Development of RNA Aptamers through Supervised Learning and Capture-SELEX strategy

Neel Arvind Shah

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2020

Committee:

James Carothers

David Beck

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Development of RNA Aptamers through
Supervised Learning and Capture-SELEX strategy

Neel Arvind Shah

Chair of the Supervisory Committee:
James Carothers
Department of Chemical Engineering

Aptamers are small oligonucleotides that are capable of binding specifically to a target, with impressive potential for analysis, diagnostics, and therapeutics applications. Aptamers are selected *in vitro* from large nucleic acid combinatorial libraries using an iterative selection process called SELEX (Systematic Evolution of Ligands by EXponential enrichment). *In vitro* selected RNA aptamers can be assembled into programmable biosensors for a broad range of synthetic biology applications requiring the detection and quantification of small molecule metabolites. Aptamers have been around for three decades now, however, sensors that work in proof-of-concept studies, mainly performed in academia, do not necessarily match the conditions, the statistical relevance, the compatibility with routine equipment found in e.g., clinical diagnostic departments, and the scale required for an industrial and/or clinical application. A major obstacle to the broader development of aptamer metabolite biosensors as tools for industrial biotechnology and translational medicine stems from uncertainty in the outcomes of *in vitro* selection and the aptamer-ligand binding affinities that can be obtained. We present, to our knowledge, the first approach for *a priori* estimation of RNA aptamer binding affinities and a virtual screening tool against small molecule targets. We illustrate the generalizability of this tool to identify

in silico metabolites and biomarkers that could be targeted with aptamer biosensing. Further, we outline an RNA Capture-SELEX strategy to generate *in vitro* structure switching aptamers which could be incorporated into genetic switches for regulating gene expression; since in classical SELEX only binding and not conformational changes are selected which is an important property of 'sensing-actuation' devices that represent an important class of genetic devices which can detect, report on and act on environmental and intracellular signals, including small molecules and proteins, to study and control cellular functions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I thank my academic advisor, Dr. James Carothers, for accepting me into his group. His friendly guidance and expert advice have been invaluable throughout all stages of the work. I would like to thank the member of my thesis committee, Dr. David Beck, for his valuable feedback and taking time to help me on my thesis work in spite of his busy schedule.

The thesis has also benefited from comments and suggestions made by David Sparkman-Yager, Jason Fontana who have read through the manuscript. I take this opportunity to thank them. I would also like to extend my gratitude to members of carothers research group Ben Tickman, Cholpisit (Ice) Kiattisewee, Diego Alba Burbano, Ian Faulkner, Ryan Cardiff, Ava Karanjia and Widi Sugianto through whom I have learned and grown my knowledge about different aspects of engineering biology during group meeting presentations and discussions.

I would like to acknowledge the Department of Chemical Engineering at UW, NSF Award MCB 1517052 (to J.M.C.) and a gift from Conagen Inc. (Bedford, MA, USA).

Nobody has been more important to me in the pursuit of this MS journey than the members of my family. I would like to thank my parents and my brother; whose love and guidance are with me in whatever I pursue. I would also like to thank all my friends (old and new) for their constant support and guidance to tackle the ups and downs of this journey, making the most memorable moments at the University of Washington

Chapter 1. Introduction and Background

1.1. Synthetic Biology

According to its name, synthetic biology seems to be only a spin-off of biology. However, the definition of synthetic biology can differ between biologists and engineers. On one hand it can be consider as: “an emerging discipline that uses engineering principles to design and assemble biological components” (1) or on the other hand as: « a new emerging scientific field where ICT, biotechnology and nanotechnology meet and strengthen each other » (2).

Based on the knowledge of molecular biology, synthetic biology unites different disciplines of science going from nanotechnology to immunology and connect existing concepts between engineering and biology. Synthetic biology split itself to classical biology by this engineering aspect to develop in living organism new functions and properties. Emerged in the late 1990’s the development of the field has greatly benefited from the advancement and the discovery of cloning strategies, automated DNA synthesis and sequencing, High-Throughput Sequencing (HTS) and the elucidation of the microbial genome. Driven by the purpose of implementing new biological functions to organism or modifying existing systems, the applicability of these artificial construct is rather extensive. Reprogramming stem-cells, development of bio sensors or the production of chemical entities in genetically modified bacteria is just a small list of what is applicable in the field. In the last years, the field went through important breakthrough thanks to the advances in the reprogramming of fundamental cellular processes that operate orthogonally across several species like RNA regulatory tools derived from the Clustered-Regulatory-Interspaced-Short-Palindromic-Repeats (CRISPR) immune system from bacteria and the expansion of the genetic alphabet by unnatural amino acids provided further opportunities in the framework of cellular engineering (3, 4). Another achievement was performed by the first complete synthesis and assembly of a mini genome in a bacterial cell (5). Yeast was also part of these progress by the first synthesis of a synthetic eukaryotic chromosome (yeast SC 2.0 project) (6). The development and application of synthetic riboswitches in synthetic biology starts to emerge. Recently, riboswitches were used to study the transient dynamics of a “logic gate-like” construct at a single cell level (7) but also to build new devices that can regulate in an ON/OFF way the expression of numerous gene in yeast upon target addition (8).

This master’s thesis focuses on the development of tools and approaches for selecting RNA aptamers and using those aptamers to build functional, responsive RNA biosensors for metabolic engineering and medical diagnostics purpose. We present, to our knowledge, the first approach for a

priori estimation of RNA aptamer binding affinities against small molecule targets. By applying a supervised learning approach, we trained models to predict aptamer-small molecule binding energies from the chemical compositions of target ligands. We illustrate the generalizability of this tool to identify in silico metabolites and biomarkers that could be targeted with aptamer biosensing. The thermodynamic modeling framework and supervised learning strategy developed here, which is available online, can drive choices of RNA aptamer molecular targets and reduce uncertainty around *in vitro* selection outcomes for metabolic engineering and medical diagnostic applications. In the second part, we discuss a strategy called Capture-SELEX for *in vitro* selection of aptamers. *In vitro* selected aptamers are not always able to regulate gene expression. It requires the aptamers undergo a conformational change during ligand binding. Also, *in vitro* selection is a time-consuming process. Due to the many rounds of selection necessary and due to the lack of feedback on enrichment during the process, selections typically take multiple months. Analyses of *in vitro* selection conditions show that the median RNA aptamer selection requires 10 rounds, with some reported RNA aptamer selections taking as many as 29 rounds (9). Due to the financial and time costs of sequencing, sequencing is generally run at the conclusion of a selection rather than after each round, which results in minimal feedback on the enrichment of sequences over the course of the selection. Although the overall success rate of running *in vitro* selections has not been characterized, anecdotes in literature sources suggest that a majority of selections are unsuccessful (10). In classical SELEX, however, only binding and not conformational changes are selected. A new technique was developed for DNA in 2012 by Stoltenburg et al, called Capture-SELEX allowing the selection of structure switching aptamers against small molecules in solution. Recently, RNA based Capture-SELEX methodologies was published for small molecule-binding aptamers by Boussebayle et al and Lauridsen et al. Adapting the strategies from Lauridsen et al, here we discuss a detailed layout of experimental method which could be used to generate *in vitro* RNA aptamers in the solution that could be incorporated into genetic switches and biosensors for regulating gene expression or sense metabolites in the field of metabolite engineering and medical diagnostics.

1.2. Nucleic Acid Aptamers

Aptamers are short synthetic RNA or DNA oligonucleotides sequences which are able to bind a broad range of targets with high affinity and specificity. The type of target is going from small molecules to proteins, viruses, or even whole cells. Based on their nucleotide composition, they are able to fold in complex three-dimensional structures allowing them to form binding pockets like their proteins counterparts.

1.2.1. Three-dimensional structure of RNA structures

RNA has the unique feature to form high complex three dimensional structures compared to DNA. This characteristic makes them able to bind their target with high specificity and affinity. There are different levels of structures in RNA, the primary, the secondary and the tertiary structure. The primary structure consists in solely of the nucleic acid sequence composition (A, C, G or U). The secondary structure of the RNA is a two-dimensional representation of the structure by showing all Watson-Crick and non-canonical base pairing. These interactions are creating more complex structures like stem-loops also called hairpin. Stem-loops consists of two regions of the same strand are complementary in the nucleotide sequence when read in contrary directions. This strand form then a double helix structure that ends with an unpaired loop. Bulges are also another secondary structure created when one side of the stem possess one or more nucleotide than its complementary side. These supplementary nucleotides are not involved in base pairing and are therefore forming single strand structure inside of a hairpin.

Secondary structure is mostly determined by free energy minimization relying on approximations of RNA motif stability depending on the input sequence (11). Secondary structure prediction tools, including the inverse-folding package MODENA and the Vienna RNAfold package are used to calculate the lowest folding energy of all sequence fragments (12, 13). Generally, this prediction gives a good overview of the secondary structure of the RNA. However, these software's are not able to predict the next level of RNA structure, the tertiary structure.

Tertiary structure involves interactions between secondary structure elements. Loop-loop interaction (also called kissing complexes) happens when single stranded part loop of one hairpin is forming Watson-Crick interaction with another loop of another hairpin. Pseudoknots are also another motif that can be formed by RNA. It consists in nucleotides of a loop that form Watson-Crick interaction with a complementary single stranded sequence. Other complexes such as G-quadruplex (4 guanines interacting with each other through Hoogsteen interaction forming a square planar structure) or three-way junctions (three stem connected together) are just other examples of folding possibilities for tertiary RNA structure (14).

1.2.2. Binding and interaction potential of nucleic acid aptamers

Each nucleobase (A, C, G and T/U), but also the ribose and the phosphate backbone of nucleic acid possess hydrogen binding donor or acceptor atom collections. Hydrogen binding acceptor are either oxygen as a ketone group on a ring or tertiary amine. Hydrogen binding donor are primary or

secondary amine groups that contains respectively one or two hydrogen atoms bound to the nitrogen. According to simple valence bond theory, a hydrogen atom should be able of forming only one chemical bond. However, in many cases hydrogen is formally two-valent. This additional bond is called “hydrogen bond”. There are two types of hydrogen bond but here we will focus only of hydrogen bonds with atoms having a higher electronegativity as hydrogen (C, N, O, F, P S, Cl…) which are the most common elements found in nature. These atoms have the capability of forming A-H... B hydrogen bonds (Π-hydrogen bonds) if B has an unshared pair of electrons. Dissociation energies spread in three orders of magnitude (about 0.2-40 kcal/mol). Within this range, the electrostatic, covalent, and dispersion contributions vary in their relative weights and makes the nature of the interaction variating. The hydrogen bond possesses wide transition regions that merge continuously with the covalent bond, the van der Waals interaction, the ionic interaction, and also the cation-Π interaction (15).

RNA or ssDNA are both polar molecules and globally negatively charged due to their phosphate backbone and therefore allow dipole-dipole interaction or ion-dipole interaction. A dipole-dipole interaction is an electrostatic interaction between a negative side of a polar molecule with a positive side of another polar molecule. To allow this interaction, they need to be close together like they are in a case of aptamer-target binding. However, ion-dipole interactions are more commonly found between aptamer and their targets mostly thanks to magnesium. Magnesium belongs to the alkaline-earth metals and is mostly found in the biological context as its ion form Mg^{2+}. Magnesium is an essential mineral nutrient (i.e., element) for life (16, 17) and is present in every cell type in every organism. For example, ATP (adenosine triphosphate), the main source of energy in cells, must bind to a magnesium ion in order to be biologically active**.**

Another bonding force involved in DNA and RNA is pi-stacking. Aromatic pi-stacking is of importance in regulating some mechanisms at the gene level. For example, the recognition of nucleic acid by proteins for regulating gene expression is largely controlled by π-stacking of aromatic amino acids and nucleobases (18, 19). Furthermore, neighboring base-pair stacking controls the gross geometrical feature of a nucleic acid macromolecule (20, 21). Local nucleotide sequences determine the rigidity of a nucleic acid and the capacity of a segment to fold, kink, or supercoil and is, therefore, responsible for creating the local geometry needed for proper enzymatic recognition (22). The probability of the intercalation of a planar molecule between two stacked nucleic acid base pairs is inversely proportional with the strength of their π-stacking (23). DNA polymer analogues where the backbone sugar residues were removed still have the possibility to form the characteristic R-double helix, with stacked base pairs in the usual Watson-Crick base-pairing scheme (24). While hydrogen

bonding plays a major role in gluing the two intertwined helices together, preserving the R-helical structure could be protected by π-stacking interaction in despite the absence of sugar residues. Some evidence suggests that the interstrand hydrogen bonding and the nearest neighbor base π-stacking interaction are not independent, as the stacking of nucleobases influences their hydrogen bonding capacity (25).

1.2.3. Advantages of nucleic acid aptamers over antibodies

All these characteristics explain the versatility of aptamers to interact with a broad amount of target in the same way as antibodies (10). However, aptamers have multiple advantages compared to antibodies. Monoclonal antibodies are laborious and expensive to produce involving high-cost processes and screening of a large number of colonies. Additionally, a large-scale synthesis is needed nowadays for antibodies, but the performance of the same antibody differs in a different batch. Unlike antibodies, aptamers can be generated with high accuracy and reproducibility de novo by chemical synthesis or by in vitro transcription at high yields, with a low batch to batch variation and free from biological contamination (26). They can be as well synthesized or modified with various chemical moieties to increase their resistance to nucleases or to fluorophores for the fabrication of biosensors (27, 28). Furthermore, it is well known that proteins are easily denatured and lose their tertiary structure at high temperature. On the other hand, oligonucleotides are more thermally stable and maintain their structures over repeated cycles of denaturation/renaturation. Hence, the greatest advantage of oligonucleotide-based aptamers over protein-based antibodies is their stability at elevated temperatures. Aptamers fold back to their original shape and still interact in the way with their target, whereas antibodies are under heat treatment irreversibly denaturized (29). The storage and the stability of aptamers is cheap and easy as they can last for long time stored at – 20°C. Finally, aptamers can bind molecules going from very small organic molecules or even ions (30), but also to higher size molecules such as peptides, proteins or even whole cells having K_D from the mM range to the low nM range (30–32). All these characteristics make aptamers ideal candidates for a wide range of application.

Aptamers are often used as sensors. In the case of fluorescence detection, the easiest way is to label the aptamer with both fluorophore and quencher. Using a FRET signal between amine-reactive quencher DABCYL moieties and fluorescein, a device detecting cocaine could be built based on the cocaine aptamer (33). Aptamers are able to change in several different conformations. This ability makes them suitable for the building of probes such as aptamer beacons. Upon target binding, hairpin

shape or other more complex rearrangement can allow these probes to detect the presence of the target in solution by using fluorophore-quenchers pairs (34–36). Furthermore, quantum dots, gold nanoparticles (NP) and other nanomaterials were studied because of their fluorescence quenching properties in combination with aptamers to replace classical quenchers (37–39). The advantage of gold nanoparticles, due to their properties can be applied for the development of colorimetric sensors, removing the necessity of expensive tools to have a readout but only with naked eyes. The aptamer placed in the surrounding of the NP prevent the gold to aggregates thanks to its highly negatively charged backbone, but upon target detection, aptamers are removed from the particles because of their structure switching capacities and cause the gold to aggregate in bigger particles that lead to a change in color (40). This method could be applied successfully to detect ATP, cocaine, Pb^{2+}, and K^{+} (41).

1.2.4. Application and potential of Aptamers

The application of aptamer is not limited to in vitro sensing but also can be applied in vivo. In 2012, the Jaffrey's lab developed a system that will lead to a broad range of aptamer-based sensors (42). Based on the GFP fluorescence mechanism, a synthetic ligand called DFHBI (or spinach) was synthesized and an aptamer was developed against this molecule. It appeared that upon RNA binding, the fluorescence of the molecule is activated. This result made the development of hybrid aptamers possible, where the folding of the first aptamer (in this case, the sensor part) will allow the folding of the spinach aptamer and make it glow, creating in vivo producible sensor (43–45).

Application of aptamer is though not only limited to the sensing field. Considering their numerous advantages compared to antibodies, the development of aptamer applied in health became a major axe of research in the aptamer community. The very first aptamer used as a drug, Macugen, is already commercially available for age-related macular degeneration on the form of a single strand nucleic acid pegylated aptamer specific to VEGF165 which has a major role in angiogenesis and permeability (46). Following the example of the Macugen, several aptamer-based drug such as AS1411 (a nuclein-specific aptamer) for acute myeloid leukemia (47), ARC1779 (a von Willebrand factor-specific aptamer) for carotid artery disease (48), and NU172 (a thrombin-specific aptamer) for anticoagulation, are currently tested in clinical trials (49). Aptamers as drug delivery system were also developed based on the selection of candidates that could bind to internalized cell surface receptors. One of the examples is an aptamer binding the prostate-specific membrane antigen was 11 coupled to doxorubicin, an anti-cancer drug, onto the aptamer strand. This drug-loaded aptamer complex was able to successfully carry doxorubicin into prostate cancer cells (50).

1.3. SELEX – establishment, optimization, and diversification

1.3.1. Development and concept of SELEX

As aptamers are synthetic nucleic acid sequences, they have to be developed in the laboratory. The standard process to develop aptamers is called SELEX. SELEX stands for Systematic Evolution of Ligands by EXponential enrichment. In the SELEX, aptamers are selected from a synthetized library containing a random region between 20 to 60 random nucleotides flanked by two primer binding sites on each side for reamplification after the selection. In this work, the size of the RNA library used is 123 nt using a randomization of 50 nt, achieving a 4^{50} theoretical maximal diversity. This number is equivalent to approximatively $1.2*10^{30}$ different sequences or 2 mmol (10^6) of RNA. With a molecular weight of around 40,000 g/mol the total weight of nucleic acid would be around 80,000 metric tons or $2*10^{10}$ liter of oligonucleotide at 100 µM (standard delivery concentration when order to a furnisher). In practice, it has been proved that around 1 nmol of RNA per selection is enough (51, 52). This shows that not all the 50 nucleotides are involved in the binding to the target. A major part of the sequence is responsible for the formation of the secondary structure of the aptamer meaning that these nucleotides can be variable as long as the folding does not change, and the base-pair are kept. The advantage of long motif such as 50 nt is that a small binding motif can be presented in different folding possibilities, such as stem loop, bulge, or three-way junction.

The original protocol was developed in the early 90's by three different working groups in parallel (53–55). The selection process is split in several step. The first one is the synthesis of the library. For the first round the library is synthesized chemically in an amount big enough to achieve a theoretical diversity of at least 10^{15} different sequences. The library is afterwards amplified to generate enough copies of each sequences and afterwards converted in single-strand oligonucleotides (either by separating the two strands for DNA or by *in vitro* transcription for RNA). Once the library is generated, the DNA/RNA are folded by denaturation/renaturation and incubated with the target molecule which is immobilized on a matrix. During this step, aptamers able to recognize the target will be able to bind and form stable interaction. Using different type of methods, depending on the nature of the matrix used, the unbound nucleic acids are washed away by using selection buffer and removed from the pool. Aptamers with a good enough affinity to the target will go through the washing steps and remain on the target. These aptamers are afterwards recovered by a non-specific elution from the matrix (heating, chelation of the ions necessary for the interaction or harvesting of the matrix) or a specific elution (addition of an excess of target in solution to force aptamers to undock from the originally bound target). Once recovered, these aptamers are amplified and used again in another

selection cycle.

1.3.2. Optimization and diversification of SELEX strategies

Through the years, SELEX never stop to evolve in order to create a more complete and reliable method. Some basics ameliorations were made at the beginning such as the introduction of the Counter-SELEX (56, 57). This technique was established to segregate in the library non-specific binders to the target that could also bind an unwanted closely related target. The counter-elution step happens as follow: The counter-target is implemented into the matrix and will fish out the non-specific binders. Furthermore, ameliorations on the libraries were performed, such as the using modified bases (27) and nuclease resistant backbone (58–60). These ameliorations were making to aptamers more resistant to degradation on one hand, and on the other hand allowed aptamers to have more interaction possibilities thanks to the introduction of chemical moieties not originally present on the nucleic acids. The drawback of these modification is that no organism is able to produce these sequences in vivo as they are fully artificial. Also, special enzymes are necessary to treat with these modifications during the SELEX (61). Finally, SELEX method expended to a wide range of support and the method got diversified. Cell-SELEX is one of the most known example. This technique is not using anymore a matrix to bind the target protein but directly living cells that are expressing the protein of interest directly on the surface of the cell (62). Thanks to this technique, aptamers could be selected for membrane proteins (63) which are very difficult to produce in vitro and if succeeded, are probably not in the folding they should be at the surface of a cell (64). Capillary Electrophoresis (CE) SELEX is another alternative to perform SELEX. This technique has the advantage to realize the selection without any immobilization of the target. Indeed, based on the mobility shift observed when there is a formation of a complex aptamer-target, the binding sequences are migrating at a different rate allowing the separation and collection of high-affinity aptamers from the non-binding sequences (65). This process was the first step towards a faster SELEX protocol. Actually, all the protocols mentioned beforehand have the drawbacks of being longer and time consuming. However, CE-SELEX requires a specific equipment and an adapted formation to be performed.

Chapter 2. Applying supervised learning to predict aptamer affinities for biosensing applications in metabolic engineering and medical diagnostics

2.1. Introduction

Small-molecule biosensing plays a critical role throughout synthetic biology. In metabolic engineering, biosensing is useful for effecting metabolite-responsive changes in gene expression, as well as detecting and quantifying metabolite production in engineered cells (66–69). The ability to quantitatively or qualitatively sense metabolites is a key aim in medical diagnostics and could allow for dynamic drug dosing and tailored therapeutics in the near future. However, many small molecules that are production targets for metabolic engineering or diagnostic targets in medicine lack known sensors and regulators. RNA aptamers selected *in vitro* exhibit high affinity and specificity due to their stable three-dimensional structure (70), making them well-suited for many applications that require binding. Aptamers are easier to generate than protein binders for molecules lacking known sensors (71, 72) and they can be incorporated into genetic devices for controlling gene expression or reporting on metabolite concentration through design methods utilizing RNA secondary structure prediction tools (13, 73, 74). Aptamers have been incorporated into fluorogenic biosensors for enzyme evolution (75), assembled into Aptazyme for genetic control (76, 77), assembled into riboswitches for translational control (74), encapsulated in droplets to provide single-cell readouts for high-throughput screening of secreted metabolites (78), incorporated into electrochemical sensors for real-time small molecule measurements in animals (79), utilized to control an infusion pump to actuate closed-loop control of drug dosing (80), and incorporated into sgRNAs to allow for small molecule control of CRISPR-Cas9 gene editing and regulation (81). Collectively, these efforts demonstrate the wide utility of RNA aptamers as sensors in metabolic engineering, drug detection, and diagnostics (**Fig. 2.1a**). Despite the wide range of possible applications, uncertainty about generating useful aptamers has hampered broader efforts in aptamer engineering.

One of the central goals of synthetic biology is to make engineering biology predictable enough that trial-and-error experiment is minimized (82, 83). Synthetic aptamers are generated through enrichment from diverse nucleic acid pools via *in vitro* selection (or the systematic evolution of ligands through exponential enrichment, SELEX), a process of iteratively partitioning binding from non-binding sequences (53, 54). However, *in vitro* selections can typically take several months to complete and there is no guarantee of generating functional aptamers (10). The binding affinities for aptamers derived from successful selections have spanned more than seven orders of magnitude

introducing further uncertainty into the overall process (**Fig. 2.1b**) (10). Finally, the resultant aptamer-small molecule binding affinities may be insufficient for the intended biosensing application, especially as most RNA aptamer biosensor devices are only responsive at concentrations two or more orders of magnitude above the aptamer binding affinity (84, 85).

While the binding properties of small molecule aptamers have been previously probed, practically speaking, there are currently no methods to screen which metabolic pathways are amenable to RNA aptamer biosensing, nor to assess which metabolic intermediates make better targets within a pathway (**Fig. 2.1c**) (86–88). Furthermore, for applications of aptamers in medical diagnostics, there is no screening approach to identify biomarkers that are more amenable to detection through RNA aptamer biosensing. *De novo* generation of RNA biosensors for these applications is thus undertaken blind to the likely utility of such work. Therefore, generating *a priori* estimates for aptamer binding affinities to small molecule targets will make possible *in silico* virtual screening to evaluate the suitability of RNA aptamer-based biosensors for a given application. To address this gap, here we present, to our knowledge, the first approach for estimating binding affinities for RNA aptamers against small molecule targets *a priori*. We leverage prior work on group contribution methods, which express thermodynamic properties of small molecules as an aggregate of thermodynamic contributions from their parts, to build a functional group-based approach (89–91). Existing work has highlighted relationships between RNA aptamer binding affinity and small molecule target properties such as molecular weight and rotational degrees of freedom (88). We used existing small molecule-RNA aptamer pairs and their characterized binding affinities to develop a group contribution-based method reached through a supervised learning approach with feature selection. With a thermodynamic model of aptamer-small molecule interactions, we can make *a priori* predictions of RNA aptamer-small molecule ligand binding affinities for virtual screening. By coupling binding affinity predictions with existing knowledge of small molecule concentrations necessary for intended applications, we introduce the Metabolite Utility Factor: a predictive, quantitative measure of the usefulness of RNA aptamers for biosensing applications.

Machine learning is increasingly used in synthetic biology to drive systems-level engineering through data-enabled DBTL cycles (92). Machine learning has been used to optimize riboswitch performance (93), but it has otherwise been generally underappreciated that machine learning approaches could be used to inform experimental choice at several different levels, including in identifying whether parts-level engineering goals are feasible. As there now exists a non-trivial number of RNA aptamers for small molecules generated by *in vitro* selection, we reasoned that it should be possible to generate quantitative *a priori* estimations of aptamer-small molecule energetics. If so, this

would permit virtual screens of industrially- and medically important metabolites, providing a tool that can be refined and improved as the number of aptamer-small molecule ligand pairs increases. More generally, this work shows that there is utility in applying machine learning to drive the choice of experimental efforts, allowing us to identify aptamer engineering efforts that are more likely to succeed. Using these strategies, the suitability of RNA aptamer biosensing for a given application in metabolic engineering or medical diagnostics can be assessed before committing to a time-consuming *in vitro* selection.

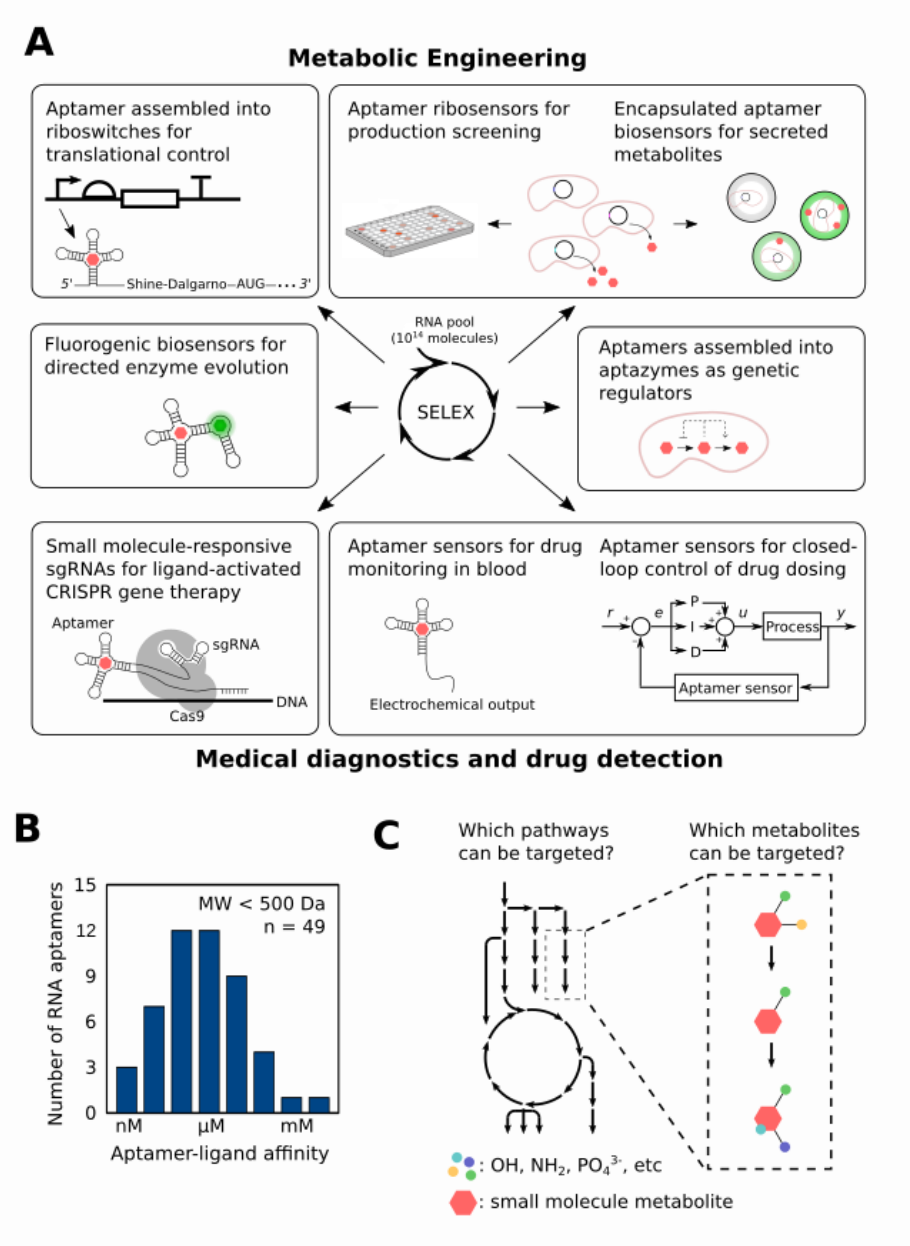


Figure 2.1. RNA aptamer biosensors for metabolic engineering and medical diagnostics. **A**. *In vitro* selected RNA aptamers have been incorporated into biosensors for a wide array of applications for

metabolic engineering, medical diagnostics, and drug detection. **B**. Binding affinities, as measured by aptamer-ligand dissociation constant (K_D) for the 49 *in vitro* selected RNA aptamer-small molecule pairs in the literature, span 7 orders of magnitude. This widespread in binding affinities leaves uncertainty in outcome of any new aptamer selection. **C**. Virtual screening approaches that predict aptamer-ligand Kds for candidate metabolites can identify small molecule targets suitable for RNA aptamer biosensing.

2.2. Methods

2.2.1. Modeling interactions between RNA aptamers and small molecules

Dissociation constants for RNA aptamer-small molecule binding can be interpreted as an equilibrium constant useful for determining the Gibbs free energy of reaction between the two species, where:

$$\Delta G_{bind} = -RTln(K_D) \qquad \text{Eq. 1}$$

Here, we assumed this binding energy to be primarily a property of the small molecule and that contributions to this binding energy correlated linearly with the chemical characteristics of small molecule targets. This allowed us to adapt a group contribution approach towards building a linear regression model to estimate the binding energies of small molecules for RNA aptamers in which contributions towards electrostatic interactions are assigned exclusively to the small molecule targets.

In order to construct a group contribution method, we made three major assumptions about the relationship between RNA aptamers and their small molecule targets. The first assumption was that the binding energy between an RNA aptamer and a small molecule was mostly dependent on the small molecule and was derived from the characteristics of a small molecule that would allow the molecule to form electrostatic interactions with RNA. Based on this assumption, we built a linear regression model that assigns the contributions towards these electrostatic interactions exclusively to the small molecule targets. Secondly, we assume that the dissociation constants reported for each aptamer of our data set reflect binding energies of the small molecules for RNA sequences that are likely to occur in RNA libraries of diversity less than 10^{15}. Thirdly, we assume an additive relationship between thermodynamic contributions of individual functional groups in a molecule and the overall binding energy of a molecule, an assumption that underlies most group contribution methods (89–91)

2.2.2. Collate a dataset to train the model

To date, about 60 small molecules possess characterized cognate RNA aptamers generated by *in vitro* selection (10, 94, 95). Reasoning that individual functional groups may no longer provide direct contributions to binding energy in larger small molecules, we excluded all molecules greater than 500 Da from this aptamer data set. Remaining small molecules in the set were then evaluated for outlier functional groups; those molecules with unique functional groups were removed as well. This left a dataset of 31 small molecules with characterized RNA aptamers (**S.I. Table 1**). Characterized dissociation constants between RNA aptamer-small molecule pairs were converted to binding energies with a blanket assumption of T = 298 K; for each small molecule, the RNA aptamer with the tightest reported binding affinity (lowest dissociation constant, K_D) was used to estimate binding energy.

Previous group contribution methods have addressed the identities of functional groups with varying levels of granularity. Some group contribution methods consider as many as 70 different functional groups by, for example, treating a secondary carbon and a secondary carbon in a ring as two different features (96). Other group contribution approaches establish binned functional groups, such as considering all sp^3 carbons as a single feature (91). High granularity approaches can distinguish among the varying effects of related but different functional groups (primary amines, for instance, might have a contribution quite different from that of secondary amines). Due to the relative paucity of small molecule RNA aptamers, we avoided such high granularity approaches to prevent overfitting. We adopted a functional group binning strategy similar to that presented in Andrews et al. that resulted in nine separate classes of functional groups (91). Prior work identified correlations between molecular weight and binding affinity and suggested a negative relationship between rotatable bonds and binding affinity, while biochemical analysis of aptamer-target crystal structures suggested a major role for stacking interactions in aptamer-target binding (88, 97). Accordingly, molecular weight, rotatable bonds, and the presence of aromatic rings were introduced as additional features in the analysis performed here that were not presented in Andrews et al.

Not considering every feature meant that we would invariably ignore actual contributions from some functional groups. To work around this, we introduced molecular weight as a variable into our model. We assumed that each functional group would contribute to the overall binding energy of a small molecule in several ways and that beyond polarity and charge, the mass of each group was important as something to bind to. The molecular weight term that we introduced would account for the mass contribution of all the groups in each molecule, thus requiring us only to account for binding energy contributions from polarity and charge.

2.2.3. Supervised learning for predicting aptamer binding energy towards small molecules

Generating the five-feature aptamer-ligand KD prediction model

Models were constructed using Python 3.7.3 and the Pandas 0.24.2 dataframe. SMILES strings for each of the 31 small molecule ligands in the training set of in vitro selected RNA aptamers were obtained from PubChem. Rdkit version 2019.03.1 and Pandas were used to generate a matrix of the 12 functional group features by enumerating the following features in each small molecule ligand: presence of an aromatic ring, molecular weight, nitrogens, phosphates, rotatable bonds, sulfurs, sp^2 hybridized carbons, sp^3 hybridized carbons, hydroxyls, carboxyls, carbonyls, and ethers. The parametric model for $\triangle G_{bind}$ was analyzed via multiple linear regression with cross validation of the functional group feature matrix onto the measured aptamer-ligand binding free energies. To estimate the root mean square error (RMSE), or difference between the model-predicted aptamer-ligand binding energies and the measured binding energies, for each combination of features, a repeated random subsampling (RSS) cross validation was performed with 100 randomly drawn training and validation set pairs of aptamers and ligands. In each run, 70% of the aptamer-ligand pairs were used for training and 30% for testing. The RMSE was taken as the jackknifed prediction spread for each model calculated from the 100 RSS runs. 4095 feature combinations covering all possible combinations of the 12 features were sorted and ranked by RMSE and Bayesian Information Criterion (BIC). The BIC provides a measure of model fit that penalizes the number of parameters in the model and was evaluated here to minimize overfitting (98). The model with the lowest sum of the RMSE and BIC ranks was selected as the five-feature model described in the main text (**Fig. 2.2a**). Additionally, each of the 12 features was assessed based on the distribution of ranks of models containing the given feature to determine which features were more strongly associated with low prediction variance.

Generating MUF predictions for metabolic engineering

53 bio-based molecules and production titers were hand curated from the metabolic map assembled by Lee et al. (99). SMILES strings for each of the 53 molecules were obtained from PubChem. Rdkit and Pandas were used to generate the functional group matrix from SMILES strings with the 12 features. The selected five feature parametric model was used to calculate the free binding energy $\triangle G_{bind}$ and K_Ds for all the molecules above. MUFs were calculated using the highest reported titers in the assembled metabolic map.

Generating MUF predictions for medical diagnostic targets

To demonstrate virtual screening of potential biomarkers for aptasening, we selected serum and urine metabolites from the HMDB version 4.0. The XML files were scrapped using ElementTree XML API and yielded 25372 serum and 4237 urine metabolites. Normal condition titers for serum metabolites were calculated by taking the mean of the decadic logarithm of the reported and quantified blood biospecimen titers in the database. Similarly, normal condition titers for urine metabolites were calculated firstly by multiplying the reported creatinine normalized urine titers by conversion factor 10.4 mmol creatinine (100) and secondly, taking mean of the decadic logarithm of the reported and quantified urine biospecimen titers in the database. Further, preprocessing for relevant molecules fitting the model criterion such as molecular weight < 500 Da and functional group was performed, along with removing missing data. SMILES strings for both serum and urine metabolites were generated using InChI strings from Pubchem pub REST API since reported SMILES in HMDB were not coherent. After preprocessing, we ended up with 3100 and 3144 serum and urine metabolites and we calculated the $\triangle G_{bind}$ and K_Ds using the five-feature parametric model. Since all detected molecules are not quantified for their concentrations in different biospecimens, only 974 and 825 serum and urine metabolites reported have quantified normal condition concentrations and we calculated MUFs for these metabolites only. Abnormal condition MUFs were only calculated for molecules whose normal condition MUFs were >2.

Similarly, to demonstrate our virtual screening approach in metabolomics for potential small molecule biomarkers that are related to therapeutic treatment and that provide insights into biochemical pathways relating to different diseases, we selected metabolites from Worldwide Clinical Trials Bioanalytical Sciences Biomarker methods. The titers for the same were obtained from HMDB version 4.0, by taking the mean of the decadic logarithm of the reported and quantified biospecimen as per the biomarker assay methods. We calculated the $\triangle G_{bind}$ and K_Ds using the five-feature parametric model and MUFs with the above-mentioned formula.

Generating DNA aptamer-ligand K_D predictions

SMILES strings for each of the 13 small molecule ligands in a set of in vitro selected DNA aptamers were obtained from PubChem. Rdkit and Pandas were used to generate a matrix of the 12 functional group features by enumerating the features in each small molecule ligand: presence of an aromatic ring, molecular weight, nitrogens, phosphates, rotatable bonds, sulfurs, sp^2 hybridized carbons, sp^3 hybridized carbons, hydroxyls, carboxyls, carbonyls, and ethers. The selected five feature parametric model was used to calculate the free binding energy $\triangle G_{bind}$ and K_Ds.

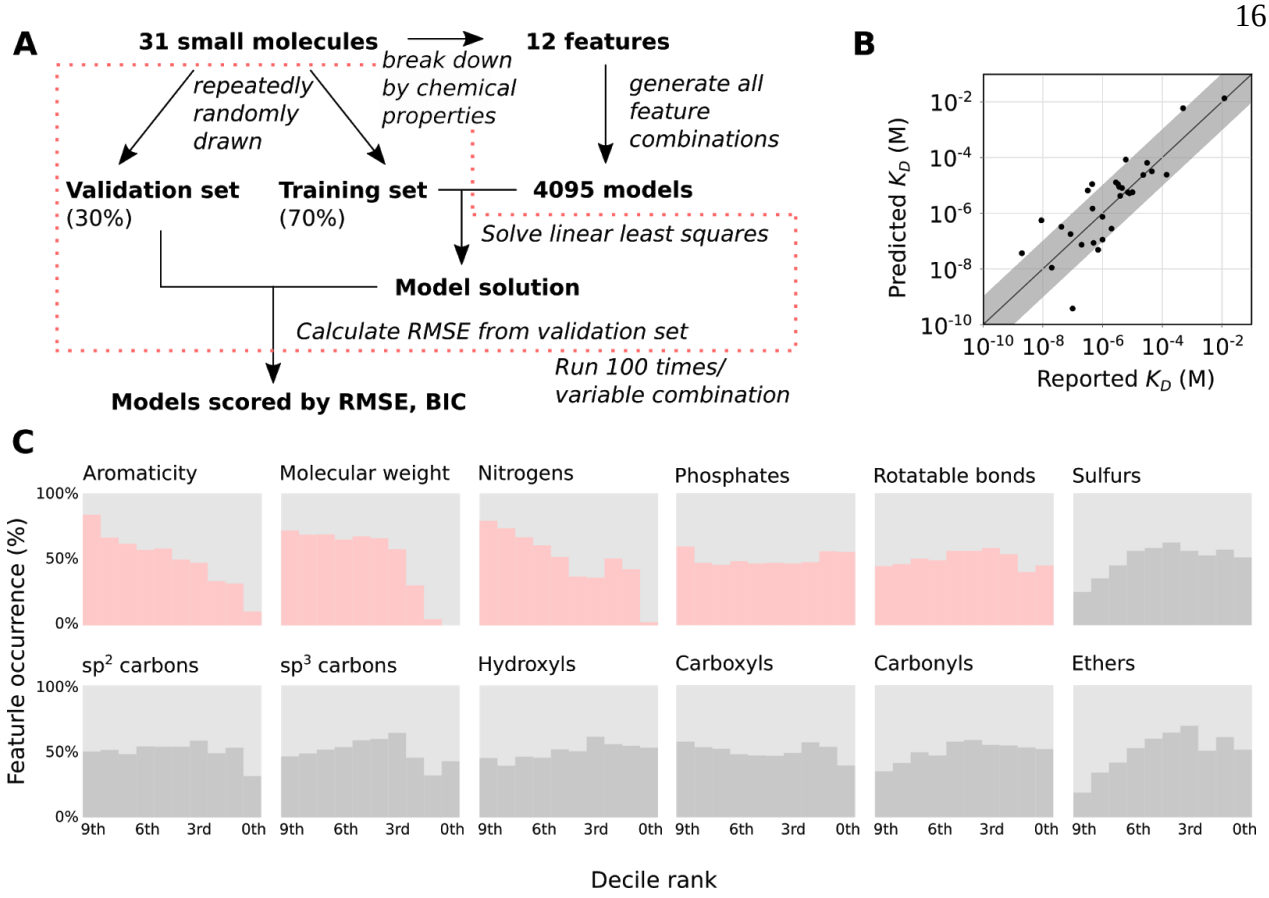


Figure 2.2 Generating a model to predict aptamer-ligand K_Ds. **A**. A supervised learning approach employing multiple regression analysis and cross-validation was formulated to generate a model for predicting aptamer-ligand K_Ds. Briefly, 31 aptamer-ligand pairs were described using 12 functional group features. Multiple regression models of measured aptamer-ligand K_Ds onto the 12 functional group features were then generated. For each combination of features, the difference between the model-predicted aptamer-ligand binding energies and the measured binding energies was estimated as the root mean square error (RMSE). The five-feature model with the lowest sum of ranks for the RMSE and the Bayesian Information Criterion (BIC, a measure of fit that penalizes overfitting) was utilized throughout the rest of this work. **B**. Plot of aptamer-ligand K_Ds predicted by the five-feature model versus the measured aptamer-ligand K_Ds. The diagonal illustrates a one-to-one correspondence; the gray shading indicates K_D values within one order of magnitude of the measured aptamer-ligand K_Ds. **C**. Feature distribution across all models, where each plot shows the distribution of models with a given feature by performance, with the 9th decile representing the most predictive 10% of models scored by RMSE and the 0th decile representing the least. Features in the five-feature aptamer-ligand K_D prediction model are highlighted in red.

2.3. Results and Discussion

Here, we lay out an approach for predicting the binding affinities of RNA aptamers for small molecule targets and for predicting the utility of RNA aptamer-based biosensing towards applications in metabolic engineering and medical diagnostics. We use supervised learning to identify a model that relates the binding energies, and thus the binding affinities, of RNA aptamer-small molecule interactions to the chemical makeup of the small molecule targets and demonstrate a relationship that agrees with our biochemical understanding of these interactions. We provide experimental support for this modeling approach with our reported binding affinity for the p-ACA aptamer. The uncertainty surrounding the outcomes of aptamer selections has long frustrated applications of RNA aptamer for biosensing; this is the first work to lay out predictions to address those uncertainties.

2.3.1. Supervised learning for predicting aptamer-ligand binding affinities

Exhaustive feature selection with repeated random subsampling cross validation in the space of multiple linear regression models yielded a five-feature model for the binding energy ΔG_{bind} of the small molecule with the form:

$$\Delta G_{bind} = 2.71A + 0.011MW + 0.59n_N - 2.49n_{PO_4^{3-}} + 0.33n_{RB} \qquad \text{Eq. 2}$$

where A indicates the presence of aromatic rings, MW is the molecular weight, n_N is the nitrogen count, n_{PO4} is the phosphate count, n_{RB} is the rotatable bonds count. We applied the model to the dataset of RNA aptamers and found that nearly three-quarters of the aptamers had characterized binding affinities within an order of magnitude of their predicted K_D (**Fig. 2.2b**). Through bootstrapping, we found the standard deviation of predictions to be 1.15 kcal/mol, corresponding to a 5-fold difference in aptamer-ligand K_D at 37°C.

The features selected in this model through supervised learning largely comport with biochemical intuition about the energetic forces driving RNA aptamer-small molecule interactions (**Fig. 2.2c**). Aromaticity was identified as a large contributor (2.7 kcal/mol) to binding energy, which is consistent with the role that pi stacking is known to play in RNA tertiary structure stability and in the observed interactions of aromatic molecules such as theophylline with the cognate aptamer (97). The calculated thermodynamic contribution of an aromatic ring is consistent with reported thermodynamic contributions of base stacking to RNA stability of 1.5 to 2.7 kcal/mol (101). Nitrogen makes large contributions to binding energy in this model, which is unsurprising given the propensity for hydrogen bonding by amines. Phosphates, which carry a negative charge in the range of physiological pH, have

an inverse relation to small molecule binding energy, which is consistent with the expectation that RNA, which is polyanionic, will tend to exhibit less favorable binding to negatively charged targets. The positive relationship between molecular weight and binding energy has been noted in the literature and is further corroborated here (88).

The positive binding energy contribution from rotatable bonds is the most surprising result from feature selection. We had expected to see a negative contribution from rotatable bonds, as increased degrees of freedom in a small molecule should correspond to an increase in conformational entropy, which would be a penalty on binding free energy (91). Assessments of small molecule-protein interactions suggest, however, that high conformational entropy and strain energy associated with multiple rotatable bonds are not correlated with lower binding affinities (102). While an increased number of rotatable bonds increases conformational entropy, our work suggests that it is offset, potentially by allowing for flexible molecules to make more contacts for association. Additionally, ligand binding incurs strain energy in proteins; as RNA has only four bases and can assume limited geometries relative to proteins, ligand flexibility might lead to lower strain energy binding conformations for RNA aptamers (103).

The model should be applicable to any molecules that are 1) composed of the 12 functional group features included in the model training set, and 2) smaller than 500 Da, as the binding affinity predictions do not appear to extrapolate well for molecules larger than 500 Da (not shown). Such molecules include the 65 metabolites identified as targets for high-value bio-based chemical production (99), and 30 small molecule biomarkers as tools for providing insights into the biochemical pathway related to various disease conditions and gear drug development process (104). We identified 3100 of 25372 small serum and 3144 of 4237 small urine metabolites, detected and quantified molecules in the Human Metabolome Database that have molecular weights < 500 Da, although some do possess functional groups not in the training set (105). Thus, although the final model uses just five features to estimate binding energy, it may have wide utility for predicting binding affinities for the metabolic targets of aptamer biosensing, whether those targets are industrially-, or medically important chemicals.

Due to the paucity of data, no test set was reserved from the existing aptamers. However, recent work and outside publications provide three test cases for an initial evaluation of model performance. We used *in vitro* selection to generate a novel aptamer to bind *p*-aminocinnamic acid (*p*-ACA), and then compared the predicted and measured binding affinities. *p*-ACA is a bio-based aromatic that can be produced using engineered microbes and polymerized into polyimides and polyamides functional as ‘super-engineered’ thermoplastics (82, 106). *p*-ACA can also be converted

into the vinyl aromatic *p*-aminostyrene and formed into optically- and mechanically-polymers for industry and medicine (107). *p*-ACA is 163 Da and contains a diverse set of functional groups, including a carboxylic acid, a vinyl group, an aromatic ring, and a primary amine. The measured aptamer-p-ACA KD of 29 ± 6 μM corresponds remarkably well with the 40 ± 32 μM KD predicted by the model. A synthetic naringenin riboswitch was recently reported with an in vitro selected naringenin RNA aptamer; although the K_D of the aptamer was not reported, the EC_{50} of the device *in vivo* in *E. coli* was reported between 100-200 μM (108). As the EC_{50} of aptamer-based devices is expected to be at a higher concentration than the K_D of the aptamer, the measured EC_{50} of the naringenin devices is consistent with the model prediction of a 37 μM RNA aptamer from an *in vitro* selection against naringenin. An RNA aptamer with a K_D of 220 nM was reported for benzylguanine; the model predicted a K_D of 260 nM (109).

2.3.2. Defining a Metabolite Utility Factor

Predictions of aptamer binding affinity alone are necessary, but not sufficient, for determining the utility of an RNA aptamer biosensor for a given metabolic engineering or diagnostic application. First, the usable dynamic range of ligand concentrations where an aptamer biosensor is sensitive depends on both the aptamer-ligand K_D, and the architecture of the biosensor (110). Second, the aptamer biosensor will only have utility if the usable dynamic range overlaps with the production titers of metabolic engineering target, or the biologically relevant concentrations ranges of a biomarker targeted by a medical diagnostic. We employ an RNA aptamer metabolite utility factor (MUF) to capture how the interplay of aptamer sensitivity and target molecule concentration determines whether RNA aptamer-based biosensing can be employed for a given application. We define MUF (a dimensionless value), as:

$$MUF = log_{10}(S) - log_{10}(K_D^*) \qquad \text{Eq. 3}$$

S is the concentration of the metabolite that needs to be detected for an intended application and K_D* is the model generated K_D prediction for a given small molecule. The MUF provides a measure of the difference between the sensitivity necessary for an intended application and the predicted binding affinity of an RNA aptamer for that metabolite. MUF will be positive when the predicted aptamer-ligand binding affinity (i.e., lower K_D) is tighter than what is needed, while a negative value indicates that the predicted aptamer-ligand binding affinity is insufficient for the application.

When biosensors are engineered through the assembly of an aptamer into an RNA architecture that couples ligand binding to a measurable biochemical output, the lower limit of detection is generally much higher than the aptamer-ligand K_D (Ricci et al.). For instance, a theophylline aptamer-controlled ribozyme (Aptazyme) presented by Win and Smolke exhibited a measured EC_{50} of 1 mM, even though the aptamer-ligand K_D is 400 nM. As a second example, a *p*-aminophenylalanine (*p*-AF) responsive biosensor has a measured EC_{50} of 300 µM, even though the parental aptamer has a K_D for *p*-AF of 3.6 µM (57, 84, 85, 88). Thus, simply having an aptamer K_D at the same concentration as the desired sensitivity is likely to be insufficient for building RNA biosensor devices. Here, we take a conservative view that MUF values >2, indicating that the aptamer-ligand K_D is two orders of magnitude higher than the targeted biosensor EC_{50}, reflecting a reasonable likelihood that an aptamer can be useful for a metabolite biosensing application.

2.3.3. Virtual screening of RNA aptamer biosensing applications

As discussed above, the utility of an aptamer for an intended biosensing application depends on whether it can be assembled into a sensor that is responsive at relevant metabolite concentrations. In this section, we illustrate how the five-feature aptamer-ligand K_D prediction model can be used as a tool for virtual screening to assess the suitability of RNA aptamer biosensing before committing to a time- and resource-consuming *in vitro* selection.

We first investigated the range of predicted aptamer-ligand K_Ds for key metabolic intermediates and final products in a set of four industrially- and medically-relevant small molecules, consisting of a vinyl aromatic (*p*-aminostyrene, MW=119 Da), a flavonoid (naringenin MW=272 Da), a phenylpropanoid (resveratrol, MW=228 Da), and an alkaloid (morphine, MW=285 Da) (**Fig. 2.3a**) (Halls 2008, Jiang 2005, Thodey 2014, Carothers 2013). The predicted aptamer-ligand K_Ds for the 'larger' final products naringenin, resveratrol, and morphine, are all similar, and about an order of magnitude lower than the aptamer-ligand K_D predicted for the smallest product, *p*-aminostyrene. Across the set of 12 selected pathway intermediates, the predicted aptamer-ligand K_Ds span more than three orders of magnitude, ranging from low mM for chorismate (*p*-aminostyrene biosynthesis pathway) to <10 µM for thebaine (morphine biosynthesis pathway). Even within individual pathways, the predicted aptamer-ligand K_Ds vary by more than an order of magnitude, underscoring the point that there can be significant differences in aptamer binding affinities even among molecules that have superficially similar structures.

The product titers biosynthesized from a metabolically engineered pathway increases throughout the course of a successful strain optimization project (66). Because of this, aptamer

production biosensors used at the beginning of an optimization process will need to be responsive to lower concentrations of target ligand than aptamer production biosensors used at the end of an optimization process. In this case, the maximum production titer previously observed for a metabolite provides a starting point to investigate whether aptamer biosensors can be developed to analyze production from engineered strains.

Recently, Lee and colleagues developed a comprehensive metabolic map to outline bio-synthetic routes for producing industrially-important chemicals through metabolic engineering (99). We calculated MUF values for 53 molecules derived from 8 different parts of the metabolic map as the difference between the highest reported biosynthetic titers and the aptamer-ligand K_Ds predicted by the five-feature model (**Fig. 2.3b**). The calculated MUF values were >2 for 10 of the 53 molecules (**S.I. Table 2**), indicating that these are good candidates for aptamer production biosensing because the predicted aptamer-ligand K_Ds are more than 100 times lower than the reported production titers. As a class of targets, molecules in pathways branching from D-erythrose-4-phosphate are predicted to have the greatest number of high affinity aptamer binders and generally exhibit the highest MUF values.

MUF values for 12 of the molecules from the metabolic map are negative, consistent with the idea that many metabolic engineering targets are not currently amenable to aptamer production biosensing. Notably, several of the molecules have low MUF values even though the predicted aptamer-ligand K_Ds are relatively low (i.e., the predicted aptamer-ligand affinities are high). For instance, although vanillin is predicted to have one of the tightest aptamer-ligand binding affinities among the molecules from this metabolic map at 193 μM, the calculated MUF value for vanillin is 0.19, indicating limited utility, because the reported production titers are just slightly higher at 296 μM (**S.I. Table 2**) (99). Thus, with currently achievable production titers, this analysis indicates that it may not be possible to create a usable aptamer production biosensor for vanillin. In contrast, although γ-aminobutyric acid is not expected to have a high affinity aptamer solution (predicted aptamer-ligand K_D > 1 mM), reported titers of >1 M could make aptamer biosensing useful for production analysis. Collectively, these results underscore the point that the feasibility of aptamer production biosensing depends on both the predicted aptamer-ligand K_D and the titer of the targeted biosynthetic product.

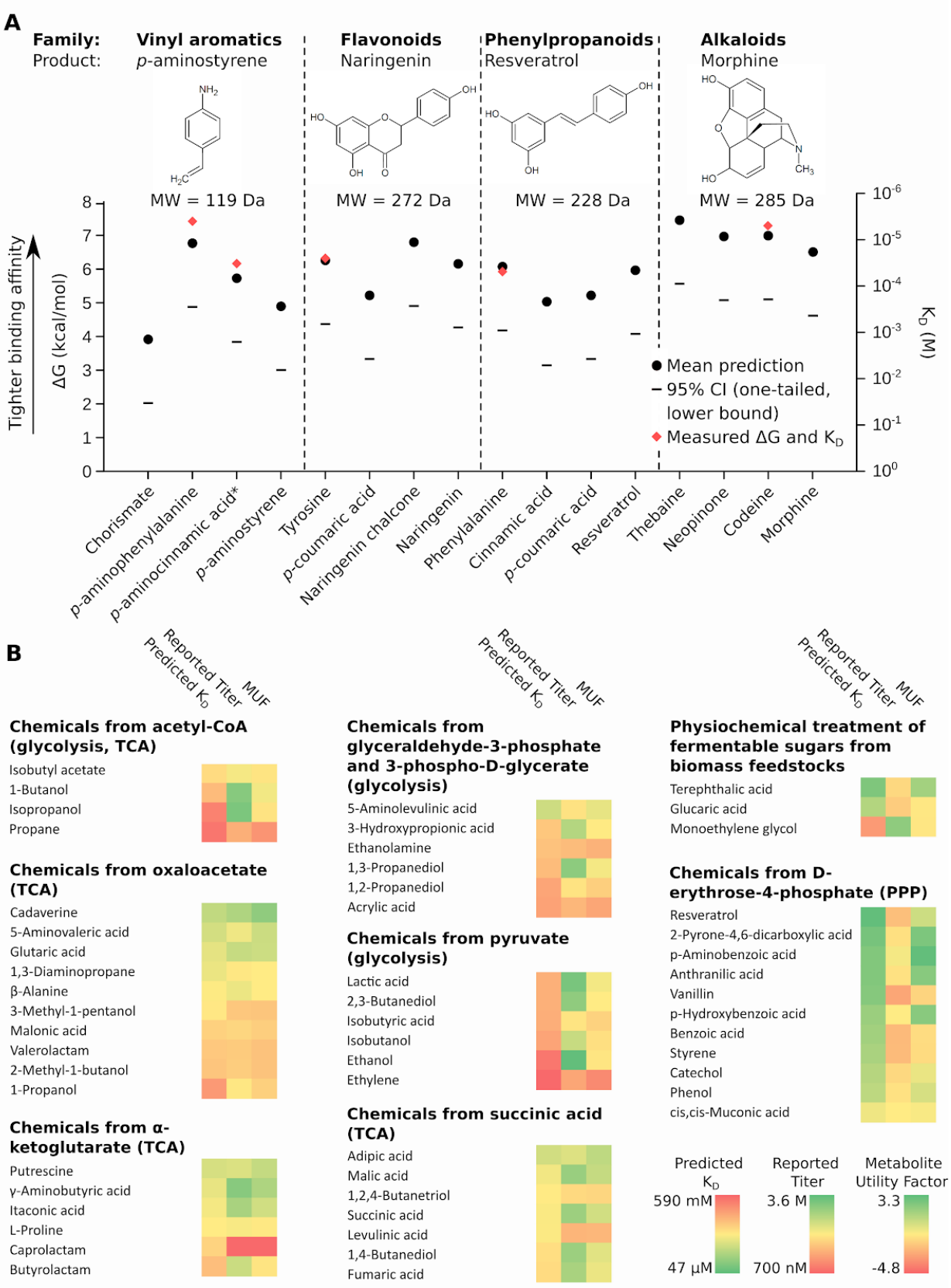


Figure 2.3. Virtual screening for RNA aptamer biosensing applications in metabolic engineering. **A.** Aptamer-ligand binding affinities predicted using the five-feature model are shown for key metabolic

intermediates and final products in four biosynthetic pathways leading to industrially- and medically-relevant small molecules. The chosen set consists of biosynthetic pathways producing a vinyl aromatic (*p*-aminostyrene, MW=119 Da), a flavonoid (naringenin MW=272 Da), a phenylpropanoid (resveratrol, MW=228 Da), and an alkaloid (morphine, MW=285 Da). The mean and one-tailed 95% confidence interval are given for each aptamer-ligand binding affinity prediction. Measured binding affinities for ligands to which aptamers have been reported are indicated with red squares.
B. Metabolite Utility Factor (MUF) values for 53 molecules derived from 8 different parts of a comprehensive metabolic map of biosynthetic products (99), were calculated as the $\log_{10}$ difference between the highest reported biosynthetic titers and the aptamer-ligand K_ds predicted by the five feature model.

A similar approach to virtual screening can be applied to identify small molecules that can be detected by aptamer biosensors for medical diagnostics. We demonstrated this approach by assessing the human serum and urine metabolome and the associated diseases available at the Human Metabolome Database (HMDB) (105). HMDB as of version 4.0 consists of 114,224 metabolites providing a comprehensive mapping of human metabolites with their biological roles, physiological concentrations, disease associations, chemical reactions, metabolic pathways, and reference spectra. We selected serum and urine metabolites as they are comparatively easy to analyze and are collected noninvasively and have shown strong precedence for use as point-of care (POC) diagnostics for quantitative or semiquantitative monitoring of different human metabolites (111). The serum metabolome and urine metabolome consist of 25372 and 4237 molecules, respectively. A panel of 3100 small serum metabolites and 3144 small urine metabolites fitting the criteria of MW < 500 and relevant functional groups were screened for further analysis to predict aptamer-ligand K_D using our five-feature model. The median K_D for serum and urine metabolites are 3 μM and 6.6 μM and the interquartile range is 94 μM and 174 μM respectively. Our predictions on serum metabolites tell us that 15.6% of 3100 have $K_D \leq 50$ nM, 32.6 % have $K_D \leq 500$ nM implying very strong targets for aptamer-ligand detection, 59.4% have $K_D \leq 100$ μM suggesting high affinity aptamer-solutions. Similarly, for urine metabolites, 13.8% of 3144 have $K_D \leq 50$ nM, 29.4 % have $K_D \leq 500$ nM and 53.2% have $K_D \leq 100$ μM (**Fig.2.4a)**. We further analyzed these molecules by their superclass and observed that nucleosides, nucleotides, and analogues show greater potential for aptasensing and organic oxygen compounds have poor aptamer-ligand binding affinities consistent with our biochemical understanding of these interactions (**S.I. Figs 1a & 1b**).

From 3100 serum metabolites and 3144 urine metabolites we calculated the MUFs using the titers for 974 serum and 825 urine metabolites (**S.I. Figs 1a & 1b**) quantified and reported in the

HMDB database. We found 166 and 9 positive MUFs for normal conditions in serum and urine, respectively, out of which 52 and 1 are MUFs > 2. We selected the molecules with MUFs > 2 in normal conditions for calculation of MUFs in disease conditions. This analysis showed there are 16 metabolites having positive MUFs under normal and abnormal conditions, associated with a variety of disease conditions, suggesting that aptamer biosensors could be engineered to detect and differentiate healthy and pathological conditions. For instance, Anserine has a MUF of 2.8 under normal conditions but an MUF of 0.8 in Alzheimer's disease, while the MUF of nervonic acid is 2.7 under normal conditions but is 4.01 in isovaleric acidemia, suggesting that RNA aptasensors would be well suited for diagnostic applications in these cases. Conditions such as *Obesity* (MUFs: trans-2-Dodecenoylcarnitine = 2.13, Oleoylcarnitine = 2.08, 2-Octenoylcarnitine = 1.97), *Pregnancy* (MUFs: LysoPC(16:0) = 4.23, LysoPC(16:1(9Z)/0:0) = 2.45, Octadecenoylcarnitine = 2.26), *Schizophrenia* (MUFs, LysoPC(16:0) = 4.23, Gamma-Tocopherol = 2.89, Oleic acid = 2.16, Linoleic acid = 2.10) (**S.I. Table 3**) can be diagnosed by integrating aptasensors in various multiplex detection platforms for the simultaneous, but independent, detection of more than one analyte in a single assay. Multiplexing has improved the overall specificity of detection by enabling the user to see a panel of test results altogether (112, 113).

To further corroborate our approach, we took real time biomarker assay methods from Worldwide Clinical trials to understand which small molecule can be readily integrated into the pipeline for aptasening. We calculated the MUFs for 30 small molecule analytes (MW < 500 Da) used to understand clinical conditions using the predicted aptamer-ligand K_D with the five-feature model and concentrations of the markers in different biospecimens available in HMDB (104). Metabolites such as Serotonin (biospecimen = plasma, K_D = 16.7 μM, MUF = -1.42), 5-Hydroxyindoleacetic acid (5-HIAA) (biospecimen = urine, K_D = 34.3 μM, MUF = -3.02) are used as biomarker for Parkinson and related disorders and 5-HIAA specific for intestinal carcinoid disorder (114–117), Hippuric acid (biospecimen = plasma, K_D = 24.4 μM, MUF = -0.34) for environmental exposure (118), Kynurenine (KYN) (biospecimen = plasma, K_D = 2.98 μM, MUF = -0.15) for Indoleamine 2,3-Dioxygenase (IDO) activation (immune stress, inflammation, cancer). Most of these biomarker molecules have negative MUF values suggesting a vast majority of human metabolites might not be amenable for aptasensing

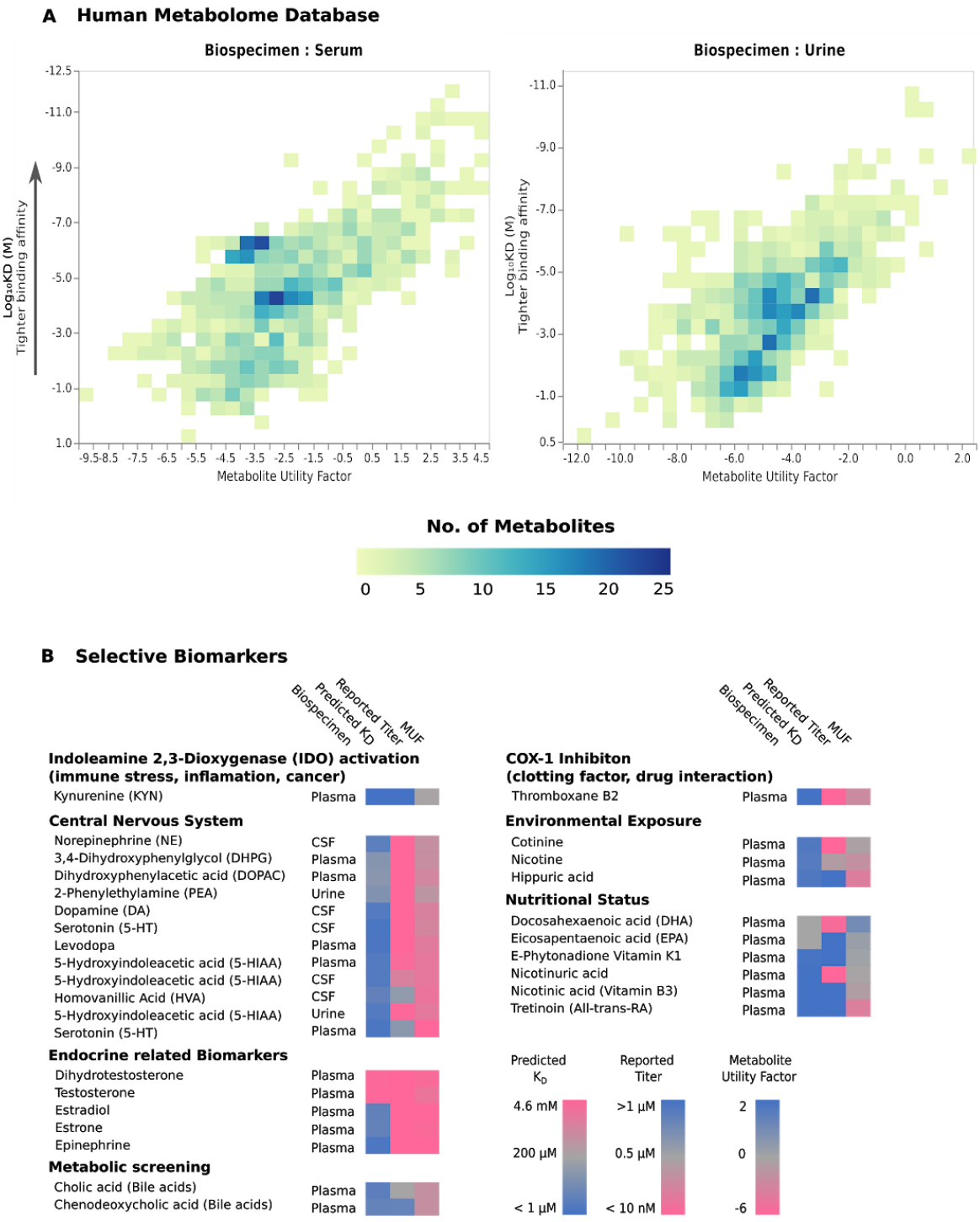


Figure 2.4. Virtual screening for RNA aptamer biosensing applications in medical diagnostics.

A. Heat map between Metabolic Utility Factor (MUF) and decadic logarithm of the aptamer-ligand binding affinities ($\log_{10}K_D$) predicted using the five-feature model from the available normal concentrations in the HMDB database serum and urine metabolites. Screening molecules fitting the

model criteria and having detected and quantified normal concentrations, we arrived finally with 974 and 825 serum and urine metabolites. **B**. Metabolite Utility Factor (MUF) values for 30 small molecule biomarkers useful for various drug development programs and diagnosis of a condition (104), were calculated as the $\log_{10}$ difference between the highest reported concentration titers in HMDB and the aptamer-ligand K_Ds predicted by the five-feature model.

---

due to their relative lower titer values consistent with the analysis of HMDB metabolites even though the predicted aptamer-ligand binding affinities are high (**Fig.2.4b**). We did find a cluster of biomarkers with positive MUF values that can be targeted to understand the nutritional state of an individual. Analytes such as E-Phytonadione Vitamin K1 (MUF = 0.13), Eicosapentaenoic acid (EPA) (MUF = 0.27) and Docosahexaenoic acid (DHA)(MUF = 1.04) may be detectable using aptasening (**S.I. Table 4)**.Thus, as with aptamer biosensing, the low predicted utility of RNA aptamer biosensing towards many metabolites and biomarkers and the high predicted utility of aptamer biosensing toward other analytes -- demonstrates how virtual screening could be employed to focus aptamer biosensor engineering efforts toward appropriate molecular targets.

2.3.4. Investigating uncertainty in the aptamer-ligand K_D prediction model

The aptamer dataset available to train the aptamer-ligand K_D prediction model was shaped by a number of experimental factors that could confound the association between aptamer-ligand binding energies and the aggregated contributions of the ligand functional groups. The aptamers in the training set were generated through *in vitro* selections carried out under varied experimental conditions, using combinatorial sequence libraries with different lengths and sequence diversities. The process of optimizing *in vitro* selected aptamers is not standardized, and often not performed, meaning that some of the reported aptamer-ligand K_Ds may be significantly suboptimal (86). Finally, measured aptamer-ligand K_Ds have been seen to vary as much as 10-fold, depending on the binding affinity characterization method employed (119).

We investigated the impact of these experimental factors on the aptamer-ligand K_D prediction model by computing the correlation between the residuals of the predicted and reported aptamer binding energies (ΔG_{bind} difference). First, we examined whether the length of the randomized region of a starting pool was correlated with ΔG_{bind} difference; for randomized regions varying from 25 to 120 nucleotides, no trend was observed (**Fig. 2.5a**). Next, we investigated whether the diversity of the

initial selection pool was correlated with the residuals between the predicted and reported aptamer-ligand binding energies. Because pools with more sequence diversity are more likely to contain rare, high-affinity solutions (86), we hypothesized that there could be a positive correlation between pool size and aptamers outperforming the model. Somewhat surprisingly, there was no observable trend between these two variables across the set of aptamers with initial pool diversities varying between 10^{12} and $5x10^{15}$ unique molecules (**Fig. 2.5b**). We also compared the performance of structured to unstructured RNA aptamer pools: seeding structures such as a stem-tetraloop within the randomized pool is a strategy expected to improve binding affinities (120). Only three of the selections considered in the set utilized this strategy and all three exhibited tighter than predicted binding affinities, although none of these differences were statistically significant.

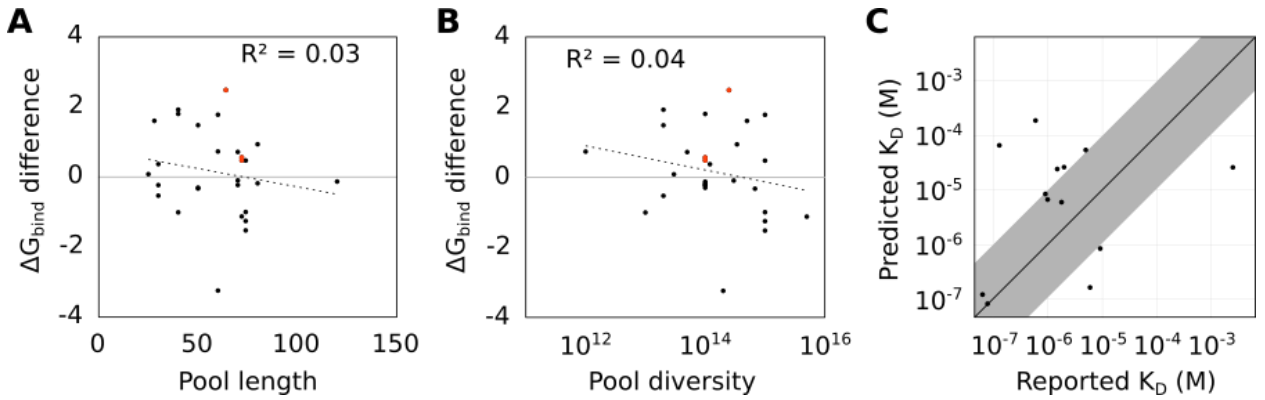


Figure 2.5. Impact of *in vitro* selection experimental designs on the aptamer-ligand K_D prediction model. The impact of variations in the experimental designs of the *in vitro* selections used to generate the aptamers on the aptamer-ligand K_D prediction model were evaluated by computing the correlation between the residuals of the predicted and reported aptamer binding energies as the ΔG_{bind} difference. **A**. ΔG_{bind} difference is plotted against the length of the randomized region of the starting *in vitro* selection pools (in nucleotides). **B**. ΔG_{bind} difference is plotted against the diversity of the starting *in vitro* selection pools (in the number of unique molecules). **C**. The five-feature aptamer-ligand K_D prediction model trained on RNA aptamers was used to predict DNA aptamer-ligand K_Ds. The predicted DNA aptamer-ligand K_Ds are plotted against the reported DNA aptamer-ligand K_Ds. Grey shading indicates K_D values within one order of magnitude of reported DNA aptamer-ligand K_Ds.

As a final analysis, we investigated whether the RNA aptamer-ligand K_D prediction model could be applied to predict the affinities of small-molecule binding DNA aptamers. We generated predictions for 13 small molecules that have existing DNA aptamers and found that, on average, the

RNA model tended to slightly underestimate binding affinity for DNA aptamers. Small molecule DNA aptamers exhibited a mean binding energy that was 0.7 kcal/mol higher than the predictions from the RNA model, which remains within the uncertainty of the model (**Fig. 2.5c**). It may be that the functional group contributions to DNA aptamer-ligand binding are quantitatively different from the functional group contributions to RNA aptamer-ligand binding. Intriguingly, differences in the observed binding affinities between RNA and DNA aptamers that bind the same small molecule are consistent with this view that contributions to the binding energy may carry different weights for DNA compared to RNA.

As the number of RNA and DNA aptamers for small molecule targets increases, supervised learning could be employed to generate more refined aptamer-ligand K_D prediction models. In the same vein, standardizing methods for aptamer isolation, optimization, and binding characterization would be expected to reduce uncertainty in the training data and further improve the prediction models.

2.4. Conclusions and Future directions

We have demonstrated here an approach for estimating binding affinities of RNA aptamers for small molecules *a priori* based on the properties of the small molecule. However, many applications of RNA aptamers depend not just on the dissociation constant of the aptamer, but on the binding kinetics- the kinetic aptamer ribosensors discussed in this work are one such application. Ultimately, this work would benefit from providing predictions for aptamer binding kinetics- especially association kinetics- in addition to thermodynamic dissociation constants.

The principal challenge to this undertaking is the lack of a large set of characterized small molecule aptamer kinetics. Most studies reporting newly selected aptamers merely report the K_D, characterized through equilibrium approaches such as in-line probing or equilibrium filtration (88, 121). A small number of small molecule RNA aptamers have had their kinetics characterized by SPR (122); otherwise there has been no systematic work for characterizing aptamer binding kinetics. To apply a similar modeling approach to predict small molecule aptamer association kinetics, there must be a systematic survey to characterize RNA aptamer binding kinetics; this would likely be best served by using SPR, which is the state of the art for characterizing RNA aptamer kinetics. Once this data is collected, a similar approach to the one detailed here can be used to calculate a regression model to predict aptamer association rates from small molecule chemical properties.

There are also many other applications of this generalized K_D prediction modeling approach. Artificial nucleic acid aptamers are an emerging field that takes advantage of non-standard

nucleobases or backbones to access different binding chemistries and conformations, which may allow for tighter binding or greater stability (58, 59, 123). Binding affinity modeling can help us understand whether certain artificial nucleic acid approaches can lead to tighter binding aptamers, and if so, what that binding energy contribution looks like and whether there are certain functional groups or properties that respond to given nucleic acid modifications.

2.5. Author Contributions

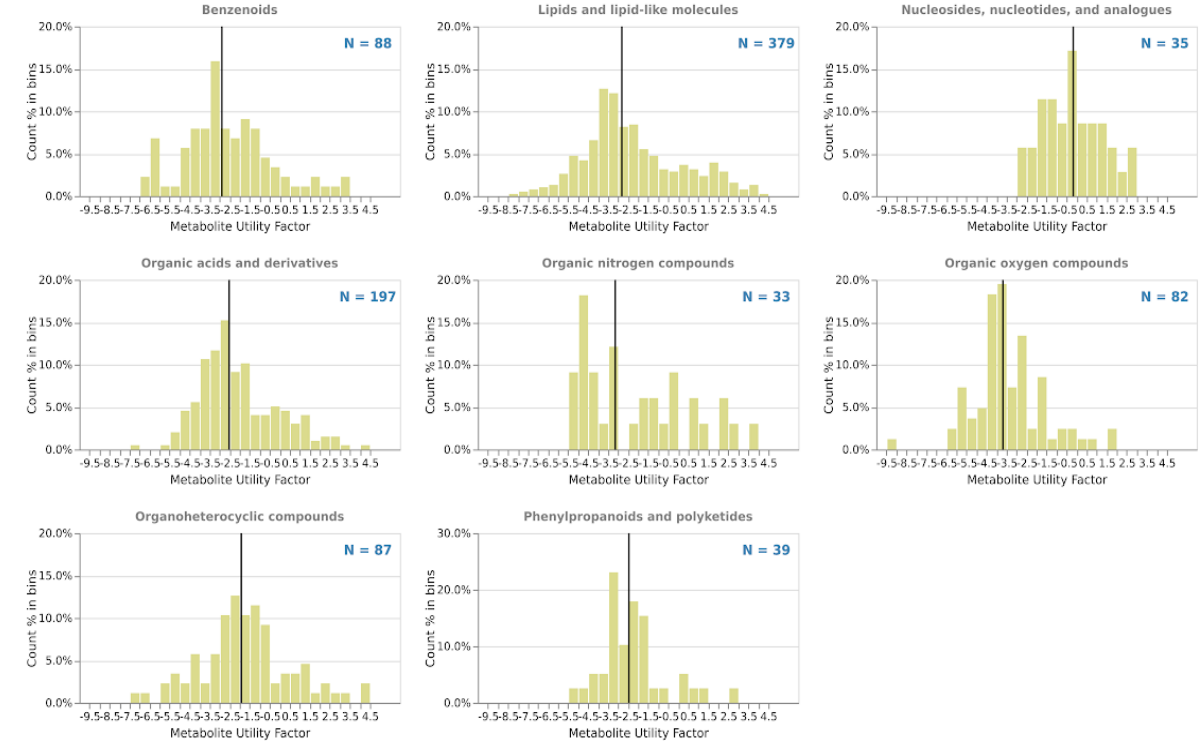
Chuhern Hwang (CH), David Sparkman-Yager (DS-Y), and James Carothers (JMC) conceived of the project. CH developed the machine learning approaches and performed the *in vitro* selection and aptamer characterization. Neel Shah (NS) conducted the metabolite virtual screening and developed the virtual screening tool. The manuscript was written with contributions of all authors. All authors have given approval to the final version of the manuscript.

2.6. Supplementary Material

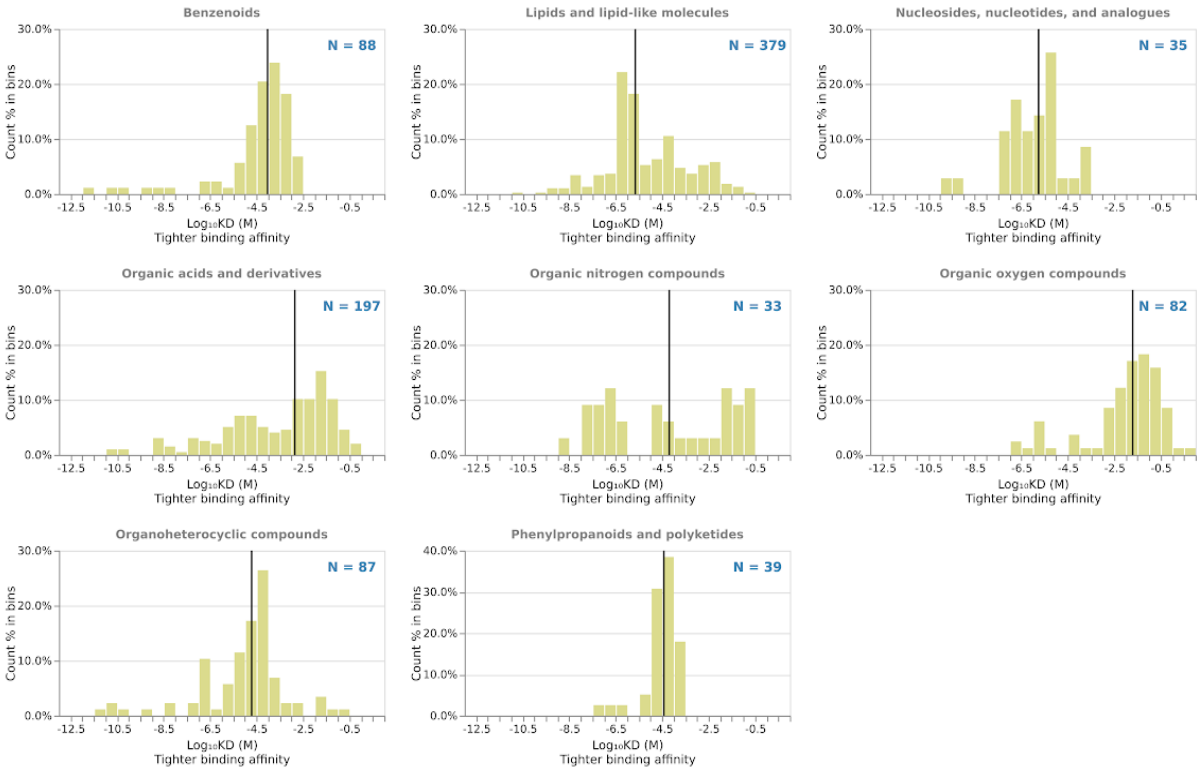
S. I. Figure 1. Distribution of MUF and tighter binding affinities of HMDB metabolites by their super classes **A**. *In silico* screening of serum metabolites resulted in 974 metabolites which have reported titers in HMDB, MUFs and tighter binding affinity ($log_{10}K_D$) were calculated and histogram plots along with median for each superclass where N > 10 is shown. **B**. *In silico* screening of urine metabolites resulted in 825 metabolites which have reported titers in HMDB, MUFs and tighter binding affinity ($log_{10}K_D$) were calculated and histogram plots along with median for each superclass where N > 10 is shown.

A Biospecimen: Blood

Metabolic Utility Factor

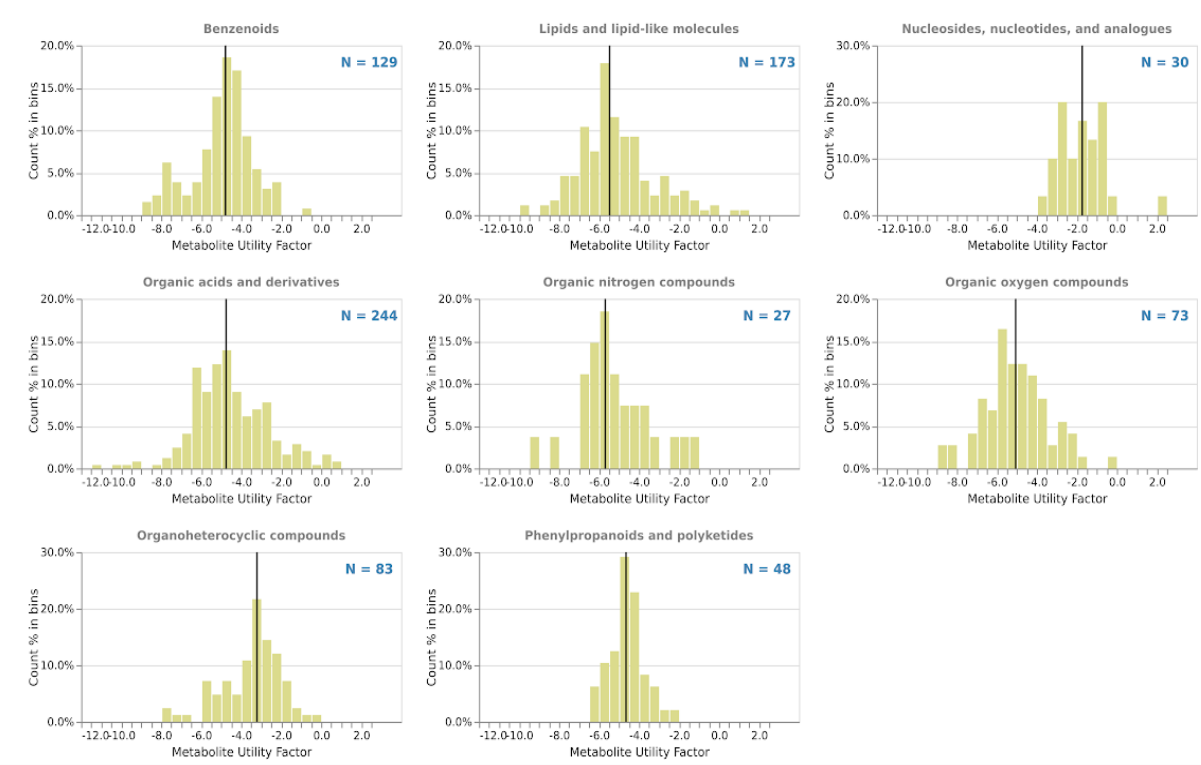


Tighter Binding Affinity

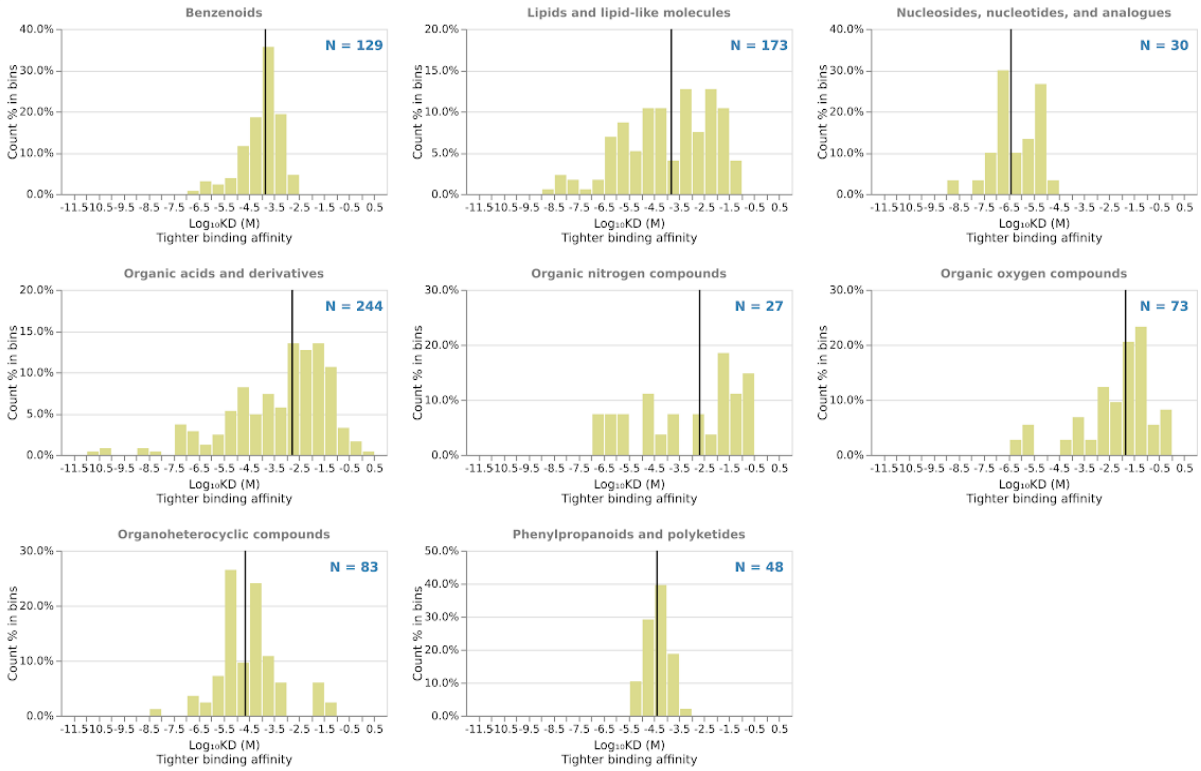


B Biospecimen: Urine

Metabolic Utility Factor



Tighter Binding Affinity



S. I. Table 1. Small molecule RNA aptamer-ligand dataset used for model training. For each molecule, we used the RNA aptamer with the tightest reported binding affinity and tallied nine different functional groups and three molecular characteristics: aromaticity, molecular weight, and the rotatable bonds count.

Molecule	K_D (M)	Rot Bonds	MW	Aromatic	sp^2	sp^3	N	OH	COOH	CO	PO_4^{2-}	S	-O-
4-aminophenylalanine	3.6E-06	3	180	1	6	2	2	0	1	0	0	0	0
Tyrosine	2.3E-05	3	181	1	6	2	1	1	1	0	0	0	0
Isoleucine	5.0E-04	3	131	0	0	5	1	0	1	0	0	0	0
Phenylalanine	4.5E-05	3	165	1	6	2	1	0	1	0	0	0	0
Theophylline	3.2E-07	0	180	1	3	2	4	0	0	2	0	0	0
Tetramethylrhodamine	8.5E-08	3	389	1	19	4	2	0	1	0	0	0	1
guanosine monophosphate	9.0E-09	4	363	1	4	5	5	2	0	1	1	0	1
adenosine monophosphate	1.0E-06	4	347	1	5	5	5	2	0	0	1	0	1
arginine	1.4E-04	6	174	0	1	4	4	0	1	0	0	0	0
codeine	4.5E-06	1	299	1	8	10	1	1	0	0	0	0	1
flavin mononucleotide	7.1E-07	7	456	1	8	7	4	3	0	2	1	0	2
xanthine	3.3E-06	0	152	1	3	0	4	0	0	2	0	0	0
L-Histidine	8.0E-06	3	155	1	3	2	3	0	1	0	0	0	0
Dopamine	2.8E-06	5	153	1	6	2	1	2	0	0	0	0	0
L-tryptophan	7.0E-06	3	204	1	8	2	2	0	1	0	0	0	0
Valine	1.2E-02	2	117	0	0	4	1	0	1	0	0	0	0
Tobramycin	2.0E-09	6	468	0	0	18	5	5	0	0	0	0	4
Kanamycin	2.0E-07	6	484	0	0	18	4	7	0	0	0	0	4
adenine	1.0E-05	0	135	1	5	0	5	0	0	0	0	0	0
Tetracycline	1.0E-06	2	444	1	10	9	2	5	0	3	0	0	0
biotin	6.0E-06	5	244	0	0	8	2	0	1	1	0	1	0
glutathione	4.2E-08	11	307	0	0	6	3	0	2	4	0	1	0
malachite green	2.0E-06	3	365	1	19	4	2	0	0	0	0	0	0
7-methylguanosine	5.0E-07	2	298	1	4	6	5	3	0	1	0	0	1
s-adenosylhomocysteine	1.0E-07	7	384	1	5	8	6	2	1	0	0	1	1
5-hydroxy-L-tryptophan	3.9E-06	3	220	1	8	2	2	1	1	0	0	0	0
cyclic adenosine monophosphate	1.0E-05	1	329	1	5	5	5	1	0	0	1	0	1
DMHBI	4.6E-07	3	276	1	9	4	2	1	0	1	0	0	1
citrulline	3.1E-05	6	175	0	0	4	3	0	1	1	0	0	0
sphingosylphosphorylcholine	2.0E-08	21	464	0	2	21	2	1	0	0	1	0	0
4,4'-methylenedianiline	4.5E-07	2	198	1	12	1	2	0	0	0	0	0	0

S.I. Table 2. Metabolite Utility Factors calculated for a panel of molecules produced through biosynthesis. For each molecule, we generate a prediction for RNA aptamer binding affinity (K_D*) using our model and comparing that affinity with the highest reported biosynthetic titers to calculate the MUF.

Molecule Name	K_D* (M)	Titer (M)	Metabolite Utility Factor
p-Aminobenzoic acid	1.66E-04	3.14E-01	3.28
2-Pyrone-4,6-dicarboxylic acid	1.06E-04	9.08E-02	2.93
Anthranilic acid	1.66E-04	1.02E-01	2.79
p-Hydroxybenzoic acid	4.41E-04	2.65E-01	2.78
Cadaverine	2.11E-03	1.02E+00	2.68
Terephthalic acid	1.49E-04	4.03E-02	2.43
γ-Aminobutyric acid	9.88E-03	1.93E+00	2.29
Adipic acid	3.88E-03	4.65E-01	2.08
Putrescine	4.82E-03	4.80E-01	2.00
Malic acid	1.50E-02	1.46E+00	1.99
5-Aminovaleric acid	4.32E-03	3.41E-01	1.90
Glutaric acid	8.86E-03	6.81E-01	1.89
Resveratrol	4.66E-05	3.56E-03	1.88
Itaconic acid	1.62E-02	1.12E+00	1.84
Succinic acid	2.03E-02	1.36E+00	1.83
Phenol	1.77E-03	1.01E-01	1.76
1,4-Butanediol	3.41E-02	1.39E+00	1.61
5-Aminolevulinic acid	3.33E-03	1.12E-01	1.53
Catechol	1.31E-03	4.06E-02	1.49
Fumaric acid	3.70E-02	1.12E+00	1.48
1-Butanol	8.10E-02	1.98E+00	1.39
Lactic acid	1.06E-01	2.50E+00	1.37
1,3-Propanediol	7.80E-02	1.77E+00	1.36
cis,cis-Muconic acid	1.29E-02	2.59E-01	1.30
2,3-Butanediol	1.06E-01	1.71E+00	1.21
β-Alanine	2.26E-02	3.63E-01	1.21

1,3-Diaminopropane	1.10E-02	1.76E-01	1.20
3-Hydroxypropionic acid	6.01E-02	9.30E-01	1.19
L-Proline	2.44E-02	2.94E-01	1.08
Glucaric acid	1.17E-03	1.23E-02	1.02
Monoethylene glycol	1.78E-01	1.74E+00	0.99
Ethanol	4.23E-01	3.60E+00	0.93
Butyrolactam	7.52E-02	6.36E-01	0.93
Isobutyl acetate	3.69E-02	3.10E-01	0.92
Isopropanol	3.26E-01	2.38E+00	0.86
Isobutanol	1.42E-01	6.75E-01	0.68
Benzoic acid	5.95E-04	2.10E-03	0.55
Styrene	8.32E-04	2.50E-03	0.48
1,2,4-Butanetriol	2.53E-02	4.81E-02	0.28
Vanillin	1.93E-04	2.96E-04	0.19
Isobutyric acid	1.10E-01	1.33E-01	0.08
1-Propanol	1.85E-01	1.80E-01	-0.01
1,2-Propanediol	1.37E-01	1.18E-01	-0.06
Malonic acid	4.63E-02	3.46E-02	-0.13
3-Methyl-1-pentanol	2.73E-02	7.77E-03	-0.55
2-Methyl-1-butanol	6.23E-02	1.42E-02	-0.64
Valerolactam	5.79E-02	1.19E-02	-0.69
Levulinic acid	2.10E-02	1.37E-03	-1.19
Ethanolamine	6.70E-02	2.60E-03	-1.41
Acrylic acid	1.48E-01	1.67E-03	-1.95
Propane	4.39E-01	7.26E-04	-2.78
Ethylene	5.92E-01	3.17E-04	-3.27
Caprolactam	4.46E-02	7.03E-07	-4.80

S.I. Table 3. Metabolite Utility Factors calculated for a panel of molecules found in serum. For each molecule, we generate a prediction for RNA aptamer binding affinity (K_D*) using our model and comparing that affinity with the highest reported titers in HMDB database for both normal and abnormal to calculate the MUF conditions and the best available molecules are presented here.

Molecule Name	K_D* (M)	Normal Metabolite Utility Factor	Abnormal Metabolite Utility Factor	Disease Condition (Biospecimen = Serum)
2-Octenoylcarnitine	1.2E-09	2.22	1.97	Obesity
alpha-Tocopherol	3.7E-09	3.82	2.99	27-hydroxylase deficiency
				Abetalipoproteinemia
				Cancer
				Colorectal adenoma
				Colorectal, breast, or neuroendocrine cancer
				Vitamin E deficiency
Anserine	7.2E-08	2.76	0.80	Alzheimer's disease
Folic acid	1.5E-11	3.14	2.58	Dimethylglycinuria
				Folate deficiency
				Hereditary folate malabsorption
				Rheumatoid arthritis
Gamma-Tocopherol	4.9E-09	2.96	2.89	Endometrial cancer
				Prostate cancer
				Schizophrenia
Hexacosanoic acid	7.9E-10	2.94	3.66	Adrenoleukodystrophy, neonatal
				Adrenomyeloneuropathy
				D-Bifunctional Protein Deficiency
				Mental retardation, enteropathy, deafness, peripheral neuropathy, ichthyosis, and keratoderma
				Neonatal adrenoleukodystrophy
				Peroxisomal biogenesis disorder
				Peroxisomal disorders, new type, liver
				Pseudo Zellweger->D-bifunctional protein deficiency

				Pseudoneonatal adrenoleukodystrophy
				Rhizomelic chondrodysplasia punctata
				X-linked adrenoleukodystrophy
				Zellweger syndrome
Linoleic acid	2.0E-06	2.10	2.10	Breast cancer
				Cirrhosis
				Colorectal cancer
				Essential hypertension
				Gestational diabetes mellitus (GDM)
				Hypertension
				Isovaleric acidemia
				Lung cancer
				Oesophageal cancer
				Schizophrenia
LysoPC(14:0/0:0)	2.8E-08	2.14	1.96	Acetaminophen overdose
				Ovarian cancer
				Pain or fever
LysoPC(16:0)	5.4E-09	4.30	4.23	Heart failure with preserved ejection fraction
				Pregnancy
				Pregnancy with fetus having congenital heart defect
				Schizophrenia
LysoPC(16:1(9Z)/0:0)	9.8E-09	2.47	2.45	Acetaminophen overdose
				Heart failure with preserved ejection fraction
				Pain or fever
				Pregnancy
				Pregnancy with fetus having congenital heart defect
Melanin	3.6E-06	3.15	3.37	Chronic renal failure
Nervonic acid	7.5E-09	2.67	4.01	Isovaleric acidemia
Octadecenoylcarnitine	1.62E-09	2.01	2.26	Glutaric aciduria II
				Heart failure with preserved ejection fraction
				Pregnancy
				Pregnancy with fetus having congenital heart defect
Oleic acid	1.07E-06	2.23	2.16	Colorectal cancer

				Gestational diabetes mellitus (GDM)
				Isovaleric acidemia
				Lung cancer
				Schizophrenia
Oleoylcarnitine	1.62E-09	2.31	2.08	Acetaminophen overdose
				Obesity
S-Adenosylhomocysteine	3.77E-10	2.58	2.42	Adenosine kinase deficiency
				Neurodegenerative diseases
S-Adenosylmethionine	2.84E-10	2.61	3.07	Adenosine kinase deficiency
				Neurodegenerative diseases
trans-2-Dodecenoylcarnitine	1.20E-09	2.18	2.13	Glutaric aciduria II
				Heart failure with preserved ejection fraction
				Obese
				Obesity

S.I. Table 4. Metabolite Utility Factors calculated for a panel of molecules used as biomarkers. For each molecule, we generate a prediction for RNA aptamer binding affinity (K_D*) using our model and comparing that affinity with the highest reported titers in HMDB database to calculate the MUF.

Molecule Name	Biospecimen	Titer (M)	K_D* (M)	Metabolite Utility Factor	Matrix
Serotonin (5-HT)	Plasma	6.39E-07	1.67E-05	-1.418	CNS
Homovanillic Acid (HVA)	CSF	6.08E-07	6.26E-05	-2.01309	CNS
5-Hydroxyindoleacetic acid (5-HIAA)	CSF	2.25E-07	3.43E-05	-2.18306	CNS
5-Hydroxyindoleacetic acid (5-HIAA)	Plasma	5.16E-08	3.43E-05	-2.82306	CNS
5-Hydroxyindoleacetic acid (5-HIAA)	Urine	3.24E-08	3.43E-05	-3.02469	CNS
Levodopa	Plasma	7.23E-09	1.74E-05	-3.38262	CNS
Dopamine (DA)	CSF	3.76E-09	3.53E-05	-3.9732	CNS
Serotonin (5-HT)	CSF	1.24E-09	1.67E-05	-4.13179	CNS
Dihydroxyphenylacetic acid (DOPAC)	Plasma	8.02E-09	1.43E-04	-4.25145	CNS
3,4-Dihydroxyphenylglycol (DHPG)	Plasma	7.56E-09	1.38E-04	-4.261	CNS
Norepinephrine (NE)	CSF	1.33E-09	5.18E-05	-4.59009	CNS
2-Phenylethylamine (PEA)	Urine	1.14E-10	1.27E-04	-6.04501	CNS

Thromboxane B2	Plasma	4.1E-09	1.13E-06	-2.43939	COX-1 Inhibition (clotting factor, drug interaction)
Epinephrine	Plasma	5.66E-10	2.27E-05	-4.60224	Endocrine related Biomarkers
Estradiol	Plasma	1.96E-10	6.32E-05	-5.50832	Endocrine related Biomarkers
Estrone	Plasma	1.6E-10	6.57E-05	-5.61377	Endocrine related Biomarkers
Dihydrotestosterone	Plasma	1.77E-09	4.41E-03	-6.39764	Endocrine related Biomarkers
Testosterone	Plasma	7.99E-10	4.58E-03	-6.75832	Endocrine related Biomarkers
Hippuric acid	Plasma	1.12E-05	2.44E-05	-0.33851	Environmental exposure
Cotinine	Plasma	4.35E-07	2.94E-05	-1.94422	Environmental exposure
Nicotine	Plasma	5.54E-09	3.82E-05	-3.72551	Environmental exposure
Kynurenine (KYN)	Plasma	2.09E-06	2.98E-06	2.87	Indoleamine 2,3-

					Dioxygenase (IDO) activation (immune stress, inflammation, cancer)
Chenodeoxycholic acid (Bile acids)	Plasma	8.31E-07	6.84E-05	-1.91563	Metabolic screening
Cholic acid (Bile acids)	Plasma	4.97E-07	5.07E-05	-2.00941	Metabolic screening
Docosahexaenoic acid (DHA)	Plasma	8.79E-06	7.98E-07	1.041729	Nutritional Status
Eicosapentaenoic acid (EPA)	Plasma	4.28E-06	2.28E-06	0.272841	Nutritional Status
E-Phytonadione Vitamin K1	Plasma	1.13E-09	8.34E-10	0.13095	Nutritional Status
Nicotinuric acid	Plasma	5.77E-06	8.84E-06	-0.18532	Nutritional Status
Nicotinic acid (Vitamin B3)	Plasma	4.1E-05	2.15E-04	-0.72036	Nutritional Status
Tretinoin (All-trans-RA)	Plasma	5.39E-08	2.17E-04	-3.60589	Nutritional Status

Chapter 3. In vitro selection of RNA aptamers using Capture-SELEX

3.1. Introduction

Capture-SELEX - an optimized method for free target selection

In 2012 a paper was published showing a new way to perform SELEX on DNA called Capture-SELEX (124). This strategy was developed to allow the selection of aptamers against target that was not suitable for classical immobilization like small molecules. Indeed, the immobilization of the target requires the formation of a covalent bond between the target and the matrix. In the most often cases, either a primary amine or a carboxyl group is used to create this bond through a chemical reaction. This way of action is suitable for most of the proteins because they have enough amino acids that contains these reactive groups on the side chain of some amino acids. Furthermore, thanks to its size, most of the protein surface will still remain available to interact with potential aptamers during the SELEX. However, this is not the case for small molecules (antibiotics, metabolites…). Indeed, small molecules have a limited amount of reactive group on their surface (sometimes none), often requiring the usage of hard chemistry to perform the immobilization (8, 125). Leading often to either a poor yield of immobilization or even in worst case the complete degradation of the target that could not undergo the immobilization conditions. In the Capture-SELEX strategy, the roles are inverted. The DNA is immobilized, and the target is free in solution. For this purpose, a special library had to be developed. The DNA pool in addition of the primer binding sites and the randomized region, contains a small conserved sequence, placed inside of the randomized part. This sequence is called the docking sequence (DS). This DS will be used has an anchor for another small oligonucleotide which is complementary to the DS, called the capture-oligonucleotide (CO). The CO is derivatized on its 3' side with a linker terminated by a biotin. Thanks to this biotin, the complex CO-DNA library can be immobilized on magnetic streptavidin-coated beads, forming a complex were the library is immobilized on a magnetic support. As all DNA are not able to interact with the CO due to different folding possibilities, some wash steps are necessary to remove unbound DNA from the complex beads-pool. After removing these sequences, the elution of the target binders has to be performed. For that the target mixed with selection buffer is added directly on the beads. Next, only aptamers that undergo a change in their structure upon target binding will be set free from the magnetic beads and can be recovered by applying a magnet on the tube to separate them and reamplify them for the next SELEX cycle.

This technique contains several advantages. The most important one is that Capture-SELEX is suitable for all kind of small molecules as long as they can be soluble in selection buffer at a high enough concentration. Indeed, the target only need to be mixed with the beads to allow the elution of aptamers. This technique is as well faster and straightforward; all the steps can be tuned very easily. Another big advantage is the selection of only structure switching aptamers. One can see this feature as a drawback. Indeed, aptamers that do not change structure upon ligand binding will stay on the beads and therefore not be eluted. Hence, some interesting candidates can be lost during the selection. However, this structure-switching characteristic is very interesting for downstream applications. When we take a look at biosensors there are different possibilities. One can be the coupling of the aptamer with another aptamer for a fluorescent dye such as spinach aptamer or mango aptamer. Indeed, these aptamers have the particularly to exhibit fluorescence only upon target binding. When couple to a structure switching aptamer for a metabolite for example, the folding of the unbound state of this aptamer will not allow the formation of the dye aptamer and therefore no fluorescence will be observed. Nevertheless, if the metabolite is detected in the medium, the metabolite aptamer will change structure and consequently will correctly fold the dye aptamer that will be able to bind its target and to fluoresce. The other possibility is the development of a synthetic riboswitch that can be exploited as internal bio sensors in a cell. Here we present a variant of Capture-SELEX, which uses an RNA library to yield RNA aptamers against small organic molecules. Indeed, RNA cannot be heated up at high temperature in the presence of magnesium and the reamplification step must be changed completely. The Capture-SELEX protocol is divided in several steps: First, the library immobilized on the streptavidin-coated magnetic beads by the through its interaction with the CO forming a complex Beads-RNA in solution. Then, the beads are washed, pelleted and the wash fraction containing non-immobilized RNA is removed. This step is repeated several times to eliminate a maximum of unbound RNA. Afterwards, the target is incubated in the presence of the beads. Aptamer that undergoes a structure switching upon ligand binding are undocked from the CO. Other sequences are still bound to the beads. Eluted aptamer is recovered and amplified using RT-PCR followed by in vitro transcription. After a purification step, the RNA is ready for the following round.

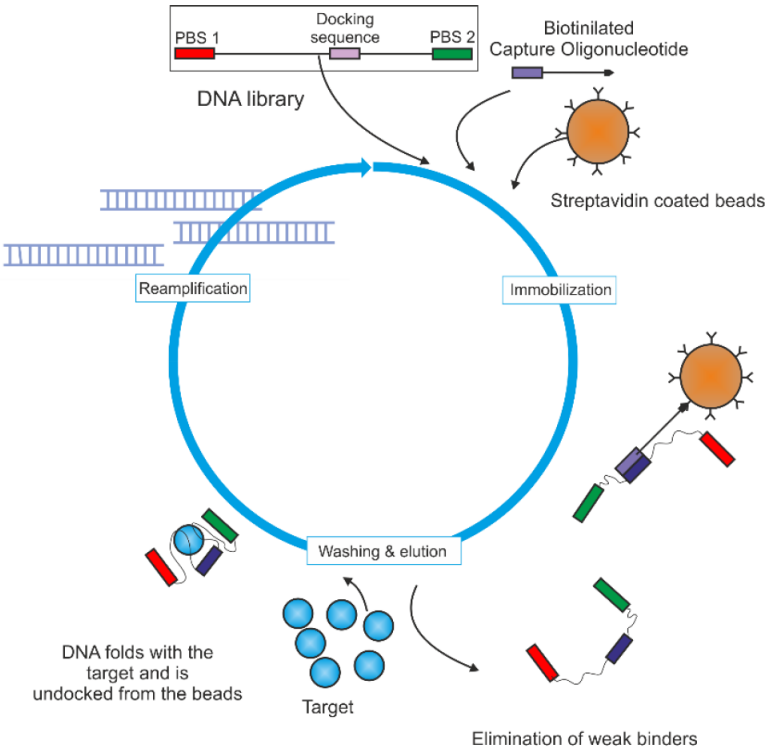


Figure 3.1. Capture-SELEX strategy adapted from Stoltenburg et al adapted (124). Schematic representing the DNA Capture-SELEX protocol. In the first step, the library is mixed with streptavidin-coated mixed and a biotinylated capture-oligonucleotide. The library gets immobilized on the capture-oligonucleotide, that is bound to the streptavidin-coated beads. After several washing steps to remove weak binders to the capture-oligonucleotide, the target is added to the beads and sequences that undergo a structure switching upon ligand binding get undock. These sequences are afterwards recovered and amplified for another cycle.

3.2. Materials and Methods

Preferably, perform all the following experiments in an area designated for RNA work, remember to wear gloves, and to all possible extent use RNase-free reagents. We recommend using filter tips to prevent RNase contamination and carry-over of library between samples. The following method has been an adaptation from a recent paper by Lauridsen et al. (126)

3.2.1. Materials

Library Preparation

10 x hybridization buffer (200 mM Tris–HCl, pH 8.0, 500 mM NaCl, 10 mM EDTA), CASE001 DNA template library with central capture probe region (desalted 5'-AGA-TTG-CAC-TTA-CTA-TCT(N)40 GAT -CGAGCC - TCA(N)10AAT-TGA-ATA-AGC-TGG-TATCCT- ATA-GTG-AGT-CGT-ATTAG-3', Integrated DNA Technologies), CASE003 reverse primer (desalted, 5'-CTA-ATA-CGA-CTCACT- ATA-GGA-TAC-CAG-CTT-ATT-CAATT-3'), TranscriptAid T7 High Yield Transcription kit (1 mL, Thermo Scientific), 0.22 μm Corning Costar Spin-X filter (Sigma-Aldrich), TRIzol (Sigma-Aldrich), Chloroform, Isopropanol, RNase-free water, DNase I (1.25 units/μg DNA), Bioanalyzer RNA Nano 6000 kit (Agilent), Qubit RNA HS kit (Life Technologies), RiboLock RNase inhibitor (40 U/μL, Thermo Fisher Scientific).

Capture-SELEX

M-270 Streptavidin Dynabeads, Magnetic rack for 1.5 mL Eppendorf tubes, 0.2 M NaOH in RNase-free water,, 0.1 M NaCl in RNase-free water, Binding and washing buffer, B&W (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, and 0.005% (v/v) Tween-20), CASE004 capture oligonucleotide (HPLC-purified 5'-biotin-GTC-[hexaethylene glycol]-GATCGAGCCTCA-3', IDT), CS buffer (20 mM Tris–HCl pH 7.4, 2 mM MgCl2, 5 mM KCl, 1 mM CaCl2, 100 mM NaCl and 0.005% (v/v) Tween-20), Target Ligands stock solution (0.5 mM).

Quantification of eluates by RT-qPCR

RevertAid reverse transcriptase (200 U/μL), EvaGreen qPCR dye, 20 x in water (Biotium), ROX reference dye, 1 μM in RNase-free water, Phusion HF II PCR enzyme (2 U/μL), Phusion HF buffer 5 x, dNTPs 10 mM, CASE003 reverse primer (desalted, 5'-CTA-ATA-CGA-CTCACT- ATA-GGA-TAC-CAG-CTT-ATT-CAATT), CASE002 forward primer (desalted, 5'-AGA-TTG-CAC-TTACTA-TCT), Bioanalyzer High Sensitivity DNA kit (Agilent), Amicon Ultra-0.5 30 K spin filters.

Cloning and Biolayer Interferometry

CloneJet PCR cloning kit, XL-Blue Subcloning-Grade Competent Cells, Agencourt AMPure XP, Qiagen Plasmid Plus 96 kit, Black 96-well microplates (Greiner Bio-One), Dip-and read Streptavidin (SA) Biosensors, Regeneration buffer: 50 mM NaOH, 1 M NaCl, Agencourt AMPure XP, Direct-zol RNA miniprep kit.

3.2.2. Methods

3.2.2.1. Preparation of RNA library for Capture-SELEX

Using filter pipet tips, add 20 µL 10 x hybridization buffer, 8 µL 1 mM (8 nmol) CASE003, 40 µL 100 µM CASE001(4 nmol) and 132 µL RNase-free water into a 1.5 mL RNase free tube and mix by vortexing. Then incubate the reaction mixture at 80 ^{0}C for 5 min and let cool at room temperature for 30 min and then place it at 5 ^{0}C for 10 min prior to setting up the T7 transcription reaction. To the hybridization reaction add 200 µL 5x TranscriptAid T7 buffer, 100 µL 100 mM rATP, 100 µL 100 mM rCTP, 100 µL 100 mM rUTP, 100 µL 100 mM rGTP, 100 µL T7 enzyme mix, and 100 µL RNase-free water. Again, incubate at 30 min at 37 ^{0}C on a heat block and till then set centrifuge to cool to 4 ^{0}C. Remove the DNA template by adding 200 µL DNase I to the transcription reaction and incubate for 15 min at 37 ^{0}C. Split the reaction into two 650 µL aliquots and remove the precipitate by passing both through a Costar Spin-X 0.22 µM spin filter (4 min at 1000 x g). Divide the filtrates into 5 RNA free tubes (260 µL each) and to each vial, add 1/10 volume of 3 M NaOAc, pH 5.2 (supplied with the T7 kit). To this add three volumes (780 µL) of Trizol and mix by pipetting and add one volume (260 µL) of chloroform and shake 15 s by hand. Spin at 12,000 x g for 2 min at 4 ^{0}C to separate phases.

Transfer the upper aqueous phase into a new RNase-free tube and add one volume (260 µL) of chloroform and repeat shaking and phase separation. Precipitate the RNA by adding one volume of 100% isopropanol to the aqueous phase and incubate at RT for 10 min. Spin down precipitate at 14,000 x g for 20 min at 4 ^{0}C and remove the supernantant. (The precipitated RNA may be stored like this at -20 ^{0}C). Carefully wash the RNA pellet with 600 µL ice-cold 70% ethanol and spin at 14,000 x g for 5 min and remove the supernatant. Repeat the ethanol washing step once more. Air-dry the RNA pellets for 10 min. Redissolve pellets in 5 x 50 µL RNase-free water and combine and store at -80 ^{0}C for future use. The resuspended RNA should be run on a 1.5 mm 8% PAGE gel at 13 W for 1.5 hours to identify the correct band with RNA ladder. Recover RNA from the cut gel slices with 0.5 x TBE using a Whatman Elutrap operating at 200 volts for 2 hours.

3.2.2.2. Bead Preparation for Capture-SELEX

Coat Beads with Docking Oligonucleotide

We took 10^9 M-270 Streptavidin Dynabeads (1538 µL) and placed it on a magnet to remove the supernatant. Repeating twice, washing steps were applied to beads in 1 mL 0.1 NaOH, 0.05 NaCl, 0.005% (v/v) Tween-20 and supernatant was removed with the help of magnet. Another washing step with 0.1 NaCl, 0.005% (v/v) Tween-20 was applied. Beads were washed 3 times with 500 µL B&W buffer and resuspended in 0.5 mL 1 x B&W buffer. To this we added 0.5 mL 12 µM CASE004 docking oligonucleotide in 1 x B&W buffer and incubated at RT for 1 hr.

Remove Unbound Docking Oligonucleotide

After incubation again wash the beads with 500 µL 1x B&W buffer 3 times and then wash with 500 µL with 1x CS buffer 3 times and resuspend and store the beads in 1 mL 1 x CS buffer at 5 ^{0}C.

Loading the RNA

Dilute 10 nmol RNA library with RNase-free water and 2 x CS buffer to a final volume of 300 µL 1 x CS buffer. Allow the folding of RNA by denaturing at 80 ^{0}C for 8 min and cool for 10 min at 5 ^{0}C. Add the folded RNA library along with 7.5 µL RiboLock to CASE004-coated beads and leave it on a thermomixer at 21 ^{0}C, 1400 RPM overnight.

3.2.2.3. RNA Capture-SELEX

Removing Unbound RNA Library

After overnight incubation place the RNA/bead mixture on a magnetic separator and remove supernatant and save the supernatant for later analysis by RT-qPCR. Wash the beads 9 times with 500 µL 1 x CS buffer with Tween and resuspend beads in 300 µL 1 x CS buffer.

Elution of Weakly Bound RNA Library (Temperature Step)

Incubate the RNA/bead mixture at 28 ^{0}C for 15 min in a thermomixer at 1200 RPM and place it on a magnetic separator and remove supernatant and save the supernatant for RT-qPCR. Wash the beads 7 times with 500 µL 1 x CS buffer with Tween and resuspend beads in 300 µL 1 x CS buffer with Tween.

Background Elution of RNA Library (Background Elution Step)

Incubate the RNA/bead mixture at 21 ^{0}C for 45 min in a thermomixer at 1200 RPM and place it on a magnetic separator and remove supernatant and save the supernatant for RT-qPCR. Wash the beads 7 times with 500 µL 1 x CS buffer with Tween and resuspend beads in 500 µL 1 x CS buffer with Tween.

Elution of RNA Library with Target Compounds (Ligand Step)

Discard the supernatant with magnetic separator and add 300 µL 0.5 mM ligand stock solution prepared in CS buffer to beads. Incubate the ligand/bead mixture at 21 ^{0}C for 45 min in a thermomixer at 1200 RPM and place it on a magnetic separator and remove supernatant and save the supernatant for RT-qPCR.

Heat Elution of Remaining RNA Library (Heat Elution Step)

Resuspend beads in 300 µL 1 x CS buffer and incubate the RNA/bead mixture at 80 ^{0}C for 5 min in a thermomixer at 1200 RPM to undock all the non-binding aptamers and place it on a magnetic separator and remove supernatant and save the supernatant for RT-qPCR. Repeat the same steps twice and save the supernatant.

3.2.2.4. Quantification of eluates by RT-qPCR

The progress of overall SELEX process can be monitored by Reverse transcription-quantitative PCR (RT-qPCR) of the eluates and also for checking DNA contaminants. To set up for reverse transcription prepare 1 µL aliquot of seven samples collected at different stages of Capture-SELEX process. In order to prepare standard curves for quantification, make a dilution series of stock solutions from unselected RNA pool containing 33 nM to 33 fM RNA. To include a positive control, use a 100 pM CASE001 DNA and a no template control (NTC) in RT-qPCR. Set up 14.5 µL pre-hybridization reactions containing 1 µL of unknown, standard or control aliquots, 20 pmol CASE002, and 10 µmol of dNTPs. Anneal the primer to the RNA by heating the mixture to 65 ^{0}C for 5 min and chill on ice. For the positive reaction mix 7.25 µL hybridization mix, 4 µL 5 x RT buffer, 20 U RiboLock and 200 U RevertAid Premium reverse transcriptase for a total volume of 20 µL. The control reactions without reverse transcriptase are assembled in the same manner, substituting reverse transcriptase with water. Incubate the reverse transcription reactions at 50 ^{0}C for 30 min and inactivate the reactions at 85 ^{0}C for 5 min. To set up PCR reactions, use Phusion DNA polymerase

in HF buffer in accord with the manufacturer's instructions containing 1 μL reverse transcription reaction, 10 pmol of primers CASE002 and CASE003, 1 x EvaGreen and 1 x ROX reference dye. Total reaction volume to be 20 μL. Run the PCR using this program: 1 min at 98 ^{0}C; 40 cycles of: 30 s at 98 ^{0}C, 30 s at 51 ^{0}C, 30 s at 72 ^{0}C, 5 min at 72 ^{0}C using appropriate settings for volume normalization using ROX and qPCR amplification using EvaGreen. Validate the PCR product of each sample by 2% 1 x TAE agarose gel. Plot Ct-values obtained from qPCR of the standard curves and use the regression curve to estimate RNA concentration in the Capture-SELEX samples.

1. Unbound RNA	Quantifies the amount of RNA loaded
2. Unbound RNA (1:999 diluted)	Quantifies the amount of RNA loaded
3. Background elution	Quantifies baseline RNA release
4. Elution	Quantifies ligand-induced RNA elution
5. Heat elution (1st of 3)	Quantifies total RNA on beads
6. Heat elution (2nd of 3)	Quantifies total RNA on beads
7. Heat elution (3rd of 3)	Quantifies total RNA on beads

3.2.2.5. Sequencing

To isolate clones for screening, the aptamer pools are blunt-end cloned into the pJET 1.2 vector (Thermo Fisher Scientific) and transformed into XL1-Blue Subcloning-Grade Competent Cells (Agilent) using standard protocols. Perform PCR with pJET1.2 Forward Sequencing Primer or pJET1.2 Reverse Sequencing Primer after picking random colonies and verify the insertion of the aptamer candidate sequences by gel electrophoresis. Perform overnight culture with 1 mL LB broth and 100 μg/mL ampicillin of the picked colonies. Purify the plasmids with Qiagen Plasmid Plus 96 kit and send it for sequencing. Resultant sequences were aligned in Jalview using the ClustalO algorithm; phylogenetic tree analysis was used to sort similar sequences into families.

3.2.2.6. Biolayer interferometry characterization of aptamer kinetics

The binding affinity and kinetics of the selected aptamer can be characterized using biolayer interferometry; DNA aptamer binding kinetics for a small molecule have been previously demonstrated (127). Aptamer DNA templates were built with a T7 promoter and a 3' tail designed to minimize interactions with the aptamer sequence. DNA templates were transcribed in the conditions indicated above to generate RNA aptamers with 3' tails, which were then annealed with 5' biotinylated DNA adapters. RNA aptamers were combined at a 2:1 ratio with DNA adapters in 10 mM Tris, 50

mM NaCl, and 1 mM EDTA. Samples were heated to 85 ºC for 1 minute and cooled -1 ºC every 30 seconds until reaching 25 ºC. Annealing reaction was cleaned up using a Qiagen PCR clean-up kit and diluted into selection buffer. Biolayer interferometry was run using a Pall Fortebio Octet Red96 with Super Streptavidin biosensors. On and off rate kinetics and dissociation constants can be calculated using a monophasic fit of both association and dissociation using Fortebio Octet Data Analysis software.

BIBLIOGRAPHY

1. Wellhausen,R. and Oye,K.A. (2007) Intellectual Property and the Commons in Synthetic Biology: Strategies to Facilitate an Emerging Technology. In *2007 Atlanta Conference on Science, Technology and Innovation Policy*.pp. 1–2.

2. de Vriend, H.C., Walhout, B. and Rathenau Instituut (2006) Constructing Life - Early social reflections on the emerging field of synthetic biology Rathenau Instituut.

3. Bandyopadhyay,S., Mehta,M., Kuo,D., Sung,M.-K., Chuang,R., Jaehnig,E.J., Bodenmiller,B., Licon,K., Copeland,W., Shales,M., *et al.* (2010) Rewiring of Genetic Networks in Response to DNA Damage. *Science*, **330**, 1385–1389.

4. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.

5. Gibson,D.G., Benders,G.A., Andrews-Pfannkoch,C., Denisova,E.A., Baden-Tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A., *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science*, **319**, 1215–1220.

6. Dymond,J.S., Richardson,S.M., Coombes,C.E., Babatz,T., Muller,H., Annaluru,N., Blake,W.J., Schwerzmann,J.W., Dai,J., Lindstrom,D.L., *et al.* (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, **477**, 471–476.

7. Schneider,C., Bronstein,L., Diemer,J., Koeppl,H. and Suess,B. (2017) ROC'n'Ribo: Characterizing a Riboswitching Expression System by Modeling Single-Cell Data. *ACS Synth. Biol.*, **6**, 1211–1224.

8. Groher,F., Bofill-Bosch,C., Schneider,C., Braun,J., Jager,S., Geißler,K., Hamacher,K. and Suess,B. (2018) Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. *Nucleic Acids Res.*, **46**, 2121–2132.

9. McKeague,M., McConnell,E.M., Cruz-Toledo,J., Bernard,E.D., Pach,A., Mastronardi,E., Zhang,X., Beking,M., Francis,T., Giamberardino,A., *et al.* (2015) Analysis of In Vitro Aptamer Selection Parameters. *J. Mol. Evol.*, **81**, 150–161.

10. McKeague,M. and Derosa,M.C. (2012) Challenges and opportunities for small molecule aptamer development. *J. Nucleic Acids*, **2012**, 748913.

11. Mathews,D.H., Turner,D.H. and Zuker,M. (2007) RNA secondary structure prediction. *Curr. Protoc. Nucleic Acid Chem.*, **Chapter 11**, Unit 11.2.

12. Taneda,A. (2010) MODENA: a multi-objective RNA inverse folding. *Adv. Appl. Bioinforma. Chem. AABC*, **4**, 1–12.

13. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neuböck,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70-74.

14. Batey, null, Rambo, null and Doudna, null (1999) Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed Engl.*, **38**, 2326–2343.

15. Steiner,T. (2002) The hydrogen bond in the solid state. *Angew. Chem. Int. Ed Engl.*, **41**, 49–76.

16. Lusk,J.E., Williams,R.J.P. and Kennedy,E.P. (1968) Magnesium and the Growth of Escherichia coZi. **243**, 8.

17. Bolanle,A.O. and Israel,M.O. (2014) Effect of Sesamum indicum L. Seed oil Supplementation on Hepatic and Renal Mineral Concentrations of Hypercholesterolemic Rats. *Am. J. Life Sci.*, **2**, 308.

18. Wilson,K.A., Holland,D.J. and Wetmore,S.D. (2016) Topology of RNA-protein nucleobase-amino acid π-π interactions and comparison to analogous DNA-protein π-π contacts. *RNA N. Y. N*, **22**, 696–708.

19. Desfrançois,C., Carles,S. and Schermann,J.P. (2000) Weakly Bound Clusters of Biological Interest. *Chem. Rev.*, **100**, 3943–3962.

20. Gorin,A.A., Zhurkin,V.B. and Olson,W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, **247**, 34–48.

21. Olson,W.K., Gorin,A.A., Lu,X.-J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 11163–11168.

22. Manalo,M.N., Pérez,L.M. and LiWang,A. (2007) Hydrogen-bonding and pi-pi base-stacking interactions are coupled in DNA, as suggested by calculated and experimental trans-Hbond deuterium isotope shifts. *J. Am. Chem. Soc.*, **129**, 11298–11299.

23. Matta,C.F., Castillo,N. and Boyd,R.J. (2006) Extended Weak Bonding Interactions in DNA: π-Stacking (Base−Base), Base−Backbone, and Backbone−Backbone Interactions. *J. Phys. Chem. B*, **110**, 563–578.

24. Ok,T.-D., Lee,J.-H. and Lee,H.-S. (2011) γ^3PNA: Peptide Analogue of Glycol Nucleic Acid. *Bull. Korean Chem. Soc.*, **32**, 2863–2864.

25. Mignon,P., Loverix,S., Steyaert,J. and Geerlings,P. (2005) Influence of the π–π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.*, **33**, 1779–1789.

26. Birch,J.R. and Racher,A.J. (2006) Antibody production. *Adv. Drug Deliv. Rev.*, **58**, 671–685.

27. Thirunavukarasu,D., Chen,T., Liu,Z., Hongdilokkul,N. and Romesberg,F.E. (2017) Selection of 2′-Fluoro-Modified Aptamers with Optimized Properties. *J. Am. Chem. Soc.*, **139**, 2892–2895.

28. Wang,R.E., Zhang,Y., Cai,J., Cai,W. and Gao,T. (2011) Aptamer-Based Fluorescent Biosensors. *Curr. Med. Chem.*, **18**, 4175–4184.

29. Mascini,M. (2008) Aptamers and their applications. *Anal. Bioanal. Chem.*, **390**, 987–988.

30. Ye,B.-F., Zhao,Y.-J., Cheng,Y., Li,T.-T., Xie,Z.-Y., Zhao,X.-W. and Gu,Z.-Z. (2012) Colorimetric photonic hydrogel aptasensor for the screening of heavy metal ions. *Nanoscale*, **4**, 5998–6003.

31. Meli,M., Vergne,J., Décout,J.-L. and Maurel,M.-C. (2002) Adenine-aptamer complexes: a bipartite RNA site that binds the adenine nucleic base. *J. Biol. Chem.*, **277**, 2104–2111.

32. Lauhon,C.T. and Szostak,J.W. (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.*, **117**, 1246–1257.

33. Stojanovic,M.N., de Prada,P. and Landry,D.W. (2001) Aptamer-Based Folding Fluorescent Sensor for Cocaine. *J. Am. Chem. Soc.*, **123**, 4928–4931.

34. Stojanovic,M.N., de Prada,P. and Landry,D.W. (2000) Fluorescent Sensors Based on Aptamer Self-Assembly. *J. Am. Chem. Soc.*, **122**, 11547–11548.

35. Nutiu,R. and Li,Y. (2005) In vitro selection of structure-switching signaling aptamers. *Angew. Chem. Int. Ed Engl.*, **44**, 1061–1065.

36. Elowe,N.H., Nutiu,R., Allali-Hassani,A., Cechetto,J.D., Hughes,D.W., Li,Y. and Brown,E.D. (2006) Small-molecule screening made simple for a difficult target with a signaling nucleic acid aptamer that reports on deaminase activity. *Angew. Chem. Int. Ed Engl.*, **45**, 5648–5652.

37. Olek,M., Büsgen,T., Hilgendorff,M. and Giersig,M. (2006) Quantum Dot Modified Multiwall Carbon Nanotubes. *J. Phys. Chem. B*, **110**, 12901–12904.

38. Chhabra,R., Sharma,J., Wang,H., Zou,S., Lin,S., Yan,H., Lindsay,S. and Liu,Y. (2009) Distance-dependent interactions between gold nanoparticles and fluorescent molecules with DNA as tunable spacers. *Nanotechnology*, **20**, 485201.

39. Huang,Y., Zhao,S., Liang,H., Chen,Z.-F. and Liu,Y.-M. (2011) Multiplex Detection of Endonucleases by Using a Multicolor Gold Nanobeacon. *Chem. – Eur. J.*, **17**, 7313–7319.

40. Zhao,W., Chiuman,W., Brook,M.A. and Li,Y. (2007) Simple and rapid colorimetric biosensors based on DNA aptamer and noncrosslinking gold nanoparticle aggregation. *Chembiochem Eur. J. Chem. Biol.*, **8**, 727–731.

41. Liu,J. and Lu,Y. (2005) Fast colorimetric sensing of adenosine and cocaine based on a general sensor design involving aptamers and nanoparticles. *Angew. Chem. Int. Ed Engl.*, **45**, 90–94.

42. Paige,J.S., Wu,K.Y. and Jaffrey,S.R. (2011) RNA mimics of green fluorescent protein. *Science*, **333**, 642–646.

43. Song,R.L.S.,Wenjiao and Jaffrey,S.R. (2014) Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nat. Protoc.*, **9**, 146–155.

44. DasGupta,S., Shelke,S.A., Li,N. and Piccirilli,J.A. (2015) Spinach RNA aptamer detects lead(II) with high selectivity. *Chem. Commun.*, **51**, 9034–9037.

45. Pothoulakis,G., Ceroni,F., Reeve,B. and Ellis,T. (2014) The Spinach RNA Aptamer as a Characterization Tool for Synthetic Biology. *ACS Synth. Biol.*, **3**, 182–187.

46. Lee,J.-H., Jucker,F. and Pardi,A. (2008) Imino proton exchange rates imply an induced-fit binding mechanism for the VEGF165-targeting aptamer, Macugen. *FEBS Lett.*, **582**, 1835–1839.

47. Bates,P.J., Laber,D.A., Miller,D.M., Thomas,S.D. and Trent,J.O. (2009) Discovery and

development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp. Mol. Pathol.*, **86**, 151–164.

48. Diener,J.L., Daniel Lagassé,H.A., Duerschmied,D., Merhi,Y., Tanguay,J.-F., Hutabarat,R., Gilbert,J., Wagner,D.D. and Schaub,R. (2009) Inhibition of von Willebrand factor-mediated platelet activation and thrombosis by the anti-von Willebrand factor A1-domain aptamer ARC1779. *J. Thromb. Haemost. JTH*, **7**, 1155–1162.

49. Tasset,D.M., Kubik,M.F. and Steiner,W. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.*, **272**, 688–698.

50. Min,K., Jo,H., Song,K., Cho,M., Chun,Y.-S., Jon,S., Kim,W.J. and Ban,C. (2011) Dual-aptamer-based delivery vehicle of doxorubicin to both PSMA (+) and PSMA (-) prostate cancers. *Biomaterials*, **32**, 2124–2132.

51. Pollard,J., Bell,S.D. and Ellington,A.D. (2001) Design, synthesis, and amplification of DNA pools for construction of combinatorial pools and libraries. *Curr. Protoc. Mol. Biol.*, **Chapter 24**, Unit 24.2.

52. Hall,B., Arshad,S., Seo,K., Bowman,C., Corley,M., Jhaveri,S.D. and Ellington,A.D. (2009) In vitro selection of RNA aptamers to a protein target by filter immobilization. *Curr. Protoc. Mol. Biol.*, **Chapter 24**, Unit 24.3.

53. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Sci. Am. Assoc. Adv. Sci.*, **249**, 505–510.

54. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nat. Lond.*, **346**, 818–822.

55. Robertson,D.L. and Joyce,G.F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.

56. Geiger,A., Burgstaller,P., von der Eltz,H., Roeder,A. and Famulok,M. (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.*, **24**, 1029–1036.

57. Jenison,R.D., Gill,S.C., Pardi,A. and Polisky,B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–1429.

58. Rangel,A.E., Chen,Z., Ayele,T.M. and Heemstra,J.M. (2018) In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Res.*, **46**, 8057–8068.

59. Karlsen,K.K. and Wengel,J. (2012) Locked nucleic acid and aptamers. *Nucleic Acid Ther.*, **22**, 366–370.

60. Eulberg,D. and Klussmann,S. (2003) Spiegelmers: Biostable Aptamers. *ChemBioChem*, **4**, 979–983.

61. Zhu,B., Hernandez,A., Tan,M., Wollenhaupt,J., Tabor,S. and Richardson,C.C. (2015) Synthesis of 2’-Fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Res.*, **43**, e94.

62. Shangguan,D., Li,Y., Tang,Z., Cao,Z.C., Chen,H.W., Mallikaratchy,P., Sefah,K., Yang,C.J. and

Tan,W. (2006) Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 11838–11843.

63. Pestourie,C., Cerchia,L., Gombert,K., Aissouni,Y., Boulay,J., De Franciscis,V., Libri,D., Tavitian,B. and Ducongé,F. (2006) Comparison of different strategies to select aptamers against a transmembrane protein target. *Oligonucleotides*, **16**, 323–335.

64. Nallani,M., Andreasson-Ochsner,M., Tan,C.-W.D., Sinner,E.-K., Wisantoso,Y., Geifman-Shochat,S. and Hunziker,W. (2011) Proteopolymersomes: In vitro production of a membrane protein in polymersome membranes. *Biointerphases*, **6**, 153–157.

65. Mendonsa,S.D. and Bowser,M.T. (2004) In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.*, **126**, 20–21.

66. Nielsen,J. and Keasling,J.D. (2016) Engineering Cellular Metabolism. *Cell*, **164**, 1185–1197.

67. Slomovic,S., Pardee,K. and Collins,J.J. (2015) Synthetic biology devices for in vitro and in vivo diagnostics. *Proc. Natl. Acad. Sci.*, **112**, 14429–14435.

68. Rogers,J.K., Taylor,N.D. and Church,G.M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.*, **42**, 84–91.

69. Smanski,M.J., Zhou,H., Claesen,J., Shen,B., Fischbach,M.A. and Voigt,C.A. (2016) Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.*, **14**, 135–149.

70. Zhou,J. and Rossi,J. (2017) Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discov.*, **16**, 181–202.

71. Tinberg,C.E., Khare,S.D., Dou,J., Doyle,L., Nelson,J.W., Schena,A., Jankowski,W., Kalodimos,C.G., Johnsson,K., Stoddard,B.L., *et al.* (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*, **501**, 212–216.

72. Lechner,H., Ferruz,N. and Höcker,B. (2018) Strategies for designing non-natural enzymes and binders. *Curr. Opin. Chem. Biol.*, **47**, 67–76.

73. Xayaphoummine,A., Bucher,T. and Isambert,H. (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.*, **33**, W605-610.

74. Espah Borujeni,A., Mishler,D.M., Wang,J., Huso,W. and Salis,H.M. (2016) Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res.*, **44**, 1–13.

75. Kellenberger,C.A., Wilson,S.C., Sales-Lee,J. and Hammond,M.C. (2013) RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc.*, **135**, 4906–4909.

76. Carothers,J.M., Goler,J.A., Juminaga,D. and Keasling,J.D. (2011) Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. *Science*, **334**, 1716–1719.

77. Win,M.N. and Smolke,C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456–460.

78. Abatemarco,J., Sarhan,M.F., Wagner,J.M., Lin,J.-L., Liu,L., Hassouneh,W., Yuan,S.-F., Alper,H.S. and Abate,A.R. (2017) RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nat. Commun.*, **8**, 332.

79. Arroyo-Currás,N., Somerson,J., Vieira,P.A., Ploense,K.L., Kippin,T.E. and Plaxco,K.W. (2017) Real-time measurement of small molecules directly in awake, ambulatory animals. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 645–650.

80. Mage,P.L., Ferguson,B.S., Maliniak,D., Ploense,K.L., Kippin,T.E. and Soh,H.T. (2017) Closed-loop control of circulating drug levels in live animals. *Nat. Biomed. Eng.*, **1**, 1–10.

81. Kundert,K., Lucas,J.E., Watters,K.E., Fellmann,C., Ng,A.H., Heineike,B.M., Fitzsimmons,C.M., Oakes,B.L., Qu,J., Prasad,N., *et al.* (2019) Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat. Commun.*, **10**, 2127–11.

82. Carothers,J.M. (2013) Design-driven, multi-use research agendas to enable applied synthetic biology for global health. *Syst. Synth. Biol.*, **7**, 79–86.

83. Engineering Biology Research Consortium. EBRC Research Roadmap. 2018.

84. Win,M.N. and Smolke,C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA*, **104**, 14283.

85. Burke,C.R., Sparkman-Yager,D. and Carothers,J.M. (In revision.) Multi-state design of kinetically-controlled RNA aptamer ribosensors. *Nat. Chem. Biol.*, 10.1101/213538.

86. Carothers,J.M., Oestreich,S.C., Davis,J.H. and Szostak,J.W. (2004) Informational complexity and functional activity of RNA structures. *J Am Chem Soc*, **63**, 57–94.

87. Carothers,J.M., Davis,J.H., Chou,J.J. and Szostak,J.W. (2006) Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *Rna- Publ. Rna Soc.*, **12**, 567–579.

88. Carothers,J.M., Goler,J.A., Kapoor,Y., Lara,L. and Keasling,J.D. (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic Acids Res.*, **38**, 2736–2747.

89. JOBACK,K.G. and REID,R.C. (1987) Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.*, **57**, 233–243.

90. Lydersen,A.L. (1955) Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions, by A.L. Lydersen.

91. Andrews,P.R., Craik,D.J. and Martin,J.L. (1984) Functional group contributions to drug-receptor interactions. *J. Med. Chem.*, **27**, 1648–1657.

92. Camacho,D.M., Collins,K.M., Powers,R.K., Costello,J.C. and Collins,J.J. (2018) Next-Generation Machine Learning for Biological Networks. *Cell Camb.*, **173**, 1581–1592.

93. Groher,A.-C., Jager,S., Schneider,C., Groher,F., Hamacher,K. and Suess,B. (2019) Tuning the Performance of Synthetic Riboswitches using Machine Learning. *ACS Synth. Biol.*, **8**, 34–44.

94. Wilson,D.S. and Szostak,J.W. (1999) In vitro selection of functional nucleic acids. *Ann Rev Biochem*, **68**, 611–647.

95. Pfeiffer,F. and Mayer,G. (2016) Selection and Biosensor Application of Aptamers for Small Molecules. *Chem. Biol.*, 10.3389/fchem.2016.00025.

96. Jankowski,M.D., Henry,C.S., Broadbelt,L.J. and Hatzimanikatis,V. (2008) Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. *Biophys. J.*, **95**, 1487–1499.

97. Hermann,T. and Patel,D.J. (2000) Adaptive Recognition by Nucleic Acid Aptamers. *Science*, **287**, 820–825.

98. Kass,R.E. and Raftery,A.E. (1995) Bayes Factors. *J. Am. Stat. Assoc.*, **90**, 773–795.

99. Lee,S.Y., Kim,H.U., Chae,T.U., Cho,J.S., Kim,J.W., Shin,J.H., Kim,D.I., Ko,Y.-S., Jang,W.D. and Jang,Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.*, **2**, 18–33.

100. Bouatra,S., Aziat,F., Mandal,R., Guo,A.C., Wilson,M.R., Knox,C., Bjorndahl,T.C., Krishnamurthy,R., Saleem,F., Liu,P., *et al.* (2013) The Human Urine Metabolome. *PLoS ONE*, **8**.

101. Brown,R.F., Andrews,C.T. and Elcock,A.H. (2015) Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.*, **11**, 2315–2328.

102. Perola,E. and Charifson,P.S. (2004) Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.*, **47**, 2499–2510.

103. Mobley,D.L. and Dill,K.A. (2009) Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Struct. Lond. Engl. 1993*, **17**, 489–498.

104. Worldwide Clinical Trials Bioanalytical Sciences Biomarker Methods-Worldwide Clinical Trials.

105. Wishart,D.S., Feunang,Y.D., Marcu,A., Guo,A.C., Liang,K., Vázquez-Fresno,R., Sajed,T., Johnson,D., Li,C., Karu,N., *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.*, **46**, D608–D617.

106. Minakawa,H., Masuo,S., Kaneko,T. and Takaya,N. (2019) Fermentation and purification of microbial monomer 4-amminocinnamic acid to produce ultra-high performance bioplastics. *Process Biochem.*, **77**, 100–105.

107. Stevens,J.T. and Carothers,J.M. (2015) Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synth. Biol.*, **4**, 107–115.

108. Xiu,Y., Jang,S., Jones,J.A., Zill,N.A., Linhardt,R.J., Yuan,Q., Jung,G.Y. and Koffas,M.A.G. (2017) Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli co-cultures. *Biotechnol. Bioeng.*, **114**, 2235–2244.

109. Xu,J., Carrocci,T.J. and Hoskins,A.A. (2015) Evolution and characterization of a benzylguanine-binding RNA aptamer. *Chem. Commun.*, **52**, 549–552.

110. Ricci,F., Vallée-Bélisle,A., Simon,A.J., Porchetta,A. and Plaxco,K.W. (2016) Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.*, **49**, 1884–1892.

111. Xiang,Y. and Lu,Y. (2011) Using personal glucose meters and functional DNA sensors to quantify a variety of analytical targets. *Nat. Chem.*, **3**, 697–703.

112. Chen,S.-J., Huang,Y.-F., Huang,C.-C., Lee,K.-H., Lin,Z.-H. and Chang,H.-T. (2008) Colorimetric determination of urinary adenosine using aptamer-modified gold nanoparticles. *Biosens. Bioelectron.*, **23**, 1749–1753.

113. Dhiman,A., Kalra,P., Bansal,V., Bruno,John.G. and Sharma,T.K. (2017) Aptamer-based point-of-care diagnostic platforms. *Sens. Actuators B Chem.*, **246**, 535–553.

114. Goldstein,D.S., Holmes,C., Bentho,O., Sato,T., Moak,J., Sharabi,Y., Imrich,R., Conant,S. and Eldadah,B.A. (2008) Biomarkers to detect central dopamine deficiency and distinguish Parkinson disease from multiple system atrophy. *Parkinsonism Relat. Disord.*, **14**, 600–607.

115. Hou,C., Jia,F., Liu,Y. and Li,L. (2006) CSF serotonin, 5-hydroxyindolacetic acid and neuropeptide Y levels in severe major depressive disorder. *Brain Res.*, **1095**, 154–158.

116. Vincent,S., Bieck,P.R., Garland,E.M., Loghin,C., Bymaster,F.P., Black,B.K., Gonzales,C., Potter,W.Z. and Robertson,D. (2004) Clinical assessment of norepinephrine transporter blockade through biochemical and pharmacological profiles. *Circ. N. Y. N*, **109**, 3202–3207.

117. Perry,H. and Keevil,B. (2008) Online extraction of 5-hydroxyindole acetic acid from urine for analysis by liquid chromatography–tandem mass spectrometry: *Ann. Clin. Biochem.*, 10.1258/acb.2007.007067.

118. Penner,N., Ramanathan,R., Zgoda-Pols,J. and Chowdhury,S. (2010) Quantitative determination of hippuric and benzoic acids in urine by LC–MS/MS using surrogate standards. *J. Pharm. Biomed. Anal.*, **52**, 534–543.

119. McKeague,M., De Girolamo,A., Valenzano,S., Pascale,M., Ruscito,A., Velu,R., Frost,N.R., Hill,K., Smith,M., McConnell,E.M., *et al.* (2015) Comprehensive analytical comparison of strategies used for small molecule aptamer evaluation. *Anal. Chem.*, **87**, 8608–8612.

120. Jonathan H. Davis and Jack W. Szostak (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. - PNAS*, **99**, 11616–11621.

121. Lévesque,D., Beaudoin,J.-D., Roy,S. and Perreault,J.-P. (2007) In vitro selection and characterization of RNA aptamers binding thyroxine hormone. *Biochem. J.*, **403**, 129–138.

122. Chang,A.L., McKeague,M., Liang,J.C. and Smolke,C.D. (2014) Kinetic and Equilibrium Binding Characterization of Aptamers to Small Molecules using a Label-Free, Sensitive, and Scalable Platform. *Anal. Chem.*, **86**, 3273–3278.

123. Shoji,A., Kuwahara,M., Ozaki,H. and Sawai,H. (2007) Modified DNA Aptamer That Binds the

(R)-Isomer of a Thalidomide Derivative with High Enantioselectivity. *J. Am. Chem. Soc.*, **129**, 1456–1464.

124. Stoltenburg,R., Nikolaus,N. and Strehlitz,B. (2012) Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *J. Anal. Methods Chem.*, **2012**, e415697.

125. Wallis,M.G., von Ahsen,U., Schroeder,R. and Famulok,M. (1995) A novel RNA motif for neomycin recognition. *Chem. Biol.*, **2**, 543–552.

126. Lauridsen,L.H., Doessing,H.B., Long,K.S. and Nielsen,A.T. (2018) A Capture-SELEX Strategy for Multiplexed Selection of RNA Aptamers Against Small Molecules. *Methods Mol. Biol. Clifton NJ*, **1671**, 291–306.

127. Gao,S., Hu,B., Zheng,X., Cao,Y., Liu,D., Sun,M., Jiao,B. and Wang,L. (2016) Gonyautoxin 1/4 aptamers with high-affinity and high-specificity: From efficient selection to aptasensor application. *Biosens. Bioelectron.*, **79**, 938–944.