nM of sensor strands (SMG3.1A and SMG3.1B) was complexed with 250 nM of its target analyte (NC-45). Inclusion of a trithymidine linker on SMG3.1A resulted in simultaneous lowering of the background fluorescence (dark grey) and elevation of the fluorescence in the ON condition (grey). The above image is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Fourth Generation Split Malachite Green Sensor (SMG4)

Having established that the presence of a trithymidine linker on the strong analyte binding arm, SMG-3.1AT, strand produces the most highly active complex, further optimization was performed to improve the signal-to-background ratio (SBR), which is defined as the ratio of ON signal (in presence of analyte) and its absence (OFF). Optimization focused on the influence of the upper and lower stem on the development of MG fluorescence with emphasis given to the stem length and GC content. Both of these factors will influence the stability of the aptamer core with longer stems or higher GC content expected to give a more stable structure to aid the core folding and allow it to remain in the active confirmation. The respective free energies, as calculated by NUPACK, for each of the core-stem complexes are –8.2, –12.7 and –9.9 kcal/mol for SMG-3, SMG-4.1 and SMG-4.2 respectively. While this increase in stability is intended to elevate the sensor's fluorescence when complexed to its target, unintended increase in background fluorescence can occur as a higher proportion of aptamer sensor strands are folded into an active confirmation. To study this effect, SMG4.1 and SMG4.2 were designed to have equivalent total binding arm lengths of 24 nucleotides, which are comparable to SMG-3's total binding length of 25 nucleotides. Since these sensors all interrogate the same portion of their target sequence, there binding arms are nearly identical sequences. This sequence conservation allows for a more direct comparison of the stem's influence on the sensor's function. As can be seen in Figure 13, shown below, the increase in stem stability resulted in higher SBR ratios between generation 3 (SMG-3.1T) and generation 4 (SMG-4.1 and SMG-4.2). Further SBR improvement was elicited by altering the GC-content the stems from 60% in SMG-4.1 to 40% in SMG-4.2, reducing the stem stability to an intermediate value between SMG3.1 and SMG4.1.

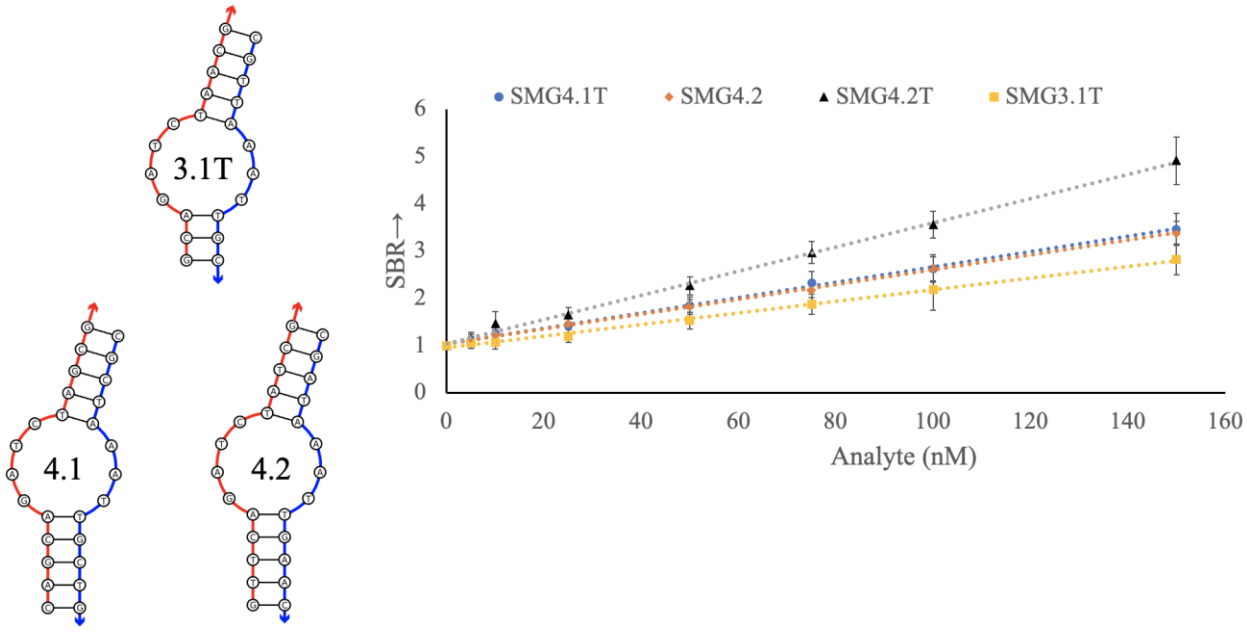


Figure 13. Structures of stem components around the MG DNA aptamers core sequence. Between generation 3 and 4 of the SMG sensor, structural features surrounding the core sequence were optimized to improve the achievable SBR in our standard assays. These generations utilized common binding arms, allowing a direct comparison between the different stem structures. Between generation 3 and 4, the lower stem was extended to provide a more stable structure. This was further optimized by changing the GC content to produce a slightly less stem, between SMG4.1 and SMG4.2.

Higher GC content is associated with higher melting temperatures a measure of stability, so the reduction in GC content lowers the stability of both the upper and lower stem between SMG-4.1 and SMG-4.2. This improvement at first appears counter intuitive, as an increase in stability demonstrated a higher SBR between SMG-3 and SMG-4. However, the noted improvement in SBR was a consequence of a reduction in background signal for SMG-4.2T as opposed to a significant increase in fluorescence magnitude. This reduction in background can be attributed to a smaller proportion of associated SMG-A and SMG-B strands in the absence of its target, which is influenced by the stems ability to form a stable complex in the absence of target. The full sequence diagrams and calibration curves for SMG41A, SMG41AT, SMG42A and SMG42AT are shown in figure 14 and 15, respectively.

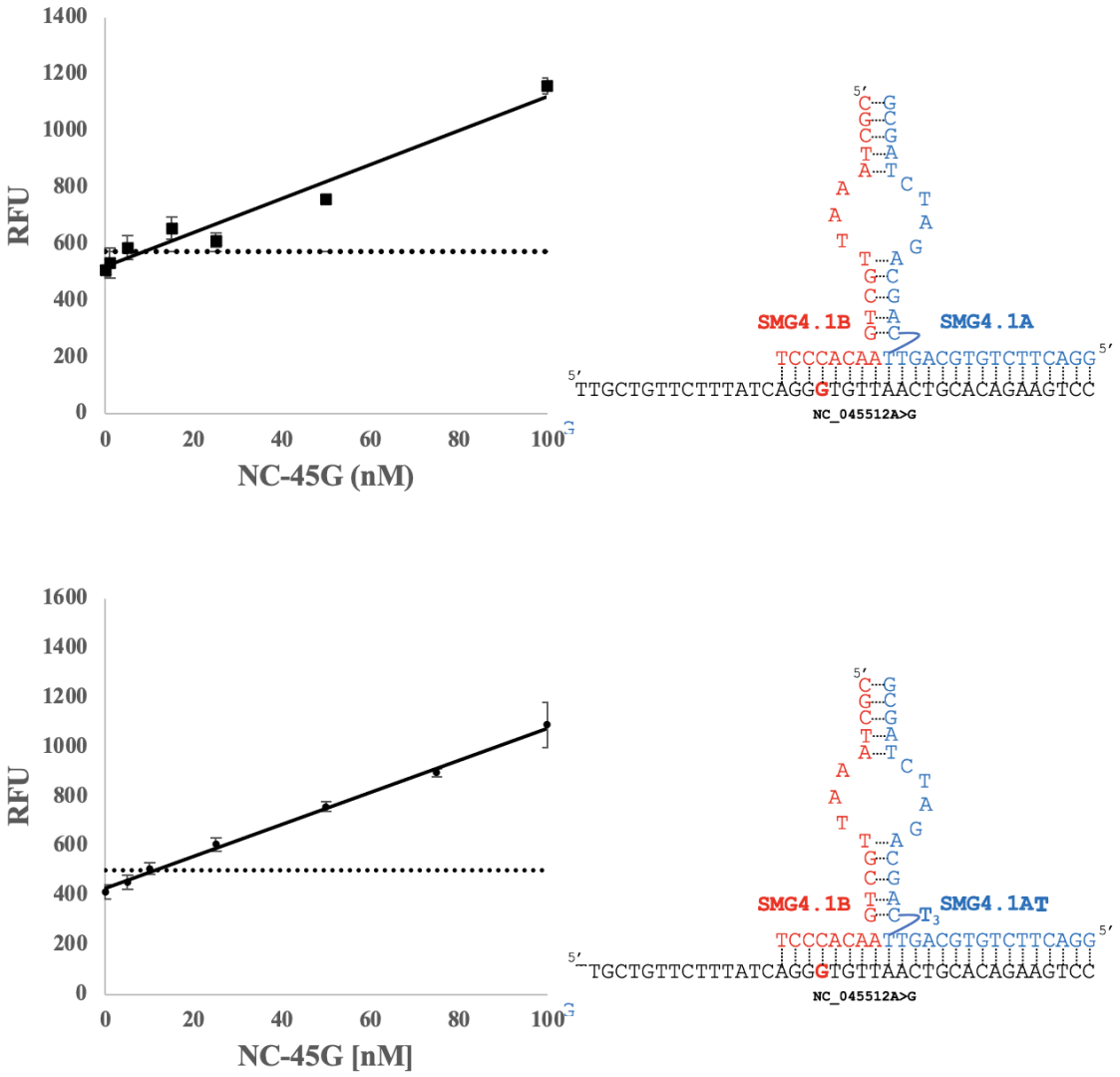


Figure 14. Sequence diagrams of generation 4 SMG sensors and their perfomance in a calibration curve assays. Sequence diagrams of SMG-4.1, SMG-4.1T, and SMG-4.2T are depicting right with their accompanying calibration curves left. to the stem elements. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$ is indicated by the horizontal solid line. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

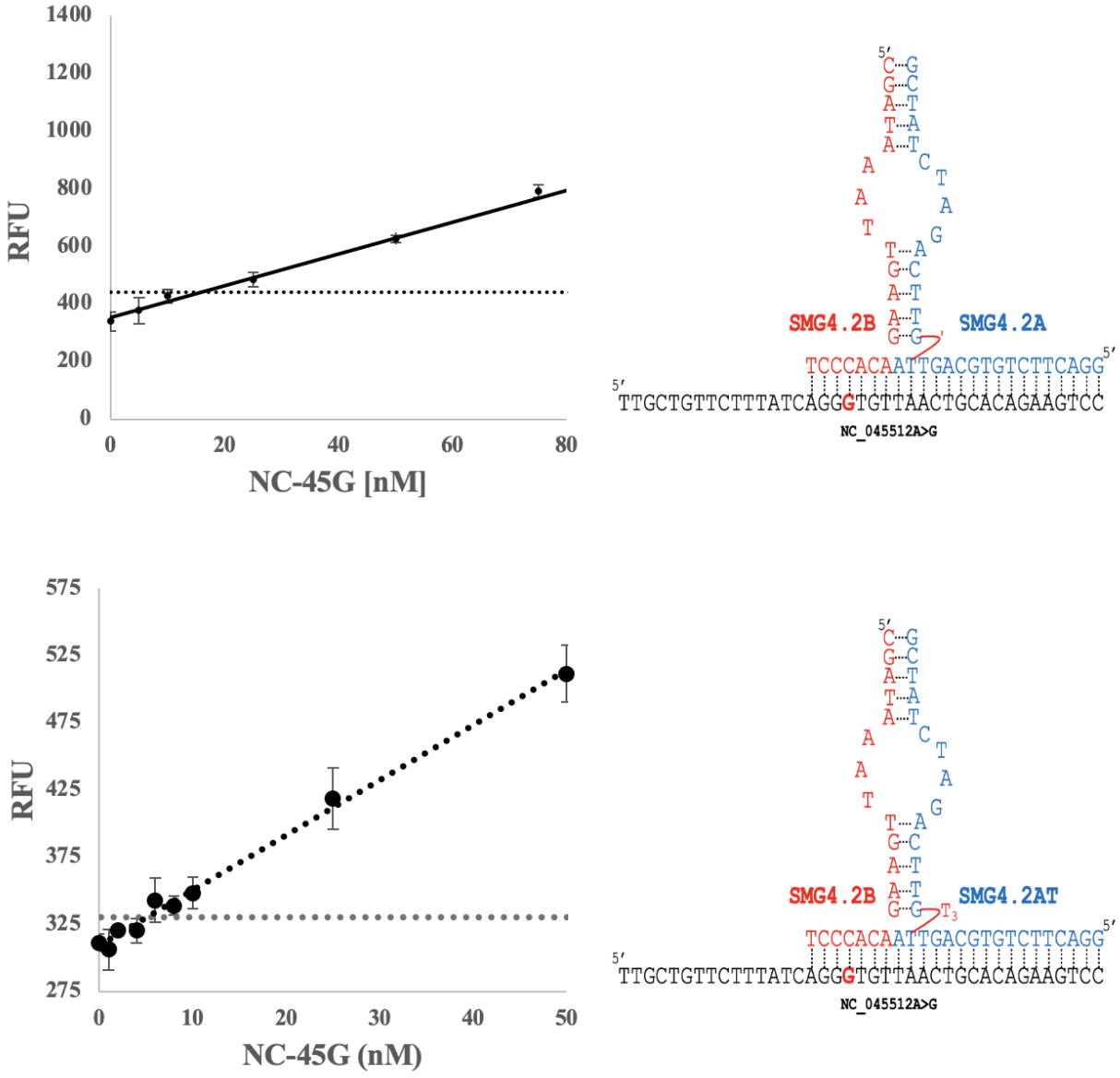


Figure 15. Sequence diagrams and calibration curves for SMG4.2 and SMG4.2T. The calibration curve for each sensor is shown left with the accompanying diagram on the right. The inclusion of the trithymidine linkers resulted in improvements to the LOD of the resulting sensor. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$ is indicated by the horizontal solid line. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

The intent of these structural optimizations is to produce a complex that can sense analytes but retains activity close to that of the original, and in some respects optimal, parent aptamer. This decouples the sensing and signaling aspect of the sensor, which ideally allows the

exchange of target specificity by exchange of the binding arms. It should be stated while these results demonstrate that this goal as achieved, the generalizability of this approach is an open question. This aptamer was relatively small and the proposed core sequence relatively simple in terms of sequence. Larger or more complex aptamers might require different optimization strategies. The result of our series of optimizations is summarized in Figure 16, shown below, which demonstrates that SMG-4.2T has activity comparable to the intact MG aptamer at equimolar concentrations of 1 µM (SBR equals 10.8 and 11.8 for SMG-4.2T and MGwt, respectively).

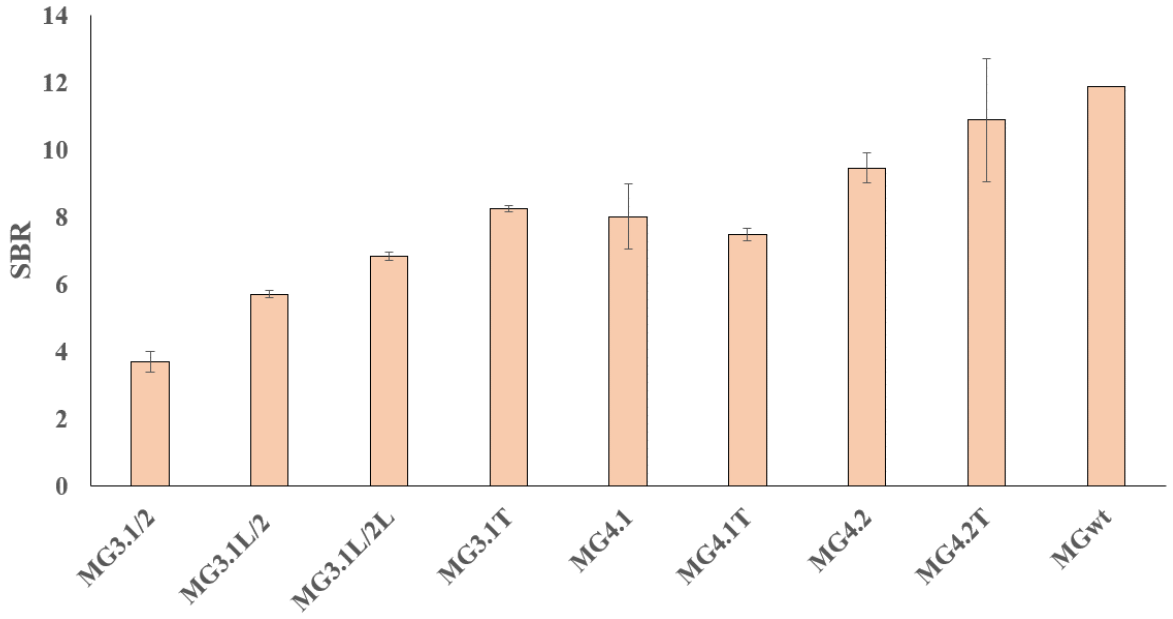


Figure 16. Overall optimization of iterative SMG sensors (SMG-3 – SMG-4.2T) as demonstrated by improvements by SBR. The overall improvement in performance, a 3-fold increase in SBR, across generations of sensors. The final iteration of the SMG sensor (SMG-4.2T) has performance comparable to the original intact aptamer at equimolar (1 µM) concentrations. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

The SDA Sensor is a Generalized Sensing Platform.

The original dapoxyl aptamer w as highly active, capable of increasing the intrinsic fluorescence of its cognate dye over 700-fold [75]. Work by Kikuchi resulted in generation of an optimized SDA sensor which relied on a combination of both hexaethylene glycol linker and dithymidine linkers to produce the most active complex with a reported LOD of 0.44 nM [85]. In this work, we create a series of derivatives of this sensor quickly and efficiently by exchange of the analyte binding arms from the original publication with two sets of COVID-19 analyte specific arms. The results of these exchanges for an SDA sensor for NC-45 and NC-14 are found in Figure 17. Of note is that the original dapoxyl aptamer was selected against a non-commercially available dye, dapoxyl sulfonyl ethylenediamine. Later work with the dapoxyl aptamer found that it exhibits promiscuous binding to a range of dyes, including commercially available Auramine O which was utilized for all experiments with SDA.

These results strengthen the argument that SDA sensing platform represents an optimal design, which can facilitate its development for other targets quickly with minimal reduction in performance. In an analogous method to sandwich based lateral flow assays, in which specificity can be altered by exchange of an antibody specific to the desired analyte in question. While not exhaustively examined, the narrow range of LODs achieved with different binding arms, instrumentation and experimenters provides evidence that the SDA platform represents an optimal and reliable design which could be quickly developed into sensors for different nucleic acid targets. The calibration curves and sequence diagrams of each generation of SDA sensor, SDA-CB, SDA3 and SDA4 are show in Figure 17.

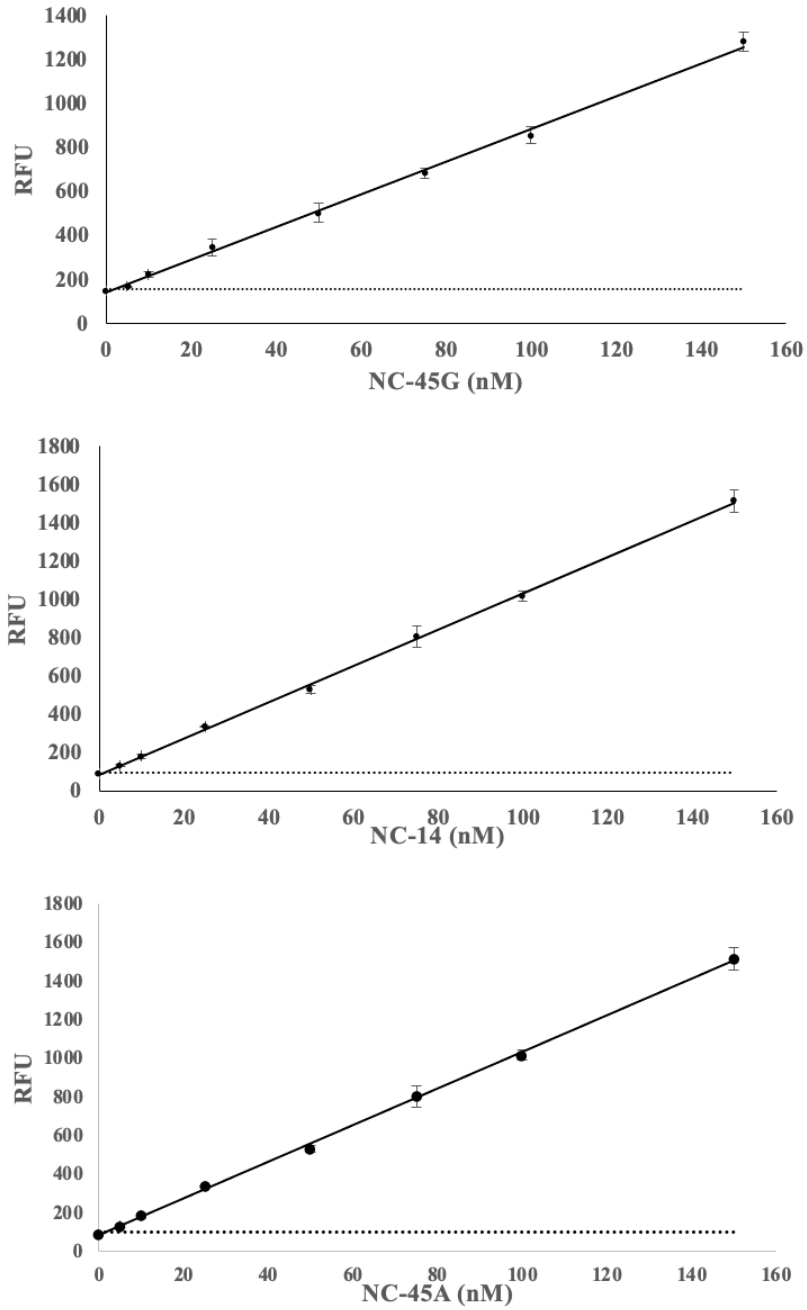


Figure 17. Calibration Curves of SDA-CB, SDA3 and SDA4 sensors with sequence diagrams. Left depicts calibration curves for SDACB (top), SDA3 (middle) and SD4 (bottom). The sequence diagrams of each sensor in complex with its target analyte (SDA4:NC-14 and SDA3 and SDACB:NC-45G). The obtained LOD for SDA-CB, SDA3 and SDA4 were 1.4, 1.8 and 1.3 nM, respectively. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$, is indicated by the horizontal solid line. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 525 nm with an excitation at 475 nm.

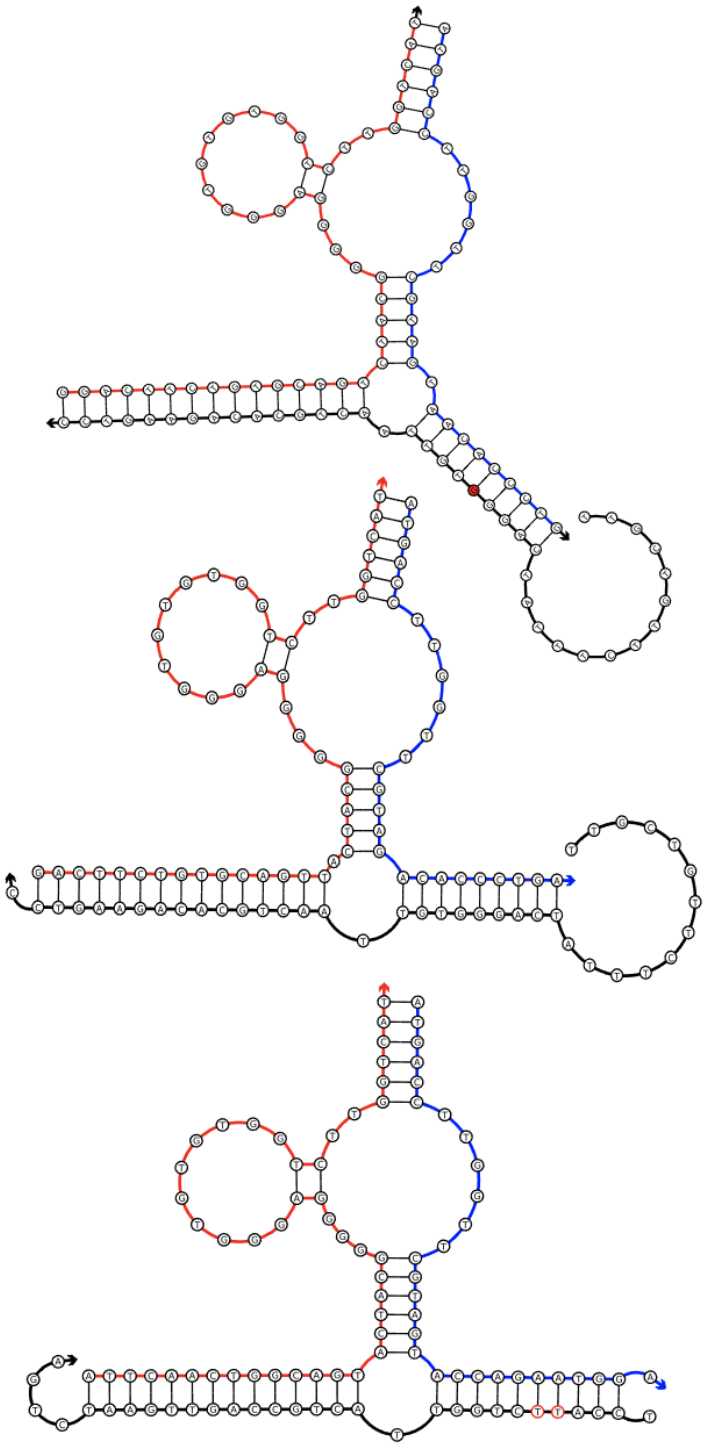


Figure 18. Sequenecs diagrams of different generations of SDA aptamers. The sequence diagrams of the various SDA sensors used in this study are shown. SDA-CB is shown top, followed by SDA3 (middle) and SDA4 (bottom).

SBR and LOD Difference in SDA3 and SMG4.2T

When comparing SBRs, all derivates of the SDA sensor far outstrip any SMG sensors when challenged by a common analyte (NC-45) in our assay conditions with SBRs of 10 and 4 (SDA3 and SMG4.2T), respectively with 500 nM of sensor strands and 150 nM of analyte. SDA3 and the SMG-4.2T sensors utilized common analyte binding arms in their designs, eliminating a difference in hybridization efficiency as an explanation for this discrepancy in performance. Instead, this performance gap may be attributable to the difference in aptamer affinity and mode of interaction with their cognate dyes. The dapoxyl aptamer exhibits a higher affinity for auramine O (K_d = 7-25 nM) [75] in comparison to the MG aptamer for MG (2.1 μM) [67]. This difference in affinity may contribute to the difference in the limits of detection between the two aptamers, 5.3 nM for SMG4.2T and 1.2 nM SDA3, as the dye complex between MG and its aptamer may not be as stable reducing its ability to induce torsional fluorescence. Another potential contributing factor could be the mode of recognition differs between the aptamers. Aptamer interactions generally progress through two modes, shape complementarity between the target and aptamer surface[89] or via incorporation of the molecule into the folded structure of the aptamer (encapsulation) [1]. Given the size difference between the aptamers and structural differences between their cognate dyes, it is plausible that the interaction mode differs with one strategy conferring a lower detection limit. Confirmation of this could be a fruitful endeavor for further development of aptamer-based biosensors. The sequence diagrams and comparative calibration curve of the sensing assay are shown in Figure 19 for comparison.

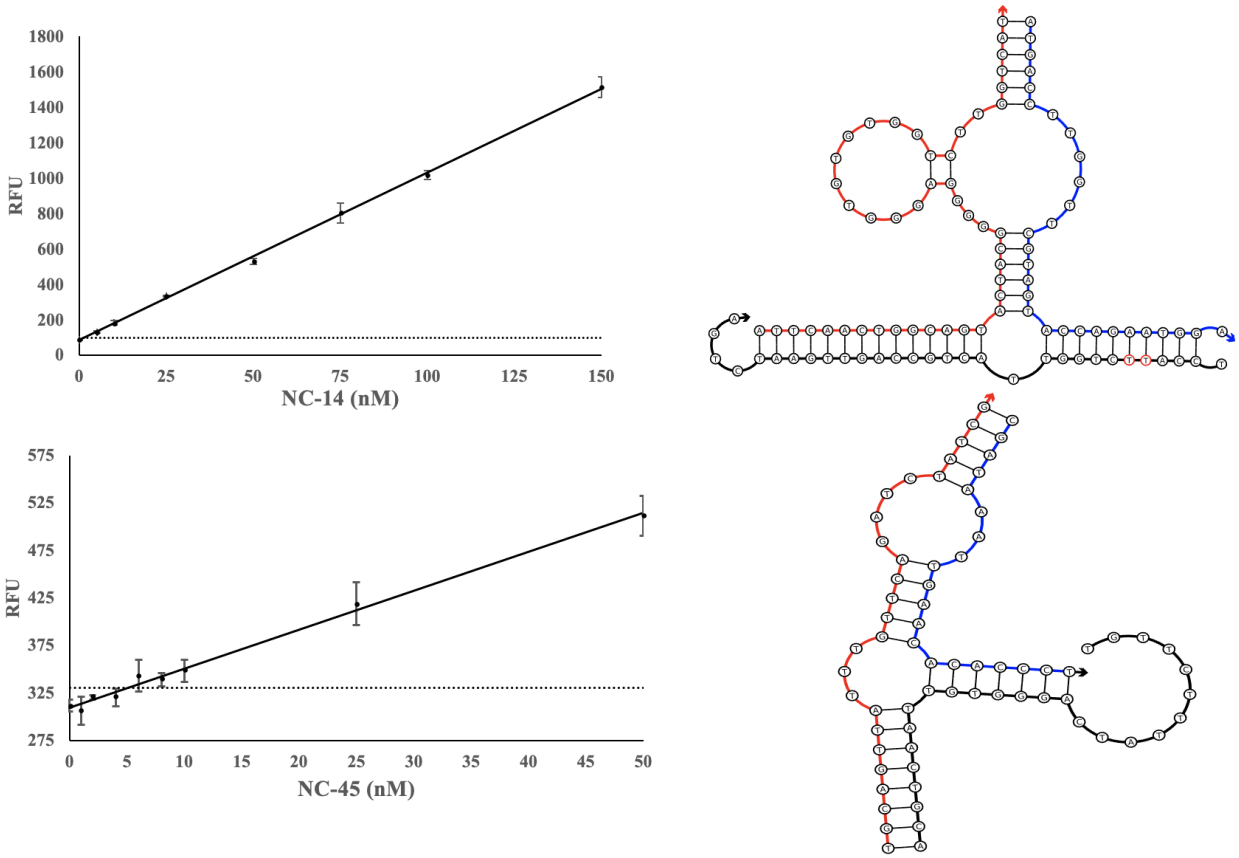


Figure 19. Comparative performance of SDA3 and SMG4.2T for the common analyte of NC-45G are shown left with accompanying sequence diagram shown right. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$, is indicated by the horizontal solid line. The SDA curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 525 nm with an excitation at 475 nm. The SMG curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Multiplex Detection

Multiplex detection assays are a workflow that allows simultaneous detection of several targets within a single sample. Since the detection is simultaneous, spectral overlap of dye pairs could lead to reduced performance. The fluorescence spectra of MG and auramine O have very little overlap, making them suitable for simultaneous detection. In comparison to traditional hybridization approaches, the aptamer-based approach is more complex to develop. As sequence

interactions between the aptamer pairs can occur unexpectedly and the aptamer pairs could possess different environmental requirements which are determined by the aptamer's initial selection. Initial results were poor as the SDA sensor demonstrated significant crosstalk with the MG dye, as shown in Figure 20. This crosstalk was present even in the absence of the analyte for SDA, as seen in the condition of SDA+MG(-) which corresponds to the SDA sensor strands with no stabilizing analyte. In comparison the SMG sensor demonstrated high selectivity for its dye and did not significantly elevate the signal of auramine o in both the absence and presence of its target.

Selection of the dapoxyl and MG aptamers utilized different buffer conditions with the dapoxyl aptamer selection incorporating Mg^{2+} and K^{+} in comparison to the salt free selection of the MG aptamer. The SDA sensor platform demonstrates a K^{+} dependence for signaling which in combination with its G-rich sequence implies that the aptamer's folded confirmation incorporates a G-quadruplex (a planar arrangement of guanine tetrads) structural element, as stabilization of G-quadruplex structures requires the presence of mono-valent cations to fold correctly [90]. Initial results when performing SMG assays with SDA buffer resulted in low fluorescence enhancement even at high target concentrations. The SDA aptamer similarly experienced a significant decrease in fluorescence when utilized in a Tris only buffer, producing no appreciable increase in fluorescence above the background. This salt dependence is well established for aptamers, as aptamer cores often incorporate cations for stabilization of binding pockets [91]. As such, aptamers demonstrate an intrinsic sensitivity to environmental factors, in particular the ionic strength of the solution [92-94], and function best at the buffer composition similar to the one used during their selection. A screening was performed in which the buffer composition was varied between the two . Selection of the final buffer composition focused on achieving SBRs > 2 for both aptamer pairs.

This resulted in a final buffer composition (Multiplex Buffer) of 20 mM Tris-HCl (pH 7.4), 10 mM KCl and 0.4 mM $MgCl_2$. The results of this initial screening and the determination of a suitable compromise buffer are shown in Figure 21.

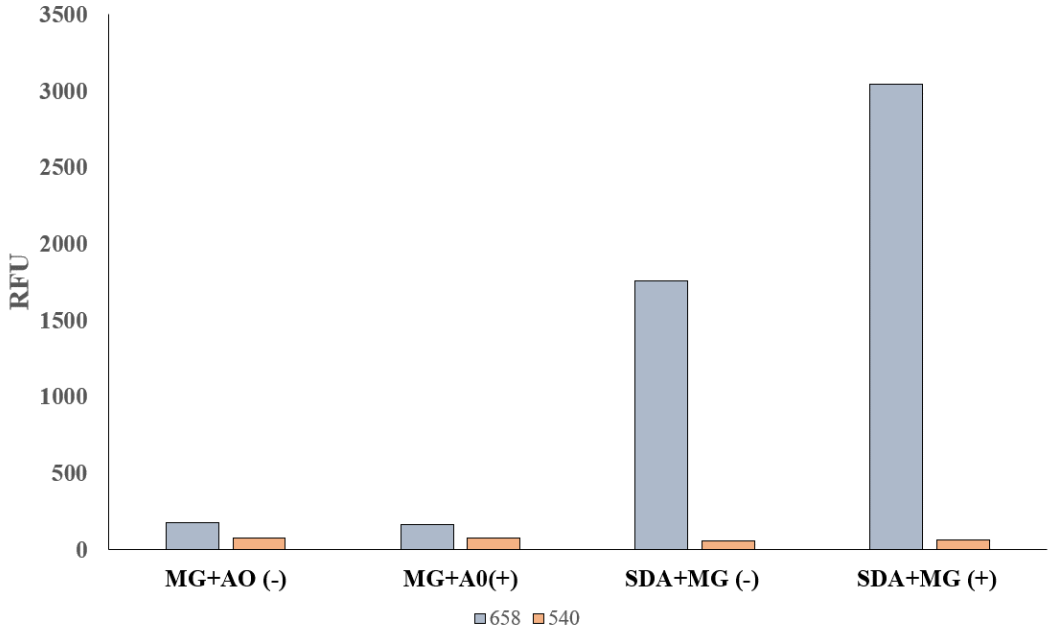
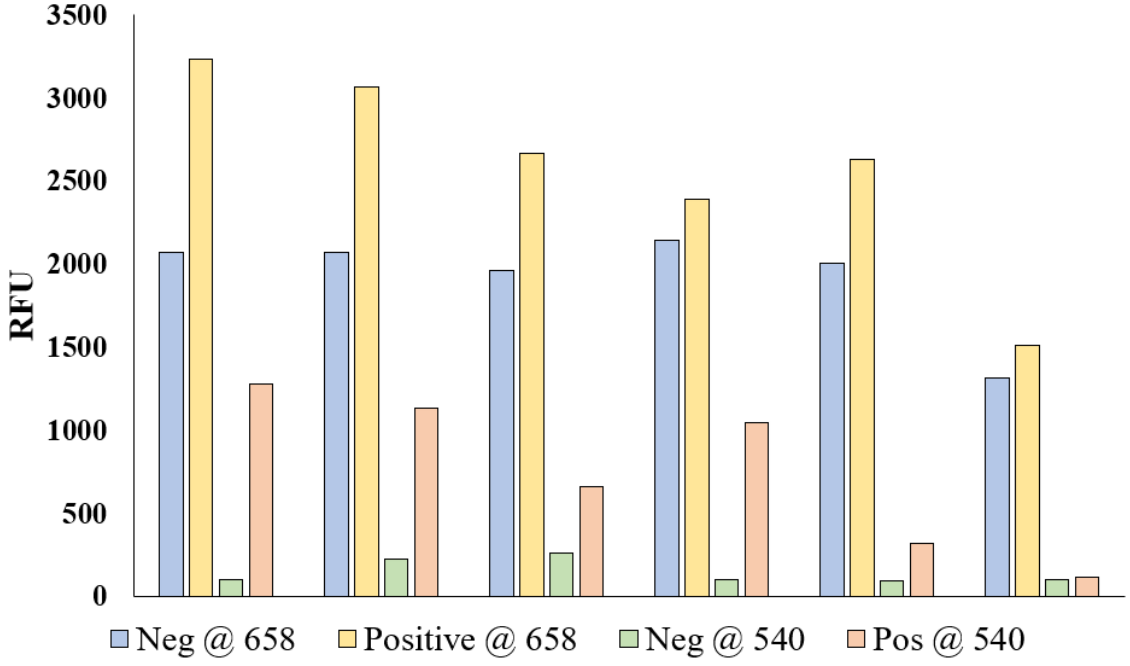


Figure 20. Demonstration of significant cross talk between SDA aptamer and MG (top) and optimization of the buffer conditions (bottom). Significant fluorescence enhancement of MG was observed when the SDA aptamer was in ON state or OFF state as indicted by the blue bars which are measurements of the fluorescence at 658 nm. MG aptamer is much more specific for its cognate dye and did not elevate Auramine O above background levels. Below each set of bars shows the conditions of the assay, with MG+AO(- and +) indicating the activity of the MG DNA aptamer in the absence (-) and presence of its target analyte (+). The same convention is utilized for the SD aptamer. The promiscuous binding of the SDA aptamer non-specifically increases the fluorescence of MG. The presence of MG fluorescence in the absence of analyte, indicates that the SDA aptamer's fold is stabilized by the presence of MG alone.



Tris-HCl (mM)	20	20	20	20	20	20
$MgCl_2$ (mM)	2.5	2.0	1.5	1	0.5	0.1
KCl (mM)	50	40	30	20	10	2

Figure 21. Optimization of multiplex buffer for the SDA Sensor and SMG Sensor multiplex assay. A range of buffer conditions, show in the bottom table below the image, were assayed to determine a useful buffer for multiplex detection. SDA was capable of increasing MG fluorescence even in the absence of its analyte, indicating a stabilization effect from the dye itself. The general trend of increased fluorescence at 540 nm, that of Auramine O, at higher concentrations of KCl and MgCl2 indicates that the structure most likely utilizes a G-quadruplex in its fold. The final conditions chosen correspond to 10 mM KCl with 1 mM of MgCl2. This retains the activity of SDA while limiting the change in MG fluorescence signal between ON and OFF states of the SDA sensor.

Once an operable buffer was found, the next stage of optimization focused on the relative concentrations of each aptamer pair. This optimization was a necessity as the dapoxyl aptamer demonstrates promiscuous dye binding, and could enhance the fluorescence of a variety of tri-phenyl or aromatic compounds, including MG [79]. At equimolar concentrations of 500 nM for both sensors, SDA4 produced a high background fluorescence for MG (crosstalk), which obscured the

enhancement from the SMG aptamer. However, this difficulty was overcome by exploiting the SDA sensor more efficient generation of fluorescence by lowering the concentration of the SDA4 strands (from 500 nM to 150 nM). Since the dapoxyl aptamer binds MG less efficiently than auramine O (K_ds of 6.9 µM and 7-25 nM, respectively [79]). This allowed a reliable increase in auramine O fluorescence while reducing the crosstalk between the SDA and SMG-4.2T sensors. Ultimately, an SBR of greater than 2 was maintained for each aptamer sensor. As seen in Figure 22, the multiplex assay was able to successfully detect the presence of one or both analytes. First giving a positive signal only in the presence of their respective analytes or in the presence of both unrelated analytes with very little increase in signal above the background in the absence of their appropriate target. This, to the author's knowledge, represents the first reported solution based multiplex assay utilizing self-assembling aptamers.

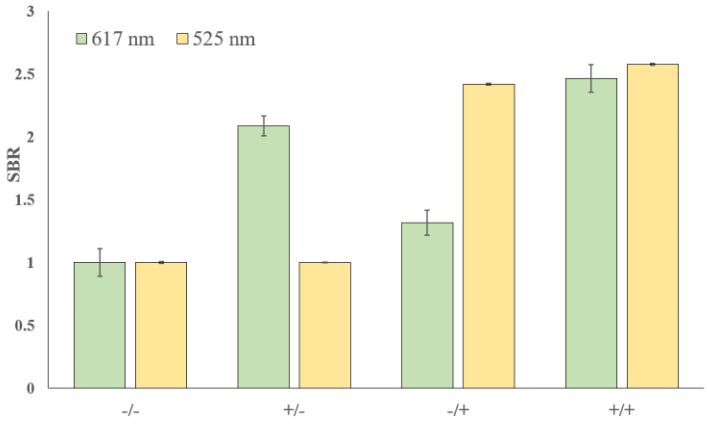


Figure 22. Multiplex Assay Utilizing the SDA4 and SMG4.2T sensors. SMG4.2T (500 nM) and SDA4 (150 nM) were incubated for 2h at room temperature (22ºC) in a buffer consisting of 20 mM Tris-HCl (pH 7.4), 10 mM KCl and 0.4 mM $MgCl_2$. The light grey bars represent fluorescence at 617 nm specific for SMG-4.2T, while dark grey bars signify fluorescence at 525 nm for SDA4. Sample +/- contained 150 nM of NC-45 (SMG4.2T's target analyte), while sample -/+ contained 150 nM of NC-14 (SDA4's target analyte). Some cross talk between the SDA4 and MG was observed (compare light grey columns in the -/- condition and -/+ condition) but both achieved an SBR > 2 for detection of 150 nM of their respective analytes. The above image is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Sensitivity of Split Aptamer Sensors

Aptamers exhibit salt dependence in unexpected, and sometimes contradictory ways, that seems individualized to each prospective aptamer. Illustrating this point is the effects of differing salt concentrations on the cocaine aptamer and ATP aptamer. Both aptamers are specific for small molecules but exhibit opposing behaviors in high salt concentrations. The cocaine aptamer binding is enhanced with increasing Na concentration while the ATP binding suffers at comparable ranges. These contradictory effects are most likely related to the mode of recognition and the final folded structure of each aptamer. The two aptamers used here, the dapoxyl and MG aptamer, were selected in conditions that significantly differed in terms of salt content. The MG aptamer selection buffer was relatively simple, just containing a buffering agent (TRIS) with no additional added salt, a highly unusual composition for selections. The lack of salt, in particular Na or K, most likely biased the selection towards sequences that had a relative low guanine content. In comparison the more standard buffer utilized for the selection of the dapoxyl buffer, such as the presence of 100 mM KCl, resulted in a selection pressures that more closely adhere to the norm. The resulting selected aptamer, rich in guanines, utilizes a G quadruplex as a structure to facilitate its ligand binding. Any changes from these set factors could impact the functioning of the aptamer, and therefore its sensing function. The influence of non-native buffer conditions on aptamer function was observed but not unexpected. To quantify the impact of these differing buffer conditions on the split aptamer sensors performances, a comparison between the achievable LOD with our sensor constructs in their native versus multiplex buffer conditions was performed. As can be seen in Figure 23, the LODs of both sensors were negatively impacted by use of the multiplex buffer. However, this impact was not evenly distributed between the aptamer pairs. The SMG-4.2T sensor

experienced an increase in the LOD of roughly 19%, from 5.3 to 6.3 nM. For comparison, LOD for SDA4 increased by 70%, from 1.3 nM to 2.2 nM. This degradation in performance by SDA4 can be attributed to the role of the cations in folding and signaling of the dapoxyl aptamer, most likely stemming from formation of the G quadruplex, as opposed to the MG aptamer which demonstrates no such salt requirement.

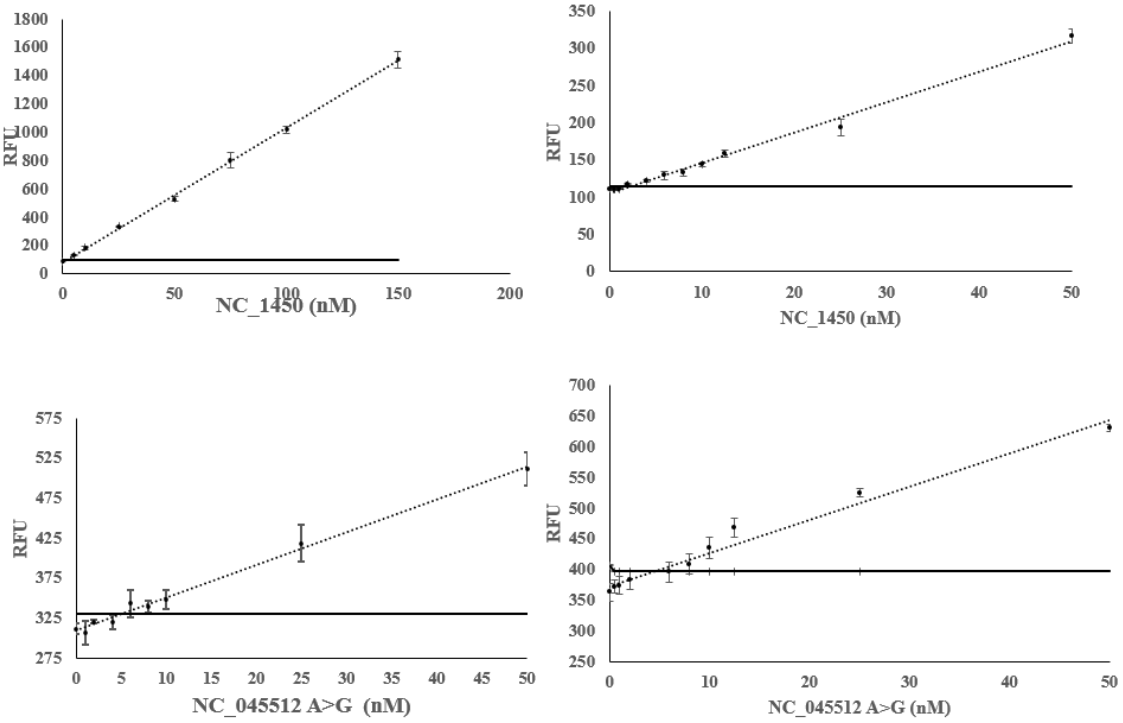


Figure 23. Comparison of SDA4 and SMG4.2T performance in native and multiplex buffer. The top two curves represent the fluorescence response of SDA4 in native (left) and multiplex (right) buffer. Calibration curves from SMG4.2T are shown bottom and follow the same convention. The sensitivity of both sensors was impacted in multiplex buffer but to different degrees with the impact on SDA4 more severe in comparison to SMG4.2T. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$, is indicated by the horizontal solid line. The above curves are the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Selectivity of Split Aptamer Sensors

With established LODs in both native and multiplex conditions, the selectivity of each sensor was examined by comparing the signal between analyte sets containing single or double nucleotide mismatches between the analyte binding arms and analyte. As can be seen in Figure 24 (left), SDA4 exhibited selectivity for its target when was challenged to differentiate both the single and double mismatch analytes. At 30 min, the SBRs obtained for the match (13.5), single mismatch (1.74) and double mismatch (0.97) demonstrate a highly selective sensor which is in line with previous reports and is suitable for qualitative determinations about the presence of absence of SNVs. In contrast, SMG4.2T demonstrated very little difference between matched or mismatched analyte with a signal difference of only 20% between matched and mismatched analytes (right). Given this lack of selectivity, earlier iterations of the SMG sensor were challenged to the same discrimination assay and demonstrated consistent inability to discriminate between the analytes NC-45 and NC-45A. This lack of selectivity was consistent throughout all generations of SMG sensors tested. This result was unexpected as discrimination was expected given the small size of the SNV sensitive analyte binding arm (8 nucleotides).

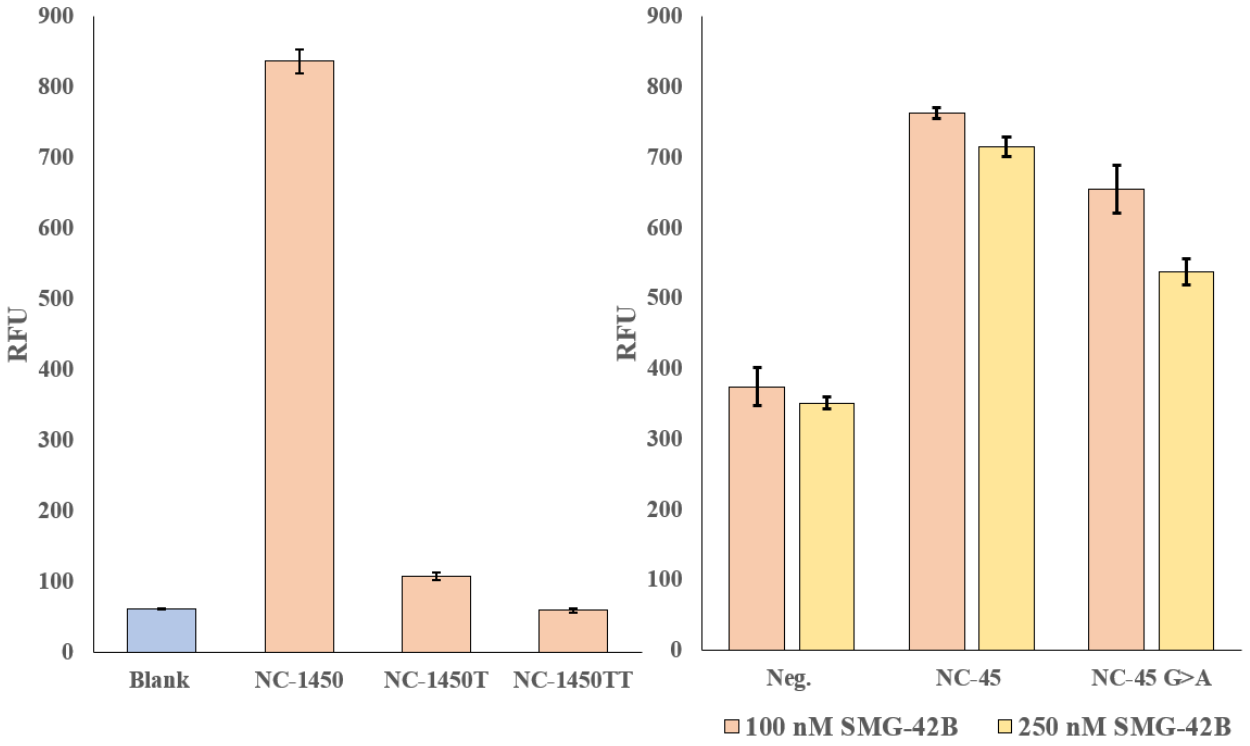


Figure 24. Selectivity of SDA4 Sensor for NC-1450 (M), NC-1450T (MM), and NC1450TT (MM) (left) and SMG42 For NC-45 (M) and NC-45G>A (MM). The conditions for the discrimination assay utilized 250 nM of the sensor strand for SDA4 and 250 or 100 nM for SMG4.2T and 150 nM of the matched (NC-1450 and NC-45G for SDA4 and SMG4.2B) or mismatched analytes.

Given the poor selectivity of the SMG sensor, a variety of derivatives were developed that explored different avenues to improve differentiation between the matched analyte (GGG) and mismatched analyte (GGA). This particular mismatch is somewhat challenging as the G rich sequence can stabilize even short helices, which may account for the observed lack of sensitivity. The main approaches to improving selectivity with these split hybridization probes can be done from a structural perspective, by such modifications as shortening the binding arm or by adding inhibitory fragments, or a functional perspective, by lowering the concentration of the SNV sensitive arm thus dissuading formation of a mismatch complex by reducing the equilibrium

concentration. Both these avenues were explored, with the former being heavily investigated by a series of modifications. The first and simplest approach is to alter the SNV sensitive binding arm, SMG-#B, by reducing the number of base pairs formed with the target analyte. To this end, a derivative designated SMG4.2B-6 was designed which reduced the SNV sensitive analyte binding arm from 8 to 6 nucleotides. Additional derivatives were made that added inhibitory fragments to the upper stem or lower stem of the weak analyte binding arm (SMG-B) or both. These inhibitory fragments result in intramolecular hybridization in the binding arm. Almost any sequence of nucleic acid will form some degree of secondary structure once it is sufficiently long, and the SMG-A binding arm is no exception. These inhibitory fragments are designed in such a way that the intramolecular folding results in obscuring the site necessary for the second half the aptamer sensor to bind, which makes the hybridization between the two more difficult to achieve and thus is expected to dissuade interactions in the absence of the target analyte. The sequence diagrams for these derivatives are show in Figure 25 while the selectivity results and secondary structures of 42B are show in Figure 26.

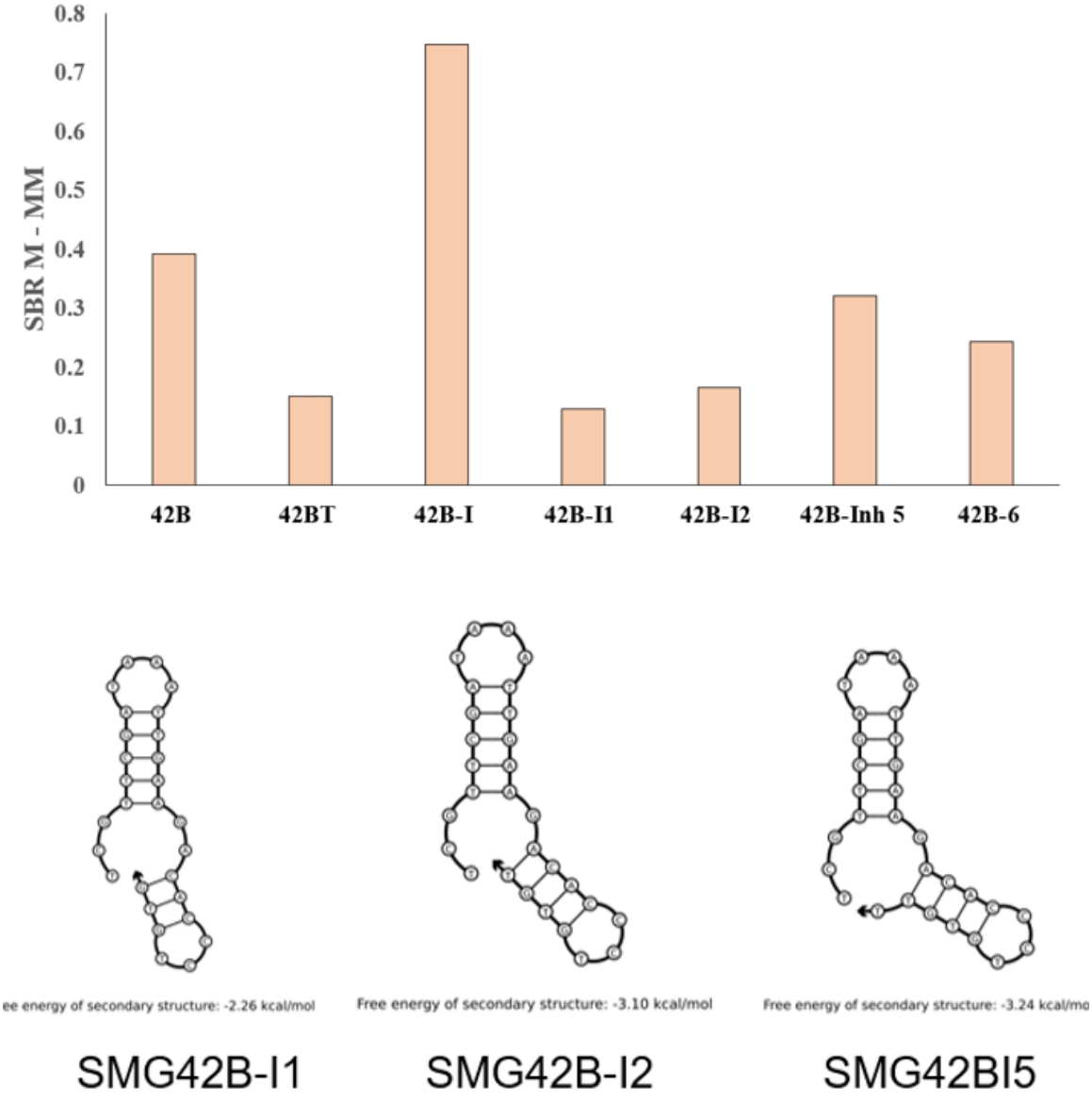


Figure 25. SBR difference obtained with various SMG sensors in discrimination assays (top) and sequence diagrams of the inhibit SNV sensitive arms for SMG-42B-I1 (left), SMG-42B-I2 (middle) and SMG-42BI5 (right). All tested iterations of the SMG sensor displayed poor differentiation between their target and mismatched analyte, with no sensor achieving an SBR difference of 1. The inhibited binding arms are shown bottom, in which intramolecular hybridization results in a lower proportion of mismatched complex. This is reflected by the increasing difference between the SBR of the Matched and Mismatched analyte seen for I1, I2, and Inh-5 which mirrors the relative stability of the folded arms corresponding to -2.26, -3.1 and -3.24 kcal/mol, respectively.

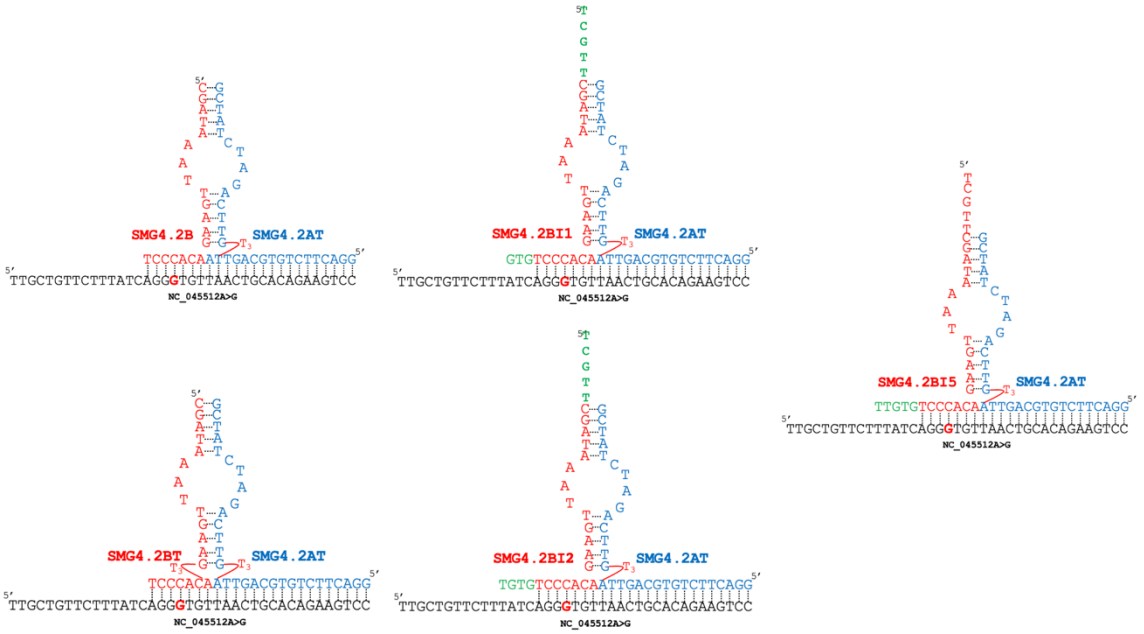


Figure 26. Sequence diagrams of inhibited SMG sensors. The sequence diagram for SMG4.2T (Top Left), SMG42BT (Top Right), SMG-BI5 (Bottom Left) and SMG-BI1 (Bottom Right) are fully depicted. The above images were made utilizing Microsoft PowerPoint.

As shown in Figure 25, these structural modifications did not produce the intend result and selectivity still remained poor with no derivative achieving a difference SBR difference of 1. To determine when this insensitivity was introduced into the construct, early iterations of the sensor (SMG3) were screened for their ability to differentiate the matched and mismatched analytes, the results of this investigation is shown in Figure 28. Every interaction of the SMG sensor displayed the same insensitivity to the presence of the mismatch, indicating a systematic issue in the design. In an effort to improve the selectivity performance of SMG4.2T, the concentration of the SNV sensitive arm (SMG4.2B) was varied from 100 – 400 nM and challenged to discriminate between its target analyte, NC-45, and mismatched analyte, NC-45A. This is a common optimization approach utilized by our group, as the formation of mismatched complex is reduced in tandem with lower strand concentrations. This naturally should increase the sensitivity as the lower

concentrations will shift the equilibrium away from formation of the mismatch complex with the caveat that the overall signal will be diminished. This lower signal would be the product of producing less overall complex in the presence of the matched target, as the SNV arm would be limiting in the binding reaction.

As can be seen in Figure 27, all concentrations of SMG4.2B demonstrated a lack of selectivity towards it target. Upon closer analysis of the results, the magnitude of the fluorescence and background signal was independent of the concentration the SNV sensitive arm (SMG4.2B).

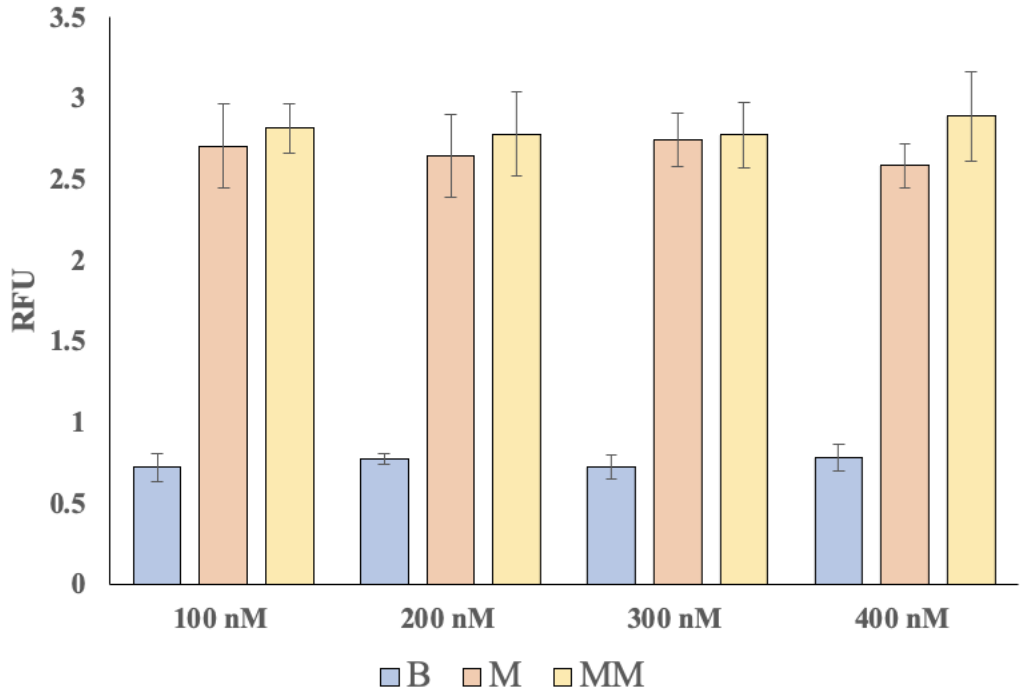


Figure 27. Discrimination assay of SMG4.2T with various concentrations of SMG4.2B. SMG4.2T demonstrates little to no selectivity for its match and mismatched analytes. This lack of selectivity was intractable even when the concentration of the SNV sensitive arm, SMG4.2B, was varied by a 5 fold concentration difference. This result is anomalous as the magnitude of fluorescence did not vary substantially, which is unanticipated if both halves of the aptamer sequence equally contribute to the development of fluorescence. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

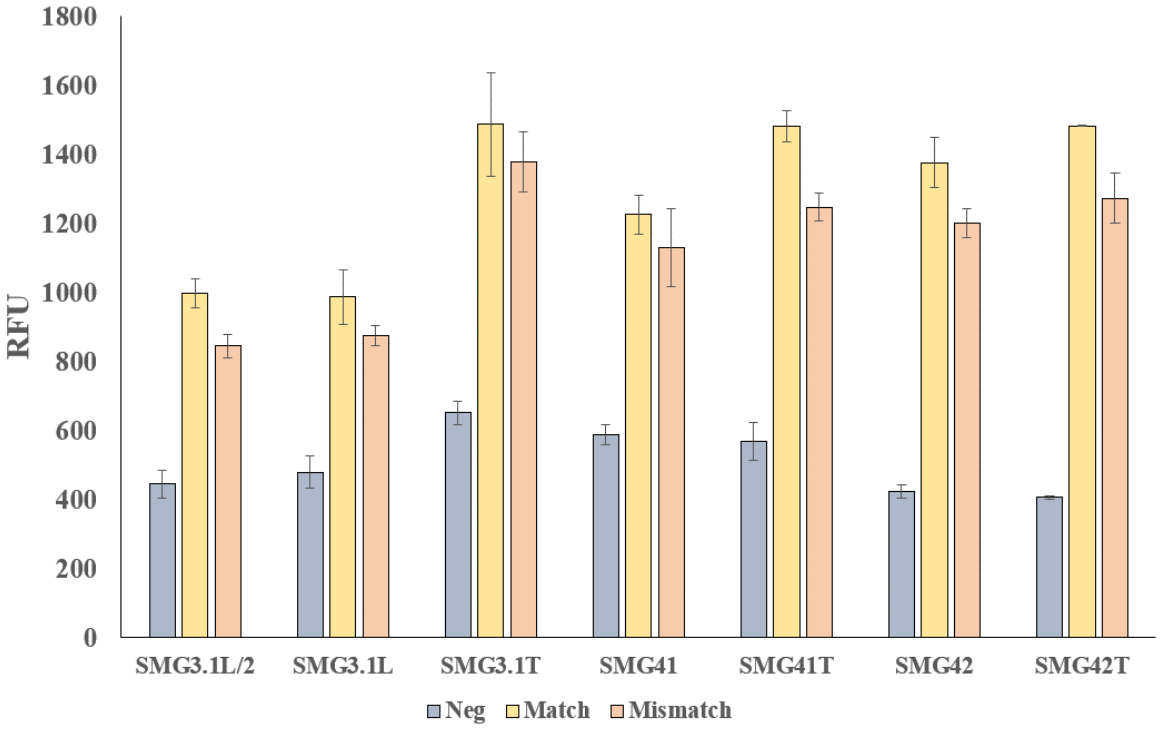


Figure 28. Discrimination Results for both SDA4 and SMG4.2T (top) and various earlier iterations of SMG sensor (bottom). As can be seen in the top left of the image, 250 nM SDA4 exhibited excellent discrimination of single or double mismatch analytes (150 nM) with the double mismatch producing no signal above background. In contrast, SMG4.2T exhibits much greater difficult in distinguishing G>A mutation (top right) a trend that is consistent throughout all generations of SMG sensors (bottom). The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

This was unexpected as the anticipated results would demonstrate improved discrimination (a decrease in the yellow bar) at the cost of fluorescence magnitude. This hypothesis relied on the assumption that both halves of the aptamer sequence encompassing the central region of free bases contribute equally to the development of fluorescence. These unexpected results led to the hypothesis that the contribution of each half of the aptamer sensor (SMG4.2AT and SMG4.2B) to developing fluorescence is not equal and seems to rely more on the presence of the SMG4.2AT. To verify this, the ability of each sensor component to induce fluorescence was checked in all

possible permutations. The results of this line of questioning are shown in Figure 29 and present a different image of the true MG aptamer sequence.

As can be seen Figure 29, the combination of SMG4.2AT and NC-45 is sufficient to generate fluorescence, with only minor differences between the partial complex (SMG4.2AT+NC-45) and the full complex (SMG4.2AT+SMG4.2B+NC-45). Structural analysis of the strong binding arm, shown in Figure 29, detailed that the hexanucleotide fragment (highlighted red) derived from the original MG aptamer is occluded in the absence of analyte. Upon addition of the target analyte, this occlusion is removed by the more complete hybridization between the analyte and analyte binding arm. When combined with the analysis presented in Figure 29, one arrives at that the hexanucleotide sequence, 5'-AGATCT-3', represents the true aptamer core. While convincing, it is possible that a portion of the analyte sequence, NC-45, could contribute nucleotides that allow for the core to fold correctly. A minimized construct, Mini-MG 0.1, was designed which contained solely the hexanucleotide sequence confined into a hairpin, as seen in Figure 29, to eliminate this possible contribution from the analyte sequence.

The sequence diagrams and fluorescence results are shown in Figure 30. The hexanucleotide sequence along with a stabilizing stem is sufficient to induce fluorescence of MG, resulting in a fluorescence increase of 5-fold above background. This small size puts the MG aptamer as one of the smaller functional aptamers and investigations continue into determining structural requirements for MG fluorescence and its exploitation as a bioanalytical tool. In the context of this work, the lack of selectivity across all generations of SMG sensors is attributed to the fact that this short sequence was present on the strong binding arm resulting in little

differentiation between matched or mismatched analyte as the SNV sensitive arm, SMG-B, was not necessary for the production of fluorescence.

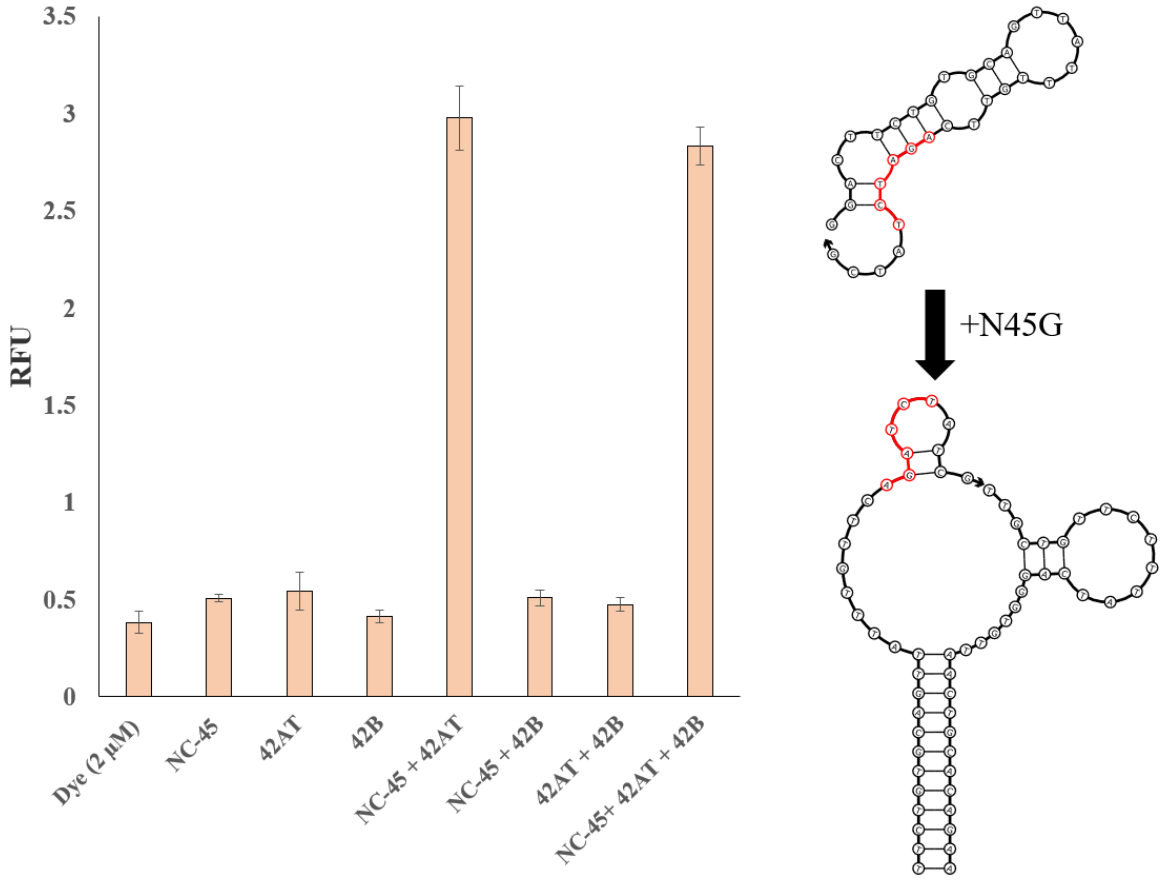


Figure 29. Fluorescence response of SMG4.2T components Assembled Piecewise. The fluorescence results of SMG4.2T (at 1 μM) encompassing the MG in the absence of any sensor strands (Dye), each component singular (NC-45 through 42B) or in combination. As can be seen, no component of the sensor induced fluorescence of MG above the background elevation that occurs via non-specific interactions with DNA. The presence of SMG4.2AT and NC-45 is sufficient to increase the fluorescence of MG by 5-fold with no observable increase when both halves of the SMG4.2T are present. Structural analysis via NUPACK of the SMG4.2AT in the absence (center) and presence of NC-45 (right) are shown on right side of the image. In the absence of NC-45, SMG4.2AT exhibits a tightly folded secondary structure with the core sequence derived from the MG aptamer (highlighted red) occluded by unintended hybridization with the analyte binding arm. In the presence of NC-45, hybridization of the analyte binding arm to the NC-45 liberates this sequence which is sufficient to produce detectable fluorescence. The above graph is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

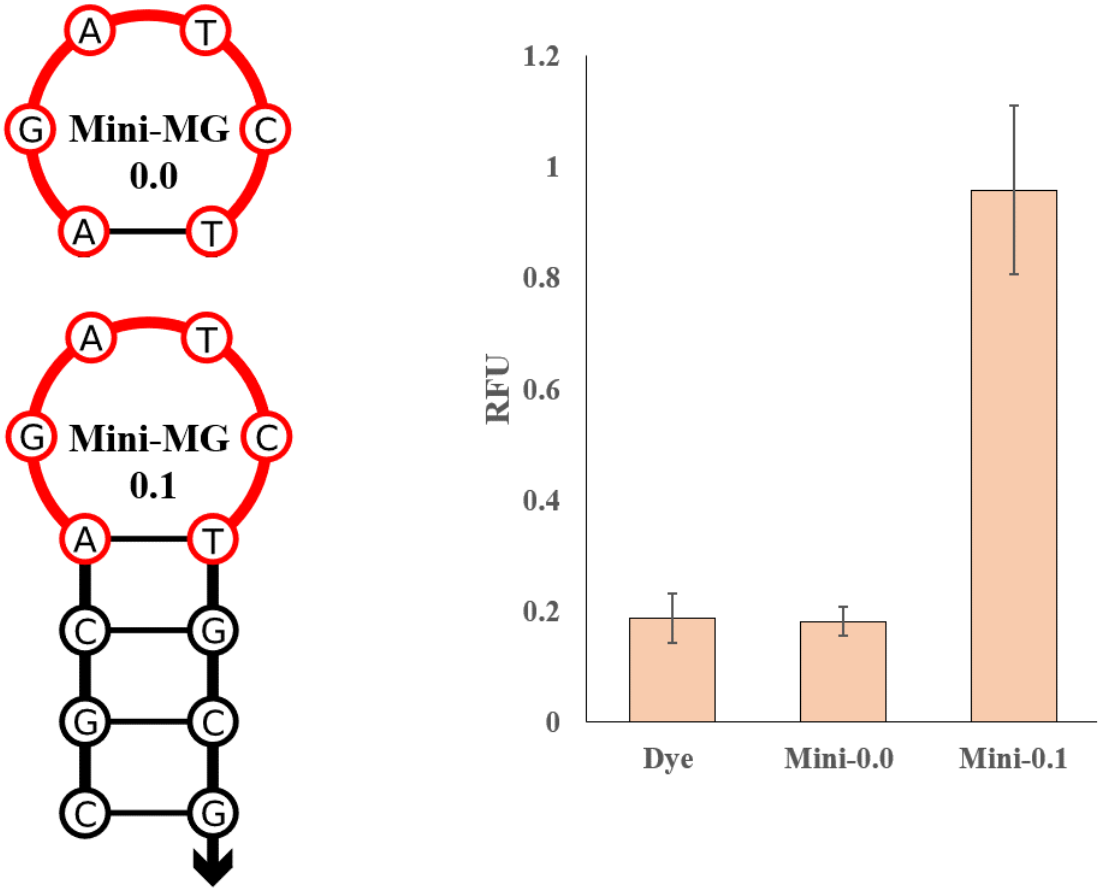


Figure 30. Sequence diagrams of Mini-MG 0.1 construct and its associated fluorescence results. The sequence diagrams of Mini-MG 0.1 is shown left with their corresponding fluorescence results shown right. Mini-MG 0.1 was able to elevate the fluorescence of MG 5-fold. The image was produced via the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

MG Aptamer Switch Sensors

With this new information, a redesign of the MG aptamer-based sensor was required. The small size of the active aptamer sequence along with its efficient inhibition via partial hybridization allowed for the generation of a single stranded switch sensor, which is fluorescent only in the presence of matched analyte, whose design is illustrated in Figure 31. This design utilizes the same

analyte binding arms as SMG4.2T but incorporates an inhibitory fragment that results in tighter binding to the active sequence of the aptamer (color-coded green in the image). Tight binding regardless of presence or absence of a mismatch occurs along the 5' side of the sensor. When the matching analyte (NC-45G) is present, the inhibitory fragment is separated from the active sequence via hybridization of the SNV sensitive arm to its target analyte by inducing a separation in space. This allows for the active sequence to fold, producing detectable fluorescence.

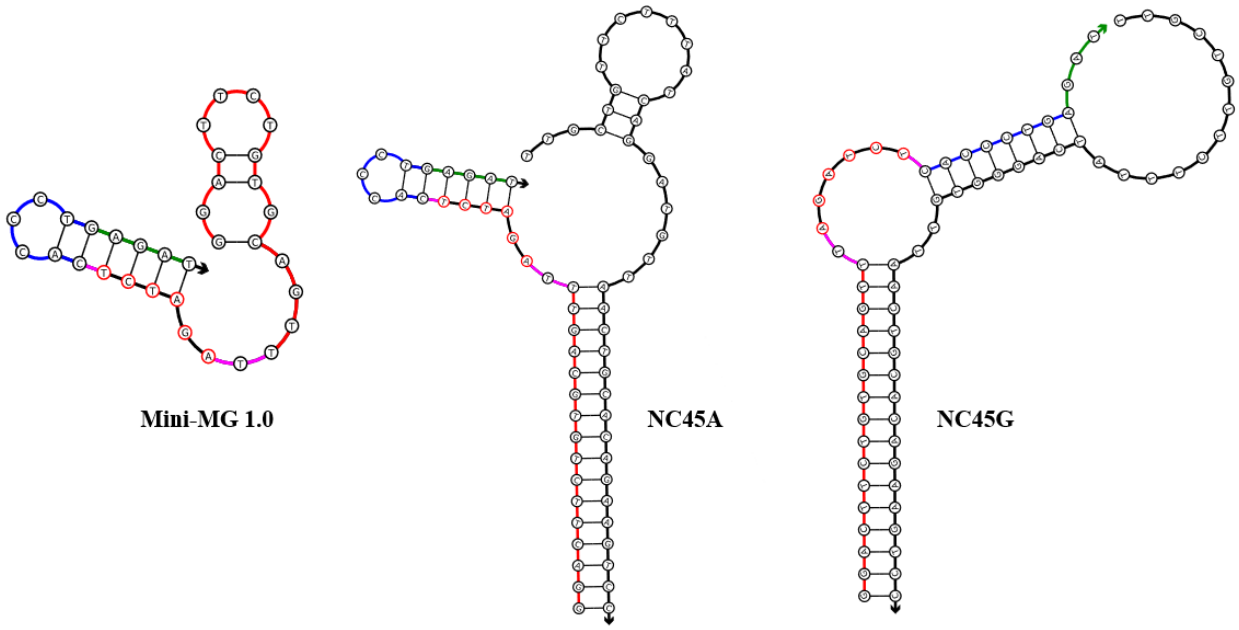


Figure 31. Design of Mini-MG 1.0 in complex with its mismatched analyte (NC-45A) and matched analyte (NC-45). The design of Mini-MG 1.0, show left, consist of a strong analyte binding portion (shown red), the hexanucleotide active sequence (highlighted red nucleotides), the SNV sensitive arm (shown blow) and blocking sequence (shown green). When hybridized with its target analyte, the blocking sequence is removed from the active sequence allowing MG fluorescence. Prospective linker site is highlighted magenta.

Similar to our initial work, a structural optimization focused on incorporation of linkers, hexaethylene glycol or trithymidine linkers, and their impact on the background and fluorescent output of the aptamer sensor. Given the reduced complexity of the construct, these linkers were

incorporated into either at the 5’ or 3’ end of the hexanucleotide sequence. As can be seen in Figure 32, the inclusion of hexaethylene glycol linkers on the 3’ side of the active sequence resulted in a lower background and elevated fluorescence in the presence of matching analyte. This increase could not be recapitulated in full via the use of the dithymidine linkers. These dual effects, lowering background and raising the magnitude of fluorescence in the ON state, could be related to the increase in length of the sensor. Incorporation of hexaethylene glycol linkers add roughly the equivalent of 2 nucleotides in length. This additional length could relieve the tension caused by the tight turn on the inhibitory fragment in the off state and relax some tension formed by the matched complex. In comparison the more rigid thymine linkers still provide the same length increase benefit but with the added rigidity due to linker composition results in a wash of effects, generating construct with the same activity as the original switch sensor. The structurally optimized sensor, Mini-MG 1.2B, exhibited a lower LOD in comparison to the SMG4.2T with an LOD of 3.6 nM (compared to 5.3 nM for SMG4.2T). This improvement likely stems from a higher equilibrium concentration of productive signaling complex. One limitation to the use of the split sensor approach is that the resulting system is more complex than single stranded sensors. This complexity can result in the formation of partial complexes that result in depletion of the analyte effective concentration, which can lower the concentration of fully formed, and fluorescent, complex in comparison to the single stranded approach at any given analyte concentration. The direct comparison between mini-MG 1.0, 1.1L, 1.2L, 1.3 and 1.3TT is shown in Figure 32.

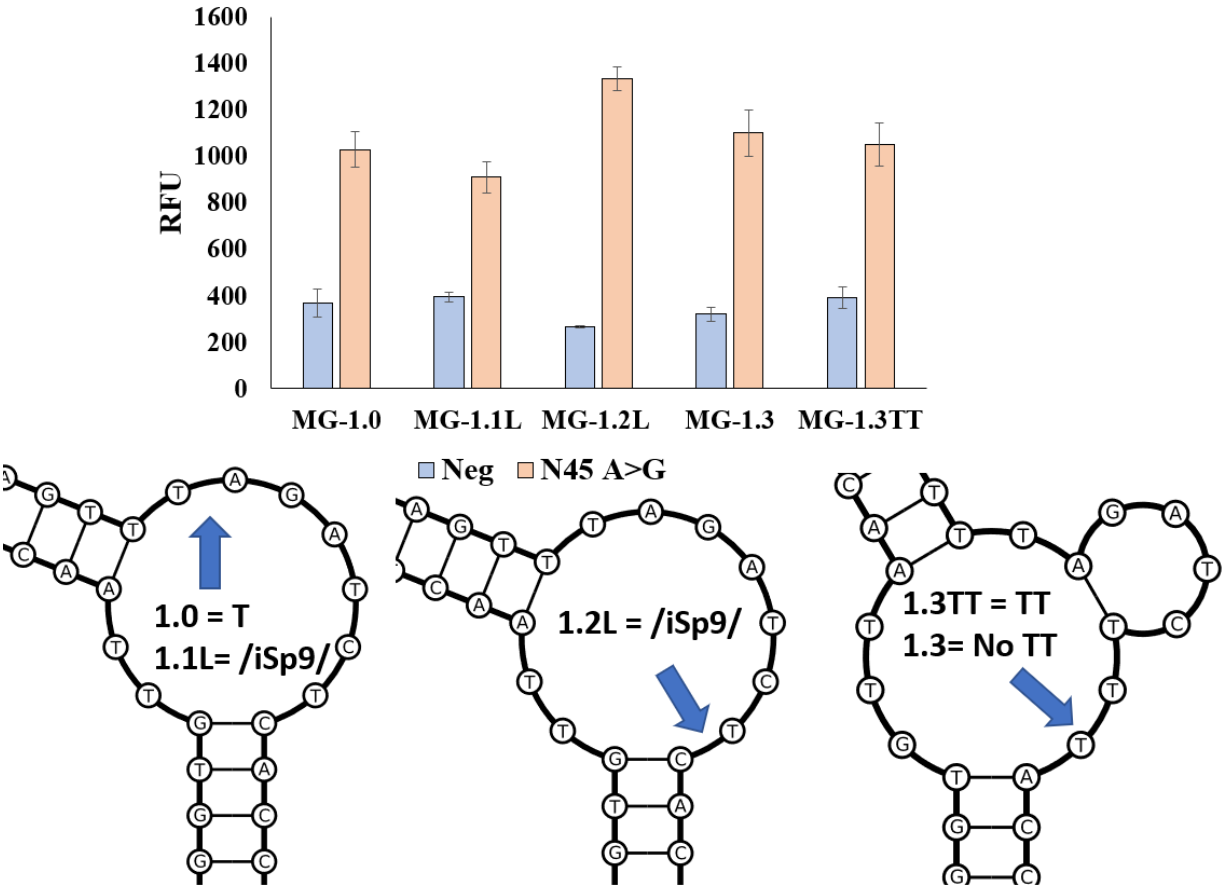


Figure 32. Optimization results of the Mini-MG 1.0-1.3TT (top) with sequence diagrams showing position of the trithymidine modification (T) or hexaethylene glycol linker (/iSp9/). Mini-MG 1.2L demonstrated the most optimal performance, which resulted from a simultaneous lowering of background and elevation of fluorescence in the ON condition. The above graph is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Even with improvements of LOD conferred by this new design, selectivity still remained an issue. To this end bases were removed from the strong analyte binding portion of the sensor, producing a construct that contains 15 and 14 nucleotides as the strong analyte binding arm, a sensor designated Mini-MG 1.4A and 1.4B respectively. This reduction did not significantly impact the LOD of the construct but improved the discrimination between matched and mismatched analytes (NC-45 and NC-45A, respectively). The resulting calibration curve and the results of its discrimination assay are shown in Figure 33.

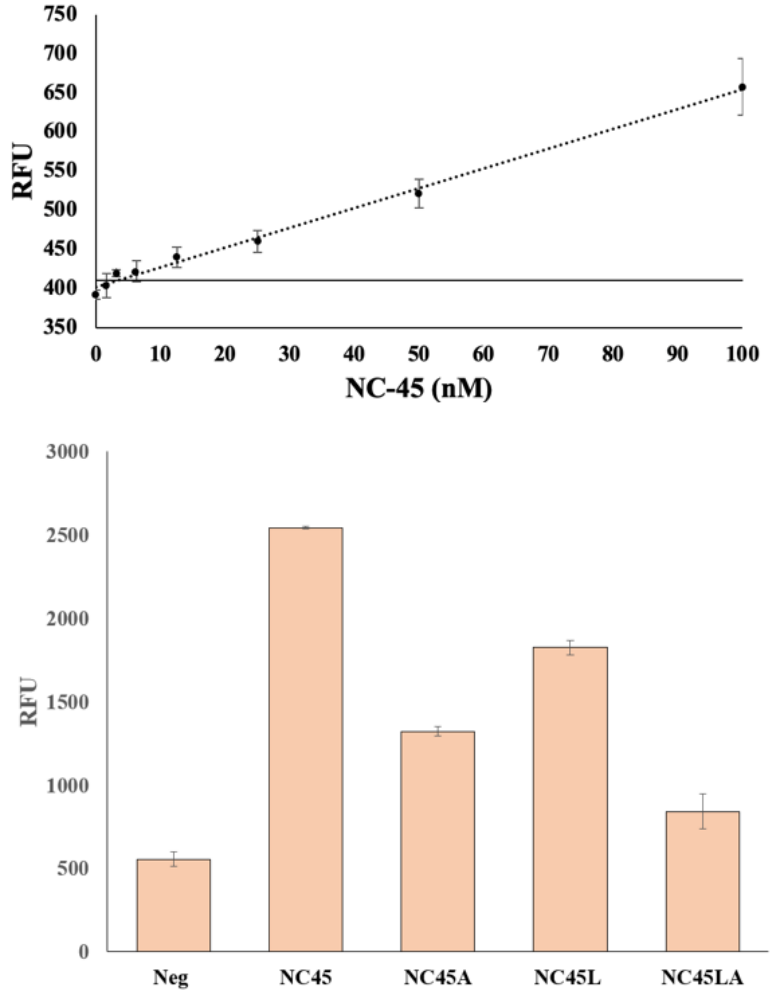


Figure 33. Calibration Curve and Discrimination Assay with Mini-MG 1.4B. Top, the calibration curve of Mini-MG1.3L is shown with a calculated LOD of 3.2 nM. The ability of Mini-MG 1.3 to discriminate between its target analyte NC-45 and mismatched analyte NC-45A and their corresponding long derivatives (NC-45L and NC-45LA) is shown on the bottom. Mini-MG 1.3 was able to differentiate a G>A SNV with roughly a two-fold difference in signal which was further improved upon when the long-mismatched analyte, NC-45A, was used. The threshold signal, defined as the $\bar{x}_{blank} + 3\sigma_{Blank}$, is indicated by the horizontal solid line. The above curve is the average of 3 independent trials and was processed on Microsoft Excel. Fluorescence was recorded at 658 nm with an excitation at 617 nm.

Conclusions:

In this work, a DNA aptamer selected for increasing the intrinsic fluorescence of MG was successfully split and utilized to develop a sensor that achieved limits of detection in the low nM range. A previously reported split dapoxyl aptamer (SDA) was utilized for a new analyte by exchange of its analyte binding arms with very little change in sensitivity and exhibiting the expected high selectivity towards SNV. These aptamers were then combined into a multiplex assay utilizing self-assembling aptamer sensors for the simultaneous detection of two SARS-COVID19 gene fragments. This represents the first demonstration of such an approach and helps develop guidance for further development of such multiplex aptamer pairs. Further development of aptameric multiplex assays should consider the following points in selection of their aptamer partners, minimization of crosstalk between the aptamer partners and selection of compatible aptamers. Minimization of crosstalk between components of the multiplex assay can be achieved by selection of aptamer pairs that bind disparate dyes. The SDA sensor is promiscuous in its binding, resulting in additional optimization steps as it was able to induce fluorescence of MG, while the SMG sensor showed no interaction with Auramine O. Additional consideration should be given to the composition requirements for each utilized aptamer. Ideally these aptamers would be selected under the same environmental conditions, thus ensuring full activity in the multiplex detection assay. Further development of multiplex aptameric assays may include the selection scheme for several dye-binding DNA under the same buffer conditions, preferably PCR and/or isothermal amplification (LAMP, NASBA, RPA) buffers. Such set of DNA aptamers will facility the multiplex assay development and may contribute the widespread use of stable and label-free inexpensive fluorescent assays for nucleic acid analysis.

During the course of developing the multiplex assay we have demonstrated the creation of a new aptamer construct, Mini-MG, which is a minimalist aptamer sequence consisting of just 6 nucleotides. Even at this reduced size and without much optimization, the sequence was able to induce a 5-fold increase in MG fluorescence. This hexanucleotide sequence was then utilized in a switch sensor format, which should better activity than the parent split aptamer. The small size of this aptamer is enticing for the development of future bioanalytical tools and should be further explored to determine the structural requirements necessary to induce high fluorescence of MG.

References:

1. Fernandez-Millan, P.; Autour, A.; Ennifar, E.; Westhof, E.; Ryckelynck, M., Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *Rna* **2017,** *23* (12), 1788-1795.
25. Dolgosheina, E. V.; Jeng, S. C. Y.; Panchapakesan, S. S. S.; Cojocaru, R.; Chen, P. S. K.; Wilson, P. D.; Hawkins, N.; Wiggins, P. A.; Unrau, P. J., RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* **2014,** *9* (10), 2412-2420.
53. Holland, P. M.; Abramson, R. D.; Watson, R.; Gelfand, D. H., Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **1991,** *88* (16), 7276-7280.
54. Tyagi, S.; Kramer, F. R., Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nature Biotechnology* **1996,** *14* (3), 303-308.
55. Kolpashchikov, D. M., An Elegant Biosensor Molecular Beacon Probe: Challenges and Recent Solutions. *Scientifica* **2012,** *2012*, 928783.
56. Smith, A. L.; Kolpashchikov, D. M., Divide and Control: Comparison of Split and Switch Hybridization Sensors. *ChemistrySelect* **2017,** *2* (19), 5427-5431.
57. Zhao, Y.-W.; Wang, H.-X.; Jia, G.-C.; Li, Z., Application of Aptamer-Based Biosensor for Rapid Detection of Pathogenic Escherichia coli. *Sensors* **2018,** *18* (8), 2518.
58. Hamula, C. L. A.; Zhang, H.; Li, F.; Wang, Z.; Chris Le, X.; Li, X.-F., Selection and analytical applications of aptamers binding microbial pathogens. *TrAC Trends in Analytical Chemistry* **2011,** *30* (10), 1587-1597.
59. Chang, Y.-C.; Yang, C.-Y.; Sun, R.-L.; Cheng, Y.-F.; Kao, W.-C.; Yang, P.-C., Rapid single cell detection of Staphylococcus aureus by aptamer-conjugated gold nanoparticles. *Scientific Reports* **2013,** *3* (1), 1863.

60. Berlina, A. N.; Zherdev, A. V.; Pridvorova, S. M.; Gaur, M. S.; Dzantiev, B. B., Rapid Visual Detection of Lead and Mercury via Enhanced Crosslinking Aggregation of Aptamer-Labeled Gold Nanoparticles. *Journal of Nanoscience and Nanotechnology* **2019,** *19* (9), 5489-5495.
61. Bala, R.; Mittal, S.; Sharma, R. K.; Wangoo, N., A supersensitive silver nanoprobe based aptasensor for low cost detection of malathion residues in water and food samples. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2018,** *196*, 268-273.
62. Ellington, A. D.; Szostak, J. W., Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **1992,** *355* (6363), 850-852.
63. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990,** *249* (4968), 505-510.
64. Hong, K. L.; Sooter, L. J., Single-Stranded DNA Aptamers against Pathogens and Toxins: Identification and Biosensing Applications. *BioMed research international* **2015,** *2015*, 419318-419318.
65. Shi, H.; Cui, W.; He, X.; Guo, Q.; Wang, K.; Ye, X.; Tang, J., Whole Cell-SELEX Aptamers for Highly Specific Fluorescence Molecular Imaging of Carcinomas In Vivo. *PLOS ONE* **2013,** *8* (8), e70476.
66. Zhou, H.; Zhang, S., Recent Development of Fluorescent Light-Up RNA Aptamers. *Critical Reviews in Analytical Chemistry* **2021**, 1-18.
67. Wang , H.; Wang, J.; Sun, N.; Cheng , H.; Chen, H.; Pei, R., Selection and Characterization of Malachite Green Aptamers for the Development of Light–up Probes. *ChemistrySelect* **2016,** *1* (8), 1571-1574.
68. Wang, H.; Wang, J.; Xu, L.; Zhang, Y.; Chen, Y.; Chen, H.; Pei, R., Selection and characterization of thioflavin T aptamers for the development of light-up probes. *Analytical Methods* **2016,** *8* (48), 8461-8465.
69. Wang, J.; Zhang, Y.; Wang, H.; Chen, Y.; Xu, L.; Gao, T.; Pei, R., Selection and analysis of DNA aptamers to berberine to develop a label-free light-up fluorescent probe. *New Journal of Chemistry* **2016,** *40* (11), 9768-9773.
70. Pothoulakis, G.; Ceroni, F.; Reeve, B.; Ellis, T., The Spinach RNA Aptamer as a Characterization Tool for Synthetic Biology. *ACS Synthetic Biology* **2014,** *3* (3), 182-187.
71. Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R., Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *Journal of the American Chemical Society* **2014,** *136* (46), 16299-16308.
72. Tsourkas, A.; Behlke, M. A.; Rose, S. D.; Bao, G., Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Research* **2003,** *31* (4), 1319-1330.
73. Luo, Y.; Yu, H.; Alkhamis, O.; Liu, Y.; Lou, X.; Yu, B.; Xiao, Y., Label-Free, Visual Detection of Small Molecules Using Highly Target-Responsive Multimodule Split Aptamer Constructs. *Analytical Chemistry* **2019,** *91* (11), 7199-7207.
74. Kolpashchikov, D. M.; Spelkov, A. A., Binary (Split) Light-up Aptameric Sensors. *Angewandte Chemie International Edition* **2021,** *60* (10), 4988-4999.
75. Kato, T.; Shimada, I.; Kimura, R.; Hyuga, M., Light-up fluorophore–DNA aptamer pair for label-free turn-on aptamer sensors. *Chemical Communications* **2016,** *52* (21), 4041-4044.

76. Plante, J. A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B. A.; Lokugamage, K. G.; Zhang, X.; Muruato, A. E.; Zou, J.; Fontes-Garfias, C. R.; Mirchandani, D.; Scharton, D.; Bilello, J. P.; Ku, Z.; An, Z.; Kalveram, B.; Freiberg, A. N.; Menachery, V. D.; Xie, X.; Plante, K. S.; Weaver, S. C.; Shi, P.-Y., Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2021,** *592* (7852), 116-121.
77. Li, Y.; Xu, S.; Wu, X.; Xu, Q.; Zhao, Y.; Lou, X.; Yang, X., Thioflavin T as a fluorescence light-up probe for both parallel and antiparallel G-quadruplexes of 29-mer thrombin binding aptamer. *Anal Bioanal Chem* **2016,** *408* (28), 8025-8036.
78. Chen, Y.; Wang, J.; Zhang, Y.; Xu, L.; Gao, T.; Wang, B.; Pei, R., Selection and characterization of a DNA aptamer to crystal violet. *Photochemical & Photobiological Sciences* **2018,** *17* (6), 800-806.
79. Connelly, R. P.; Madalozzo, P. F.; Mordeson, J. E.; Pratt, A. D.; Gerasimova, Y. V., Promiscuous dye binding by a light-up aptamer: application for label-free multi-wavelength biosensing. *Chemical Communications* **2021,** *57* (30), 3672-3675.
80. Liu, H.; Gao, Y.; Mathivanan, J.; Shen, F.; Chen, X.; Li, Y.; Shao, Z.; Zhang, Y.; Shao, Q.; Sheng, J.; Gan, J., Structure-guided development of Pb2+-binding DNA aptamers. *Scientific Reports* **2022,** *12* (1), 460.
81. VarnBuhler, B. S.; Moon, J.; Dey, S. K.; Wu, J.; Jaffrey, S. R., Detection of SARS-CoV-2 RNA Using a DNA Aptamer Mimic of Green Fluorescent Protein. *ACS Chemical Biology* **2022,** *17* (4), 840-853.
82. Katilius, E.; Flores, C.; Woodbury, N. W., Exploring the sequence space of a DNA aptamer using microarrays. *Nucleic Acids Research* **2007,** *35* (22), 7626-7635.
83. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A., NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **2011,** *32* (1), 170-173.
84. Kikuchi, N.; Kolpashchikov, D. M., A universal split spinach aptamer (USSA) for nucleic acid analysis and DNA computation. *Chem Commun (Camb)* **2017,** *53* (36), 4977-4980.
85. Kikuchi, N.; Reed, A.; Gerasimova, Y. V.; Kolpashchikov, D. M., Split Dapoxyl Aptamer for Sequence-Selective Analysis of Nucleic Acid Sequence Based Amplification Amplicons. *Analytical Chemistry* **2019,** *91* (4), 2667-2671.
86. Sah, R.; Rodriguez-Morales, A. J.; Jha, R.; Chu, D. K. W.; Gu, H.; Peiris, M.; Bastola, A.; Lal, B. K.; Ojha, H. C.; Rabaan, A. A.; Zambrano, L. I.; Costello, A.; Morita, K.; Pandey, B. D.; Poon, L. L. M., Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Isolated in Nepal. *Microbiology Resource Announcements* **2020,** *9* (11), e00169-20.
87. Gribas, A. V.; Korolev, S. P.; Zatsepin, T. S.; Gottikh, M. B.; Sakharov, I. Y., Structure–activity relationship study for design of highly active covalent peroxidase-mimicking DNAzyme. *RSC Advances* **2015,** *5* (64), 51672-51677.
88. Merkle, T.; Holder, I. T.; Hartig, J. S., The dual aptamer approach: rational design of a high-affinity FAD aptamer. *Organic & Biomolecular Chemistry* **2016,** *14* (2), 447-450.
89. Dolot, R.; Lam, C. H.; Sierant, M.; Zhao, Q.; Liu, F. W.; Nawrot, B.; Egli, M.; Yang, X., Crystal structures of thrombin in complex with chemically modified thrombin DNA

aptamers reveal the origins of enhanced affinity. *Nucleic Acids Res* **2018,** *46* (9), 4819-4830.
90. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S., Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry* **2016,** *4* (38).
91. Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A., Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **2000,** *6* (5), 659-667.
92. Bruno, J. G.; Carrillo, M. P.; Phillips, T.; Hanson, D.; Bohmann, J. A., DNA Aptamer Beacon Assay for C-Telopeptide and Handheld Fluorometer to Monitor Bone Resorption. *Journal of Fluorescence* **2011,** *21* (5), 2021.
93. Jeong, S.; Rhee Paeng, I., Sensitivity and Selectivity on Aptamer-Based Assay: The Determination of Tetracycline Residue in Bovine Milk. *The Scientific World Journal* **2012,** *2012*, 159456.
94. Bruno, J. G., Effects of Various Additives on Cancer Biomarker Aptamer-Magnetic Pull-Down in Human Serum. *Journal of Bionanoscience* **2017,** *11* (1), 45-51.

CHAPTER 3: CHARACTERIZATOIN OF THE MG DNA APTAMER

Introduction:

Aptamer function is the result of the interplay between its core sequence and environmental factors which lead to a distinct fold conformation. Target binding is often mediated by a small portion of the aptamer, which confers the target specificity and affinity, with other portions serving as stabilizing factors for the unique arrangement of these bases. These surrounding bases do directly participate in the binding event, and their function is to stabilize the proper aptamer fold which can be recapitulated by other sequences that maintain similar secondary structure. While slight differences are imparted in helices as a function of the particular sequence composition [95], these differences are slight and are only more apparent with concatemers of distinct sequences. Thus, bases not directly involved in binding can be altered to ensure sequence compatibility or utilized in the design of functional nano-scale devices.

The general approach to aptamer research is selection followed by sequence optimization often relying on empirical evidence derived from the crawling deletion approach or more intuitively from experience working with aptamers. The vast majority of aptamer publications end at sequence minimization, which is a coarse description rather than the fine details needed for engineering. This is an incomplete description of the aptamer and leaves much of the aptamers fine structure unknown, an end state for many reported aptamers. More comprehensive investigations will undergo a series of mutagenesis studies [96], in which contributions to binding and affinity are resolved on a single nucleotide level. These investigations can proceed even deeper, in which the use of modified nucleotides that are structural analogs of the natural bases

can be used to determine functional group contribution to the active site [97]. Thus, structural characterization is a "rabbit hole" of sorts, leading to a series of deeper and deeper investigations. The resulting data from these investigations can often be considered similar to a puzzle in which each aspect builds to the final image of the binding landscape of the aptamer.

Here in we report our initial work focusing on characterizing some of the sequence and structural requirements for a malachite green (MG) DNA aptamer. This aptamer belongs to a sub-class of aptamers that are capable of enabling fluorescence of normally non-fluorescent or weakly fluorescent compounds, known as Light up aptamers. These aptamers are attractive for biosensor development, as the ability to fluoresce obviates the need for costly fluorescent modifications. Through this work a more comprehensive description of the light up MG DNA aptamer will be completed, which should facilitate further downstream applications for the development of this unique sequence into functional devices.

Results and Discussion:

Determination of MG Aptamer Structural Requirements:

In this work we provide a more detailed structural analysis of the MG DNA aptamer. Particular focus will be given on determining the influence of structural aspects such as the gap size on the trans strand, the core sequence identities and the contributions of the adjacent base paired nucleotides. While these investigations have not proceeded to the level of individual functional group contributions, we do demonstrate a framework which can serve as a strong

foundation for further characterization and to facilitate modeling of this interestingly small aptamer.

Our initial work with this aptamer, as outlined in chapter 1, resulted in the discovery that the fluorescence enhancement by the MG DNA aptamer can be attributed to a small hexanucleotide sequence of 5'-AGATCT-3'. This was validated by engineering a construct which solely contained this nucleotide sequence constrained by appending self-complementarity sequences to the 5' and 3' end, forming a hairpin like construct similar to that of traditional molecular beacon technology deemed Mini-MG 0.1. A series of derivatives (Mini-MG 0.0-0.6, shown in Figure 34) of this design were constructed to answer a series of distinct questions. The first, and most pressing question, is determining whether the sequence requires the presence of some sort of structure to induce fluorescence. Most aptamers require the presence of some sort of stabilizing structure to achieve their proper fold. If the free sequence is incapable of producing elevated fluorescence, what structural features are necessary for activity to be observed? This is not immediately apparent as our previous work with the aptamer in a split-aptamer sensor format, SMG42AT in complex with its target analyte is competent to enhance MG fluorescence. Examination of the predicted secondary structure, even at high salt concentrations which favor the formation of weak structures, reveal a mostly linear conformation devoid of secondary structure in the region around the active sequence. To this end, our derivatives can be divided into several categories. The first is a series of hairpins with variable stability, in the order of Mini-MG-0.2 > 0.1 > 0.3. The least stable derivative, Mini-MG 0.3, is designed to contain a mismatch (T-T) immediately below the active sequence, which should result in disruption of the possible terminal A-T base pair that can be formed by the active sequence. These were compared

to the free sequence, Mini-MG 0.0, which is the hexanucleotide sequence with no supporting base pairs , excluding the formation of the possible terminal A-T base pair or making it much more transient in nature. The results of these fluorescence assays are show in Figure 35.

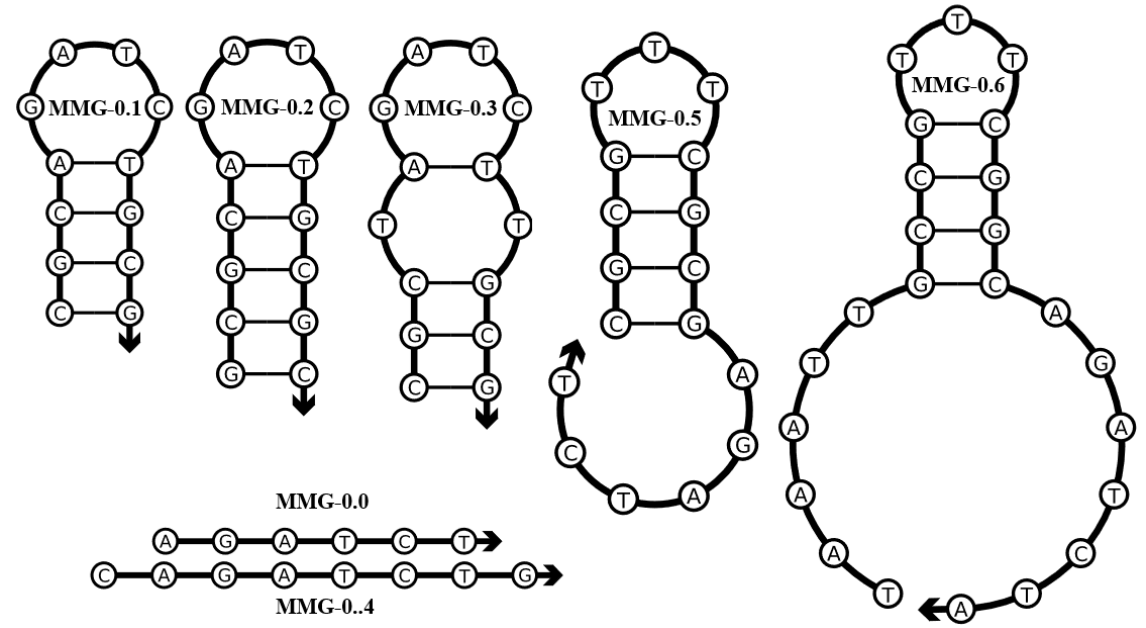


Figure 34. Sequence diagram of all Mini-MG constructs 0.0-0.6. The above constructs were designed to validate the influence of stem stability on aptamer function (Mini-MG 0.1-0.3), the capability of the free sequence to induce MG fluorescence (Mini-MG 0.0 and 0.4) or the requirement for adjacent secondary structure (Mini-MG 0.5 and 0.6). The above images were produced in Inkscape utilizing SVGs produced by NUPACK.

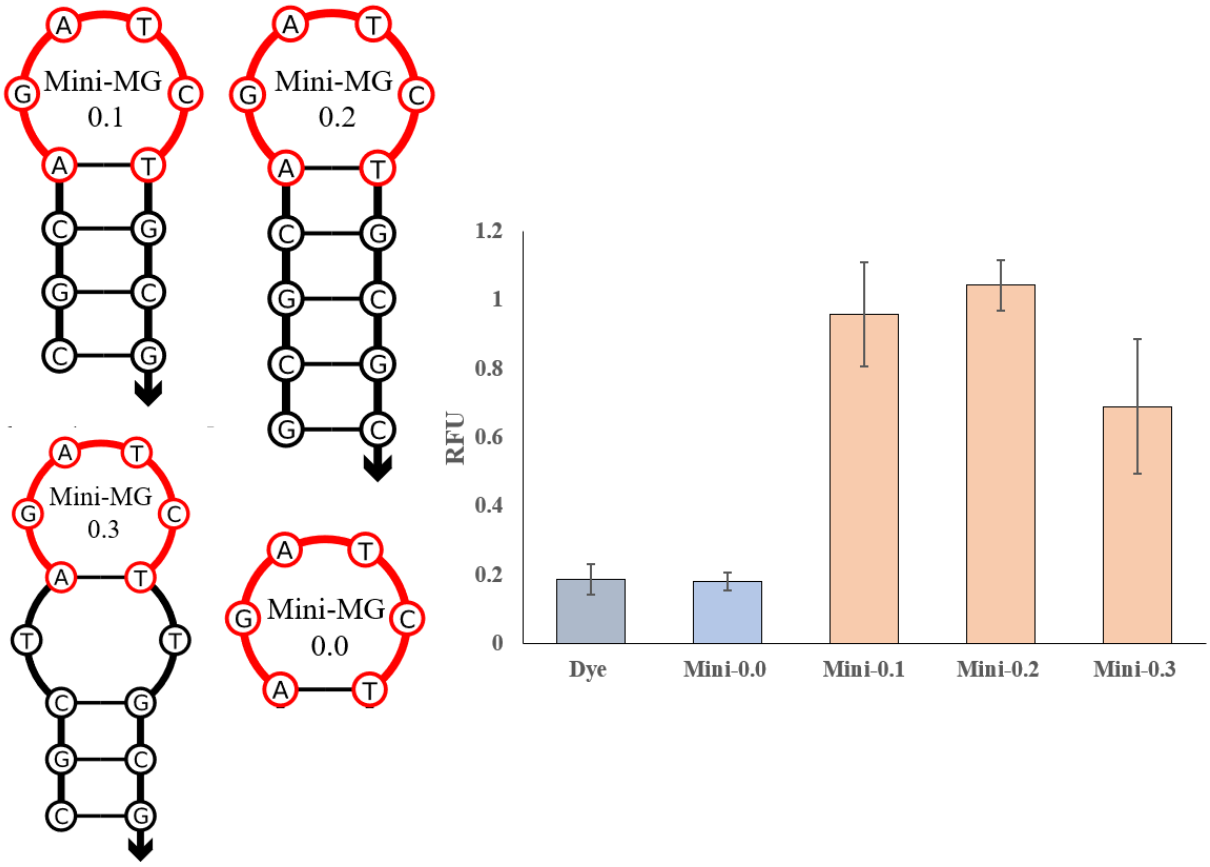


Figure 35. Fluorescence results of Mini-MG 0.0-0.3. The sequence diagrams of Mini-MG 0.0-0.03 are depicted on the left and their associated fluorescence results are shown right. Mini-MG 0.1-0.3 were able to induce fluorescence of MG between 4-5 fold above the background levels. The magnitude of this enhancement tracks with stem stability as the fluorescence arranged from highest to lowest is Mini-MG 0.2 > 0.1 > 0.3 which mirrors stem stability 0.2 > 0.1 > 0.3.

In this series of experiments the fluorescence results track well with the stability of the stem, with Mini-MG 0.2's GC rich stem of 5 base pairs producing the highest enhancement of MG fluorescence. In contrast Mini-MG 0.0, a sequence with an absence of any secondary structure, failed to produce any detectable increase in fluorescence above the background signal of MG. Based on these results the core sequence would seem to require the presence of some form of structure to stabilize the fold in the proper conformation. However, the ability of these structures to increase fluorescence utilizing a structure that is quite different than the parent

aptamer or its split derivative. The parent aptamer and split derivative contain this core sequence flanked by two stems, constituting an internal and asymmetric bulge. Given the significant difference between the structures, the active conformation may not rely on formation of a stabilizing structure but rather may result from stacking interactions in adjacent bases.

When these results are contextualized with those from SMG42AT, which has no predicted secondary structure, there is a possibility that folding does not require direct constraint by a secondary structure. The aptamer's function may not require both ends of the sequence to be restrain but rather rely on the presence of some stacking interactions in adjacent bases to obtain its fold. To this end a series of derivatives, Mini-MG 0.4-0.6, were designed and utilized in the same fluorescence assay as those for 0.0-0.3. Mini-MG 0.4 is the free sequence but incorporates an additional G-C base pair at the terminal ends which is not predicted to be stable at room temperature but could transiently form or participate in stacking interactions and is more stable than just the free sequence. This transient association may be stabilized by the free energy released as the core sequence folds, providing an overall active complex. Mini-MG 0.5 and 0.6, in contrast, are designed to contain a larger stem element that places the active sequence on the terminal end of the stem, rather than be contained within the turn of the hairpin. Mini-MG 0.6 uses the same base construction as Mini-MG 0.5 but introduces an additional extension on the opposing side of the stem which contains the trans sequence of the parent aptamer, AAT. This design would hopefully demonstrate whether the trans aspect of the aptamer facilitates the binding of MG by participating in the fold or increases the torsional restrictions, the presumed mechanism responsible for fluorescence, of MG. The results from these assays and accompanying sequence diagrams are show in Figure 36.

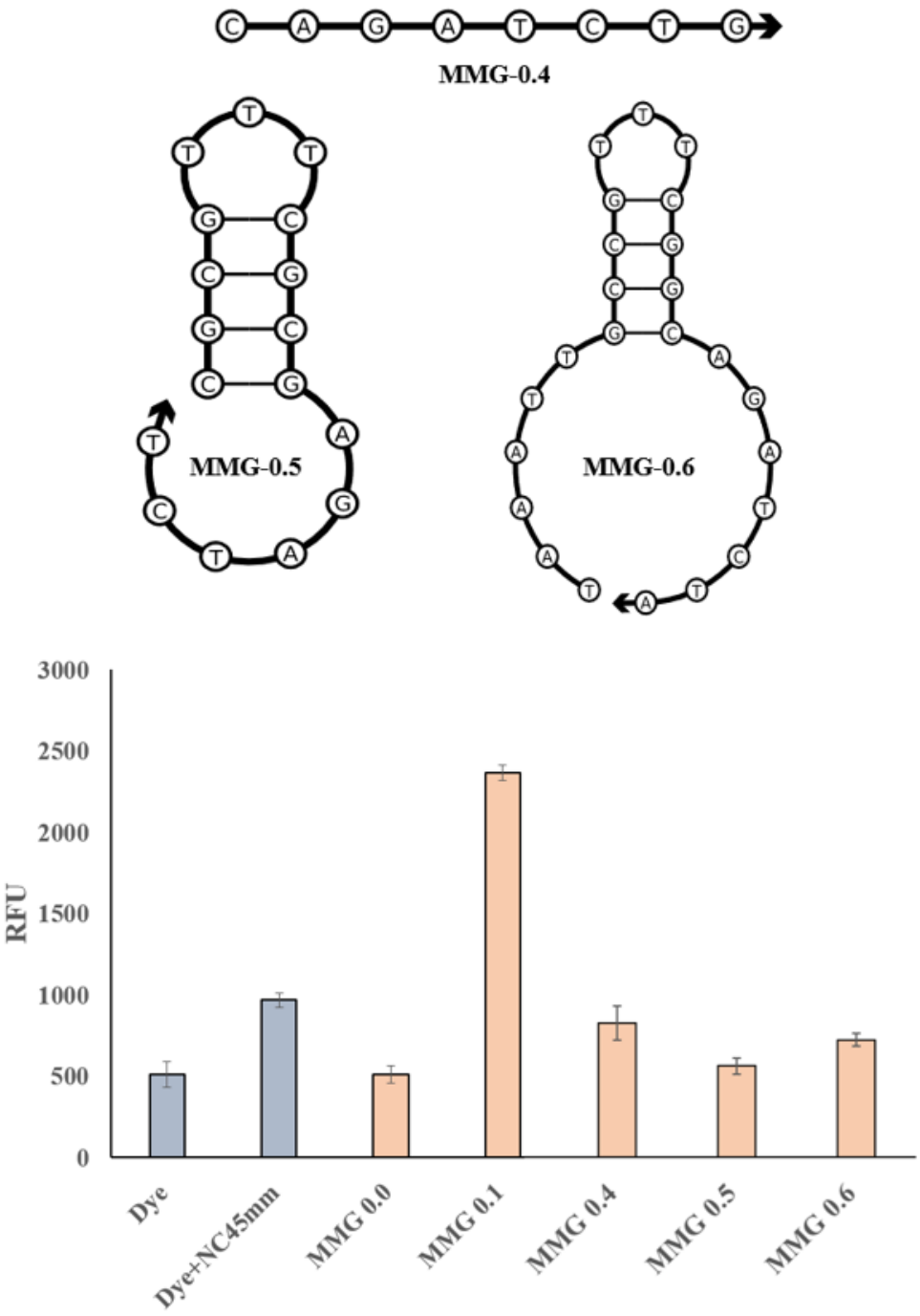


Figure 36. Sequence diagrams of Mini-MG 0.4-0.6 (top) and associated fluorescence results (bottom). Mini-MG 0.4-0.6 were incapable of elevating MG fluorescence above the background generated by non-specific interaction with DNA (shown in Dye+NC45mm). Interpretation of this would support the need for a secondary structure stabilization to occur on both sides of the hexanucleotide active sequence.

In figure 36, none of these derivatives increased the fluorescence of MG above the base line intrinsic fluorescence or the small elevation associated with non-specific interactions with DNA. This result and the results of SMG42AT are at odds with one another and our future strategy to address this discrepancy should be to remove portions of SMG42AT from the non-base paired end, to isolate the minimal sequence that retains the ability to enhance the fluorescence of MG and use it to develop a more comprehensive image of the minimal structural requirements for MG fluorescence enhancement.

Nucleic acid structures are sensitive to both cation content and temperature of their environment [98]. Mono-valent cations, such as Na and K, form a sphere around the anionic backbone due to long range electrostatic interactions [99]. These inner sphere ions neutralize the high density negative charge which facilitates the formation of secondary structure. Some divalent cations, such as Mg, possess a coordination shell which is apt to bridging the distance between adjacent oxygens on the opposing strands of the helix [100]. The resulting coordination complex is very stabilizing even in comparison to Na, with 50 mM Mg providing the same stabilization as 1 M NaCl [101]. The original aptamer was selected utilizing a buffer devoid of any stabilizing ions, which is atypical for work within this field. Utilizing high concentration of ions to stabilize weak secondary structure in the context of the MG DNA aptamer is unsuitable, as the exact mechanism of binding is unknown and such drastic changes could result in low intrinsic activity which would confound the results of determining which features are optimal or required. Thus, stabilization transient structure will have to proceed through the use of low temperatures, which stabilizes secondary structure by depressing transient melting behavior between base pairs and favors the enthalpy-driven hybridization of nucleic acids [102]. This stabilization may be able

to rescue some of the Mini-MG derivatives making them competent and enhance the activity of functional derivatives, which could open up use of the MG DNA aptamer in low temperature applications. Thus the fluorescence enhancement as a function of temperature was explored for all derivatives and is shown in Figure 37.

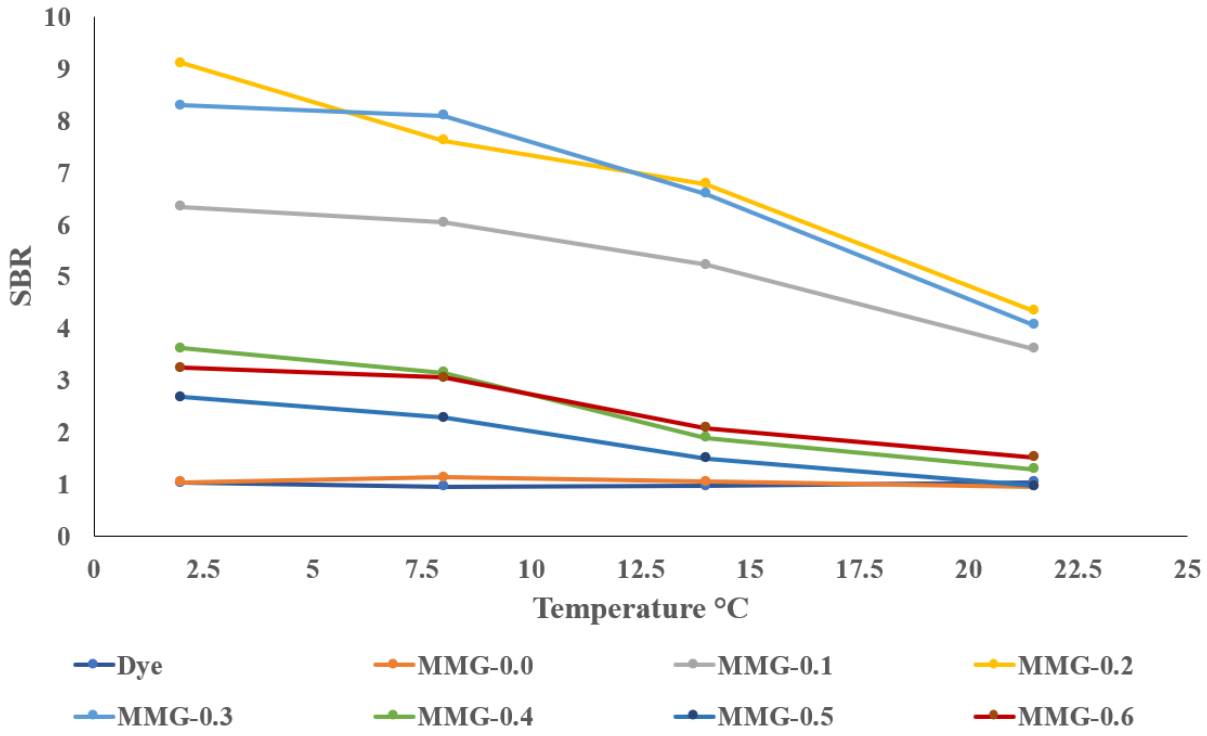


Figure 37. Fluorescence enhancement of MG by Mini-MG 0.0-0.6 as a function of temperature. Mini-MG 0.0 was unable to produce elevated fluorescence even at depressed temperatures, while all other constructs abilities were enhanced with a gradual decrease back to room temperature. This would imply that the fold of the MG DNA aptamer is relatively weak and in the absence of stabilizing effects is less efficient at induce MG fluorescence.

Here a general enhancement was detected at lower temperatures with a gradual decrease until the end point of the experiment, 22 °C. The decreased temperature did result an increase in MG enhancement even for complexes that were unproductive at room temperature, such as Mini-MG 0.4-0.6. This rescue effect would imply that the aptamer's fold is relatively weak or transient

and requires some stabilization, either in the form of facilitating secondary structures or by lower temperatures. The interpretation of this experiment should be done with caution however, as ascribing the enhancement all to the aptamer's fold neglects the potential contribution of depressed temperatures effect on MG. This enhancement could plausibly be attributed to MG becoming a more efficient fluorophore at this temperature range, or by increasing the non-specific interactions between the MG and DNA. This is particularly relevant if this non-specific interaction relies on generic structural features of nucleic acids, such as the anionic backbone, which would demonstrate a length dependent effect as each added base results an increase in these generic features. Both of these scenarios would be obscured by these results, even though the latter is less likely as Mini-0.0 fluorescence never increased above the background of MG, length of 6 nucleotides, while Mini-MG 0.4 was much more active with only the addition of two more bases.

Geometric Contributions of MG DNA Aptamer Function:

At this juncture, declaring that the MG DNA aptamer requires some structure to stabilize the fold, even if the fine intricacies of such are not obvious, is reasonable conjecture. The next series of experiments sought to determine if the aptamer's fold is sensitive to the geometry of the resulting complex between the active sequence strand and its trans compliment. This juncture between the two sequences will assume a specified angle which is the result of the hybridization between the two strands and the existence of internal unpaired nucleotides [103]. Normally this change would be accounted for by the difference in lengths between the two asymmetric strands, with the longer strand imposing a more acute angle on the junior partner, or by the number of

free bases at the juncture in which the smaller juncture is compressed. Here since the number of base pairs in the stems are equivalent, any changes in the angle should be the result of the free bases contained within the juncture. As the number of nucleotides is reduced on the trans strand, the resulting complex will progressively be reduced, with the impact being greatest at 0 nucleotides and progressively relieving with each additional base. The structure will be linear when this gap possesses 4 nucleotides, which would be equivalent, after which the change in angle will proceed in the opposing direction. To this end, a series of constructs were designed to address this line of questioning, whose sequence diagrams are shown in Figure 38. Of special note is that these constructs constitute a complete exchange of all auxiliary bases, those not in the proposed active site, to produce a more stable construct by utilizing sequences more rich in G-C base pairing. This was done to ensure that the formed complex remained stable throughout testing, as the less GC rich sequence of the split parent aptamer could result in a proportion of the strands being melted and the asymmetry in the sequence composition could result in localized partial melting. The retention of activity after this replacement is strong evidence that

these bases contribute little to the overall binding of MG.

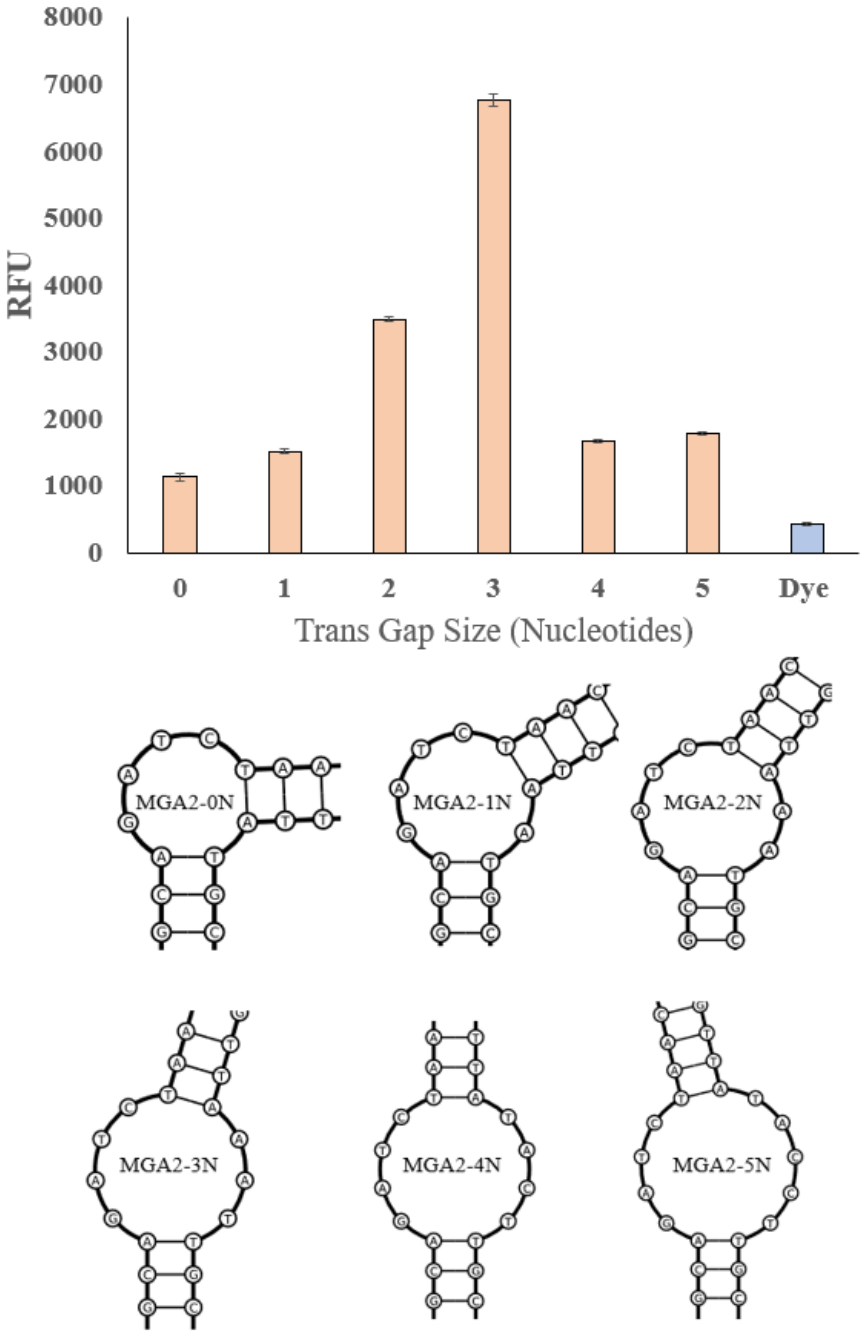


Figure 38. Investigations into the gap size on the trans strand with the sequence diagrams showing the changing angle as a function of gap size (top) and associated fluorescence results (bottom). The optimal gap size was 3 nucleotides, which is found in the parent aptamer structure. Increasing the gap from 0 to 3 results in a rapid rise of observed fluorescence, followed by a stable depressed value for gap sizes of 4 and 5.

The results of this gap investigation are shown in Figure 38. Each of the construct was able to enhance the fluorescence of MG above that of the intrinsic value of the dye, which is shown as the blue bar. There is significant increase up to the 3-nucleotide gap, which corresponds to that of the original aptamer, with a decrease and plateau in signal at 4 and 5 nucleotides. This would seem to indicate that there is some sensitivity to the angle formed in the complex between the two strands, with 3-nucleotides being optimal. This interpretation, however, disregards the influence of the bases in the gap and their contribution to MG binding or participation in torsional restriction. This caveat is especially applicable to the 0,1 and 2 nucleotide gaps, as each subsequent base could provide additional contact points for MG restriction. The plateau at 4 and 5-nucleotides is also confounded by the alteration to the gap sequence, from 5'-AAT-3' to 5'-TACT-3' and 5'-TACCT-3'for the original sequence, 4-nucleotide and 5-nucleotide gap respectively. This sequence change was necessary as the addition of bases resulted in internal hybridization, which as established in our previous work, results in efficient inhibition of the active sequence. A more thorough approach to determining if the geometry has any appreciable effect on the generation of MG fluorescence would require the use of abasic, those lacking any nucleobase, sites. With these strands, the changes in geometry would be retained but eliminate the influence of the base contribution to fluorescence development.

Determining Sequence Requirements of the Trans 3 Nucleotide Gap:

The increase in fluorescence between 0-3 nucleotides, disregarding the geometric contribution, can be interpreted as each base's individual contribution to the overall fluorescence. As the raw RFU values for 0, 1 and 2 gap constructs summate to 6125 RFU. This value is comparable to that of the 3-nucleotide gap construct which is found in the parent aptamer, with an RFU of 6763. To determine if the trans strand's sequence content is relevant and to more clearly describe the contributions of each of base within the sequence of 5'-AAT-3', a series of derivatives were made in which all possible combinations, as listed in Table 2, were designed and accompanying representative sequence demonstrating no internal hybridization are show in Figure 39.

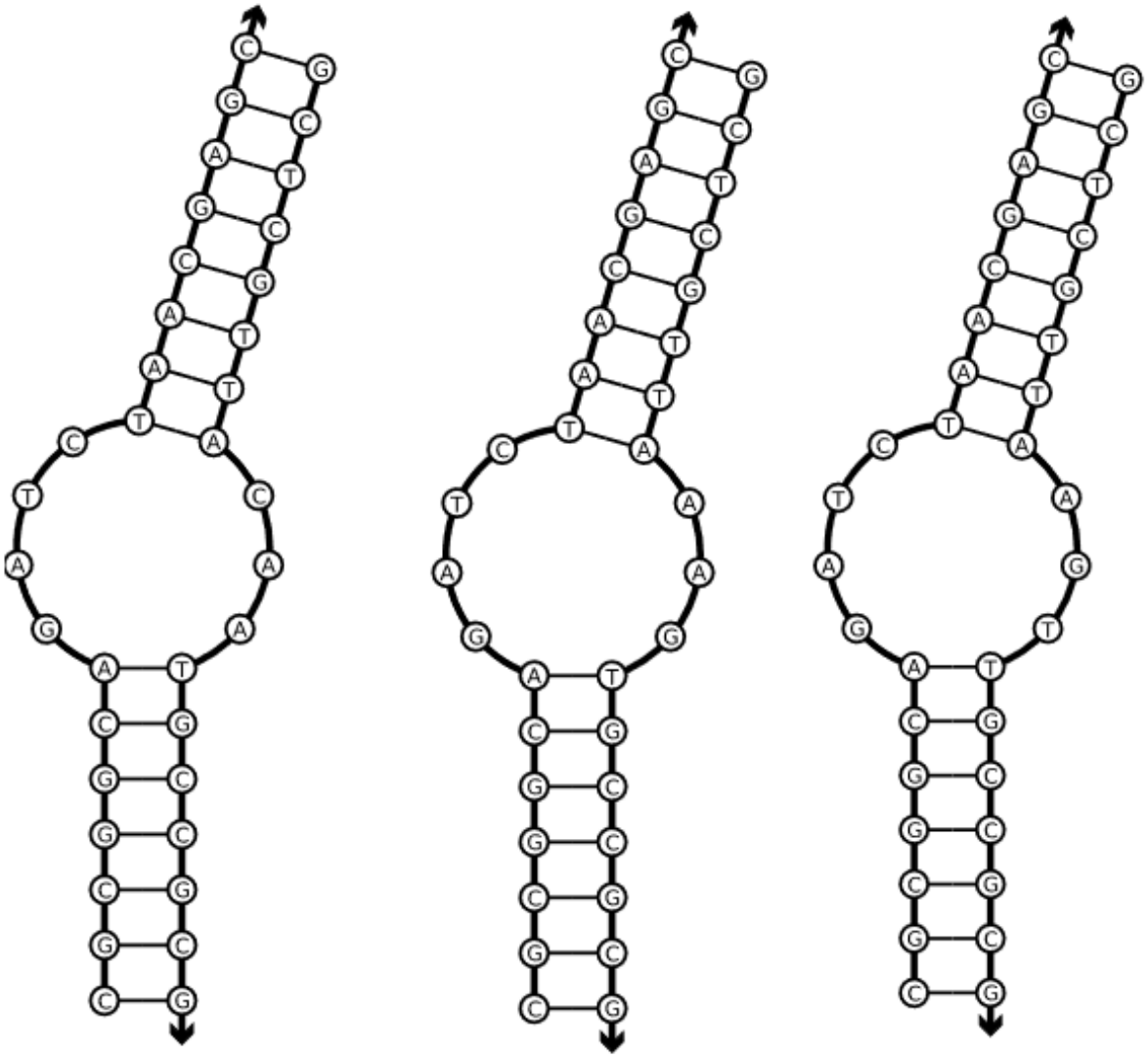


Figure 39. Representative sequence diagrams of the differing 3 nucleotide junctions for determining the influence of sequence on MG fluorescence. These diagrams depicted the 3 nucleotide gap being mutated from AAT to CAA (left), AAG (middle) and AGT (right). No unintended hybridization occurs between the two strands, allowing for the sequence influence to be determined without the confounding variable of partial inactivation by hybridization of the active sequence.

Table 1. Potential 3 nucleotide junctions that do not produce any hybridization with the hexanucleotide active sequence.

Potential 3 Nucleotide Junctions
ACA, ACT, ACC, AGA, AGT, AGG, AAA, AAT, AAG, CCA CCG, CGT, CGG, CAA, CAT, CAG, TGA, TGT, TGG

These 18 sequences are all possible combinations nucleotides that are permissible to be utilized with this hexanucleotide core sequence. Permissible here is defined as sequences which allow retention of the original structure by not forming any base pairing with the opposite strand. This is particularly important as seen in chapter 1 with SMG42AT, as partial hybridization occludes the sequence and results in complete inactivation. Thus sequences that would naturally hybridized with the opposite sequence were discarded, as the difference in signal would be difficult to ascribe to whether it is a sequence effect or a consequence of hybridization-based reduction. The available sequences, though, do allow significant coverage in most positions, in that each of the four bases is represented in each position and that each combination of purine and pyrimidine arrangements are represented. These were analyzed in much the same way as all other of our constructs and the overall results are shown in figure 40.

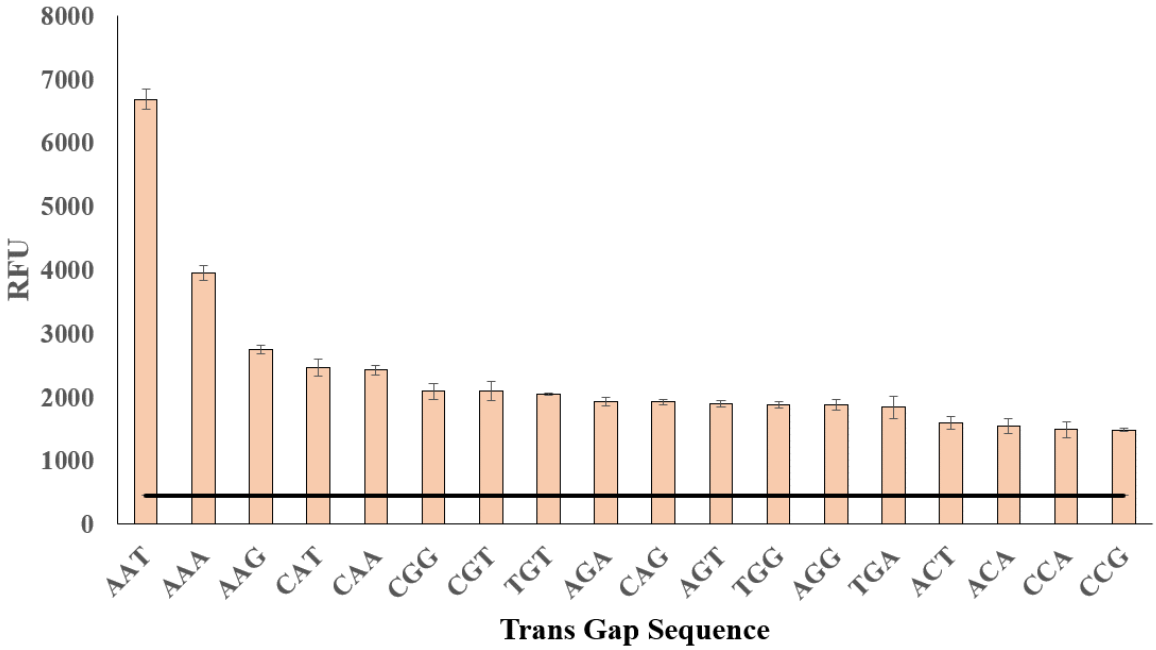


Figure 40. Sequence effects of the 3 nucleotide gap. The wild type sequence, AAT, is the most effective in inducing MG fluorescence with the closely related sequences AAA and AAG being the next most effective. A central adenine (A) is associated with the highest levels of relative activity, while a centrally located cytosine is associated with much lower relative activity levels.

As can be seen in Figure 40, the most optimal sequence is that found in the original aptamer, AAT, with next most active sequences being AAA and AAG with relatively intensities of 0.6 and 0.4 respectively with the rest of sequences having a relative activity between 0.36 (CAT) down to 0.22 for the least active construct (CCG). Even so all the tested sequences still elevated MG fluorescence above baseline, with all constructs obtaining an SBR > 3. A general observation from these results is that the central position of the 3-nucleotide gap is most active when it contains an adenine and least active when it contains cytosine.

The original sequence, consisting of 5'-AAT-3', has a purine-purine-pyrimidine arrangement. The importance of the central position being occupied by adenine is shown by the

derivatives of AAA and AAG, which are the next two most active sequences, and by AGT, in which the mutation of A > G results in a significant decrease in relative activity by roughly 70%. While guanine and adenine are both purines, the functionalities of the ring systems differ. Guanine possesses a higher capacity for hydrogen bonding than adenine, with 7 vs 5 donor and acceptor sites respectively. The significant difference in activity between these two sequences would seem to indicate that the central adenine's contribution is specific for the base rather than incorporation of the generic ring attributes, such as pi-stacking capability, or the nitrogen's contained within the ring system. Rather the presence of the exocyclic amine in adenine maybe necessary for MG fluorescence. Similarly, the mutation from T > G is a cross class mutation but retains the presence of the ketone functionality and the cyclic nitrogen's hydrogen bond donating capacity. The mutation of A > C in the first position also reduces activity, roughly by 60%, which would provide evidence that the exocyclic amine is not utilized heavily in MG binding as it is retained by both nucleotides for this position.

These results indicate that while the gap sequence has a significant impact on the development of fluorescence, there role can be considered auxiliary as the hexanucleotide sequence is still capable of enhancing the fluorescence of MG with a mutated gap sequence or its absence. This allows some conjecture in which the binding pocket for MG is primarily composed by the hexanucleotide sequence. The role of the trans gap sequence could be stabilization of the pocket, by contributing to a hydrogen bonding network that relies on the cyclic nitrogen of adenine in the first position, and the exocyclic amine in the second adenine position. Alternatively, this sequence could serve to make the pocket more restrictive by contributions of specific functional groups which may interact with portions of MG, providing additional

restriction by via H bonding contacts. The exact functional contributions for these trans nucleotides could be determined through the use of structurally related nucleotides, such as uracil for thymine which would elucidate the function of the C5 methyl group. Similarly, adenine structural contribution could be determined through the use of purine base, which would elucidate the contribution of the exocyclic amine. Since these groups do not seem to be solely responsible for the development of fluorescence, their contribution might not be the pocket itself but rather as additional contact sites to further restrict malachite green in the binding pocket. To determine the influence of the second scenario, the K_d of these respective derivatives can be compared to that of the parent aptamer. This is plausible as the K_d of these derivatives can be determined, even if the overall efficiency of fluorescence generation is lowered. If the affinity of the aptamer is not seriously impaired by the usage of other bases, this would tend support the second scenario in which their contributions to binding of MG is not in the formation of the pocket but rather enhancing the torsional restriction of MG.

Determination of Terminal Base Pair Requirements:

All of these designs have focused almost exclusively on the 3-nucleotide gap on the trans strand. This is mostly due to the relatively small size, 6 nucleotides, of the active sequence on the other. Of these 6 nucleotides, 4 exist in an unpaired state, GATC, while the terminal ends are in a paired condition A:T and T:A for the 5’ and 3’ ends respectively. As base paired nucleotides contribution to binding is lessened by the occupancy of their hydrogen bonding capacity, these bases may not be a necessary requirement for the aptamer’s function. This is exceptionally relevant from a statistical standpoint for the analysis of genomic targets, as the probability of a

sequence containing AAT, the most active sequence, is 16 times higher than the corresponding sequence of TAATA. The next series of experiments sought to determine the influence of these two terminal pairs. In particular whether the sequence identity of the these paired bases matters.

To address this question, a series of constructs were designed that either singly or doubly mutated the 5’ and 3’ terminal base pairs with a reciprocal mutation in the other strand to maintain the base paired state. Representative sequence diagrams and the results of this investigation are shown in Figure 41.

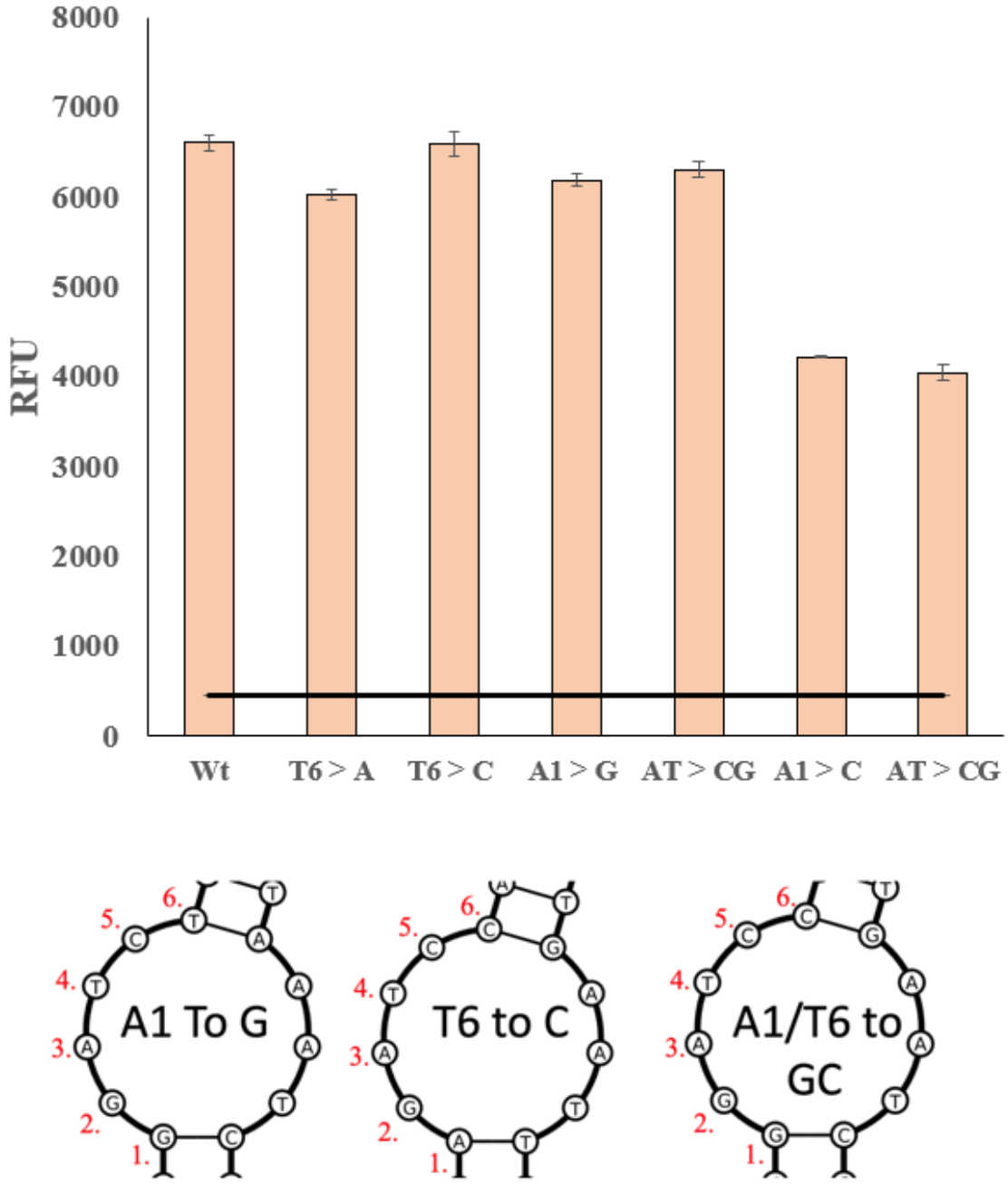


Figure 41. Mutational analysis of the terminal base paired nucleotides in the hexanucleotide active sequence. Here mutations in which the nucleotide class is preserved (purine to purine or pyrimidine to pyrimidine) results in maintenance of activity regardless of nucleotide activity. Mutation of A1 from a purine to a pyrimidine (A1>C) results in a complex less efficient than the wild type or preserved mutants. This would indicate that a structural feature of purines is optimal in the 1st position of the active sequence.

Here the mutations were chosen to either maintain the same class of nucleotide, for example a purine-to-purine mutation of A>G, or to result in class switch from, purine to pyrimidine such as A>T. T6 was very tolerant of both mutation types, with very little difference in activity when mutated to a C or A showing that this particular base pair is flexible in its identity. The A1 position tolerates mutations within the same class of nucleotide, purine to purine, as can be seen by A1>G. However, when the mutation cross classes, as in the instance of A1>C, the activity decreases by roughly 30% which would indicate that this position has a purine preference. This purine preference and the near equivalency between adenine or guanine in this position would seem to indicate that the aptamer utilizes a common feature of the purine ring system. Namely the purine edge, corresponding to C5-N7-C8, which are preserved between both adenine and guanine. This edge would be available toward the interior of the pocket, and its absence in the thymine mutation could account for the loss activity.

Conclusions:

This series of investigations, while not exhaustive, provide a more detailed map of MG DNA aptamers sequence requirements. In particular we have shown through a series of mutated constructs that the hexanucleotide sequence of 5'-AGATCT-3' is tolerate of mutation in both terminal position but with a distinct purine preference in the first position. This produces a more accurate consensus sequence of Pur-GATC-N, Pur indicates the presence of either A or G and N any nucleotide. The contributions of the trans strand are more complex. The wild type sequence AAT is the most active complex. The aptamer still retains the ability to enhance MG fluorescence even in the presence of a very disparate sequence, such as CCG, but to a smaller

degree. This characterization is incomplete and further study should focus on appropriate experimental design to more convincingly determine geometric sensitivity, by use of abasic site modified oligonucleotides. Once a more comprehensive analysis of the sequence and structural requirements is complete, the MG DNA aptamer presents an attractive bioanalytical tool given its small size. The development and optimization of this aptamer in a beacon format would also be a fruitful endeavor.

References:

95. Dickerson, R. E., Base sequence and helix structure variation in B and A DNA. *J Mol Biol* **1983,** *166* (3), 419-41.
96. Gao, S.; Zheng, X.; Tang, Y.; Cheng, Y.; Hu, X.; Wu, J., Development of a Fluorescently Labeled Aptamer Structure-Switching Assay for Sensitive and Rapid Detection of Gliotoxin. *Analytical Chemistry* **2019,** *91* (2), 1610-1618.
97. Wachowius, F.; Höbartner, C., Probing Essential Nucleobase Functional Groups in Aptamers and Deoxyribozymes by Nucleotide Analogue Interference Mapping of DNA. *Journal of the American Chemical Society* **2011,** *133* (38), 14888-14891.
98. Einert, Thomas R.; Netz, Roland R., Theory for RNA Folding, Stretching, and Melting Including Loops and Salt. *Biophysical Journal* **2011,** *100* (11), 2745-2753.
99. Gebala, M.; Giambaşu, G. M.; Lipfert, J.; Bisaria, N.; Bonilla, S.; Li, G.; York, D. M.; Herschlag, D., Cation–Anion Interactions within the Nucleic Acid Ion Atmosphere Revealed by Ion Counting. *Journal of the American Chemical Society* **2015,** *137* (46), 14705-14715.
100. Serra, M. J.; Baird, J. D.; Dale, T.; Fey, B. L.; Retatagos, K.; Westhof, E., Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *Rna* **2002,** *8* (3), 307-23.
101. Fischer, N. M.; Polêto, M. D.; Steuer, J.; van der Spoel, D., Influence of Na+ and Mg2+ ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res* **2018,** *46* (10), 4872-4882.
102. Dubins, D. N.; Lee, A.; Macgregor, R. B.; Chalikian, T. V., On the Stability of Double Stranded Nucleic Acids. *Journal of the American Chemical Society* **2001,** *123* (38), 9254-9259.

103. Schreck, J. S.; Ouldridge, T. E.; Romano, F.; Louis, A. A.; Doye, J. P., Characterizing the bending and flexibility induced by bulges in DNA duplexes. *J Chem Phys* **2015,** *142* (16), 165101.

APPENDIX A: CHAPTER 2 SUPPLEMENTAL MATERIALS

Table 2: List of sequences used in Chapter 2

Target Analyte Strands	
NC-45A	TTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCC
NC-45	TTGCTGTTCTTTATCAGG**G**TGTTAACTGCACAGAAGTCC
NC-45L	TACTTCTAACCAGGTTGCTGTTCTTTATCAGGGTGTTAACTGCACAGAAGTCCCTGTTGC
NC-14	TCCATTCTGGTTACTGCCAGTTGAATCTGA
NC-14C	TCCAT**C**CTGGTTACTGCCAGTTGAATCTGA
NC-14CC	TCCA**CC**CTGGTTACTGCCAGTTGAATCTGA
Malachite Green Aptamer Sequences (SMG)	
SMG2A	GGACTTCTGTGCAGTTAACAGCAGATCTAACG
SMG2B	CGTTAAATTGCTCCTGATAAAGAACAGCAA
SMG3.1A	GGACTTCTGTGCAGTGCAGATCTAACG
SMG3.1AL[2]	GGACTTCTGTGCAGT**/iSp18/**GCAGATCTAACG
SMG3.1AT	GGACTTCTGTGCAGTTTTGCAGATCTAACG
SMG3B	CGTTAAATTGCTAACACCCTG
SMG3BL	CGTTAAATTGC**/iSp18/**TAACACCCTG
SMG4.2A	GGACTTCTGTGCAGTTA**/iSp18/**GTTCAGATCTATCG
SMG4.2AT	GGACTTCTGTGCAGTTA **TTT** GTTCAGATCTATCG
SMG4.2B	CGATAAATTGAACACACCCT
SMG4.1A	GGACTTCTGTGCAGTT **/iSp18/** CAGCAGATCTAGCG
SMG4.1AT	GGACTTCTGTGCAGTT **TTT** CAGCAGATCTAGCG
SMG4.1B	CGCTAAATTGCTGAACACCCT
Split Dapoxyl Aptamer Sequences (SDA)	
SDA2A	GACTTCTGTGCAGTTA CTACGGGGGAGGGTGTGTGGTCTTGGTCAT
SDA2B	ATGACCTTGGTTCGTAG **/iSp18/**ACACCCTGA
SDA4A[3]	ATTCAACTGGCAGTACTACGGGGGAGGGTGTGTGGTCTTGGTCAT
SDA4B[3]	ATGACCTTGGTTCGTAG**/iSp18/**ACCAGAATGGA

Mini-MG Sequences	
Mini-MG 0.0	AGATCT
Mini-MG 0.1	CGCAGATCTGCG
Mini-MG 0.2	GCGCAGATCTGC GC
Mini-MG 0.3	CGCTAGATCTTG CG
Mini-Mg 1.0	GGACTTCTGTGCAGTTTAGATCTCACCCTGAGAT
Mini-MG 1.1L	GGACTTCTGTGCAGTT/iSp9/AGATCTCACCCTGAGAT
Mini-MG 1.2L	GGACTTCTGTGCAGTTTAGATCT/iSp9/CACCCTGAGAT
Mini-MG 1.3	GTGCAGTTTAGATCT/iSp9/ACCCTGAGATC

Experimental Section

General Fluorescence Assay for SMG Sensor:

Split malachite green aptamer (SMG#.A/B) strands were added to 5× SMG buffer (100 mM Tris-HCl, pH 7.4) and diluted to a final volume of 99 µL by addition of water and the appropriate amount of analyte (NC-45) strand obtaining a final concentration of 500 nM for the aptamer strands. All samples were incubated at room temperature (22.5 ºC) for 90 min before addition of 1.25 µL of 100 µM malachite green solution (final concentration of 1.25 µM) followed by additional incubation of 30 min at room temperature. Eighty µL fractions of these samples were transferred to a Thermo Fisher Scientific-Nunclon 96 Flat Bottom Transparent Polystyrene plate and the fluorescence recorded on a Tecan Infinite 200Pro at 658 nm with an excitation of 617 nm. The data of three independent trials was processed with Microsoft Excel.

General Fluorescence Assay for SDA Sensor:

Split Dapoxyl aptamer (SDA#.A/B) strands were added to 5× SDA buffer (100 mM Tris-HCl, 1 M KCl, 50 mM $MgCl_2$,pH 7.4) and diluted to a final volume of 99 µL by addition of water and the appropriate amount of analyte (NC-14) strand obtaining a final concentration of 500 nM for the aptamer strands. All samples were incubated at room temperature (22.5 ºC) for 90 minutes before addition of 1.0 µL of 200 µM Auramine O solution (final concentration of 2 µM) followed by additional incubation of 30 minutes at room temperature. Eighty µL fractions of these samples were transferred to a Thermo Fisher Scientific-Nunclon 96 Flat Bottom Transparent Polystyrene plate and the fluorescence recorded on a Tecan Infinite 200Pro at 540 nm with an excitation of 475 nm. The data of three independent trials was processed with Microsoft Excel.

Discrimination Assays:

All discrimination assays followed the same procedure but utilized 150 nM of the match or mismatched analyte and 250 nM of the sensor strands.

General Multiplex Assay:

Split malachite green aptamer strands (SMG42AT and SMG42B) and split Dapoxyl aptamer (SDA) strands (SDA4A and SDA4B) were added to 5× Multiplex buffer (100 mM Tris-HCl, 50 mM KCl and 2 mM $MgCl_2$, pH 7.4) and diluted to a final volume of 98 µL by addition of water and the appropriate amount of matched or mismatched analyte (NC_045512 A>G or NC1450) strand obtaining a final concentration of 500 nM for SMG42AT/B and 150 nM SDA4A/B. All samples were incubated at room temperature (22.5 ºC) for 90 min before addition of 1.0 µL of 100 µM of malachite green (final

concentration of 1 µM) and 1.0 µL of 200 µM Auramine O solution (final concentration of 2 µM) followed by additional incubation of 30 min at room temperature. 80 µL fractions of these samples were transferred to a Thermo Fisher Scientific-Nunclon 96 Flat Bottom Transparent Polystyrene plate and the fluorescence recorded at 658 nm with an excitation of 617 nm to detect malachite green fluorescence and at 540 nm with an excitation of 475 nm on a Tecan Infinite 200Pro. The data of three independent trials was processed with Microsoft Excel.

Limit of Detection Calculations:

The LOD of split aptamer assays (SMG and SDA) were calculated by determining the linear trendline of the curves (EXCEL) and calculating the concentration in which the fluorescence surpassed the threshold value (defined as 3 SD of the blank sample).

Structure Prediction and Schematic Generation:

All structural predictions utilized NUPACK which were exported as .svg files and edited in InkScape

APPENDIX B: CHAPTER 3 SUPPLEMENTAL MATERIALS

Experimental Section:

General Fluorescence Assay:

1 uM of MGA and 1 uM of MGB are incubated for 105 minutes prior to the addition of MG to a final concentration of 2 uM and final volume of 100 uL with a buffer composition of 20 mM Tris-HCl (pH 7.4) followed by additional 15 minutes of incubation. Eighty µL fractions of these samples were transferred to a Thermo Fisher Scientific-Nunclon 96 Flat Bottom Transparent Polystyrene plate and the fluorescence recorded on a Tecan Infinite 200Pro at 658 nm with an excitation of 617 nm. The data of three independent trials was processed with Microsoft Excel.

Structure Prediction and Schematic Generation:

All structural predictions utilized NUPACK which were exported as .svg files and edited in InkScap

Table 3: List of Sequences Used in Chapter 3

Mini-MG Sequences:	
Mini-MG 0.0	AGA TCT
Mini-MG 0.1	CGC AGA TCT GCG
Mini-MG 0.2	GCG CAG ATC TGC GC
Mini-MG 0.3	CGC TAG ATC TTG CG
MINI-mg 0.4	CAG ATC TG
mini-mg 0.5	CGC GTT TCG CGA GAT CT
mini-mg 0.6	TAA ATT GCC GTT TCG GCA GAT CTA
Mini-MG 1.0 Sequences	
Mini-Mg 1.0	GGA CTT CTG TGC AGT TTA GAT CTC ACC CTG AGA T
Mini-MG 1.1L	GGA CTT CTG TGC AGT T/iSp9/AG ATC TCA CCC TGA GAT
Mini-MG 1.2L	GGA CTT CTG TGC AGT TTA GAT CT/iSp9/C ACC CTG AGA T
MG1.3	GGA CTT CTG TGC AGT TTA GAT CTA CCC TGA GAT C
MG1.3TT	GGA CTT CTG TGC AGT TTA GAT CTT TAC CCT GAG ATC
Mini-MG 1.4A	TGC AGT TTA GAT CT/iSp9/A CCC TGA GAT C
Mini-MG 1.4B	GTG CAG TTT AGA TCT /iSp9/ACC CTG AGA TC
Gap Interrogation Sequences:	
MGA2	CGC GGC AGA TCT AAC GAG C
MGB0N	GCT CGT TAT GCC GCG
MGB-1N;	GCTCGTTAATGCCGCG
MGB-2N;	GCTCGTTAAATGCCGCG
MGB3N	GCT CGT TA AAT TGCC GCG
MGB4N	GCT CGT TAT AAT TGC CGC G
MGB-5N;	GCTCGTTATACCTTGCCGCG
Sequence Identity Sequences:	
MGB-AGG	GCT CGT TAA GGT GCC GCG
MGB-CCA	GCT CGT TAC CAT GCC GCG
MGB-CGG	GCT CGT TAC GGT GCC GCG
MGB-CAT	GCT CGT TAC ATT GCC GCG
MGB-CAG	GCT CGT TAC AGT GCC GCG
MGB-TGA	GCT CGT TAT GAT GCC GCG
MGB-TGT	GCT CGT TAT GTT GCC GCG
MGB-TGG	GCT CGT TAT GGT GCC GCG
MGB-PPY	GCT CGT TAA GTT GCC GCG
MGB-PYP	GCT CGT TAA CAT GCC GCG
MGB-PYY	GCT CGT TAA CTT GCC GCG
MGB-PPP	GCT CGT TAA AGT GCC GCG
MGB-YPP	GCT CGT TAC AAT GCC GCG
MGB YPY	GCT CGT TAC GTT GCC GCG
MGB-YYP	GCT CGT TAC CGT GCC GCG
MGB-AGA	GCTCGTTA AGA TGCCGCG
MGB-AAA	GCTCGTTA AAA TGCCGCG
Base Pair Identity Sequences:	
MGA2-A>G	CGC GGC GGA TCT AAC GAG

MGA2-T>C	CGC GGC AGA TCC AAC GAG C
MGA2-AT>GC	CGC GGC GGA TCC AAC GAG
MGB3NA>G	GCT CGT TAA ATC GCC GCG
MGB3NT>C	GCT CGT TGA ATT GCC GCG
MGB3NAT>CG	GCT CGT TGA ATC GCC GCG
MGA2-T6>A	CGC GGC AGA TCA AAC GAG C
MGB-T6>A	GCT CGT TTA AT TGCC GCG
MGA2-A>G+C	CGC GGC GGA TCT AAC GAGC
MGA2-AT>GC+C	CGC GGC GGA TCC AAC GAGC