DEVELOPMENT AND CHARACTERIZATION OF HIGH-THROUGHPUT METHODS AND TECHNOLOGIES FOR IN VITRO RNA APTAMER SELECTIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kylan Szeto

May 2014

DEVELOPMENT AND CHARACTERIZATION OF HIGH-THROUGHPUT METHODS AND TECHNOLOGIES FOR IN VITRO RNA APTAMER SELECTIONS

Kylan Szeto, Ph. D.

Cornell University 2014

Aptamers are an emerging class of molecules that are valued for their ability to bind with high affinity and specificity to a desired target. These DNA or RNA ligands can be synthesized easily in the lab making them much more stable, reproducible, and cost-effective than other affinity reagents such as antibodies. These molecules possess a diverse and versatile set of abilities upon binding, making them useful not only in biotechnology and diagnostics, but also in therapeutics.

Aptamers are typically generated through an *in vitro* process called SELEX (Systematic Evolution of Ligands by EXponential Enrichment). SELEX is an iterative process whereby a library of random sequences is exposed to a target, and bound sequences are selected and amplified making a pool enriched for improved binding over the original library. This process can be repeated until the best binding sequence dominates the pool. In this way, aptamers can be generated to nearly any conceivable target or molecule, overcoming the immunological limitations in generating antibodies. By manipulating the conditions within this selection process, aptamers can also be generated to bind in non-physiological environments, giving them the ability to perform in diverse applications.

To help accelerate the discovery of aptamers, we developed a modular microcolumn technology that decreases reagent consumption, while permitting

multiplex SELEX, allowing for more sophisticated selection schemes. By characterizing aptamer enrichments for all the available parameters, we have significantly increased the selection efficiency, while illustrating the failures of simple binding theories to non-classical modes of selection. We have also demonstrated the power of high-throughput sequencing for early identification of aptamers before they fully converge. This was used to validate a new method which optimizes selection time over individual selection cycle efficiencies. Surprisingly, our results also show the new method significantly improves selection efficiency, bringing into question some common beliefs derived from SELEX theory. Finally, we scaled our microcolumn technology to an automatable 96-well microplate format, and demonstrated its utility by characterizing and validating specific, non-specific, and background binding behavior of several RNA aptamers. The results reveal binding behaviors that fundamentally limit the performance and sensitivity of aptamer selections.

BIOGRAPHICAL SKETCH

The author was born in Claremont, New Hampshire, USA on July 17, 1986 to parents John Szeto and Ruth Roylance. He grew up with two older brothers, Jimmy Szeto and Marvin Szeto; an older sister, Erna Szeto; and a younger sister, Kaylene Szeto. From a young age all the way through high school, he worked at his father's restaurant, The Royal Dragon, with the rest of his family. In 2004, he graduated third in his class from Stevens High School and matriculated at Rensselaer Polytechnic Institute in Troy, New York. There, he pursued research in two laboratories including an NNIN summer internship in the Cornell Nanophotonics Group with Dr. Michal Lipson at Cornell University where he learned micro- and nanofabrication. Working for Dr. Peter Persans at RPI, he learned to build optical equipment for characterizing PbS quantum dots and semiconductor GalnN quantum well diode lasers. In 2007, he was the recipient of the Robert Resnick Physics Prize for outstanding academic achievement in physics and in 2008 he graduated *summa cum laude* from Rensselaer Polytechnic Institute with a Bachelor of Science in Physics. He was subsequently admitted into Cornell University as a graduate student in the School of Applied and Engineering Physics. After two semesters of course work and teaching, he joined Dr. Harold Craighead's laboratory to complete his Doctoral training.

To Laura and my family

And to the physicist who showed me

how the world goes ‘round

Mr. David Stockwell

ACKNOWLEDGMENTS

Countless people have contributed to the successful completion of my PhD here at Cornell, but none have played a larger and more supporting role than my Doctoral Advisor, Dr. Harold Craighead. With his guidance I have developed the ability to apply my skills and knowledge in physics and engineering towards meaningful biological problems. Without this, I would not have been able to make such unique and hopefully valuable contributions in a field dominated by biologist. In addition, he has been a significant source of encouragement and patience in my research, allowing me the freedom and resources to carefully and systematically troubleshoot numerous problems that I've encountered in my work.

Throughout my years at Cornell, I've had the pleasure of working with many other students in the Craighead Group (Dr. Ben Cipriany, Dr. Christine Tan, Dr. Paul Zhu, and Dr. Rob Barton) who not only took the time to train me when I first joined the lab, but who have also allowed me to contribute to some of their work. Dr. Phillip Wagoner and Dr. Darren Southworth and Daniel Joe were also great friends and helped make transitioning into research within our multidisciplinary lab much less painful than it could have been on my own. Recently, I have also enjoyed working with Chris Wallin and Harvey Tian, as well as Sarah Reinholt who not only helped me develop MEDUSA, but has taken over my work and continues to address similar biological problems with an engineer's mindset. I have also worked closely with a number of postdoctoral associates within the lab such as Dr. Seung-Min Park who was my mentor for the first year, and got me started on my long journey into biochemistry and devices for *in vitro* selections. I also have tremendous gratitude for Dr. Chris Kelly and Dr. David Latulippe, both of which I have worked closely with and who have contributed significantly to the work contained within this dissertation. I have also enjoyed the company and advice from Dr. Madhukar Varshney, Dr. Tiberiu

Onuta, Dr. Steve Levy, Dr. Yoav Linzon, Dr. Aline Cerf, Dr. Thomas Alva, Dr. Juraj Topolancik, and most recently Dr. Jia-Wei Yeh, Dr. Vivek Adiga and Dr. Jaime Benitez who have been great for conversations and have helped diffuse my frustrations as I conclude my work at Cornell. I have also worked with a number of undergraduates, but Nikita Taparia and Cher (Xuexiang) Zhang have contributed to my own work. Cher has been an especially dedicated and versatile researcher in the lab and has been enormously appreciative and thoughtful to all of us (Perhaps more than we deserve!)

Without our groups extensive collaborations and access to expertise outside the lab, much of my work would have been impossible, and I have been fortunate to collaborate with both of my Committee members. I have to thank Dr. Warren Zipfel as well as his student (and my classmate) Avtar Singh for years of interesting and helpful discussions about aptamers and technologies, especially their practical and straightforward approach to research problems. I also have to thank Dr. John Lis and his group for taking me in and treating me as one of their own. They have been enormously generous and patient by giving me space to work in their lab and teaching me all of the biochemistry necessary for executing my experiments. A special thanks is owed to Dr. Abdullah Ozer who is central to all of the aptamer work and the many collaborations within it. He has been a terrific mentor and a great friend to me since the very beginning of my training, and has played a critical role in all of my research. The two of us have struggled and succeeded together from the very beginning of this grand aptamer project, and together we have been fortunate to see our collective stubbornness and persistence pay off. I've also had the pleasure working with Lis lab members Fabiana Duarte, Dr. John Pagano, Li Yao and Jacob Tome, and have become good friends with all of them. I must also thank Dr. David Shalloway and Dr. Brian White who have contributed valuable theory and analysis tools to my research, and

have been a refreshing source of quantitative and theoretical discussions which are normally absent in this field of research. Finally, I have to thank Dr. Holger Sonderman and his former student (and my fiancé!) Dr. Laura Byrnes who provided me with an opportunity to exercise more physics and programing skills than my own research has allowed.

My success here at Cornell would not have been possible without the support of my family. My mother Ruth Roylance has always instilled in me an interest in science (although apparently I initially wanted to be an OB/GYN, and then a Dentist, and then an Anesthesiologist) and my father John Szeto has always pushed me (in typical Chinese fashion) to be the best that I possibly can and is pleased that I'll still be a (although a different kind of) Doctor. My parents have always been proud of me and my accomplishments, and this has helped me to achieve what I have today. I also have to give special thanks to my little sister Dr. Kaylene (Mei mei) Szeto, who has followed a similar path as me into science. She has been a healthy source of competition, driving both of us to try and outdo the other (she finished her doctorate before me…). Through our common struggles, she has been a great person to vent to and find reassurance and regain confidence in myself. Also, her husand and my Bro-in-Law, Keyvan Asdigha is incredibly humorous and has been a great friend over the years. In addition, my family and I all share a passion for music and art. My family has been supportive of these other interests which has enabled me to continue to cultivate these skills through college and graduate school, and I have been fortunate to play French Horn and Trumpet semi-professionally in previous years. Also, drawing and playing piano have played tremendous roles in helping me balance my research with my life. These all have been important emotional outlets for me as well as a means to connect with family and friends beyond my scientific research.

Successful completion of my degree at Cornell would have been impossible

without the support of my best friend and fiancé Dr. Laura Byrnes. She's has been there for me since the very beginning of my journey into science when she tutored me in Introductory Chemistry my first semester of college at RPI (without which I would surely have flunked out!). She graciously followed me here to Cornell University, turning down opportunities to study crystallography at other amazing universities so that we could stay together. She has made many sacrifices over the years for me, and so have I for her, in order for the two of us to complete PhD's together. I'm incredibly fortunate to have had her unconditional love and understanding as we struggled together through the difficult trials of earning our degrees. She's been the most patient person in my life, tolerating many late nights, seemingly endless complaining, and shouldering so much of my frustrations. In addition, I have to thank my two cats Spike and (Mr.) Bubbles, who Laura and I generally refer to as Mere and Berbs. They too have exhibited some form of unconditional love and understanding as we've moved them about and abandoned them day after day in pursuit of research (that's what we like to think anyways…maybe it's just that we feed them!). They have always been a great distraction after a bad day, and always seem to be excited to see me when I come home (again, this could be because I feed them…) which is a welcomed response after a hard day at work. Finally, over the years, I've grown quite close with Laura's family who have been equally supportive in my and Laura's endeavors as my own family has been. So, many thanks to Joe and Carol Byrnes, Matt and Nikki, Megan and Dan, and Jenny (who has a unique appreciation for the difficulties of earning a PhD)!

Of course over the years, you can't help but make friends along the way that you don't have a regular work relationship with. These include my classmates (Dr. Hui Liu, Taige Hou, Dr. David Hutchinson, Pinshane Huang, Dr. Julia Mundy, Jun-Bo Park, John Mergo, Robin Havener, Arthur Barnard, Dr. Nikhil Fernandez, and Avtar Singh) who I was very close with in early years and still enjoy catching up with

whenever possible. Thanks for all the fun times, office pranks and memories guys! I also have to thank Dr. David Lee and Dr. Linda (Shih Lin) Goh who have been my closest friends at Cornell, as well as Dr. Huimin Chen, David Ackerman, John O'Donnell and Haley for many fun times. I'm going to miss hotpot, dumpling nights and all our other (many) food related activities! Also Dr. Srikant Iyer who always had great physics questions for me, and made me feel incredibly smart when working with; and Dr. Yvonne (Hsien-Wei) Meng for our informal collaboration and attempts to modernize Cell-SELEX and conversations about life and graduate school. I also need to thank some old friends who have stuck with me since I left my small New Hampshire town. They have kept me connected to my home, and I'm so appreciative for their continued friendship and support. So thank you Amy Hughes, Lindsay Hartzell, and Stephanie Wyman for not abandoning me when I fled Claremont!

I also want to acknowledge the critical people and experiences that have led me to this point in my scientific career. First I have to give tremendous thanks to my physics teacher at Stevens High School, Mr. David Stockwell. I never had an interest in physics until I took his class. Even then I was less interested in the subject as much as I respected and appreciated the way he taught and treated students in his classroom. In his classroom, we were men and women despite our young age, and treated as adults. His stories, lessons, and energy (he was always swinging a yard stick around, occasionally hitting objects within his reach) were what drew me more into the subject, and it is because of this amazing physicist and teacher that I decided to pursue my passion for mathematics in the context of theoretical physics while in college. Similarly, while at RPI, I had the pleasure of learning from, teaching for, and eventually doing research with Dr. Peter Persans. He is also an amazing teacher (which he eventually won a prestigious award for) and took a personal interest in many students studies and future ambitions. He is incredibly grounded and relatable

and gave me the confidence (and experience) to pursue a PhD in physics. It is because of his advice that I eventually applied to Cornell's prestigious Applied and Engineering Physics program. While at RPI, I was also fortunate enough to get accepted as an NNIN Summer REU in the Lab of Dr. Michael Lipson here are Cornell. Working closely with Dr. Carl Poitras, I got my first experience with micro- and nanoscale fabrication. This was an incredible learning experience and ultimately played a critical role in my decision to come to Cornell for a PhD and join the Craighead Group which is known for its work in nanoscience.

In addition to having world class faculty and research, Cornell has an amazing network of user-based facilities. This has enabled me and countless other researchers access to instruments and expertise that cannot be found anywhere else. I have benefitted enormously from use of CNF and the expertise and friendship of Dr. Rob Ilic and Edward Camacho who were also helpful during my NNIN REU internship. I also regularly used the facilities of NBTC and learned much from Penny Burke who is a very helpful and very energetic scientist. I have also made occasional use of the CCMR facilities. Finally, I have made great use of the LASSP Machine Shop for custom in-house fabrication of needed hardware for the lab. In addition, I have utilized the LASSP Student Machine Shop where I was fully trained and advised by Nathan Ellis, and who always has practical advice and is surprisingly patient with students and their fabrication problems.

Finally, my research would not have been possible without the generous funding I have received from the National Institute of Health Grants R01 GM090320 and DA030320.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

(c)DNA	(Complimentary) DeoxyriboNucleic Acid
(h/d)HSF	(human/drosophila) Heat Shock Factor
(q)PCR	(quantitative) Polymerase Chain Reaction
(t)RNA	(transfer) RiboNucleic Acid
AFM	Atomic Force Microscopy
BBS	Background Binding Sequence
CAD	Computer-Aided Design
CE	Capillary Electrophoresis
CHK2	Checkpoint Kinase 2
CLADE	Closed Loop Aptameric Directed Evolution
CMACS	Continuous-flow Magnetic Activated Chip-based Separation
DBD-TD	DNA Binding Domain and Trimerization Domain
DEPC	Diethylpyrocarbonate
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
FACS	Fluorescence Activated Cell Sorting
F-EMSA	Fluorescence-EMSA
FP	Fluorescence Polarization
GFP	Green Fluorescent Protein
GSH	Glutathione (reduced)
GST	Glutathione S-Transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HiTS-FLIP	High-Throughput Sequencing-Fluorescent Ligand Interaction Profiling
HiTS-RAP	High-Throughput Sequencing-RNA Affinity Profiling
MARAS	Magnetic-Assisted Rapid Aptamer Selection
MEDUSA	Microplate-based Enrichment Device Used for the Selection of Aptamers
MMLV-RT	Moloney Murine Leukemia Virus Reverse Transcriptase
MMS	MicroMagnetic Separation
M-SELEX	Microfluidic-SELEX
NELF-E	Negative Elongation Factor E
Ni-NTA	Nickel-nitriloacetic acid
PBS	Phospate Buffered Saline
PMMA	Poly(methyl methacrylate)
QPASS	Quantitative Parallel Aptamer Selection System
QSAS	Quantitative Selection of Aptamers through Sequencing
RAPID	RNA APtamer Isolation via Dual-cycles SELEX
RNA-MaP	RNA on a Massively Parallel array
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SPR	Surface Plasmon Resonance
UBLCP1	Ubiquitin-Like domain containing CTD Phosphatase 1
VDC	Volume Dilution Challenge

CHAPTER 1

DEVICES AND APPROACHES FOR GENERATING SPECIFIC HIGH-AFFINITY NUCLEIC ACID APTAMERS[‡]

ABSTRACT

High-affinity and highly specific antibody proteins have played a critical role in biological imaging, medical diagnostics, and therapeutics. Recently a new class of molecules called aptamers has emerged as an alternative to antibodies. Aptamers are short sequences of nucleic acids that can be generated and synthesized in vitro to bind to virtually any target in a wide range of environments. They are in principal less expensive and more reproducible than antibodies, and their versatility creates possibilities for new technologies. Aptamers are generated using libraries of random nucleic acid sequences that are subjected to affinity selections for binding to selected target molecules. This is commonly done through a process called Systematic Evolution of Ligands by EXponential enrichment, or SELEX, in which target-bound sequences are isolated from the pool, amplified to high copy numbers, and then reselected against the desired target. This iterative process is continued until the highest affinity sequences dominate the pool of nucleic acid molecules. Traditional selections require a dozen or more laborious cycles to converge on and isolate aptamers, and they can takes months to complete as well as consume large quantities of reagents. However, new devices and insights from engineering and the physical sciences have contributed to a reduction in the time and effort needed to generate

[‡] The following article was written for submission to the journal: Applied Physics Reviews. Szeto, K. and Craighead, H.G. Devices and approaches for generating specific high-affinity nucleic acid aptamers.

aptamers. As the demand for these new molecules increases, more efficient and sensitive selection technologies will be needed. These new technologies will need to use smaller samples, exploit a wider range of chemistries and techniques for manipulating binding, and integrate and automate the selection steps. Here, we review new methods and technologies that are being developed towards this goal, and we discuss their roles in accelerating the availability of novel aptamers.

INTRODUCTION

Antibodies are a class of proteins that are highly selective and have high binding affinity for foreign and potentially harmful antigens that are encountered in the body. In an organism's immune response, proteins mutate into a vast library of antibody candidates and are screened in the immune system for their ability to bind to the antigen. Candidates that cannot bind the antigen are eliminated, while candidates that can recognize the antigen are retained and form an enriched pool of good binders. The surviving proteins can then be amplified and further mutated, screened and enriched to allow the binding affinity to mature until a small group of high affinity antibodies are generated. Harvested from organisms, these antibodies have become indispensable affinity reagents in research and medicine, used for applications such as imaging specific proteins and their structures in cells and tissues, diagnostics to detect and quantify the presence of molecules, and even in therapeutics. However, antibodies have a number of weaknesses that limit their technological application. For example, antibodies must be produced *in vivo* and harvested from a host organism, which results in problems with reproducibility such as the total yield and variability of antibody

types between batches. In addition, new antibodies require significant investments in time and cost from researchers, and can only be generated to target molecules that elicit an immune response. In contrast, a simpler and more generalized class of tight binding molecules based on nucleic acids could address the limitations of antibodies and enable new generations of affinity reagents and biotechnologies to emerge.

Nucleic acids have long been recognized as the simple molecular template on which all of life and its processes are encoded. Despite having only four constituent bases, the information content and complexity that can be achieved through their unique combinations is astronomical. Although dwarfed by the potential sequence and structural diversity of proteins of a similar length, which can utilize twenty different amino acids, sequences of nucleic acids can also form complex three dimensional structures and acquire comparable functionalities to proteins (a known feature of naturally occurring RNA sequences such as tRNA and rRNA). These intrinsic properties make nucleic acids the perfect molecular analogues to amino acids for generating functional biomolecules as alternatives to antibodies.[1-3] These nucleic acid molecules called aptamers, can be generated through *in vitro* selections. This is often done through an iterative process called Systematic Evolution of Ligands by EXponential enrichment (SELEX). SELEX involves generating a library of DNA or RNA molecules with random sequences (~10^{15} different sequences), screening the library for sequences that bind to the target of interest, partitioning the bound sequences from the unbound sequences, and amplifying the bound sequences into a new pool enriched for good binders (see Figure 1.1). This process can be continued until the strongest binding sequences enrich and dominate the pool. If the enriched

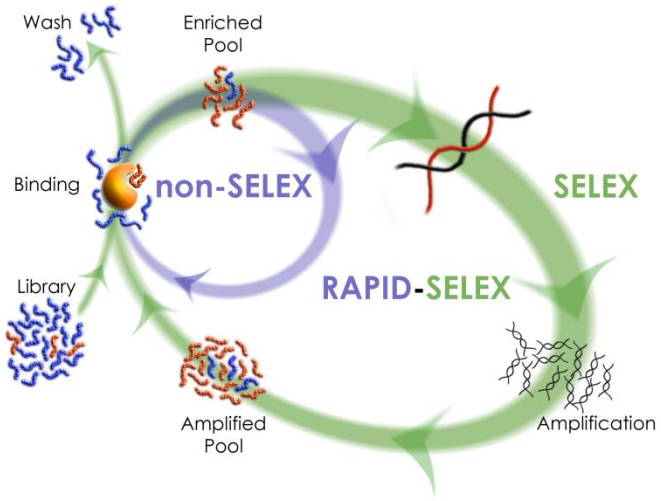


Figure 1.1. Simple process diagram illustrating and relating the different in vitro selection methods. The outer cycle (Green) is the classical SELEX cycle which iterates through binding, partitioning and amplifying target-bound sequences (represented in Red). The inner cycle (Violet) is non-SELEX in which cycles of binding and partitioning are performed without amplification. Combining both methods, RAPID-SELEX (Green and Violet) systematically implements non-SELEX to save time where possible, but incorporates classical SELEX cycles with amplifications to replenish sequence copy numbers or to drive up concentrations.

pool has converged on an aptamer, a sample of a few dozen candidates can be sequenced, and the dominant or consensus sequence identifying the aptamer can be determined.

Although the basic combinatorial chemistry of the aptamer selection process conceptually resembles that of antibodies generated *in vivo*, *in vitro* aptamer selections provide researchers with much more freedom and control in designing their desired affinity reagent. Not only can researches choose from an array of natural or modified nucleic acids and sequence lengths for their initial library, but aptamers can be generated to bind to targets that would be toxic or non-immunogenic to an organism.[4] In fact, aptamers can be generated to virtually any target from individual

metal ions to whole live cells.[5,6] Because aptamer selections do not take place inside of an organism, the environmental conditions for binding need not be physiological. Instead, these conditions can be tailored to better accommodate the ionic strength/content, pH and even temperature required for the desired application. Most importantly, these selections can make use of the tremendous advancements for amplifying, sequencing, and synthesizing nucleic acids, many of which are now commonplace protocols. This not only guarantees reproducibility between synthesis batches (the sequence is known), but aptamers are also much faster and cheaper to synthesize and modify compared to antibodies, and makes sharing or acquiring new reagents as simple as reporting the aptamer's sequence. Nucleic acids are also much more stable than antibodies and can be denatured into a linear sequence and reversibly folded back into its three dimensional structure, resulting in a much longer shelf life. For example, lyophilized nucleic acid molecules can last almost indefinitely even at room temperature, whereas proteins and antibodies must be frozen for long term storage. Finally, it is possible to link multiple aptamers to generate multivalent aptamers that can bind multiple identical targets or bind multiple sites on a single target to enhance target recognition, or even to create aptamers that recognize novel combinations of target molecules.[7-12]

The potential impact for aptamers and their resultant technologies is just beginning to be understood. However, as the demand for these powerful new molecules increases, so has the need for more powerful selection methods. In this review, we discuss the evolution of *in vitro* selections from simple bench-top techniques, to more efficient and informative technologies and ultimately into

miniaturized and fully automated chip-based devices. The basic principles and unique contributions of each technology toward improved aptamer selections are highlighted. Finally, we conclude with a discussion on the general advantages and disadvantages common among many of the technologies and propose roles that engineering and physics can play in further advancing this developing field of science.

CLASSICAL APTAMER SELECTIONS

Basic selection principles

The basic principles for *in vitro* evolution and selection of aptamers were first described over twenty years ago.[1-3] This conceptually simple optimization problem requires the determination of the total concentration for both the library and target molecules that maximizes the enrichment of the highest affinity aptamer. Given n types of sequences, the enrichment E_i for a sequence of type i can be expressed as the ratio of its fractional representation among all bound sequences to its starting fraction:

$$E_i = \frac{[T{:}S_i]/\sum_{i=1}^{n}[T{:}S_i]}{([T{:}S_i]+[S_i])/[S]_T} \quad (1)$$

where $[T{:}S_i]$ and $[S_i]$ are the concentration of bound and unbound sequences of type i, respectively. This must be maximized for sequences with the highest affinity (lowest dissociation constant K_d) given the set of n equilibrium binding equations:

$$[T{:}S_i] = \frac{[S_i][T]}{K_{d,i}+[T]} \quad i = 1 \ldots n \quad (2)$$

where $[T]$ is the unbound target concentration and $K_{d,i}$ is the dissociation constant for sequences of type i. Using the conservation equations:

$$[T]_T = [T] + \sum_{i=1}^{n} [T{:}S_i] \quad (3)$$

$$[S]_T = [S] + \sum_{i=1}^{n} [T{:}S_i] \quad (4)$$

where [*S*] is the concentration of all unbound sequences, the optimum total concentrations for the target $[T]_T$ and the library $[S]_T$ that maximize Equation 1 for high affinity sequences can be determined. This problem cannot be solved analytically for $n > 2$ and must be solved separately for each selection cycle as the relative frequency of sequences changes after every binding step, thus minimizing the total number of selection cycles needed to converge the pool onto an aptamer sequence. A number of theoretical efforts have attempted to address this optimization problem using numerical methods and/or approximations.[13-24] However, these theories assume equilibrium solution binding since time dependent models double the number of parameters (K_d is the ratio of the kinetic binding off-rate and on-rate: k_{off}/k_{on}) and generally require a significant amount of prior information about the initial library, such as the distribution of sequences ($[S_i]$) and affinities ($K_{d,i}$) to the target. In contrast, many selections are not performed in true equilibrium or with free molecules in solution, and little to no information is known about the initial library or its interaction with the target. In addition, the molecules cannot always be considered point particles as selections are often effected by molecule size and orientation.[25,26] Although some theoretical work has been done to fill these information gaps,[27-31] most parameters remain experimentally unknown, and researchers have adopted simple intuitive schemes to enforce competitiveness and stringency during selections, such as

gradually increasing the ratio of library to target molecules (i.e. the fold-excess of library to target) and/or reducing the quantity or concentration of target molecules. These selections are typically performed in one of three ways. (1) Filtering target molecules out of solution and a thus retrieving bound aptamers. (2) Filtering aptamers out of solution through a stationary phase of immobilized target molecules. (3) Spatially resolving and isolating target-aptamer complexes from unbound molecules by electrophoretic mobility differences.

Filtering targets: nitrocellulose filter binding

One of the first methods of selection was nitrocellulose membrane filtration, which adheres most closely to the theoretical models mentioned above (see Figure 1.2.A). Generally, the target molecules and the nucleic acid library are mixed together in solution and allowed to approach equilibrium. Target molecules and any sequences bound to them can then be partitioned from the solution by rapidly filtering the mixture over a nitrocellulose membrane.[3,32-34] The membrane concentrates target-bound sequences and allows unbound sequences to pass through. This process is fast and straight forward, and it is one of the most common partitioning methods used for selecting aptamers. However, since the non-specific affinity of nitrocellulose to amino acids is central to this technique, it only works well with protein targets. In addition, the large surface area of the filter enables a significant amount of free sequences to non-specifically adsorb to the membrane. This has been known to result in nitrocellulose binding sequences that can significantly hinder the enrichment of target-binding aptamers, or completely dominate the enriched pool. Extensive washing, or

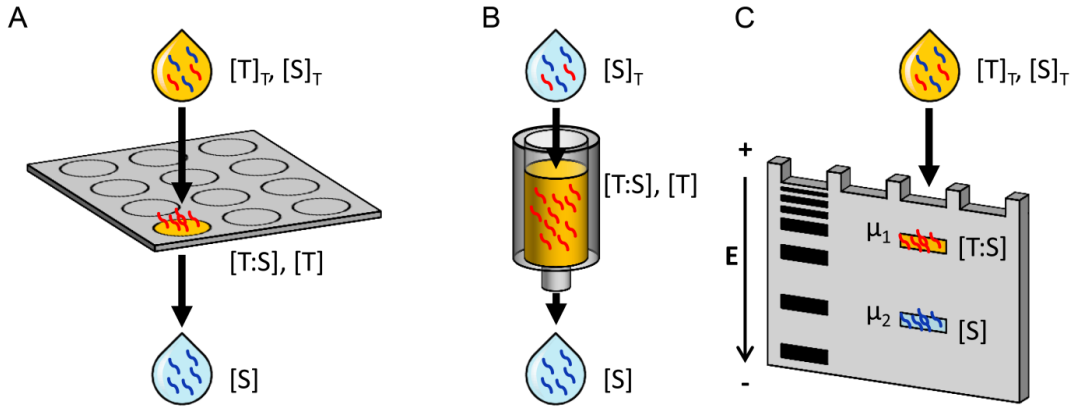


Figure 1.2. Schematics for the classical technologies used to partition binding sequences (Red) from non-binding sequences (Blue) when selected against target molecules (Orange). (A) Membrane filtration is used to non-specifically capture target molecules along with any bound sequences from an equilibrium mixture. Unbound sequences pass freely through the filter. (B) Affinity chromatography utilizes a column of backed resin which is functionalized with target molecules, and a library or pool of sequences which is passed through the column. Binding sequences are captured on target molecules while unbound sequences pass freely through the column. (C) Electrophoretic Mobility Shift Assays (EMSA) use gels to separate equilibrium mixtures containing bound and unbound sequences and targets. By applying an electric field, bound and unbound molecules, which have different sizes and charges, will migrate at different rates though the gel, which prevents the isolated populations from mixing.

negative selections that remove filter-binding sequences, could be used to improve aptamer enrichments.[35]

Filtering aptamers: affinity chromatography

Another method that was initially applied to aptamer selections is affinity chromatography (see Figure 1.2.B). This technology is traditionally used to separate and purify components from a mixture of molecules, some of which have a specific affinity to or interaction with a stationary phase (resin packed into a column) through which the mixture flows. By immobilizing target molecules to an affinity resin, the injected nucleic acid library will become enriched for target-binding sequences, which

are retained within the column. Non-binding sequences simply flow through as waste. Sequence-bound targets can then be eluted off of the resin chemically. In contrast to nitrocellulose membrane filtration, affinity chromatography can be used to immobilize proteins as well as small molecule targets.[1] Given its simplicity and familiarity among many laboratories, affinity chromatography has become the dominant method for small molecule selections.[35-39] However, this method generally requires large quantities of target to achieve sufficiently high loading onto the entire column, and can suffer from non-specific binding of the nucleic acids to the resins requiring extensive washing or negative selections.[35] In addition, this method requires the incorporation of an affinity tag to target molecules for immobilization; and although this is commonplace for proteins, it can be difficult to achieve for small molecules and restricts the modes for aptamer binding.

Isolating bound complexes: electrophoretic mobility

Early selection methods also included Electrophoretic Mobility Shift Assays (EMSA) (see Figure 1.2.C). Like the older nitrocellulose filtering method, these selections allow target-sequence mixtures to equilibrate together in solution. However here, the bound and unbound populations can be spatially separated by adding the mixture to a gel and applying an electric field *E*.[40,41] Depending on their shape, size and charge, each population has a different mobility μ that causes unbound target, unbound sequences, and bound complexes to migrate through the gel with different velocities *v*.

$$v = \mu_e E \quad (5)$$

The gel prevents any mixing or significant dispersion of the isolated populations that would otherwise take place in solution. The band containing bound sequences can then be imaged with radioactivity (or fluorescence), cut out, and crushed to allow the sequences to easily diffuse back into solution. This selection method almost completely eliminates background binding as well as the need for washing or negative selections. However, the modifications required, especially radioactivity, are generally undesirable, and selections with different targets can have differing results and make separations difficult to resolve, especially with small targets. In addition, the separation step is slow and can be far from equilibrium, providing opportunities for bound sequences to dissociate during the lengthy process.

IMPROVING CLASSICAL SELECTIONS

Filtering targets: magnetic beads

One of the key advantages of filter binding is its ability to easily and rapidly concentrate sequence-bound target molecules, while separating them out from a solution of unbound sequences. However, due to the nature of the filters, selections suffer from background binding and are limited to proteins. To resolve some of these limitations, techniques utilizing magnetic beads were devised (see Figure 1.3.A). [42,43] Magnetic bead-based SELEX allows selections to be performed to any target that can be immobilized onto the beads. In addition, selections can be performed in smaller volumes with much less target, and the bead-bound target is rapidly concentrated and isolated from the bulk solution simply by using a permanent magnet. The magnetic beads can then be aggressively washed and concentrated again if

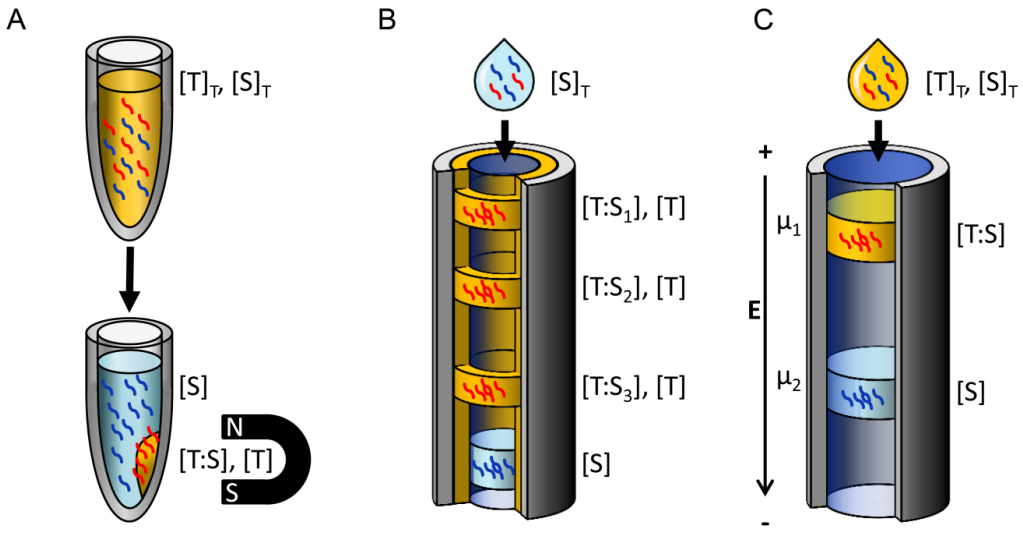


Figure 1.3. Schematics for improvements on the classical technologies used to partition binding sequences (Red) from non-binding sequences (Blue) when selected against target molecules (Orange). (A) Magnetic beads are used to capture target molecules along with any bound sequences from an equilibrium mixture into a small localized pellet. Unbound sequences remain in the bulk solution. (B) Capillary affinity chromatography utilizes a capillary with its inner wall functionalized with target molecules, and a library or pool of sequences which is passed through the capillary. Binding sequences are captured on target molecules while unbound sequences pass freely through the capillary. This flow regime can achieve such high resolutions, that individual binding species can be resolved into separated bands and isolated. (C) Capillary electrophoresis uses the high resolution and non-mixing regime of capillaries to separate equilibrium mixtures containing bound and unbound sequences and targets. By applying an electric field, bound and unbound molecules, which have different sizes and charges, will migrate at different rates though the capillary, which prevents the isolated populations from mixing.

needed, and used directly in PCR amplifications. A variation on this technology uses magnetic beads in a slightly different manner. In a method called Aptamer Selection Express (ASExp), target molecules and the library are mixed together in solution just like in filter binding or EMSA. However, in this case, the library consists of double stranded sequences that must dissociate in order to reveal a single stranded aptamer. This imposes a binding threshold on potential aptamers and allows distinguishing single stranded target-binding sequences from double stranded non-binding sequences.

This difference is exploited by recovering single stranded sequences using magnetic beads coated with long, random, single-stranded sequences. Aptamers have been selected in a single round using this technique.[44] In another variation of this technique, the library is fluorescently labeled in order to quantify the level of binding after each round of SELEX. This method, called FluMag-SELEX, allows sensitive measurements of binding to be made due to the efficient concentration of target bound sequences on the magnetic beads, and eliminates the need for radioactivity-based measurements.[45]

Magnetic bead-based techniques are of interest because the ability to manipulate the beads externally using magnetic fields has great potential for more sophisticated and automated technologies. A recent example of this, called Magnetic-Assisted Rapid Aptamer Selection (MARAS), utilizes a magnetic field to not only isolate bead-bound sequences, but also to actively remove weakly bound sequences.[46] This additional sequence discrimination is achieved by placing an equilibrated mixture within a solenoid and applying an alternating current. This results in an alternating magnetic field whose strength and frequency can be adjusted such that lower affinity sequences begin to dissociate from the target molecules. This is due to the viscous dissipative forces imparted on bound sequences as the beads oscillate within the field, enabling selections to be successfully performed in only a single cycle and has been shown to generate aptamers whose affinity increased with increasing field strength and frequency.

Filtering aptamers: miniaturized affinity chromatography

There have been many new technologies developed towards advancing affinity chromatography. One example has been the miniaturization of affinity chromatography through the use of microcolumns, which are orders of magnitude smaller than conventional columns. By slowly injecting sequences into the packed microcolumns, the entire pool is efficiently sampled and high affinity aptamers are retained within the small column. The use of microcolumns was optimized for maximum enrichments of aptamers, revealing critical target concentrations that could be explained by geometric constraints for steric hindrance.[47,48] This concentration as well as the small column volume reduces the amount of target needed by several orders of magnitude. Furthermore, tests that varied the flow rates resulted in enrichment trends that scaled oppositely from the time-dependent kinetic binding model where the concentrations of unbound and bound molecules are also dependent upon their position *x* along the column:

$$\frac{d[T{:}S_i](x)}{dt} = k_{on,i}[T(x)][S_i(x)] - k_{off,i}[T{:}S_i](x) \quad (6)$$

Also using the microcolumns, a modified SELEX method, called RAPID-SELEX, was demonstrated, which generalized the SELEX method to permit the systematic skipping of amplification steps.[49,50] This is advantageous as it significantly reduces the time and cost of performing selections. However, despite producing improved aptamer enrichments, RAPID-SELEX is viewed traditionally as less competitive than conventional SELEX due to diminishing sequence concentrations between infrequent amplifications. Together, these results highlight the importance of empirical

characterization and optimization of selection technologies. Interestingly, multiple microcolumns can be connected together allowing for simultaneous in-line negative selections, multiplexing, and parallelization. Recently, this technology was scaled up to a Microplate-based Enrichment Device Used for the Selection of Aptamers (MEDUSA), and has a 96-well microplate format to allow high-throughput and potentially automated plate-based selections and processing to be performed.[48]

In a similar modification of affinity chromatography, MonoLEX utilizes a narrow affinity column in the form of a capillary.[51] As is typical of chromatographic separations, the aptamer library can resolve itself along the capillary into different populations of binders (see Figure 1.3.B). In a manner analogous to connecting and disconnecting microcolumns in series, by physically cutting the capillary column into small fragments (i.e. ~40 segments), aptamer populations can be isolated and recovered from each fragment individually and characterized with the highest affinity binders generally residing in earlier segments. The capillary columns' efficient separation of subpopulations into narrow and well-resolved distributions allows MonoLEX to isolate aptamer candidates in a single round.

Affinity chromatography has also been taken to the absolute lower limit of miniaturization by performing SELEX to targets immobilized on a single bead.[52] With this method, aptamers are generated by incubating a single target-functionalized microbead with a fluorescently labeled library. By significantly reducing the number of target molecules, it is assumed that only the highest affinity sequences can be bound, due to competition with low affinity sequences for the few available binding

sites. This can be seen in the limit where the concentration of unbound target [T] approaches zero:

$$\langle K_d \rangle = \frac{[T][S]}{[T{:}S]} \approx \left[\sum_{i=1}^{n} \frac{\left(K_{d,i}\right)^{-1}[S_i]}{[S]_T} \right]^{-1} \quad (7)$$

[T:S] is the total concentration of bound sequence (and target) molecules, so as the concentration of unbound sequences [S_i] goes up, the average of the dissociation constant <K_d> of all bound sequences decreases. Once washed, the bead can be collected and subjected to PCR amplification. The fluorescently labeled library used in this method allows the level of binding to the bead to be monitored under a microscope during/after each cycle, and the bulk binding affinity to target molecules to be determined quickly via Fluorescence Anisotropy. Using this technique, high affinity aptamers were generated in just two rounds of SELEX, but only a single round may be necessary. Although not demonstrated, this simple process could be scaled up to include multiple single-target beads or even automated. However, identifying, isolating and handling single beads may limit this technology's capacity for more streamlined processing.

Isolating bound complexes: capillary electrophoretic mobility

As discussed earlier, target-binding sequences can be distinguished from non-binding sequences by a change in mobility when bound to target molecules. This is achieved primarily due to differences in the net charge between bound and unbound molecules. EMSA-SELEX has the advantage of equilibrium binding and (non-equilibrium) separations on a gel, which eliminates the need for washing and negative

selections. However, the use of radioactivity and the need to cut and process gels makes this method tedious and difficult to adapt for higher throughput or more automated selections. This limitation has been overcome through the application of Capillary Electrophoresis (CE). CE-SELEX utilizes integrated fluidics with an electric field applied across the system (see Figure 1.3.C). Using a capillary with a small internal diameter allows separations to occur in solution rather than requiring a gel (see Eq. 5).[53-57] This is made possible by the laminar flow regime (low Renyolds number) of the capillary, which is largely free of turbulent flow and mixing (other than diffusion). In addition, sensitive UV detectors can be used to identify the band of bound complexes as populations of molecules migrate through the capillary, eliminating the need for radioactivity. These selections are fast and work well with large target molecules and result in efficient and high resolution separations. However, selections using different targets can vary greatly and the optimal binding and running conditions may need to be determined beforehand. In addition, separations with targets that only result in modest shifts in mobility can be nearly impossible to resolve. Resolution can be particularly problematic if the capillary is overloaded with sequences that can bury the desired peak. Therefore, CE-SELEX can only handle small volumes (~100 nL) and use starting libraries at incredibly high concentrations and several orders of magnitude less diverse (10^{12}-10^{13}) than conventional selections in order to prevent overloading and achieve good separations.

With appropriate modifications, CE-SELEX has been used to estimate binding affinity during selections. This is because information can be obtained about the populations of unbound target, unbound sequences, as well as the bound complexes

and their gradual dissociation as they migrate past the detector .[58] In addition, distinguishing the various populations can be made easier by incorporating fluorescence capabilities into the selection. A novel CE-SELEX method based on these modifications, called Non-SELEX, was demonstrated, which completely eliminates all amplification steps allowing aptamers to be generated in a single round.[59] This method involves collecting the band of bound sequences and re-injecting them into the CE system for additional cycles of purification. However, due to the small volume constraints, only a tiny fraction of the collected pool can be re-injected for the next cycle, significantly limiting the number of sequences that can be sampled. Some recent optimization has been done using larger capillaries and multiple pool injections to improve sequence sampling after each cycle.[60]

Automation and parallelization

In addition to developing new and more efficient selection methods, researches have also accelerated aptamer discovery through robotic technologies. For example, semi-automated and parallel selections are possible using target-functionalized magnetic beads and microplates. Using an array of magnetic wands, magnetic beads and bound sequences can be captured, removed from solution and transferred to fresh wells on the microplate, allowing the automation of most of the selection steps apart from amplifications.[61,62] Completely automated selections have also been demonstrated that use magnets to retain beads in the microplate wells as solutions are exchanged instead,[63] although these protocols were later changed to incorporated membrane filtration to capture and wash the magnetic beads.[64-67] Affinity columns

have also been used in automated selections to filter out affinity-tagged targets and bound sequences.[68] However, an example that does not require any filtering uses simple immobilization of targets directly to the surface of microplates.[69]

As an alternative to generally low-throughput automation, substantial scaling up though massive parallelization of selections allows the average time per target to be reduced proportionally, assuming steps for each target can be performed simultaneously. Using a similar target-functionalized 96-well microplate, multiplexed and massively parallelized SELEX has been demonstrated through simultaneous processing and analysis.[70] In addition, the 96-well microplate-based affinity microcolumns discussed earlier were also used to performed simultaneous processing and analysis.[48] These highly parallelized technologies are ideal compliments to automated microplate-based protocols and combining these two approaches may allow researchers to achieve automated and massively parallelized selections using similar robotics systems.

INSTRUMENTATION FOR PARTITIONING AND DIRECT READOUT OF BINDING

Sorting sequences with flow cytometry

As interest in aptamers has grown, researchers with expertise outside of classical SELEX have begun to recruit sophisticated systems and instruments to help perform selections. This has been particularly fruitful with techniques that not only separate populations, but also provide information about the bulk binding behavior of the aptamer pools, as with CE-SELEX. One of these technologies, called flow

cytometry, is usually used to count or measure properties of cells at very high rates by rapidly interrogating individual cells in a flow stream (thousands per second). A specialized form of flow cytometry, called Fluorescence-Activated Cell Sorting (FACS), can be used to separate different populations. These systems have been applied to aptamer selections against whole cells using a method called FACS-SELEX, which is useful not only for quantifying and sorting aptamer-bound cells from the bulk solution of unbound sequences, but can also be used to separate specific populations of cells, such as living cells from dead cells.[71,72]

A key advantage of flow cytometric systems is the ability to probe one cell at a time. Therefore, a natural modification of these systems involves replacing cells with beads (see Figure 1.4.A). Selections with beads that are each bound to only a single type of sequence have been demonstrated where high affinity beads are identified by incubating them with their target and imaging the level of binding with fluorescently-labeled targets[73] or antibodies (see Figure 1.4.B).[74] This method is unique in that the highest affinity aptamers are assayed and identified directly by the brightest beads. These beads can be picked up via a micropipette and sequenced, allowing selections to be completed in a single round. Using multiple colors, additional or alternative binding requirements (which are assayed fluorescently), can be imposed. A simple two-component test mixture of beads was used to demonstrate this kind of multicolor FACS detection. Recently, a true single sequence/bead selection called Particle Display utilizing the sorting capabilities of FACS was fully demonstrated.[75] Currently FACS technologies are restricted to detecting and sorting about 10^8 beads/cells which significantly reduces the number of sequences that could be

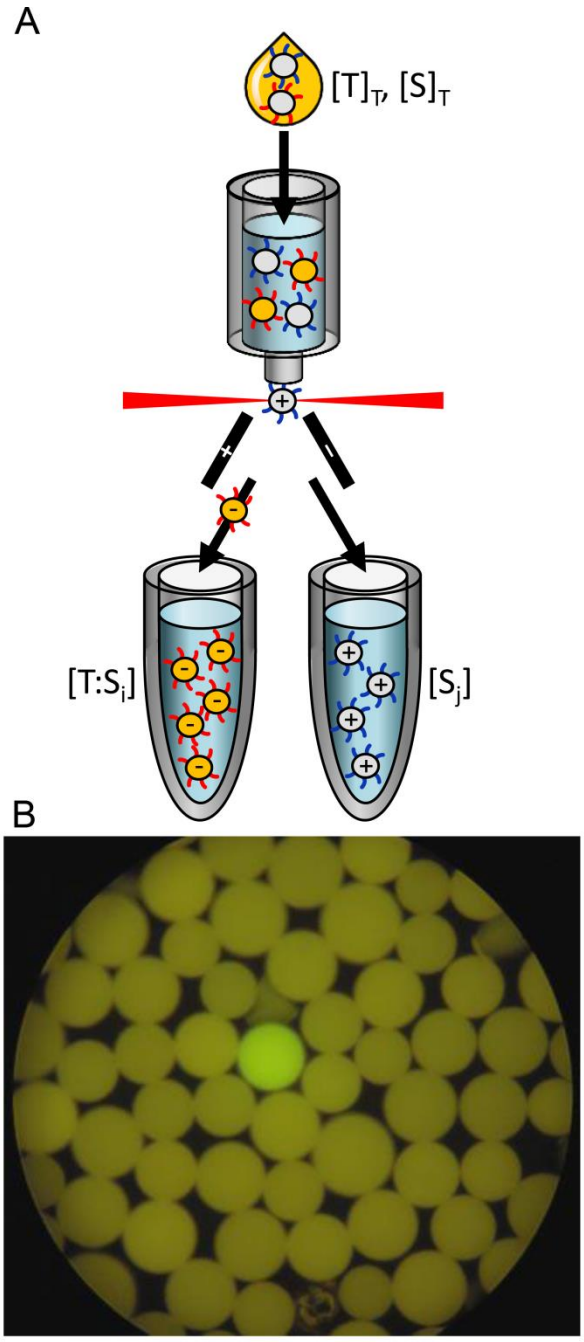


Figure 1.4. Schematic for bead-based selections utilizing Fluorescence Activated Cell Sorting (FACS). (A) Beads, each of which are coated with many copies of a unique sequence, are partitioned into binding sequences (Red) and non-binding sequences (Blue) based on binding to fluorescently labeled target molecules (Orange). Fluorescence measurements of each bead is used to identify brightly labeled beads which contain tight binding sequences and are sorted by imparting a charge on the bead and deflecting it in an electric field. (B) An example of a brightly labeled bead indicating the presence of a tight binding sequence as a potential aptamer candidate (Reproduced from Yang et al, Immunofluorescence assay and flow-cytometry selection of bead-bound aptamers, Nucleic Acids Res, 2003, 31,10, by permission of Oxford University Press)

screened in this format. Particle Display resolves this sampling limitation by performing the first selection cycle without FACS so that a full size library can be screened, generating a partially enriched pool which can be much more efficiency utilized via FACS. Interestingly, the fluorescent signature of target binding to the beads not only allows a proportional readout of binding affinity, but the near bulk binding characteristics of ~10^5 copies of each sequence on its bead also overcomes the stochastic binding of single complexes and allows for unparalleled sensitivity and confidence in discriminating and sorting between high and low affinity sequences. Currently, FACS systems can also separate mixtures into six subpopulations or distribute individual sorted objects into the wells of microplates. This highlights FACS-based methods scalability, where the ability to analyze and sort using multiple colors and sorting channels can be advantageous for multiplexing or high-throughput parallelization through downstream microplate processes, especially for selections directly to fluorescent targets such as fluorescent proteins or dyes.[76]

Imaging and detection with chips: surface-bound targets

The ability to directly observe and image interactions between sequences and target molecules has tremendous advantages over a typically blind selection strategy. This can be as simple as observing binding events under an optical microscope, such as with the FACS-based methods. However, it is often difficult to fluorescently label target molecules, and this can have undesirable consequences in aptamer-target recognition. Furthermore, adding fluorescent antibodies complicates the selection process and is not always a possible alternative. A more general and straight forward

approach involves fluorescently labeling the nucleic acid library, as in the single-bead affinity selection discussed earlier. In addition, imaging can be done easily by immobilizing target molecules on a flat substrate. Selections are done by incubating a target-functionalized coverslip with a fluorescently labeled library. The coverslip can then be extensively washed and imaged under a microscope to find bright and highly localized spots indicating aptamer binding, and the aptamers recovered thermally through heat elution in solution. Using this simple method, an inexpensive and rapid one-step selection was demonstrated.[77] Simply monitoring selections this way is useful to ensure that binding is taking place and to assess the degree of background binding which may require additional washing or negative selections. In addition, the use of fluorescently labeled target molecules with the fluorescent library can be used to image both simultaneously and find co-localized spots of aptamers binding specifically to target molecules on the surface. This ability to visually discriminate between target-aptamer interactions from non-specific background binding can have significant advantages for improving aptamer enrichments.

A similar selection method, called NanoSelection, utilizes an Atomic Force Microscope (AFM).[78] This method incorporates fluorescently labeled sequences attached to beads using a similar single sequence/ bead scheme as the FACS-based methods. Using a fluorescence microscope-AFM hybrid system, beads that bind to a target-functionalized chip are imaged using the fluorescence microscope. The AFM is then used to generate a local image of the bead on the chip, and then the AFM's tip used to physically "spear" or displace the bead from the surface for retrieval and analysis. Similarly to a single sequence/bead method discussed earlier, NanoSelection

was demonstrated successfully using a two-component text mixture of aptamers with non-aptamers and is best suited to small libraries on beads. In addition to locating bound beads, it should be possible to observe their dynamics while in solution, using their bound-state fluorescence intensity and duration to pinpoint the most tightly bound beads as well as their mobility when unbound, to determine relative binding affinities between candidate beads.

In another method called AFM-SELEX, the library is biotinylated and bound directly to a streptavidin-functionalized AFM tip instead of to beads (see Figure 1.5.A).[79] This method is particularly interesting because as the tip probes the target functionalized chip, sequences in the library that have an affinity for the target are forced to compete against the biotin-streptavidin interaction. This interaction is quite strong and acts as a binding threshold that sequences must surpass in order to detach from the tip and remain bound to the chip. In doing so, images of the chip surface as well as thousands of force curves are generated (see Figure 1.5.B). Successful selections using AFM-SELEX showed gradual increases in the average force exerted on the AFM tip using adhesion force analysis during each selection cycle. These results clearly indicate the enrichment of higher affinity aptamers to the target. Although this method provides a competitive binding threshold and data regarding the bulk binding behavior of the library, this technique has significant limitations in the number of library molecules that can be bound to the AFM tip and probed, and is currently not easily scalable for higher throughput aptamer selections.

A technique that is naturally sensitive to surface bound molecules is Surface Plasmon Resonance (SPR), which excites plasmons in thin metal films and uses their

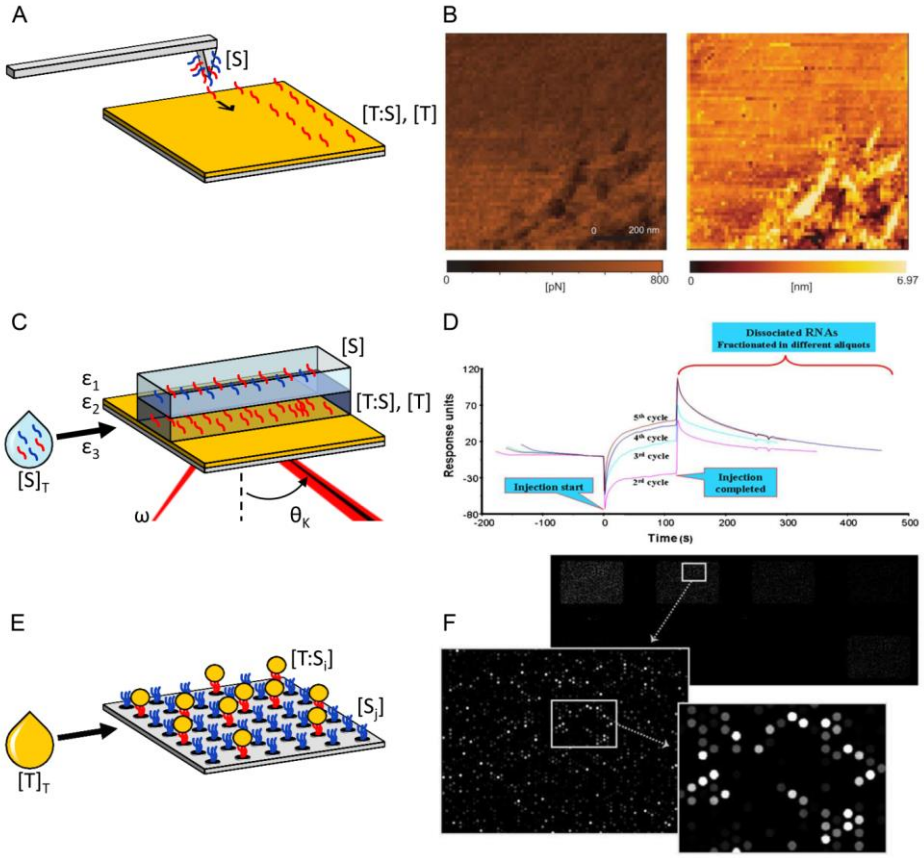


Figure 1.5. Schematics for chip-based technologies used to partition binding sequences (Red) from non-binding sequences (Blue) when selected against target molecules (Orange). (A) Atomic Force Microscopy (AFM) utilizes a library or pool of sequences bound to the AFM's probe tip. By running the tip over a surface of target molecules which are immobilized onto a chip, tight binding sequences which can overcome their binding energy to the AFM tip can detach and bind to target molecules on the surface. (B) Example force and height AFM images acquired using a sequence-coated probe tip on a target coated chip surface. (Reproduced from Miyachi et al, Selection of DNA aptamers using atomic force microscopy, Nucleic Acids Res, 2010, 38,4, by permission of Oxford University Press). (C) Surface Plasmon Resonance (SPR) utilizes target molecules which are immobilized onto a chip. By flowing a library or pool of sequences over the chip, tight binding sequences can bind to target molecules on the surface while non-binding sequences flow past. This can be imaged and quantified through proportional changes in the surface plasmon conditions. (D) Example SPR response curves for various enriched pools showing the binding and unbinding of sequences. (Reprinted from Analytical Biochemistry, 342/10, Misono TS, Kumar PKR, Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance, 312-317, Copyright (2005), with permission from Elsevier). (E) Microarrays utilize a library of predetermined sequences which are then addressed and synthesized in situ on specific elements of the chip array. By exposing fluorescently labeled target molecules to the array, tight binding sequences can bind to target molecules on the chip and individually identified through fluorescence imaging. (F) Example fluorescence image of individual sequences of a microarray binding florescent targets. (Reproduced from Fischer NO, Tok JBH, Tarasow TM, Massively Parallel Interrogation of Aptamer Sequence, Structure and Function. PloS one 3 (7). Copyright (2008) the Authors)

sensitivity to conditions near their surface to monitor binding kinetics (see Figure 1.5.C). This is typically achieved by reflecting a p-polarized laser with angular frequency ω (and speed of light c), in a medium with dielectric constant of ε_3, of off of a thin film of metal (i.e. gold) and monitoring changes in the low intensity reflection angle θ_K that satisfies the changing plasmon resonance condition of wavenumber $K(\omega)$. As molecules bind to and unbind from the surface, it changes the effective dielectric constant ε_1 of the medium in contact with the metal film with dielectric constant ε_2.

$$K(\omega) = \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_{1+}\varepsilon_2}} = \frac{\omega}{c}\sqrt{\varepsilon_3}\sin\theta_K \quad (8)$$

One key advantage of SPR systems is that they integrate fluidics and flow cells with surface-bound target selections and can have several flow channels, which potentially allow for multiplex or parallelized aptamer selections. By injecting and exposing sequences with concentration $[S]$ to the chip surface, the pool's bulk on-rate k_{on} for binding can be estimated:

$$R(t) = \frac{R_{MAX}[S]}{K_{d+}[S]}\left[1 - e^{-(k_{on}[S]+K_d)t}\right] \quad (9)$$

where the $R(t)$ is the time-dependent response of the SPR resonance angle and R_{MAX} is maximum signal. Since this measurement is performed under flow, the concentration of unbound sequences is constant making the binding trends independent of the target concentration. In addition, washing/elution of sequences off the chip allows the bulk off-rate k_{off} for unbinding to be estimated as sequences slowly dissociate and are collected for future selection cycles.

$$R(t) = R(0)e^{-k_{off}t} \quad (10)$$

SPR in SELEX was initially demonstrated as a clean-up selection following many cycles of nitrocellulose filter binding and was used to determine the bulk binding affinity of the final pool.[80] However, SPR has also been utilized as the sole SELEX platform, and showed steady increases in the time-dependent binding rates after each cycle, indicating the enrichment of higher affinity binders (see Figure 1.5.D).[81] In addition, multiple fractions were collected during the dissociation step. Although not demonstrated, the highest affinity sequences are expected to take longer (on average) to dissociate due to a smaller k_{off} and hence reside primarily in the later fractions. Fractionating the dissociation phase could be used as an off-rate threshold for aptamers in SPR-SELEX. Although off-rates can be determined easily in SPR, accurate determination of on-rate and the equilibrium binding affinity requires multiple concentrations of the library/pool to be probed in order to fit all the parameters of Eq. 9. In addition, since the determination of the on-rate requires the concentration of unbound sequences to remain constant at all times, limits on the possible flow rates that can be used have to be imposed, which must be low enough for efficient sequence sampling but high enough to overcome mass limited transport.

Imaging and detection with chips: surface-bound sequences

Performing selections with surface-bound molecules has resulted in a range of varying chip-based methods. With the emergence of technologies such as microarrays, precise localization and addressing of many different molecules onto a single chip can be achieved. This has been applied to multiplexing and parallelizing chip-based

selections with multiple surface-bound target molecules and concentrations.[82-85] Although initially only a few dozen isolated spots were placed manually on chip surfaces, automated and robotic systems have enabled much higher density arrays of molecules to be generated. Using modern lithographic techniques for microscale patterning, microarrays can now be fabricated with nearly 10^5 individual spots. This high density configuration has been adopted as a means to probe individual sequences of a nucleic acid library with sensitivities and sampling depths comparable to Particle Display using FACS (see Figure 1.5.E). For aptamer selections, random and unique sequences are individually synthesized *in situ* and addressed on predetermined spots on the microarray. Fluorescently labeled target molecules can then be exposed to the array and imaged to identify potential aptamer candidates via bright localized spots (see Figure 1.5.F). In a selection method called CLADE (Closed Loop Aptameric Directed Evolution), a set of the brightest spots (~1%) are identified and their associated sequences are controllably mutated and synthesized onto a new chip for additional selection cycles.[86,87] This is comparable to performing a pre-selection to compensate for the lower sampling capabilities of the microarray. By mutating the top performing sequences, tighter binding variants can be identified and further matured.

Although microarray-based selections such as CLADE have tremendous potential, they require a small starting library of individually synthesized and predetermined sequences. For this reason, microarrays have also been used post-selection to screen enriched pools which contain significantly fewer unique sequences and even fewer aptamer candidates, which can be identified through sequencing and chosen based on population metrics such as multiplicity or enrichment.[88-90]

Aptamers isolated from selections can also be studied and optimized through mutations and assayed on microarrays to identify key features which are critical to the highest affinity sequence such as the minimal aptamer structure, structure-function relationships, as well as conserved sequence motifs.[90-92] The majority of these selection microarrays rely on the imaging and detection of fluorescently-labeled molecules (target or library). However, since microarrays share a chip format with a number label-free detection methods, arrays of sequences have been generated and screened without fluorescent modifications to either the target or sequence. Using an array of electrodes, sequences on the array which are bound to gold nanoparticles can generate measurable current and/or voltage signals when bound to target molecules via electrochemical detection.[93] Although this technique can be used to screen multiple sequences simultaneously and is label-free, this method can only work with target molecules that are electroactive and will generate a sufficiently large signal upon binding to the electrode. As an alternative, arrays of sequences have also been generated and screened using SPR which does not require fluorescent labels or special properties of target molecules.[94]

INTEGRATED SELECTIONS ON MICROFLUIDIC CHIPS

Many researchers are interested in miniaturizing other selection technologies onto chips, as well as their possible automation. Initial efforts aimed to miniaturize conventionally large robotic systems used in automated SELEX.[95] This was done through the integration of a fluidic microchip, and was used to successfully select an aptamer through an automated process that incorporates all the steps from binding to

amplification. Although the binding step takes place off-chip in an affinity capillary and valving and other fluidic controls are performed using large external fluidic instruments, this microfluidic SELEX platform is a significantly smaller and simpler system than other automated technologies, and represents an important step toward more integrated chip-based solutions for fluidic selections. Recently, significant research efforts have focused on capitalizing even more on the micro- and nanoscale processes used in the fabrication of functional chips. This has led to the development of a number of devices that can be used to integrate and perform a number of selections steps. These new selections benefit not only from the reduced volumes and the unique phenomena that occur at the microscale, but also from the inherent scalability of the fabrication methods, putting chip-based multiplexed and automated aptamer selections within reach.

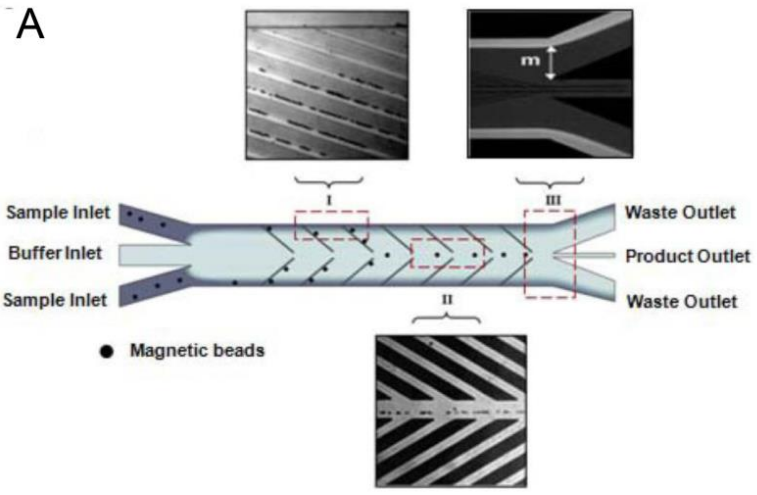
Filtering targets: magnetic beads

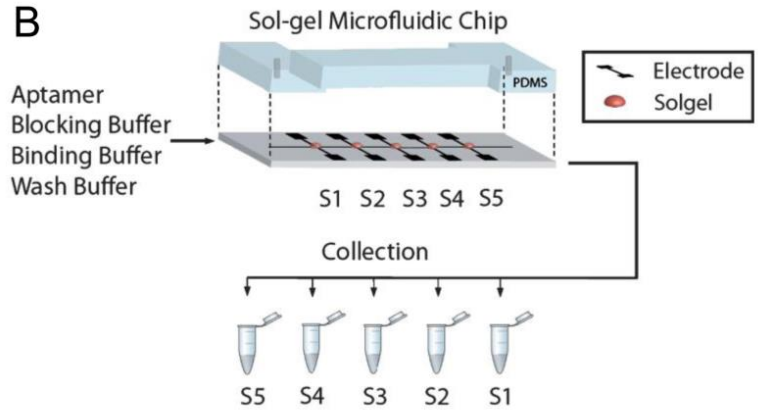
Due to their easy manipulation, magnetic beads have been incorporated into microfluidic systems to further improve the efficiency of aptamer selections. One of the first such devices is described in a method called Microfluidic SELEX (M-SELEX) (see Figure 1.6.A). Here, a typical incubation with a library and target-functionalized magnetic beads is allowed to bind and then is injected into a Continuous-flow Magnetic Activated Chip-based Separation (CMACS) device.[96] Magnetic beads and bound sequences are then magnetophoresed out of laminar input streams, into a non-mixing center stream for collection. This is done through a magnetic field gradient ∇B generated by integrated ferromagnetic structures.

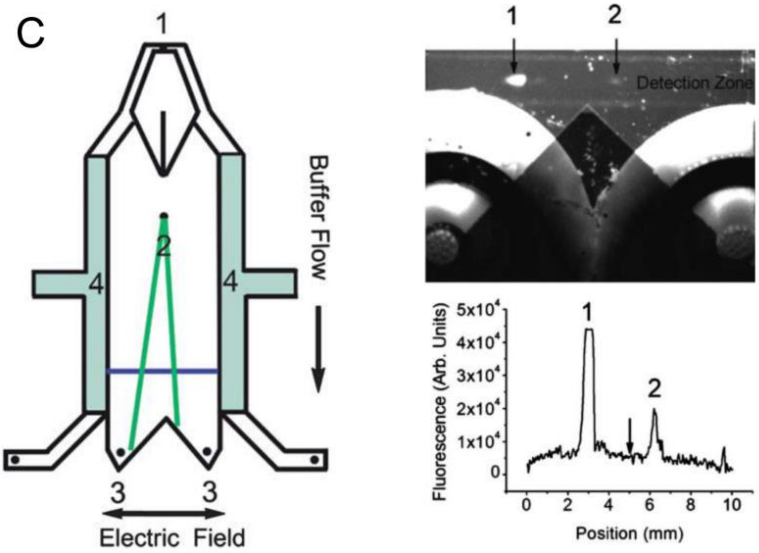
$$F = m\nabla B \quad (11)$$

Together, the magnetic beads and the CMACS device allow highly stringent selections with very low concentrations of bead-bound targets to be efficiently performed (see Eq. 7) and have resulted in an aptamer being isolated in a single cycle. A simpler MicroMagnetic Separation (MMS) chip was also demonstrated that does not limit the magnetic beads' residence time in the microfluidic chip. In this case, ferromagnetic grids are used to capture and retain magnetic beads as they flow into the device, allowing a wider range of flow rates as well as long and extensive washes to be performed in the channel.[97,98] In a technique utilizing dilutions[99] called Volume Dilution Challenge (VDC), VDC-MSELEX uses the MMS chip to capture and re-concentrate magnetic beads following dilutions of up to one thousandth of the equilibrated solution, which helps to irreversibly dissociate aptamers with poorer off-rates.[100] M-SELEX has also been coupled to high-throughput sequencing in a method called Quantitative Selection of Aptamers through Sequencing (QSAS), to analyze the copy numbers and enrichments of millions of aptamer candidates in multiple pools.[101] Recently, this has been further improved through a Quantitative Parallel Aptamer Selection System (QPASS), which utilizes a microarray of thousands of the top multiplicity aptamer candidates to measure each of their binding affinities in parallel.[89]

Figure 1.6. Examples of simple microfluidic devices which integrate the three basic methods of partitioning binding sequences from from non-binding sequences. (A) A microfluidic device which can capture and partition target-functionalized magnetic beads out of a solution equilibrated with a library or pool of sequences. Microfabricated ferromagnetic structures capture and direct magnetic particles as they are injected into the device and diverted into an isolated collection stream free from unbound sequences (Reproduced from Lou X, Qian J, Xiao Y, Viel L, Gerdon AE, Lagally ET, Atzberger P, Tarasow TM, Heeger AJ, Soh HT, Micromagnetic selection of aptamers in microfluidic channels. Proceedings of the National Academy of Sciences of the United States of America 106 (9):2989-2994, Copyright (2009), the Authors). (B) A microfluidic device which utilizes silica sol-gel matrix to immobilize and isolate multiple target molecules in a single channel. A library or pool of sequences is passed over the target sol-gel spots allowing binding sequences to diffuse and bind to targets within the sol-gel. Unbound sequences pass freely out of the device (Reproduced from Park SM, Ahn JY, Jo M, Lee DK, Lis JT, Craighead HG, Kim S, Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. Lab on a chip 9 (9):1206-1212, Copyright (2009) with permission of The Royal Society of Chemistry) (C) A microfluidic device which integrates electrodes to achieve lateral electrophoresis. A continuously injected equilibrium mixture is separated orthogonally to the bulk flow of bound and unbound molecules into different collection channels through differences in their electrophoretic mobility (Reproduced from Jing M, Bowser MT, Isolation of DNA aptamers using micro free flow electrophoresis. Lab on a chip 11 (21):3703-3709, Copyright (2011), with permission of The Royal Society of Chemistry)


A
m
I
III
II
Sample Inlet
Buffer Inlet
Sample Inlet
Waste Outlet
Product Outlet
Waste Outlet
Magnetic beads


B
Sol-gel Microfluidic Chip
PDMS
Electrode
Solgel
Aptamer
Blocking Buffer
Binding Buffer
Wash Buffer
S1 S2 S3 S4 S5
Collection
S5 S4 S3 S2 S1


C
1
2
4
4
3
3
Buffer Flow
Electric Field
1
2
Detection Zone
Fluorescence (Arb. Units)
5x10⁴
4x10⁴
3x10⁴
2x10⁴
1x10⁴
0
0 2 4 6 8 10
Position (mm)

Filtering aptamers: sol-gel target immobilization

A chip-based technology with similar characteristics of affinity chromatography was developed using sol-gel arrays as a solid support for target immobilization (see Figure 1.6.B).[102-104] Selections are achieved by trapping target molecules within a small drop of highly porous 3D silica matrix. Sequences are gently flowed across a sol-gel spot allowing them to enter and diffuse within the sol-gel and bind to the target in its native state. A microfluidic chip is used so that the injections of the library are confined close to the sol-gel spots thereby increasing the efficiency of sampling. This chip also contains integrated microheaters, which allow bound sequences to be thermally eluted. Serialized multiplexed selections using this sol-gel device were demonstrated owing to the highly localized and addressable nature of each sol-gel spot and microheater. Recently, valving was integrated into the device to allow changes to the fluidic network such that parallelized multiplex sol-gel selections can occur while minimizing cross-contamination between spots during elution.[105,106] These integrated multiplex selections along with the significantly reduced amount of target is very attractive, however the binding and elution of sequences in the sol-gel spots is diffusion limited and may reduce the number of sequences that can be effectively sampled.

Isolating bound complexes: micro free flow electrophoresis

Mobility separations have also been integrated into microfluidic chips, taking advantage of the highly efficient selections that can be achieved by CE-SELEX. By Integrating electrodes, bound and unbound sequences are separated through their

different mobilities (Eq. 5). However, in normal CE-SELEX the amount of library that can be injected is significantly limited by the separation resolution of the capillary. This restriction can be eliminated by using micro Free Flow Electrophoresis (μFFE), which utilizes a lateral electric field to continuously separate an equilibrated sample sideways as its being injected forward under hydrostatic pressure (see Figure 1.6.C).[107] Furthermore, the output of the device is partitioned into two streams allowing fluorescently labeled bound and unbound sequence populations to be imaged and continuously directed into separate collection fractions by adjusting the fluid flow rate and the electric field strength. μFFE also eliminates the need to time collections for desired fractions as in CE and was used to successfully isolate an aptamer in only a single cycle from a library over 100-fold larger than CE.

Automation

A key advantage for chip-base selections is the ability to fully integrate the entire selection process into a single device and ultimately automate it. Towards this goal, selection devices have been designed to incorporate many of the advantages from various technologies into a single microchip. For example, one such microfluidic device has two chambers connected by a channel filled with gel (see Figure 1.7.A).[108-110]. One chamber contains integrated microheaters and is packed with target immobilized onto microbeads forming a microscale affinity column. Fluorescently labeled sequences are flowed through the column and captured onto the target-functionalized beads. Bound sequences are then thermally eluted and transported through the gel into the adjacent chamber via electrophoresis. This process

can be fluorescently monitored, optimized, and repeated many times if necessary, to further select, store and concentrate aptamers in the adjacent chamber free from contamination and other processes that may be taking place in the main chamber. The isolation chamber can be used to further integrate and automate amplification steps by capturing sequences as they are electrophoretically transported through the gel onto new beads for amplification. Integrated heaters and temperature sensors can then be used to thermocycle the chamber in the presence of amplification reagents.[110] This device can also be used to perform selections by capturing cells. A similar device was designed to integrate cell culture and simple valving onto the microfluidic chip to fully integrate the selection process and allow multiple cycles of selection to be performed entirely on-chip.[111] Interestingly, temperature-dependent selections at 37°C generated aptamers which showed maximum binding at this temperature[112] and integrating cell culture into the chips selection chamber utilized the integrated heaters and sensors to maintain a temperature of 37°C at all times. This demonstrates the ability to generate aptamers which may function optimally at physiological temperatures.

Another microfluidic device was demonstrated that integrates all of the steps required for SELEX. This device utilizes magnetic beads to retain and manipulate target-bound sequences, and uses integrated microscale pumps, mixers and valves to introduce and direct fluids (see Figure 1.7.B). In addition, integrated microheaters and microtemperature sensors are used to perform on-chip thermocycling for PCR amplifications of enriched sequences, and all external controllers for the flow, valves, and other automated protocols are integrated onto a small hand-held system.[113,114]

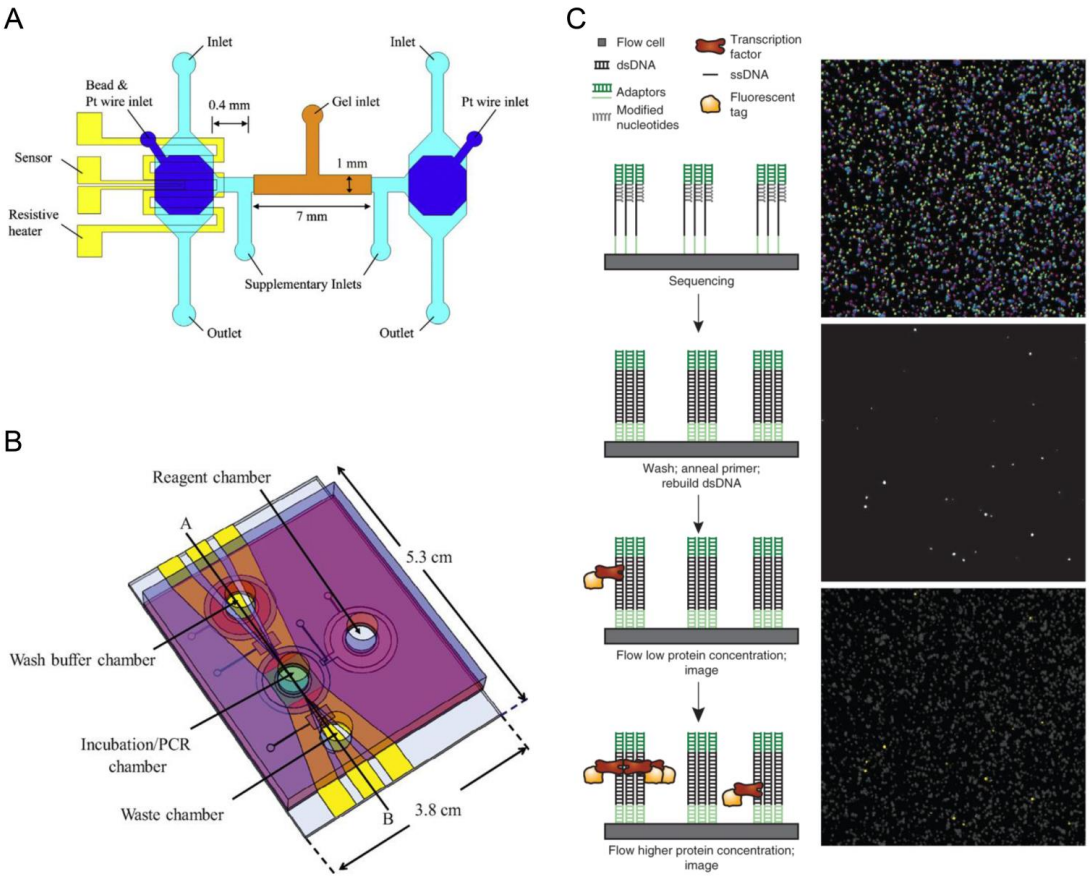


Figure 1.7. Examples of fully integrated microfluidic devices which can perform and automate all of the selection steps. (A) A microfluidic device which uses target immobilized onto beads to capture injected sequences. Bound sequences can be heat eluted with electrodes and then electrophoretically transported through a gel-filled channel into an isolated chamber for further processing (Reprinted from Sensors and Actuators A: Physical, 195, Kim J, Hilton JP, Yang KA, Pei RJ, Stojanovic M, Lin Q, Nucleic acid isolation and enrichment on a microchip, 183-190, Copyright (2013), with permission from Elsevier). (B) A microfluidic device which utilizes and manipulates target-bound magnetic beads via integrated micro-valves, mixers and pumps. Target-bound sequences are similarly controlled and then amplified using micro- heaters and temperature sensors for thermocycling. (Reprinted from Biosensors and Bioelectronics, 25/7, Huang CJ, Lin HI, Shiesh SC, Lee GB, Integrated microfluidic system for rapid screening of CRP aptamers utilizing systematic evolution of ligands by exponential enrichment (SELEX), 1761-1766, Copyright (2010), with permission from Elsevier). (C) A high-throughput sequencing technique which can identify tight binding sequences. A microfluidic flow cell and automated fluidic and optical components allow a library of sequences to be sequenced. Flow vary concentrations of fluorescently labeled target molecules through the flow cell and imaging the bound-target intensity allows a binding affinity to be determined and assigned to each sequence (Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology, Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, Li R, Zhang L, Schroth GP, Burge CB, Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument, 29 (7):659-664, copyright (2011)).

Recently, this device was improved to accommodate the transport of enriched sequences and additional reagents allowing multiple cycles of SELEX to be performed continuously and automatically, completely eliminating manual procedures.[115] This device has also been similarly adapted to perform Cell-SELEX by integrating separate reagent chambers for target and control cells which are coated with magnetic beads. A two-selection process (positive followed by negative SELEX) utilizing this microfluidic device allows for almost complete automation of the entire SELEX process for cell specific aptamers on-chip.

Recently, high-throughput sequencing technologies have been modified to perform affinity measurements with fluorescently labeled targets (see Figure 1.7.C). Although not demonstrated as a selection technology, second generation sequencing systems already integrate programmable and automated fluidic systems that can provide various reagents to a microfluidic chip where sequences are assembled on the chip surface and determined fluorescently. This is achieved through integrated optics and detectors that can distinguish different bases using unique fluorescent labels. After sequencing, fluorescently labeled target molecules are injected at various concentrations and the binding intensity is measured and co-localized with each sequencing read.[116-118] These methods called HiTS-FLIP for DNA (High-Throughput Sequencing-Fluorescent Ligand Interaction Profiling) and HiTS-RAP (-RNA Affinity Profiling) or RNA-MaP (RNA on a Massively Parallel array) for RNA have been used to perform hundreds of millions of independent affinity assays, measuring the binding affinity for every read and completely eliminating the trial and error identification of high affinity aptamers. Although similar to microarrays, these

methods can sample orders of magnitude more sequences and do not require prior knowledge or design of the library or an enriched pool to be screened. In addition, the sequencing platforms contain multiple channels in each chip allowing for the possibility to multiplex with additional targets. A possibility for these systems is to use them to identify aptamers to targets without performing affinity selections at all (i.e. 0 cycles of SELEX) as these sequencing methods can already probe similarly sized if not larger libraries than some of the selection technologies discussed in this review. In cases such as these, the elimination of non-binding sequences may be unnecessary, and the starting library may be assayed directly for potential aptamers. In this case having not been biased and enriched for any particular target, the sequenced library can be used to measure affinity profiles for multiple targets in serial injections. Ultimately, these assays can be used to characterize the SELEX process itself, and the usefulness of other quantitative methods for identifying aptamers, such as multiplicity or enrichment, as well as model their dynamics between rounds. Perhaps most interestingly, they provide a window into the (effective) affinity distribution of libraries and pools, which is a tremendously useful tool for probing the accuracy of theoretical SELEX models that require this information. As sequencing technologies improve and become even more sensitive, the utility of technologies such as HiTS-FLIP and HiTS-RAP will become even more apparent.

CONCLUSIONS AND FUTURE PROSPECTS

A significant amount of progress has been achieved in the last two decades resulting in much more routine and efficient aptamer selections. New technologies are

continuing to improve these efficiencies and reduce the overall investment needed to select for new aptamers. Future work will no doubt continue to integrate, miniaturize and automate as many of the selection steps as possible, hopefully resulting in a variety of different technologies from which researchers can choose. Ultimately, no single technology is superior as each has its limitations that must be considered along with the intended application of the desired aptamer. For example, modifications to sequences and target molecules, such as affinity tags or fluorescent labels, can allow more selection techniques to be used, which result in highly stringent selections, significantly reduced reagent consumption, and the ability to manipulate and monitor binding interactions during ongoing selections. However, modifications can also introduce bias into the binding, and modifications for immobilizing molecules can result in steric hindrance or other unintended binding consequences, such as enhanced non-specific or even specific binding which are not present in solution-based selections. Combining the advantages of several technologies discussed may be a simple method for further improving selection efficiencies, while minimizing each technology's unique sources of background binding, sequence bias, or other undesirable artefacts. In addition, other than using stringent or long washing conditions, incorporating thermodynamic thresholds for binding and dissociation is not yet commonplace, but may allow better discrimination between higher and lower affinity molecules. For example, using salt or thermal gradients to elute bound sequences decreases their effective affinity (see Eq. 12) and preferentially releases more weakly bound sequences. As the temperature T increases, or the salt concentration increases (causing a decrease in the free energy of binding ΔG), the

effective dissociation constant increases such that only the tightest bound molecules remain.

$$K_d = e^{-\Delta G/k_b T} \quad (12)$$

These methods have been used in simple selection modes to successfully isolate better binding aptamers in as little as one cycle of SELEX.[119-122]

Many technologies are now accomplishing successful aptamer selections in only a single round, completely eliminating the majority of the tedious cyclic processes of the traditional SELEX method. These are not only fast and efficient, but also allow selections to be performed with modified or non-natural nucleic acid libraries that cannot be amplified. In addition, minimizing amplifications also minimizes amplification-related bias from pools.[123,124] However, unless highly degenerate libraries are used, "enriched" aptamers are likely represented in the final pool with the same or similar frequency (i.e. copy numbers of 1) as lower affinity or even background binding sequences. This limitation can be partly alleviated by replacing the typical cloning and Sanger sequencing with high-throughput sequencing. This allows millions of sequences to be sampled so that bioinformatic techniques can find consensus sequences and conserved secondary structures with great statistical power. In contrast, basic cloning/sequencing typically allows only a few dozen sequences to be sampled, and is most reliable when individual aptamers have copy numbers that dominate the pool. However, much more information and confidence can be obtained from sequenced pools that have sufficiently diverse copy numbers and allow population metrics, such as multiplicity or enrichment, to help identify the highest affinity aptamers. This places the single-round selection methods at odds with

newly developed bioinformatic tools, and future methods (where possible) may need to consider maximizing the efficiency of selections with the utility of high-throughput sequencing methods.

One particular opportunity for current and emerging technologies is the thorough characterization of their selection parameters. Recent studies have shown that in certain technologies, many of the core and intuitive assumptions of the SELEX model and basic binding kinetics are not well supported by the binding results. This is likely due to the added complexity of the more sophisticated technologies and immobilization schemes used. Many of the existing technologies might show significantly improved performances through simple characterization and optimization, revealing specific binding behaviors that can be modeled, to aid the development of more efficient protocols and more accurate SELEX theories. This will also allow for fairer comparisons between different methods and technologies that are currently evaluated primarily on the number of cycles needed and/or the binding affinity of the resulting aptamer. However, the majority of parameters such as times, flow rates, volumes, and concentrations are not typically discussed or justified in many technologies, and there is currently no system by which each can be calibrated to assess its true efficiency or effectiveness. Fair comparisons among all the available technologies is hindered due to differences in the selected targets; the library type, size and origin; as well as the sequencing and analysis methods used. These variations obscure the true advantages and disadvantages of each selection technology. A standardized target-library system would allow a thorough and fair comparison between methods, and high-throughput sequencing and analysis would allow for more

statistically sound metrics to be generated compared to a successful selection/reselection of a single high affinity aptamer, which is susceptible to the stochastics of binding.

Although *in vitro* aptamer selections are used primarily in the biological sciences, they have been playing an increasingly important role in engineering new sensing technologies. Recent progress in aptamer selections for clinically relevant targets have demonstrated the utility of aptamers for imaging[125] and therapeutics[126,127] much like their protein analogues. However, their applications as affinity reagents for diagnostic purposes has contributed to a new field of research in aptamer-based biosensors and integrated microfluidic and point-of-care diagnostic devices.[128,129] Many of the technologies discussed in this review can be as easily used for screening and sensing target molecules as they can be used for performing selections and characterizing aptamer binding. However, new aptamer-based devices with enhanced functionality are still needed, and similar efforts toward developing valuable sensing platforms, as with the development of selection technologies, are required. Ultimately, novel sensor technologies are limited by the selection processes and the quality of the aptamers that are generated from them. Together, this represent an engineering challenge which can be tacked two-fold through technical innovations which stem from the physical sciences.

Already, significant advancements have been made through clever engineering and the application of non-traditional instrumentation, providing researchers with access to a wider range of physical principles by which to affect the binding and enrichment of high affinity molecules. New sophisticated technologies and improved

analysis methods have helped reduce aptamer selections to only a few cycles or less. However, despite their tremendous reduction in time, many of these techniques and devices are costly and require resources that may not be generally available to most researchers. A significant amount of work may be necessary to improve the reliability and accessibility of many of these techniques before their selections can become more routine. Here too, there continues to be a growing role for engineering and physical sciences to further advance these technologies and continue to integrate the various selection steps. However, even among proven technologies, much of the selection dynamics is not known or well understood. Valuable contributions can be made through detailed characterizations to optimize these existing technologies, revealing the mechanics and true operating principles of their methods. Developing a common platform in which detailed comparisons can be made will also help to explain critical or subtle differences between technologies, and new theories and analytical methods will help enable more robust aptamer identifications in the future.

REFERENCES

1. Ellington AD, Szostak JW (1990) Invitro Selection of Rna Molecules That Bind Specific Ligands. Nature 346: 818-822.

2. Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. Gene 82: 83-87.

3. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.

4. Kuwahara M (2014) Progress in Chemically Modified Nucleic Acid Aptamers. In: Erdmann VA, Markiewicz WT, Barciszewski J, editors. Chemical Biology of Nucleic Acids: Springer Berlin Heidelberg. pp. 243-270.

5. Ciesiolka J, Gorski J, Yarus M (1995) Selection of an Rna Domain That Binds Zn2+. Rna-a Publication of the Rna Society 1: 538-550.

6. Homann M, Goringer HU (1999) Combinatorial selection of high affinity RNA ligands to live African trypanosomes. Nucleic Acids Research 27: 2006-2014.

7. Xiao SJ, Hu PP, Wu XD, Zou YL, Chen LQ, et al. (2010) Sensitive discrimination and detection of prion disease-associated isoform with a dual-aptamer strategy by developing a sandwich structure of magnetic microparticles and quantum dots. Anal Chem 82: 9736-9742.

8. Shi H, Fan X, Sevilimedu A, Lis JT (2007) RNA aptamers directed to discrete functional sites on a single protein structural domain. Proc Natl Acad Sci U S A 104: 3742-3746.

9. Ikebukuro K, Kiyohara C, Sode K (2005) Novel electrochemical sensor system for protein using the aptamers in sandwich manner. Biosensors & Bioelectronics 20: 2168-2172.

10. Gong Q, Wang JP, Ahmad KM, Csordas AT, Zhou JH, et al. (2012) Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. Anal Chem 84: 5365-5371.

11. Mallikaratchy PR, Ruggiero A, Gardner JR, Kuryavyi V, Maguire WF, et al. (2011) A multivalent DNA aptamer specific for the B-cell receptor on human lymphoma and leukemia. Nucleic Acids Research 39: 2458-2469.

12. Di Giusto DA, Knox SM, Lai Y, Tyrelle GD, Aung MT, et al. (2006) Multitasking by multivalent circular DNA aptamers. Chembiochem 7: 535-544.

13. Irvine D, Tuerk C, Gold L (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. J Mol Biol 222: 739-761.

14. Sun F, Galas D, Waterman MS (1996) A mathematical analysis of in vitro molecular selection-amplification. J Mol Biol 258: 650-660.

15. Forst CV (1998) Molecular evolution: A theory approaches experiments. Journal of Biotechnology 64: 101-118.

16. Levitan B (1998) Stochastic modeling and optimization of phage display. J Mol Biol 277: 893-916.

17. Vant-Hull B, Payano-Baez A, Davis RH, Gold L (1998) The mathematics of SELEX against complex targets. J Mol Biol 278: 579-597.

18. Vant-Hull B, Gold L, Zichi DA (2000) Theoretical principles of in vitro selection using combinatorial nucleic acid libraries. Curr Protoc Nucleic Acid Chem Chapter 9: Unit 9 1.

19. Djordjevic M, Sengupta AM (2006) Quantitative modeling and data analysis of SELEX experiments. Phys Biol 3: 13-28.

20. Chen CK (2007) Complex SELEX against target mixture: stochastic computer model, simulation, and analysis. Comput Methods Programs Biomed 87: 189-200.

21. Chen CK, Kuo TL, Chan PC, Lin LY (2007) Subtractive SELEX against two heterogeneous target samples: numerical simulations and analysis. Comput Biol Med 37: 750-759.

22. Levine HA, Nilsen-Hamilton M (2007) A mathematical analysis of SELEX. Computational Biology and Chemistry 31: 11-35.

23. Aita T, Nishigaki K, Husimi Y (2012) Theoretical consideration of selective enrichment in in vitro selection: Optimal concentration of target molecules. Math Biosci 240: 201-211.

24. Wang J, Rudzinski JF, Gong Q, Soh HT, Atzberger PJ (2012) Influence of target concentration and background binding on in vitro selection of affinity reagents. PLoS One 7: e43940.

25. Ozer A, White BS, Lis JT, Shalloway D (2013) Density-dependent cooperative non-specific binding in solid-phase SELEX affinity selection. Nucleic Acids Research 41: 7167-7175.

26. Daniel C, Roupioz Y, Gasparutto D, Livache T, Buhot A (2013) Solution-Phase vs Surface-Phase Aptamer-Protein Affinity from a Label-Free Kinetic Biosensor. PLoS One 8: e75419.

27. Berg OG, von Hippel PH (1987) Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J Mol Biol 193: 723-750.

28. Gevertz J, Gan HH, Schlick T (2005) In vitro RNA random pools are not structurally diverse: a computational analysis. Rna-a Publication of the Rna Society 11: 853-863.

29. Chushak Y, Stone MO (2009) In silico selection of RNA aptamers. Nucleic Acids Research 37: e87.

30. Montgomery-Smith SJ, Schmidt FJ (2010) Statistical methods for estimating complexity from competition experiments between two populations. J Theor Biol 264: 1043-1046.

31. Aita T, Nishigaki K, Husimi Y (2014) Estimation of statistical binding properties of ligand population during in vitro selection based on population dynamics theory. Math Biosci 247: 59-68.

32. Hall B, Arshad S, Seo K, Bowman C, Corley M, et al. (2001) In Vitro Selection of RNA Aptamers to a Protein Target by Filter Immobilization. Current Protocols in Nucleic Acid Chemistry: John Wiley & Sons, Inc.

33. Conrad RC, Giver L, Tian Y, Ellington AD (1996) In vitro selection of nucleic acid aptamers that bind proteins. Combinatorial Chemistry 267: 336-367.

34. Giver L, Bartel DP, Zapp ML, Green MR, Ellington AD (1993) Selection and Design of High-Affinity Rna Ligands for Hiv-1 Rev. Gene 137: 19-24.

35. Ellington A, Szostak JW (1992) Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. Nature Biotechnology 355: 850-852.

36. Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-Resolution Molecular Discrimination by Rna. Science 263: 1425-1429.

37. Berens C, Thain A, Schroeder R (2001) A tetracycline-binding RNA aptamer. Bioorg Med Chem 9: 2549-2556.

38. Win MN, Klein JS, Smolke CD (2006) Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. Nucleic Acids Research 34: 5670-5682.

39. Holeman LA, Robinson SL, Szostak JW, Wilson C (1998) Isolation and characterization of fluorophore-binding RNA aptamers. Fold Des 3: 423-431.

40. Tsai RY, Reed RR (1998) Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. Molecular and Cellular Biology 18: 6447-6456.

41. Goodman SD, Velten NJ, Gao Q, Robinson S, Segall AM (1999) In vitro selection of integration host factor binding sites. J Bacteriol 181: 3246-3255.

42. Bruno JG (1997) In vitro selection of DNA to chloroaromatics using magnetic microbead-based affinity separation and fluorescence detection. Biochem Biophys Res Commun 234: 117-120.

43. Bruno JG, Kiel JL (2002) Use of magnetic beads in selection and detection of biotoxin aptamers by electrochemiluminescence and enzymatic methods. Biotechniques 32: 178-+.

44. Fan M, McBurnett SR, Andrews CJ, Allman AM, Bruno JG, et al. (2008) Aptamer selection express: a novel method for rapid single-step selection and sensing of aptamers. J Biomol Tech 19: 311-319.

45. Stoltenburg R, Reinemann C, Strehlitz B (2005) FluMag-SELEX as an advantageous method for DNA aptamer selection. Anal Bioanal Chem 383: 83-91.

46. Lai J-C, Hong CY (2014) A Novel Protocol for Generating High-Affinity ssDNA Aptamers by Using Alternating Magnetic Fields. Journal of Materials Chemistry B.

47. Latulippe DR, Szeto K, Ozer A, Duarte FM, Kelly CV, et al. (2013) Multiplexed microcolumn-based process for efficient selection of RNA aptamers. Anal Chem 85: 3417-3424.

48. Szeto K, Reinholt SJ, Duarte FM, Pagano JM, Ozer A, et al. (2014) High-throughput binding characterization of RNA aptamer selections using a microplate-based multiplex microcolumn device. Anal Bioanal Chem.

49. Szeto K, Latulippe DR, Ozer A, Pagano JM, White BS, et al. (2013) RAPID-SELEX for RNA Aptamers. PLoS One 8: e82667.

50. Pagano JM, Kwak H, Waters CT, Sprouse RO, White BS, et al. (2014) Defining NELF-E RNA Binding in HIV-1 and Promoter-Proximal Pause Regions. PLoS Genetics 10: e1004090.

51. Nitsche A, Kurth A, Dunkhorst A, Panke O, Sielaff H, et al. (2007) One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. Bmc Biotechnology 7.

52. Tok JBH, Fischer NO (2008) Single microbead SELEX for efficient ssDNA aptamer generation against botulinum neurotoxin. Chemical Communications: 1883-1885.

53. Mendonsa SD, Bowser MT (2004) In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. Anal Chem 76: 5387-5392.

54. Mendonsa SD, Bowser MT (2004) In vitro evolution of functional DNA using capillary electrophoresis. J Am Chem Soc 126: 20-21.

55. Mosing RK, Mendonsa SD, Bowser MT (2005) Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase. Anal Chem 77: 6107-6112.

56. Jing M, Bowser MT (2013) Tracking the Emergence of High Affinity Aptamers for rhVEGF165 During Capillary Electrophoresis-Systematic Evolution of Ligands by Exponential Enrichment Using High Throughput Sequencing. Anal Chem 85: 10761-10770.

57. Yang J, Bowser MT (2013) Capillary Electrophoresis-SELEX Selection of Catalytic DNA Aptamers for a Small-Molecule Porphyrin Target. Anal Chem 85: 1525-1530.

58. Berezovski M, Drabovich A, Krylova SM, Musheev M, Okhonin V, et al. (2005) Nonequilibrium Capillary Electrophoresis of Equilibrium Mixtures: A Universal Tool for Development of Aptamers. J Am Chem Soc 127: 3165-3171.

59. Berezovski M, Musheev M, Drabovich A, Krylov SN (2006) Non-SELEX selection of aptamers. J Am Chem Soc 128: 1410-1411.

60. Ashley J, Ji KL, Li SFY (2012) Selection of bovine catalase aptamers using non-SELEX. Electrophoresis 33: 2783-2789.

61. Wochner A, Cech B, Menger M, Erdmann VA, Glokler J (2007) Semi-automated selection of DNA aptamers using magnetic particle handling. Biotechniques 43: 344-+.

62. Schutze T, Wilhelm B, Greiner N, Braun H, Peter F, et al. (2011) Probing the SELEX process with next-generation sequencing. PLoS One 6: e29604.

63. Cox JC, Rudolph P, Ellington AD (1998) Automated RNA selection. Biotechnol Prog 14: 845-850.

64. Cox JC, Ellington AD (2001) Automated selection of anti-protein aptamers. Bioorg Med Chem 9: 2525-2531.

65. Cox JC, Hayhurst A, Hesselberth J, Bayer TS, Georgiou G, et al. (2002) Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nucleic Acids Research 30: e108.

66. Cox JC, Rajendran M, Riedel T, Davidson EA, Sooter LJ, et al. (2002) Automated acquisition of aptamer sequences. Comb Chem High Throughput Screen 5: 289-299.

67. Goertz PWCCJE, A. D. (2004) Automated selection of aminoglycoside aptamers. Journal of the Association for Laboratory Automation 9: 150-154.

68. Eulberg D, Buchner K, Maasch C, Klussmann S (2005) Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonist. Nucleic Acids Research 33: e45.

69. Drolet DW, Jenison RD, Smith DE, Pratt D, Hicke BJ (1999) A high throughput platform for systematic evolution of ligands by exponential enrichment (SELEX). Comb Chem High Throughput Screen 2: 271-278.

70. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, et al. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20: 861-873.

71. Raddatz MSL, Dolf A, Endl E, Knolle P, Famulok M, et al. (2008) Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. Angewandte Chemie-International Edition 47: 5190-5193.

72. Mayer G, Ahmed MS, Dolf A, Endl E, Knolle PA, et al. (2010) Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. Nat Protoc 5: 1993-2004.

73. Yang X, Bassett SE, Li X, Luxon BA, Herzog NK, et al. (2002) Construction and selection of bead-bound combinatorial oligonucleoside phosphorothioate and phosphorodithioate aptamer libraries designed for rapid PCR-based sequencing. Nucleic Acids Research 30: e132.

74. Yang X, Li X, Prow TW, Reece LM, Bassett SE, et al. (2003) Immunofluorescence assay and flow-cytometry selection of bead-bound aptamers. Nucleic Acids Research 31.

75. Wang J, Gong Q, Maheshwari N, Eisenstein M, Arcila ML, et al. (2014) Particle Display: A Quantitative Screening Method for Generating High-Affinity Aptamers. Angew Chem Int Ed Engl.

76. Paige JS, Wu KY, Jaffrey SR (2011) RNA Mimics of Green Fluorescent Protein. Science 333: 642-646.

77. Lauridsen LH, Shamaileh HA, Edwards SL, Taran E, Veedu RN (2012) Rapid one-step selection method for generating nucleic acid aptamers: development of a DNA aptamer against alpha-bungarotoxin. PLoS One 7: e41702.

78. Peng L, Stephens BJ, Bonin K, Cubicciotti R, Guthold M (2007) A combined atomic force/fluorescence microscopy technique to select aptamers in a single cycle from a small pool of random oligonucleotides. Microscopy Research and Technique 70: 372-381.

79. Miyachi Y, Shimizu N, Ogino C, Kondo A (2010) Selection of DNA aptamers using atomic force microscopy. Nucleic Acids Research 38.

80. Pileur F, Andreola ML, Dausse E, Michel J, Moreau S, et al. (2003) Selective inhibitory DNA aptamers of the human RNase H1. Nucleic Acids Research 31: 5776-5788.

81. Misono TS, Kumar PKR (2005) Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. Analytical Biochemistry 342: 312-317.

82. Aminova O, Disney MD (2010) A microarray-based method to perform nucleic acid selections. Methods Mol Biol 669: 209-224.

83. Aminova O, Paul DJ, Childs-Disney JL, Disney MD (2008) Two-dimensional combinatorial screening identifies specific 6'-acylated kanamycin A- and 6'-acylated neamine-RNA hairpin interactions. Biochemistry 47: 12670-12679.

84. Childs-Disney JL, Wu ML, Pushechnikov A, Aminova O, Disney MD (2007) A small molecule microarray platform to select RNA internal loop-ligand interactions. Acs Chemical Biology 2: 745-754.

85. Disney MD, Labuda LP, Paul DJ, Poplawski SG, Pushechnikov A, et al. (2008) Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. J Am Chem Soc 130: 11185-11194.

86. Knight CG, Platt M, Rowe W, Wedge DC, Khan F, et al. (2009) Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. Nucleic Acids Research 37: e6.

87. Platt M, Rowe W, Wedge DC, Kell DB, Knowles J, et al. (2009) Aptamer evolution for array-based diagnostics. Analytical Biochemistry 390: 203-205.

88. Collett JR, Cho EJ, Lee JF, Levy M, Hood AJ, et al. (2005) Functional RNA microarrays for high-throughput screening of antiprotein aptamers. Analytical Biochemistry 338: 113-123.

89. Cho M, Soo Oh S, Nie J, Stewart R, Eisenstein M, et al. (2013) Quantitative selection and parallel characterization of aptamers. Proc Natl Acad Sci U S A 110: 18460-18465.

90. Fischer NO, Tarasow TM (2011) Identification and optimization of DNA aptamer binding regions using DNA microarrays. Methods Mol Biol 723: 57-66.

91. Fischer NO, Tok JBH, Tarasow TM (2008) Massively Parallel Interrogation of Aptamer Sequence, Structure and Function. PLoS One 3.

92. Katilius E, Flores C, Woodbury NW (2007) Exploring the sequence space of a DNA aptamer using microarrays. Nucleic Acids Research 35: 7626-7635.

93. Zhu Y, Chandra P, Ban C, Shim YB (2012) Electrochemical Evaluation of Binding Affinity for Aptamer Selection Using the Microarray Chip. Electroanalysis 24: 1057-1064.

94. Li Y, Lee HJ, Corn RM (2006) Fabrication and characterization of RNA aptamer microarrays for the study of protein-aptamer interactions with SPR imaging. Nucleic Acids Research 34: 6416-6424.

95. Hybarger G, Bynum J, Williams RF, Valdes JJ, Chambers JP (2006) A microfluidic SELEX prototype. Anal Bioanal Chem 384: 191-198.

96. Lou X, Qian J, Xiao Y, Viel L, Gerdon AE, et al. (2009) Micromagnetic selection of aptamers in microfluidic channels. Proc Natl Acad Sci U S A 106: 2989-2994.

97. Oh SS, Plakos K, Lou X, Xiao Y, Soh HT (2010) In vitro selection of structure-switching, self-reporting aptamers. Proc Natl Acad Sci U S A 107: 14053-14058.

98. Oh SS, Qian J, Lou X, Zhang Y, Xiao Y, et al. (2009) Generation of highly specific aptamers via micromagnetic selection. Anal Chem 81: 5490-5495.

99. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.

100. Oh SS, Ahmad KM, Cho M, Kim S, Xiao Y, et al. (2011) Improving aptamer selection efficiency through volume dilution, magnetic concentration, and continuous washing in microfluidic channels. Anal Chem 83: 6883-6889.

101. Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, et al. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. Proc Natl Acad Sci U S A 107: 15373-15378.

102. Ahn JY, Jo M, Dua P, Lee DK, Kim S (2011) A sol-gel-based microfluidics system enhances the efficiency of RNA aptamer selection. Oligonucleotides 21: 93-100.

103. Bae H, Ren S, Kang J, Kim M, Jiang Y, et al. (2013) Sol-Gel SELEX Circumventing Chemical Conjugation of Low Molecular Weight Metabolites Discovers Aptamers Selective to Xanthine. Nucleic Acid Therapeutics.

104. Park SM, Ahn JY, Jo M, Lee DK, Lis JT, et al. (2009) Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. Lab Chip 9: 1206-1212.

105. Kim TK, Lee SW, Ahn JY, Laurell T, Kim SY, et al. (2011) Fabrication of Microfluidic Platform with Optimized Fluidic Network toward On-Chip Parallel Systematic Evolution of Ligands by Exponential Enrichment Process. Japanese Journal of Applied Physics 50.

106. Lee S, Kang J, Ren S, Laurell T, Kim S, et al. (2013) A cross-contamination-free SELEX platform for a multi-target selection strategy. Biochip Journal 7: 38-45.

107. Jing M, Bowser MT (2011) Isolation of DNA aptamers using micro free flow electrophoresis. Lab Chip 11: 3703-3709.

108. Kim J, Hilton JP, Yang KA, Pei R, Ennis K, et al. A microchip for nucleic acid isolation and enrichment; 2012 Jan. 29 2012-Feb. 2 2012. pp. 765-768.

109. Kim J, Hilton JP, Yang KA, Pei RJ, Stojanovic M, et al. (2013) Nucleic acid isolation and enrichment on a microchip. Sensors and Actuators a-Physical 195: 183-190.

110. Kim J, Hilton JP, Yang KA, Pei R, Zhu J, et al. Electrokinetically integrated microfluidic isolation and amplification of biomolecule-and cell-binding nucleic acids; 2013 20-24 Jan. 2013. pp. 1007-1010.

111. Zhu J, Olsen T, Pei R, Stojanovic M, Lin Q. A microfluidic device for isolation of cell-targeting aptamers; 2014 26-30 Jan. 2014. pp. 242-245.

112. Hilton JP, Jinho K, ThaiHuu N, Barbu M, Renjun P, et al. Isolation of thermally sensitive aptamers on a microchip; 2012 Jan. 29 2012-Feb. 2 2012. pp. 100-103.

113. Huang CJ, Lin HI, Shiesh SC, Lee GB (2010) Integrated microfluidic system for rapid screening of CRP aptamers utilizing systematic evolution of ligands by exponential enrichment (SELEX). Biosensors & Bioelectronics 25: 1761-1766.

114. Huang CJ, Lin HI, Shiesh SC, Lee GB (2012) An integrated microfluidic system for rapid screening of alpha-fetoprotein-specific aptamers. Biosensors & Bioelectronics 35: 50-55.

115. Chen YH, Lin HI, Huang CJ, Shiesh SC, Lee GB (2012) An automatic microfluidic system that continuously performs the systematic evolution of ligands by exponential enrichment. Microfluidics and Nanofluidics 13: 929-939.

116. Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, et al. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. Nature Biotechnology 29: 659-664.

117. Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, et al. (Submitted) Comprehensive analysis of RNA-protein interactions by high throughput seqencing-RNA affinity profiling. Nature Methods.

118. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, et al. (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. Nature Biotechnology.

119. Calik P, Balci O, Ozdamar TH (2010) Human growth hormone-specific aptamer identification using improved oligonucleotide ligand evolution method. Protein Expr Purif 69: 21-28.

120. Arnold S, Pampalakis G, Kantiotou K, Silva D, Cortez C, et al. (2012) One round of SELEX for the generation of DNA aptamers directed against KLK6. Biol Chem 393: 343-353.

121. Tao JS, Frankel AD (1994) Rna Structure Prediction by in-Vitro Selection of Arginine-Binding Rnas. Journal of Cellular Biochemistry: 139-139.

122. Liu Y, Wang C, Li F, Shen S, Tyrrell DL, et al. (2012) DNase-mediated single-cycle selection of aptamers for proteins blotted on a membrane. Anal Chem 84: 7603-7606.

123. Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. PLoS One 5: e9169.

124. Thiel WH, Bair T, Thiel KW, Dassie JP, Rockey WM, et al. (2011) Nucleotide Bias Observed with a Short SELEX RNA Aptamer Library. Nucleic Acid Therapeutics 21: 253-263.

125. Wang AZ, Farokhzad OC (2014) Current progress of aptamer-based molecular imaging. J Nucl Med 55: 353-356.

126. Xing H, Hwang K, Li J, Torabi S-F, Lu Y (2014) DNA aptamer technology for personalized medicine. Current Opinion in Chemical Engineering 4: 79-87.

127. Keefe AD, Pai S, Ellington A (2010) Aptamers as therapeutics. Nature Reviews Drug Discovery 9: 537-550.

128. Zhou W, Huang P-JJ, Ding J, Liu J (2014) Aptamer-based biosensors for biomedical diagnostics. Analyst.

129. Hong P, Li W, Li J (2012) Applications of aptasensors in clinical diagnostics. Sensors (Basel) 12: 1181-1193.

CHAPTER 2

MULTIPLEXED MICROCOLUMN-BASED PROCESS FOR EFFICIENT SELECTION OF RNA APTAMERS[‡]

ABSTRACT

We describe a reusable microcolumn and process for the efficient discovery of nucleic acid aptamers for multiple target [1]molecules. The design of our device requires only microliter volumes of affinity chromatography resin—a condition that maximizes the enrichment of target-binding sequences over non-target-binding (i.e., background) sequences. Furthermore, the modular design of the device accommodates a multiplex aptamer selection protocol. We optimized the selection process performance using microcolumns filled with green fluorescent protein (GFP)-immobilized resin and monitoring, over a wide range of experimental conditions, the enrichment of a known GFP-binding RNA aptamer (GFPapt) against a random RNA aptamer library. We validated the multiplex approach by monitoring the enrichment of GFPapt in de novo selection experiments with GFP and other protein preparations. After only three rounds of selection, the cumulative GFPapt enrichment on the GFP-loaded resin was greater than 108 with no enrichment for the other nonspecific targets. We used this optimized protocol to perform a multiplex selection to two human heat shock factor (hHSF) proteins, hHSF1 and hHSF2. High-throughput sequencing was used to

[‡] The following sections are reprinted with permission from: Latulippe, D.R.*, Szeto, K.*, Ozer, A., Duarte, F.M., Kelly, C.V. Pagano, J.M., White, B.S., Shalloway, D., Lis, J.T., and Craighead, H.G. 2013. Multiplexed microcolumn-based process for efficient selection of RNA aptamers. Anal. Chem. 85(6) pp3417-3424, doi: 10.1021/ac400105e, Copyright 2013 American Chemical Society, with modifications to conform to the required format.
* DRL and KS are co-first authors.

identify aptamers for each protein that were preferentially enriched in just three selection rounds, which were confirmed and isolated after five rounds. Gel-shift and fluorescence polarization assays showed that each aptamer binds with high-affinity (K_D < 20 nM) to the respective targets. The combination of our microcolumns with a multiplex approach and high-throughput sequencing enables the selection of aptamers to multiple targets in a high-throughput and efficient manner.

INTRODUCTION

Nucleic acid aptamers are short (100 nucleotide, nt) structured oligonucleotides that have been selected from large sequence-diverse libraries and shown to display high-affinity and specificity for a wide range of targets ranging from simple metal ions [1] to complex surface proteins on living cells [2].This combination of properties has led to growing interest in applications of aptamers in fields including therapeutics, chemical analysis, biotechnology, chemical separations, and environmental diagnostics [3]. Aptamers are identified from large libraries of random nucleic acid sequences via an iterative in vitro process called SELEX (systematic evolution of ligands by exponential enrichment) [4-6]. A typical SELEX round includes the following three steps: (i) binding, incubation of the library with the target; (ii) partitioning, separation of target-bound library sequences from unbound ones; and (iii) amplification, generation of a new pool of nucleic acids by making multiple copies of the sequences that bound to the target. These steps are then repeated in an iterative fashion to obtain an enriched pool and the target binding aptamers are identified via cloning and sequencing processes.

Different selection strategies have been developed to separate or partition the free and target-bound sequences, a critical step to ensure the successful identification of only the strongest binding aptamers. Affinity chromatography is one such partitioning strategy that uses specific binding onto resin-immobilized targets to purify macromolecular solutes from dilute solutions [7]. It is a well-documented aptamer selection technique, given that it can achieve a level of purification greater than 95% in a single step and that numerous types of resin media are available to bind a wide variety of targets. However, there is limited understanding of the relationship governing the process parameters (target loading, resin volume, etc.) and the selection quality, and thus, many selection rounds (typically 12–15) are required to identify aptamers with the desired specificity and affinity for the target. This approach is particularly challenging when working with RNA libraries because it takes approximately 2 days just to complete the amplification step. Some work has been done to parallelize the use of libraries and the selection process to multiple targets in order to save time and reagents [8,9].

Affinity chromatography-based selections are typically done in two different modes of operation. In batch mode, a small amount of target-immobilized resin (20–200 μL) is incubated with the nucleic acid library, or alternatively, target-free resin is incubated with a mixed suspension of target and nucleic acid library [10-12]. Any unbound sequences in the supernatant are removed, and the target-bound sequences remaining on the resin are then exposed to other solutions for the subsequent processing steps. For this approach, the entire selection process is quite laborious, because each step must be done manually and repeated several times. This is

especially true when multiple targets are considered for selection. The second mode of operation, flow mode, uses small columns (0.5–3 mL) packed with resin [1,13,14]. The primary advantage of this approach is that the resin is physically confined within the column, allowing all of the selection steps to be automated by use of pumps and/or centrifuges and thus completed more efficiently than the batch strategy. This approach was used in one of the landmark papers on aptamers; Ellington and Szostak [6] used a 3.5 mL column filled with dye-immobilized resin. However, there are a few limitations to this approach. The standard columns that have been used previously are not practical for the simultaneous selection of aptamers for multiple targets. These columns require more resin than the batch mode, as well as more of the immobilized target (e.g., protein), which can be both limiting and expensive. Thus, with the current affinity chromatography-based strategies, there is a noticeable lack of means to rapidly screen for aptamers to multiple targets in a high-throughput and efficient manner.

To address these limitations, we developed a process utilizing reconfigurable microcolumns of varying capacity for selecting RNA aptamers. The microcolumns require only microliter volumes of affinity chromatography resin (2–50 μL), they can be easily assembled in various configurations to accommodate multiple targets, and they can be easily integrated with common laboratory equipment. In addition, these microcolumns are not restricted to RNA and other nucleic acid aptamer selections but are also suitable for other affinity chromatography needs where small column volumes are desired. The assembly of microcolumns and aptamer selection process are shown in Figure 2.1 with the experimental details provided below.

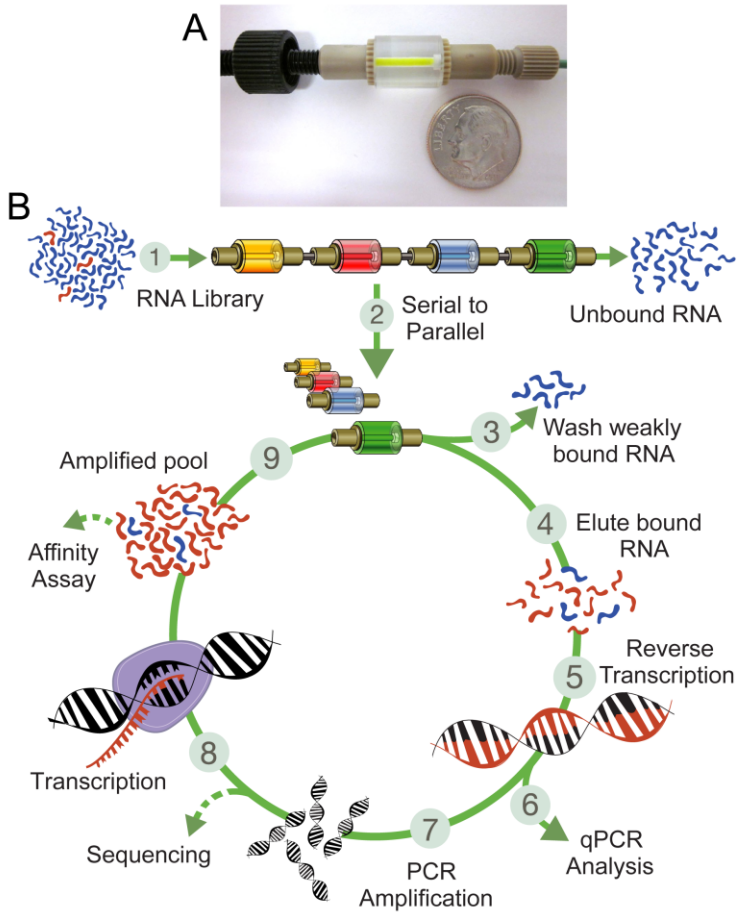


Figure 2.1. Aptamer selection workflow for multiple targets by use of microcolumns. (A) Microcolumn filled with 10 μL of GFP-immobilized chromatography resin. (B) Multiplexed selection of RNA aptamers: (1) The starting RNA library is dynamically loaded onto multiple microcolumns that are connected in a serial configuration. (2) The devices are rearranged into a parallel configuration and the subsequent cycles in the process are done independently but simultaneously. (3) Unbound and weakly bound RNAs are washed away. (4) The remaining bound RNAs are eluted from each column separately. (5) The RNA molecules are reverse-transcribed into cDNA and (6) a small fraction is analyzed via qPCR. (7) The remaining cDNA is PCR-amplified and then (8) transcribed back into RNA to make a new amplified pool for (9) the next selection round. The steps shown with dashed arrows are optional and are not necessarily done in each round.

We first evaluated the design space of our process using computer simulations of the binding kinetics of a model library over a wide range of experimental conditions. Next, we evaluated the performance of our microcolumns at those same conditions by monitoring the behavior of a known RNA aptamer (GFPapt) that binds

tightly to green fluorescent protein (GFP) and its derivatives [13]. Our results show that predictions based on simple kinetics fail to reproduce the behavior of low-affinity binders under flow, suggesting that typical SELEX processes based on theoretical kinetics are likely to be far from optimal. Furthermore, we observed the best performances at protein concentrations 100 times less than the capacity of the resin. We also empirically validated a multiplex approach by monitoring the enrichment of GFPapt in de novo selection experiments with GFP, two nonrelated protein preparations, and an empty microcolumn. To examine the utility of our device, multiplex and in-line negative selections were performed on two human heat shock factor (hHSF) proteins, hHSF1 and hHSF2. High-throughput sequencing was used to identify enriched candidate aptamer sequences for hHSF1 and hHSF2. Fluorescence electrophoretic mobility shift assays (F-EMSA) as well as fluorescence polarization (FP) confirmed the selection of novel high-affinity aptamers for hHSF1 and hHSF2.

MATERIALS AND METHODS

Selection Simulations

Computational simulations were performed with a custom-made Matlab routine to test a wide variety of experimental parameters. The differential equations of association and dissociation kinetics (first-order with respect to each species) were numerically integrated with respect to time, distance along the microfluidic channel, and each aptamer type within the modeled library. Simulation conditions were set to mimic experimental conditions wherever possible including: (i) concentration of target, (ii) microfluidic dimensions, (iii) flow-rate during library loading, (iv) volume

and concentration of the library, and (v) flow-rate and duration of washing. A discrete distribution of K_D within the starting library was assumed to be log-normal [15] with only 0.00011% of the aptamers being the strongest binders. However our simulation results proved to be very insensitive to these initial conditions due to the significantly greater concentration of target than aptamers. The on-rate (k_a) and off-rate (k_d) constants for each aptamer were approximated to match select experimental conditions: $k_d * k_a = 0.04$. Further, the simulations included the presence of a background binding possibility for all aptamers. Regardless of the binding strength of the aptamers to the target, each aptamer was capable of binding to a non-specific binding site with an equilibrium binding constant of 10 μM with a non-specific binding site concentration of 100 μM; these parameters were determined to sufficiently match the background binding experimentally observed.

Preparation of recombinant protein targets

Recombinant proteins were expressed in BL21(DE3)-RIPL E. coli cells (Agilent Technologies) transformed with plasmids that encode for hexahistidine-tagged GFP, CHK2, UBLCP1, and dHSF, or GST-tagged hHSF1 and hHSF2. Two or four liter LB cultures supplemented with 100 μg/ml ampicillin were inoculated with starter LB culture derived from a single colony and grown at 37 °C until the OD600 reached approximately 0.6. Protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After an additional incubation, bacteria were collected by centrifugation and the resulting pellet was processed according to the manufacturer's instructions for Ni-NTA Superflow (Qiagen) or Glutathione-agarose

(Thermo Scientific) resins. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used to verify the purity and quality of the final protein product. Resulting protein preps were dialyzed against 1× PBS (supplemented with 5 mM 2-mercaptoethanol and 0.01% Triton X-100) and stored in small aliquots after addition of glycerol to a final concentration of 20%.

Preparation of Protein-Immobilized Resins

For each selection round, a fresh batch of protein-bound resin was prepared. Nickel–nitrilotriacetic acid (Ni-NTA) Superflow or glutathione–agarose resins were extensively washed with binding buffer [Ni-NTA binding buffer contained 25 mM tris(hydroxymethyl)aminomethane (Tris), pH 8.0, 100 mM NaCl, 25 mM KCl, and 5 mM MgCl2; glutathione binding buffer contained 10 mM N-2-hydroxyethylpiperazine-N-ethanesulfonic acid (HEPES)–KOH pH 7.6, 125 mM NaCl, 25 mM KCl, 1 mM MgCl2, and 0.02% Tween-20] to remove any residual storage components. Hexahistidine- or GST-tagged proteins were prepared as described and immobilized onto the washed resin at 4 °C with constant mixing. The protein-bound resin was then degassed in a vacuum desiccator for approximately 20 min and carefully pipetted into the device (Figure 2.1.A)

Nucleic Acid Library and GFP-Binding Aptamer

A random library containing 5×10^{15} sequences of 120-nucleotide (nt) DNA templates, hereafter referred to as N70 library, was chemically synthesized by GenScript. It consists of a 70-nt random region flanked by two constant regions as

described elsewhere [16]. The GFP-binding RNA aptamer used in this work was previously identified as AP3-1 and characterized elsewhere [13].

The 120-nt DNA library consists of a 70-nt random region flanked by two constant regions: 5'-AAGCTTCGTCAAGTCTGCAGTGAA-N70-GAATTCGTAGATGTGGATCCATTCCC-3'. The single-stranded DNA template library was converted to double-stranded DNA while introducing the T7 promoter using Klenow exo- (NEB) and Lib-FOR oligo: 5'-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3'. The resulting library was later amplified in a 1 L PCR reaction using Taq DNA polymerase, Lib-FOR oligo, and Lib-REV oligo: 5'-AAGCTTCGTCAAGTCTGCAGTGAA-3'. A single aliquot covering the complexity of the entire library was transcribed with T7 RNA polymerase in a 88 mL reaction yielding 1200-fold amplification. An aliquot of this RNA library corresponds to an average of 4 to 6 copies of each sequence in the 5×10^{15} library was used as the starting pool for selections.

GFPapt, the 84-nt GFP binding RNA aptamer, has the following sequence: 5'-AGCUUCUGGACUGCGAUGGGAGCACGAAACGUCGUGGCGCAAUUGGGUGGGGAAAGUCCUUAAAAGAGGGCCACCACAGAAGCU-3'. The forward and reverse oligos used for qPCR analyses were GFPapt-FOR: 5'-GCTTCTGGACTGCGATGGGAGCA-3' and GFPapt-REV: 5'-GCTTCTGTGGTGGCCCTCTTTTAAGGACT-3'.

All of the oligos used in this work were obtained from Integrated DNA Technologies.

Microcolumn Selection and Amplification Protocol

All of the solutions were degassed prior to use and introduced into the microcolumns via a standard syringe pump (Harvard Apparatus). First, yeast tRNA (Invitrogen) in binding buffer was introduced to block any possible nonspecific RNA binding sites. For each loading step, the RNA library was diluted in 1 mL of RNase-free binding buffer, heat-denatured at 60 °C for 5 min, renatured by cooling down to room temperature while degassing, and then spiked with 200 units of RNase inhibitor (Invitrogen). A 10 μL aliquot was collected and used as a standard for the quantitative polymerase chain reaction (qPCR) analysis. Each device was then washed with 3 mL of binding buffer (supplemented with 10 mM imidazole for selections with Ni-NTA resin) to remove unbound RNA. Finally, the RNA–protein complexes were eluted from individual microcolumns by flowing elution buffer [binding buffer + 50 mM ethylenediaminetetraacetic acid (EDTA)] at a rate of 50 μL/min for 6 min. Eluted RNA and the input samples were phenol/chloroform-extracted and ethanol-precipitated together with 1 μL of GlycoBlue (Ambion) and 40 μg of yeast tRNA (Invitrogen), and the resulting pellet was resuspended in RNase-free water.

Both the resuspended pools and standards were reverse-transcribed with Moloney murine leukemia virus reverse transcriptase (MMLV-RT). For the optimization experiments, 10% of the selected pool and the input sample were reverse-transcribed in a separate reaction for GFPapt quantitation. Residual RNA was eliminated by treating the samples with RNase H (Ambion). A small amount (less than 5%) of the cDNA product was analyzed on a LightCycler 480 qPCR instrument (Roche) to determine the amount of RNA library and GFPapt that was retained on

each device. Two different sets of oligos were used to independently evaluate the amount of the RNA library/pool and GFPapt, respectively. The cDNA samples from each round were PCR-amplified and then subjected to phenol/chloroform and chloroform extractions and a final purification step on DNA Clean & Concentrator (Zymo Research) spin columns. A fraction of the purified PCR product was used to make the RNA pool for the next round of SELEX. A typical 72 μL transcription reaction consisted of 500 ng of DNA, 546 pmol of each ribonucleoside triphosphate (rNTP, Sigma), T7 RNA polymerase, 72 units of RNase inhibitor (Invitrogen), and 0.12 unit of yeast inorganic pyrophosphatase (New England Biolabs). The reactions were incubated at 37 °C for various times depending on the desired amount of RNA for the next selection round. The resulting RNA pool was treated with DNase I (Invitrogen) to remove the template DNA, verified by denaturing polyacrylamide gel electrophoresis (PAGE) for length and purity, and finally quantified by Qubit BR RNA assay (Invitrogen).

High-Throughput Sequencing data filtering, and clustering analyses

A small amount of the purified PCR product from each target pool for various selection rounds (e.g., hHSF1 round 5) was PCR-amplified by use of primers that contain a unique 6 nt barcode and the necessary adapters for the HiSeq 2000 (Illumina) sequencing platform. Sequences of the primers and the barcodes are available upon request. The barcoded adapter-ligated PCR products were PAGE-purified, phenol:chloroform and chloroform extracted, ethanol precipitated, and then re-suspended in 10 mM Tris-HCl pH 7.5 buffer. High-throughput sequencing (100 nt

single-end reads) was performed by the sequencing core facility at Life Sciences Core Laboratories Center, Cornell University. Up to twelve barcoded samples were successfully sequenced in a single lane of the HiSeq 2000 flow-cell.

Sequences that contained any ambiguity or had a quality score less than three (Illumina 1.8 encoding) were removed from analysis. Remaining sequences were separated based on the barcodes using the 'FASTX' toolkit (hannonlab.cshl.edu/fastx_toolkit) by requiring a perfect match for barcode sequences. The forward constant regions were stripped via semi-global alignment using 'cutadapt' [17]; given the length of the reads, the reverse constant region and Illumina sequencing adaptors were not present and did not need to be removed. Remaining sequences containing between 64 and 72 nucleotides of the random region (inclusively) and identical in sequence were collapsed using the FASTX tool kit. A two-stage clustering approach was applied to the collapsed sequences to account for sequencing error that may induce apparently distinct reads from a single aptamer. Sequences with multiplicity at most one in all pools (i.e., singletons) that are dissimilar from all other non-singleton sequences in their respective pools, as determined by using 'USEARCH' [18] to align each to a database of non-singleton reads, are likely to be spurious or present at relatively low levels. For improved computational efficiency these were excluded from subsequent analysis. Remaining sequences were aggregated across pools and those with 80% sequence identity were clustered using USEARCH. The highest multiplicity read within each cluster, the cluster representative, was identified as the true aptamer sequence. The multiplicity of this cluster was defined as the sum of multiplicities within the cluster. Representative

sequences, multiplicities and enrichment factors (defined as the ratio of the fraction of a given cluster in a given pool/ fraction of the same cluster in the starting library) for each cluster were tabulated and sorted based on the multiplicity or the enrichment factor across different pools to select candidate aptamers for each target. The analysis pipeline is available upon request.

Amplification of select aptamer sequences from pools

Candidate aptamer sequences were PCR amplified from the final Round 5 pool for each target using an aptamer specific oligo and the Lib-REV oligo with Phusion Polymerase (NEB), according to manufacturer's instructions. Aptamer specific oligos span the forward constant region and ~30 nt of the random region (Tm of 60 °C) of each aptamer. The resulting PCR product was double-digested with EcoRI or BamHI, and PstI or HindIII, and ligated into a similarly cut pGEM3Z-N70Apt plasmid, which has been obtained by cloning a random aptamer sequence together with the T7 promoter into the pGEM3Z vector (Promega) between NarI and HindIII sites. Multiple clones, typically 3 to 10, were sequenced to obtain a consensus for the full-length sequence of the candidate aptamer. Note, high-throughput sequencing only yielded the first 68 nt sequence of the 70 nt random region (100 nt sequencing – 26 nt forward constant region – 6 nt barcode). Therefore, all of the candidate aptamers that were used in subsequent assays were prepared from the sequence verified constructs to ensure the purity of the intended aptamer in each preparation.

Fluorescence Electrophoretic Mobility Shift Assay

Candidate aptamer sequences were amplified from the final round 5 pool and prepared from sequence-verified plasmid constructs. The candidate aptamers were 3′-end-labeled with fluorescein 5-thiosemicarbazide (Invitrogen) as described previously [19]. Binding reactions (50 μL) were prepared with 2 nM fluorescently labeled RNA and decreasing amounts of protein (2000 to 0 nM) in binding buffer containing 0.01% IGEPAL CA-630, 10 μg/mL yeast tRNA, and 3 units of RNase inhibitor (Invitrogen). Reactions were incubated at room temperature for 2 h, spiked with 6× loading dye, and then loaded into the wells of a refrigerated 1.5% agarose gel prepared with 0.5× Tris–borate–EDTA (TBE) buffer with 1 mM MgCl2. The gel was run for 80 min at 100 V in refrigerated 0.5× TBE. Images were acquired at the fluorescein scan settings on a Typhoon 9400 imager (GE Healthcare Life Sciences). The resulting bands were quantified with ImageQuant software and the data were fit to the Hill equation by use of Igor (Wavemetrics) to estimate the equilibrium dissociation constant (K_D).

Fluorescence polarization (FP) assay

Binding reactions were prepared by mixing 2 nM of 3'-end labeled RNA aptamer with protein concentrations that varied from 0.2 to 2000 nM in 1.5-fold increments in binding buffer containing 0.01% IGEPAL CA630, 10 μg/ml tRNA, and 3U of SUPERase•In RNase Inhibitor (Invitrogen). The reactions were prepared in black 96-well half area microplates (Corning) and then incubated at room temperature for 2 hours. The plates were read on a Synergy H1 Microplate Reader (BioTek);

fluorescence polarization was measured as ($F_{\parallel}$-$F_{\perp}$)/($F_{\parallel}$+$F_{\perp}$) using the Ex: 485/20 Em: 528/20 filter set.

RESULTS AND DISCUSSION

Fabrication of Microcolumns for Selection of Aptamers

Our microcolumns were assembled from both custom-fabricated and commercially available parts. The column consists of a transparent biocompatible plastic rod fitted with a porous frit that retains the resin in the device (Figure 2.1.A). By varying both the length and internal diameter of the column, we were able to fabricate columns with a range of volume capacities from 2 to 50 μL. NanoPorts (IDEX Health and Science) that accommodate standard tubing connectors were bonded to either end of the column. Overall, this design has a number of important features: simple union connectors can be used to arrange multiple microcolumns into various configurations, the dead volume between the devices is minimized and is generally less than 1 μL, and they can be connected to common laboratory equipment to automate the selection steps. To perform the multiple target aptamer selection, we developed a workflow process wherein a set of microcolumns, each prefilled with a different target-immobilized resin, are arranged into a serial configuration (Figure 2.1.B). With our pump system, a typical arrangement could contain up to 10 parallelized assemblies of devices, but the general approach can easily be scaled up for a larger number of parallel processes. A single aliquot of the starting random library is then flowed through these devices, allowing the target binding aptamers to be captured on the resin within each individual column. The library molecules that do not bind to

any target are discarded and then the individual microcolumns are disconnected and reorganized into a parallel configuration (Figure 2.1.B). This arrangement allows for specific elutions from each target and thus separate processing of only the bound sequences to create target-specific amplified pools for the next selection round. Note that, following reverse transcription of the RNA aptamer into cDNA, we used quantitative PCR to determine the amount of nucleic acid recovered from each device. This information was used to determine the optimal number of PCR cycles and thus minimize the chance of amplification artifacts [20]. Although Figure 2.1.B shows the microcolumns arranged exclusively in a parallel configuration after the first selection round, it is also possible to use a serial configuration in later rounds. This arrangement would allow for negative or counter selections to be done simultaneously with the selection step to enhance specificity for the target.

The size of our microcolumns was chosen for primarily two reasons. First, they are small enough that they require only small amounts of material for each selection round, yet their internal dimensions are sufficiently large enough that they could be easily filled with a variety of different resins. Second, we simulated the binding kinetics of a model library binding to a model target molecule within our device and discovered that aptamers with strong binding affinities for the target (i.e., equilibrium dissociation constant $K_D = 0.5$ nM) were preferentially retained at the input end (i.e., in the first few microliters) of the microfluidic column (Figure 2.2). Aptamers with weaker binding affinities for the target ($K_D \geq 50$ nM) were distributed almost uniformly throughout the microfluidic column with concentrations nearly identical to the input concentration. Therefore, smaller columns increase the mean density of

strong binders, a condition that would require fewer selection rounds to identify aptamer candidates. A previous study used an affinity capillary column that was physically cut into smaller pieces to isolate the highest affinity aptamers in the earliest column segments [21].

We evaluated experimentally the binding distribution of an RNA aptamer, hereafter referred to as GFPapt, on microcolumns that were filled with various amounts of GFP-immobilized resin. It had been previously shown that this aptamer has a strong binding affinity (K_D 5 nM) for GFP [13] and thus serves as a model molecule for the high-affinity target-binding aptamers that are presumed to be present in random libraries. In order to determine the amount of nonspecific, low-affinity (i.e., "background") binding within the microcolumn, a small amount of the random RNA library (5 pmol) was included in addition to the GFPapt (0.064 pmol). We used 0.6 μg of GFP/μL of resin, binding and washing flow rates of 100 μL/min, and qPCR assays to quantify the amount of both the GFPapt and the random library captured on the microcolumn. The experimental results are shown in Figure 2.2 as a percentage density of the amount of both the GFPapt and the N70 random library loaded onto the device. The experimental results are well fit by the simulations for both high-affinity (K_D = 5 nM for GFPapt) and low-affinity ($K_D \geq 10$ μM for the N70 random library) binding.

Optimization of Aptamer Selection Conditions

To optimize the performance of our devices and to further test our simulation predictions, we used microcolumns filled with GFP-immobilized Ni-NTA resin to

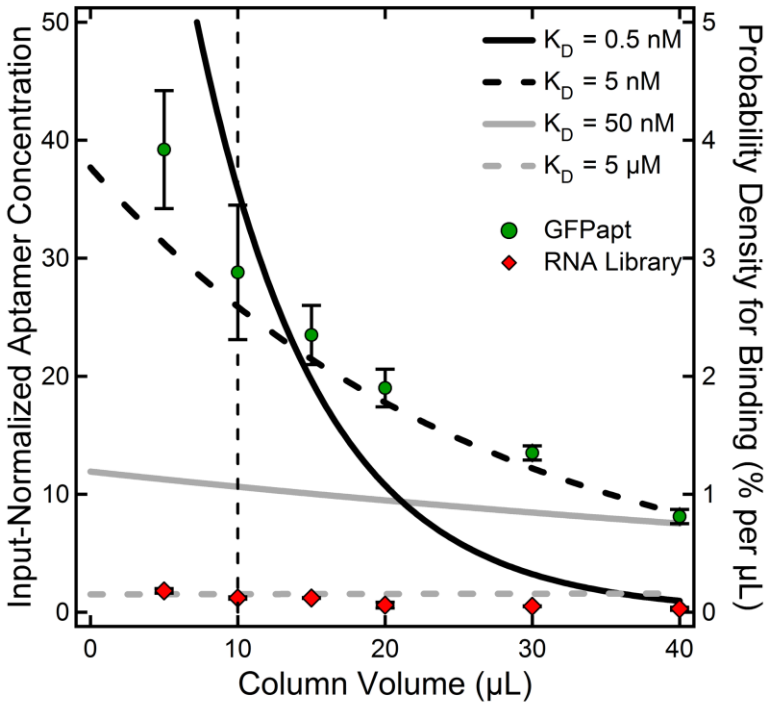


Figure 2.2. Dependence of aptamer recovery on microcolumn volume. Solid and dashed lines are the simulation results for the input-normalized concentrations of aptamers with various binding affinities (K_D ranging from 0.5 nM to 5 µM) as a function of microcolumn volume. Data points are the experimentally measured binding densities, given as percent of the total, of high-affinity GFPapt (green solid circles) and low-affinity, nonspecific-binding N70 random library (red solid diamonds) as a function of microcolumn volume. Input-normalized aptamer concentrations are defined as the bound aptamer concentration at a point along the column divided by the initial aptamer concentration. Probability densities for binding are defined as the probability per unit volume for molecules to bind in the vicinity at a point along the column. The two y-axes are related by the total loaded sample volume. The dashed vertical line at 10 µL marks the volume of the devices used for the optimization experiments.

evaluate the binding behavior of GFPapt and the random N70 library over a wide range of experimental conditions. The results are shown as a percentage of the amount loaded onto the device (Figures 2.3.A and 2.3.C). For each condition, the GFPapt enrichment was calculated as the ratio of the percent amounts of GFPapt to random library (Figures 2.3.B and 2.3.D).

Affinity chromatography resins were developed primarily for protein purification applications and thus are capable of binding relatively large amounts of target molecules; the reported binding capacity of the resin used in our experiments is 20–50 μg of protein/μL of resin. Despite the widespread use of these resins for aptamer selections, there is no information available on how this parameter affects the selection performance, since none of the previous studies optimized the resin binding conditions [7]. In order to determine the optimum amount of bound target, we prepared five different batches of resin with varying amounts of immobilized GFP, from 0.024 to 15 μg of protein/μL of resin in 5-fold increments. The amount of GFPapt captured within the device was strongly dependent on the amount of GFP target on the resin (Figure 2.3.A). Interestingly, we found that the highest recovery (40%) was obtained at the intermediate value of 0.6 μg of protein/μL of resin (15 μM). Library recovery was essentially independent of the amount of target, and so the GFPapt enrichment was also maximized at the same intermediate value (Figure 2.3.B). We had initially hypothesized that saturating the resin with GFP would maximize the number of target binding aptamer molecules while also minimizing nonspecific binding sites on the resin surface, yielding the highest GFPapt enrichment. Kinetic simulations done by a similar approach to that described for determining the optimal device size, except with a higher fraction of strong-binding aptamers (1%) to match the experimental conditions, correctly predicted that there was no effect of target amount on recovery of the random library (Figure 2.4.B). However, the simulations predicted that the highest recovery of the strong-binding aptamer would occur at the conditions with the highest amount of bound target (Figure 2.4.A). We believe that

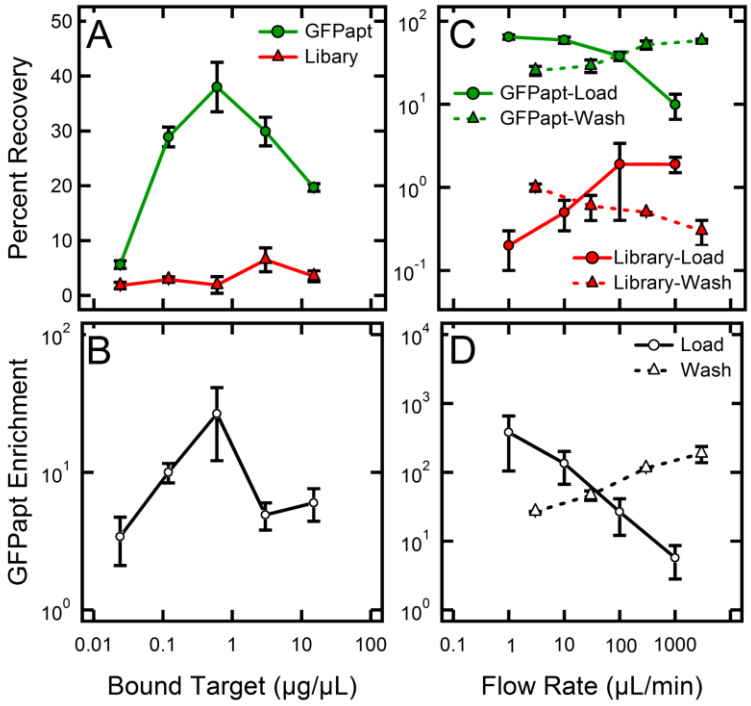


Figure 2.3. Optimization of microcolumn-based selection. (A) Percent recovery of GFPapt (green) and N70 random library (red) and (B) GFPapt enrichment for different amounts of GFP immobilized onto Ni-NTA resin in a 10 µL microcolumn; loading and washing flow rates were 100 µL/min. (C) Percent recovery of GFPapt (green) and N70 random library (red) and (D) GFPapt enrichment for different flow rates; the amount of GFP immobilized onto Ni-NTA resin was 0.6 µg/µL. Solid lines show the effect of loading flow rate at a fixed washing flow rate of 100 µL/min; dashed lines show the effect of washing flow rate at a fixed loading flow rate of 100 µL/min. Error bars represent the standard deviation of triplicate experiments and measurements at each condition.

this discrepancy between our experimental observations and simulation results is primarily due to steric hindrance, where macromolecular crowding effects on the resin surface decrease the binding affinity of the specific aptamer sequence; a similar phenomenon has been reported for the binding of soluble proteins to surface lipid vesicles [22]. Formation of the aptamer–target complex could also hinder the binding of other aptamer molecules. Our results support the use of concentrations below a critical packing density that is determined by the larger of the two biomolecules. All of

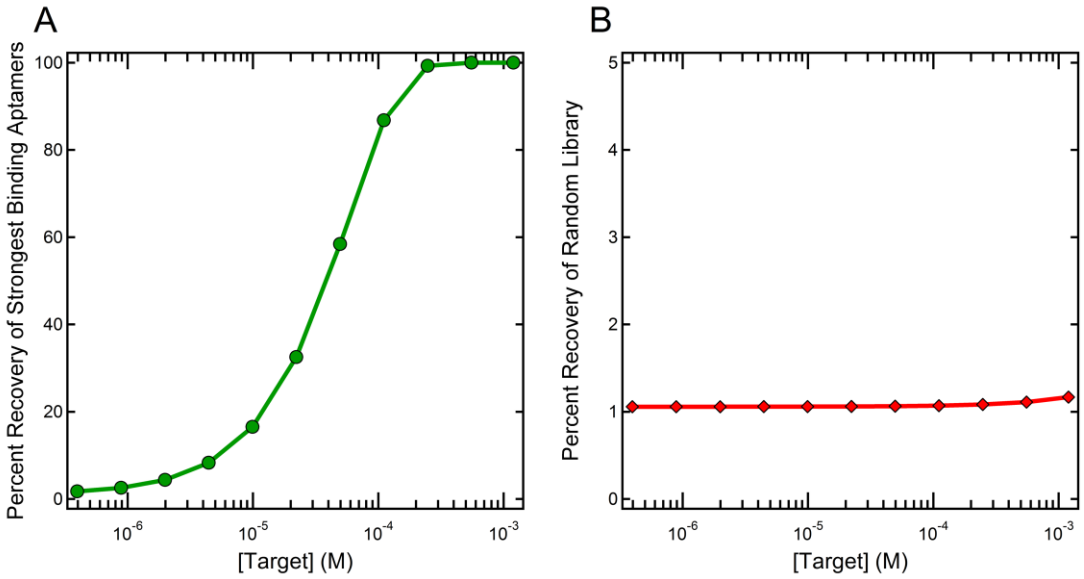


Figure 2.4. Simulation results for the effect of target protein concentration on percent recovery of (A) a strong binding aptamer (K_D = 5 nM) and (B) random library molecules. The other simulation parameters were total resin volume = 10 µL, loading flow rate = 100 µL/min, and washing flow rate = 100 µL/min.

the subsequent optimization experiments were conducted with resin prepared at the optimum value of 0.6 µg of protein/µL of resin.

It is well known that the best affinity chromatography performance is realized when the loading step is operated at the lowest possible flow rates to approach equilibrium conditions. However, there is a practical limitation to the experiment time. In order to determine the optimum condition while also keeping within a practical experimental timeline, we varied the loading flow rate from 1 to 1000 µL/min. The recoveries of GFPapt and library were both strongly dependent on the flow rate but with considerably different trends (Figure 2.3.C). At the lowest flow rate, we observed the highest GFPapt recovery, the lowest library recovery, and thus the best GFPapt enrichment. By operating the process at higher flow rates, we observed a gradual

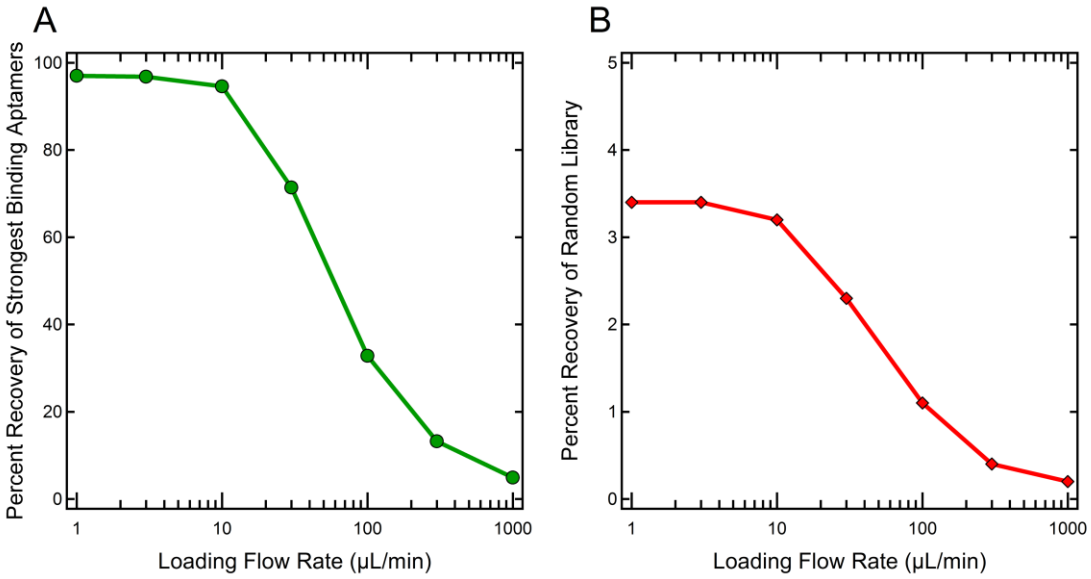


Figure 2.5. Simulation results for the effect of loading flow rate on percent recovery of (A) a strong binding aptamer (K_D = 5 nM) and (B) random library molecules. The other simulation parameters were total resin volume = 10 µL, concentration of immobilized target protein = 22 µM, and washing flow rate = 100 µL/min.

decrease in performance as measured by a decrease in GFPapt enrichment (Figure 2.3.D). Our kinetic simulations correctly predicted that the highest recovery of the strong-binding GFPapt molecule would be obtained at the lowest flow rate (Figure 2.5.A). However, the simulations also predicted the same trend of decreasing recovery with increasing loading flow rate for the random library (Figure 2.5.B). This disagreement between simulation and experimental results for the library is discussed below.

After completion of the loading step, a fixed-volume washing step was employed to improve the separation performance by removing any unbound or weakly bound sequences. We varied the washing flow rate from 3 µL/min to 3 mL/min. We observed similar results to those seen before in that the recoveries of both GFPapt and

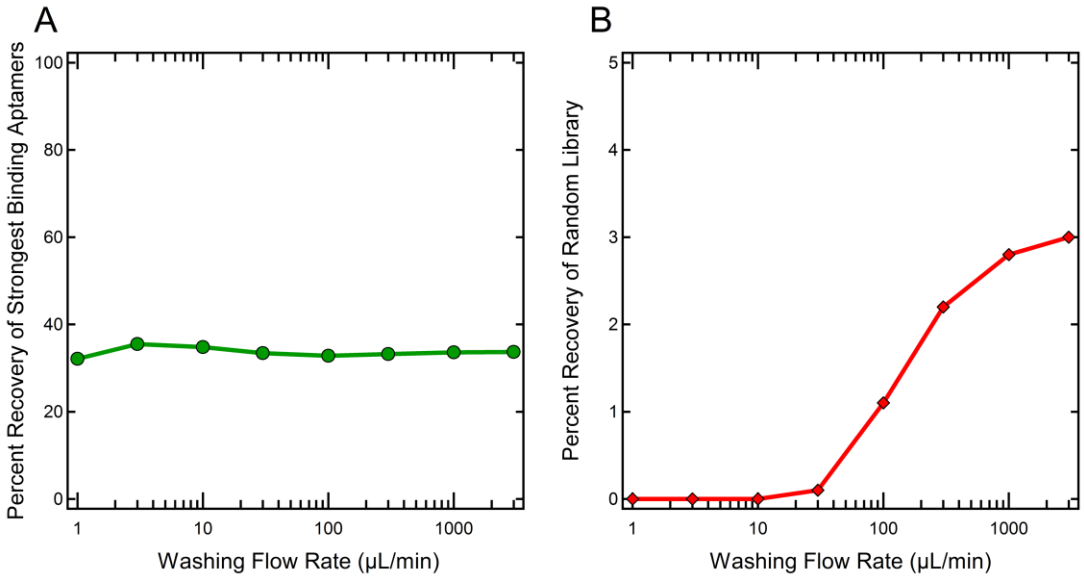


Figure 2.6. Simulation results for the effect of washing flow rate on percent recovery of (A) a strong binding aptamer (K_D = 5 nM) and (B) random library molecules. The other simulations parameters were total resin volume = 10 μL, concentration of immobilized target protein = 22 μM, and loading flow rate = 100 μL/min.

library had different trends with increasing flow rate. Whereas the GFPapt recovery increased at higher flow rates, the library recovery decreased (Figure 2.3.C). Thus, the best GFPapt enrichment was obtained at the highest washing flow rate (Figure 2.3.D). Together, all of the optimization experiments enabled us to choose the optimal experimental conditions that maximize the enrichment of strong-binding aptamers for each selection round, while keeping within practical experimental and time constraints. These conditions are particularly important for the earliest selection rounds when there are only a few copies of each aptamer sequence. Our results also revealed the importance of empirical validation and characterization of different selection conditions, since the kinetic simulations were unable to properly predict all the experimental trends. For example, simulation results for the washing step predicted a gradual increase in the recovery of random library (Figure 2.6)—the exact opposite

trend to that seen experimentally. In our simulations, the behavior of each species was determined by the on- and off-rate kinetic constants. They do not include other phenomena such as shear or pressure-related flow effects that could preferentially affect the behavior of weak binders and the ultimate separation performance of our microcolumns.

After fully characterizing the operating conditions to maximize the enrichment of strongly binding aptamers in our microcolumns, we then proceeded to validate our selection strategy empirically by monitoring the enrichment of GFPapt molecules in de novo selections with multiple protein targets, including GFP. For each round, three microcolumns were filled with Ni-NTA resin that had been preimmobilized with either GFP or two unrelated proteins, CHK2 and UBLCP1. The latter two were chosen because they have similar size and charge properties as the specific target GFP and were readily available with hexahistidine affinity tags (Table 2.1). An empty microcolumn was also included as a control to enable us to discriminate target-binding from device bias.

Table 2.1. Properties of the target proteins

Protein	Molecular Weight (kDa)	Isoelectric Point	Affinity tag
GFP	27	5.5	Hexahistidine (N-terminus)
UBLCP1	37	6.1	Hexahistidine (C-terminus)
CHK2	61	5.6	Hexahistidine (N-terminus)
hHSF1	86	5.3	GST* (N-terminus)
hHSF2	88	4.9	GST *(N-terminus)
dHSF	80	4.9	Hexahistidine (N-terminus)

*GST tag ~ 30 kDa

For the first selection round, four devices were arranged in a serial configuration in the following order: empty, UBLCP1, CHK2, GFP. The GFP target was put at the end so that all of the GFPapt molecules that were combined with the random RNA library would have an opportunity to bind to the other targets; the loading solution for the first selection round contained 40 nmol of the random RNA library (5 copies of each sequence in the 5×10^{15} library) and 6.4 fmol of GFPapt in 1 mL of binding buffer. Thus, the initial molar ratio of library to GFPapt molecules was greater than 6 million to one. Enrichment of GFPapt was determined for each microcolumn from qPCR analyses of the eluted samples (Figure 2.7). Only the microcolumn filled with GFP-loaded resin showed significant GFPapt enrichment. In the first round, the recoveries of the GFPapt and the random library were 20% and 0.022% of the corresponding inputs, respectively, yielding a GFPapt enrichment of approximately 900-fold. However, for the other three devices, the GFPapt enrichments were all near unity, indicating that there was no affinity for GFPapt over the random library. To decrease the material used in the selections, for the next two rounds the amount of the amplified RNA pool loaded onto each device was decreased by 20-fold from the amount used in the previous round. An appropriate amount of GFPapt was "spiked-in" with the amplified RNA pool to maintain the same molar ratio that was recovered in the previous round. After round 3, the amounts of GFPapt that were recovered on the three non-GFP devices were well below the qPCR detection limit, and therefore, no GFPapt enrichment results could be obtained. However, for the GFP device the GFPapt enrichments were much greater than 100-fold per selection round,

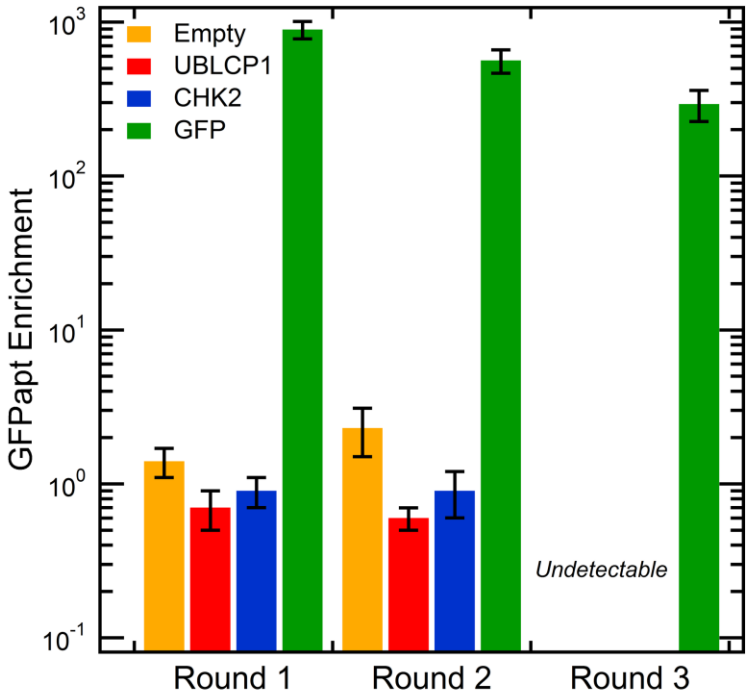


Figure 2.7. Validation of specific aptamer-target enrichment for multiplex SELEX. UBLCP1, CHK2, and GFP were preimmobilized on Ni-NTA resin at a density of 0.6 μg of protein/μL of resin and an empty microcolumn was also included as a control. The loading flow rate for all the rounds was 1 μL/min. The enrichment of GFPapt was monitored on all four microcolumns for three consecutive selection rounds. Results are presented in the same order that the four devices were arranged in the serial configuration used in round 1. GFPapt enrichment was calculated as the ratio of the percent amounts of GFPapt to random library. The error bars represent the standard deviation of triplicate measurements for each microcolumn.

giving rise to a cumulative enrichment of over 108-fold. At the end of round 3, approximately 95% of the GFP selected pool was composed of GFPapt.

Microcolumn SELEX for Human Heat Shock Factor Proteins

To conclusively demonstrate the multiplex utility of the microcolumns, we performed multiplex SELEX for a set of GST-tagged proteins: hHSF1, hHSF2, and four other related proteins. hHSF1 and hHSF2 are transcription factors that regulate

stress response, including heat shock in human cells, and hHSF1 plays a critical role during tumor formation and maintenance [23]. However, important questions remain unanswered regarding their molecular interactions and the specific mechanisms that are used to execute their functions. An RNA aptamer against the Drosophila melanogaster HSF (dHSF) has been selected previously and used to inhibit the binding and recruitment of dHSF to the promoters of heat shock genes [24,25]. However, only moderate cross-reactivity with the hHSF proteins limits the use of this aptamer in functional studies in human cells [26]. Given the biological importance of HSF, we decided to select for aptamers directly to hHSF1 and hHSF2, which can be used as inhibitory tools to dissect the mechanisms of actions of these proteins.

Each protein was preimmobilized onto glutathione–agarose resin at approximately 1 μg of protein/μL of resin. The GST tag was also included as a control and thus seven 20 μL microcolumns were arranged in a serial configuration for round 1. Following the procedure outlined above, another aliquot of the starting RNA library (5 copies of each sequence in the 5×10^{15} library) was loaded onto the seven-microcolumn assembly and then separated into a parallel arrangement for all the subsequent rounds.

A total of five selection rounds were completed. For rounds 2 through 5, in-line negative selections were done by connecting a 10 μL microcolumn filled with GST-immobilized resin to the inlet of each of the six microcolumns for the GST-tagged proteins. These precolumns were removed after the loading step for the subsequent wash and elution of the target protein-bound aptamers from each microcolumn.

We analyzed the sequences from the RNA pools from rounds 3 and 5 for both hHSF1 and hHSF2 by high-throughput sequencing [27,28]. The total number of sequencing reads per pool ranged from 6 million to 9 million. For both proteins, there was a noticeable shift toward higher multiplicity values from round 3 to round 5. In round 3, the top 20 highest multiplicity sequences for each protein represented only 0.04% of the total pool. However, in round 5, the top 20 sequences represented 85.0% and 76.5% of the hHSF1 and hHSF2 selected pools, respectively (Table 2.2). In addition, of the top 20 highest multiplicity sequences in round 3, between a quarter and a half of them were also among the top 20 highest multiplicity sequences in round 5 for hHSF2 and hHSF1. The detection of enriched candidate aptamer sequences in earlier selection rounds was possible because of the high-throughput sequencing, which allowed us to select candidate aptamers for subsequent analysis.

We decided to investigate the potential binding of one of the top candidate sequences for hHSF1 and hHSF2. We chose the highest ranked sequences from the round 5 pools that were also highly ranked in round 3. For hHSF1, this was the first-ranked sequence in round 5, hereafter referred to as hHSF1-R5-1, which was also the 11th-ranked sequence in round 3 (Table 2.2). For hHSF2, this was the second-ranked sequence in round 5, hereafter referred to as hHSF2-R5-2, and also the sixth-ranked sequence in Round 3 (Table 2.2). The full-length sequences and predicted structures of these two aptamer candidates are in Figure 2.8. These two candidates represented 13.4% and 6.8% of the corresponding round 5 pools. The full-length aptamer candidates (including the constant regions) were PCR-amplified from their respective pools by

Table 2.2. The top twenty highest multiplicity candidate aptamers from Round 5 pools for hHSF1 and hHSF2 (total number of sequencing reads = 5,930,382 and 8,644,268 respectively) and the corresponding Round 3 ranks and multiplicities (total number of sequencing reads = 8,072,400 and 9,350,602 respectively). To compare multiplicity values from different pools, the results are presented as multiplicity per 10 million reads.

hHSF1 Sequencing Results (per 10 Million Reads)			
ID	Round 5 Multiplicity	Round 3 Rank	Round 3 Multiplicity
hHSF1-R5-1	1340479	11	133
hHSF1-R5-2	1274151	9	152
hHSF1-R5-3	1261696	6	190
hHSF1-R5-4	621764	22	98
hHSF1-R5-5	550234	54	55
hHSF1-R5-6	530607	24	93
hHSF1-R5-7	469695	3	164
hHSF1-R5-8	405537	30	79
hHSF1-R5-9	358828	7	165
hHSF1-R5-10	272611	27	84
hHSF1-R5-11	271480	10	147
hHSF1-R5-12	191659	1	500
hHSF1-R5-13	170782	5	190
hHSF1-R5-14	156627	41	66
hHSF1-R5-15	143621	26	89
hHSF1-R5-16	127020	2	347
hHSF1-R5-17	98130	40	68
hHSF1-R5-18	97392	8	154
hHSF1-R5-19	84796	35	73
hHSF1-R5-20	76493	43	64

hHSF2 Sequencing Results (per 10 Million Reads)			
ID	Round 5 Multiplicity	Round 3 Rank	Round 3 Multiplicity
hHSF2-R5-1	1627399	51	87
hHSF2-R5-2	682671	6	205
hHSF2-R5-3	660708	47	93
hHSF2-R5-4	569105	33	103
hHSF2-R5-5	567516	511	27
hHSF2-R5-6	558955	49	89
hHSF2-R5-7	485019	342	33
hHSF2-R5-8	384008	9	197
hHSF2-R5-9	370668	69	73
hHSF2-R5-10	304918	93	64
hHSF2-R5-11	225959	295	36
hHSF2-R5-12	212557	116	59
hHSF2-R5-13	176490	162	50
hHSF2-R5-14	146975	41	98
hHSF2-R5-15	139564	650	22
hHSF2-R5-16	123854	45	97
hHSF2-R5-17	121634	108	61
hHSF2-R5-18	117221	10	189
hHSF2-R5-19	87884	1	242
hHSF2-R5-20	86977	11	161

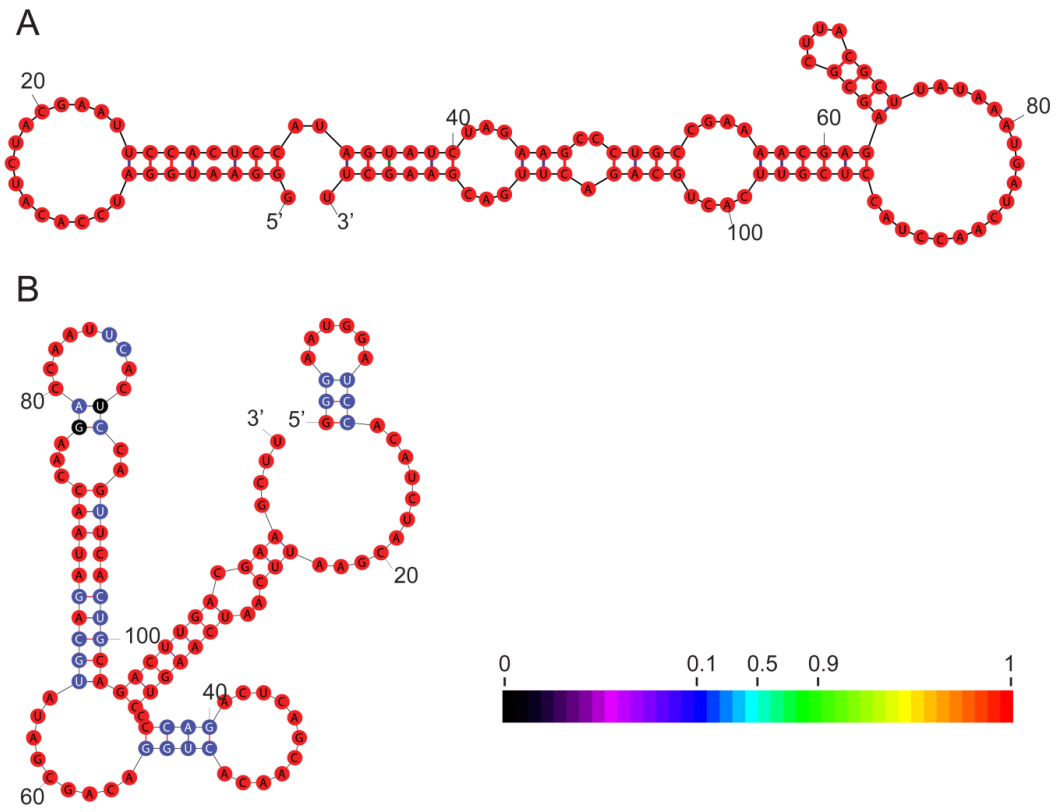


Figure 2.8. Mfold predicted secondary structures of (A) hHSF1-R5-1 and (B) hHSF2-R5-2; the full sequences for these putative RNA aptamers were 5'-GGGAAUGGAUCCACAUCUACGAAUUCCACUCCAUAGUAUCUAGAAGCCCUGCC GAAAACGAGAGCGCUUACGCUUAUAAAUGAUCAACCUACCUCGUUCACUGCAG ACUUGACGAAGCUU-3' and 5'-GGGAAUGGAUCCACAUCUACGAAUUCAAUCAAG UCCCCAGACUCAGCAACACUGGACAGCGAUAUGCAGAUAACCAAGACCAAUUC ACUCCAGUUCACUGCAGACUUGACGAAGCUU-3' respectively. The two constant regions corresponding to the library design are denoted by underlines. The structures are annotated using p-num with the color representing the probability according to the color legend.

use of a candidate-specific forward oligo and the reverse constant region oligo, and then cloned into plasmid vectors to obtain a pure template.

The putative RNA aptamers were fluorescently end-labeled and then tested for binding to their hHSF targets by F-EMSA and FP assays [19]. An image of a typical F-EMSA result is shown in Figure 2.9.A for hHSF1-R5-1 aptamer binding to hHSF1 protein. The fraction of bound aptamer was calculated as a function of protein

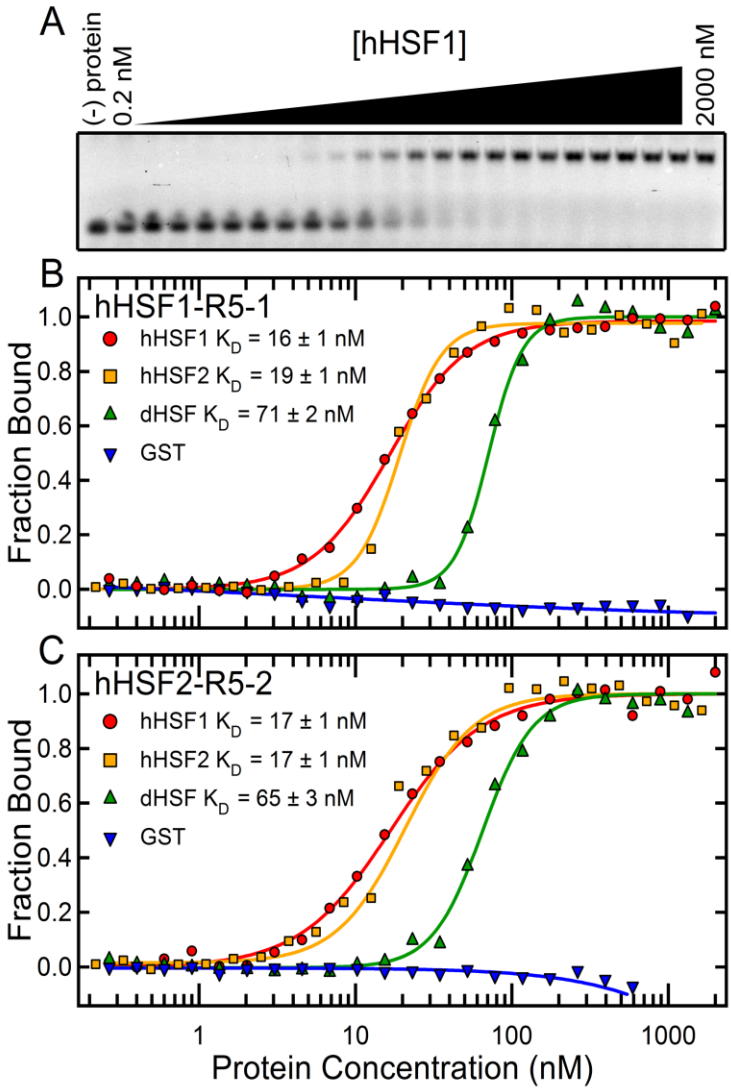


Figure 2.9. Evaluation of candidate aptamers binding to target proteins. (A) Typical F-EMSA results for hHSF1-R5-1 aptamer binding to a two-thirds dilution series (from 2000 nM to 0.2 nM) of hHSF1 protein. (B, C) Binding curves measured by F-EMSA for hHSF1-R5-1 and hHSF2-R5-2 aptamers to hHSF1, hHSF2, dHSF, and GST tag. The same dilution series in panel A was used in panels B and C. The solid lines are the best fits of the Hill equation to the experimental data for each aptamer–target pair with the appropriate K_D values given in the figure legends.

concentration and then plotted as shown in Figures 2.9.B and 2.9.C for various aptamer–protein pairings; K_D values were determined by fitting each data set to the Hill equation. Overall, both aptamers showed high-affinity binding to hHSF1 and hHSF2 (K_D < 20 nM). Interestingly, both aptamers also bound to hexahistidine-tagged dHSF, although slightly more weakly (K_D ~ 70 nM), and no binding was observed to the GST-tag alone. The F-EMSA results were confirmed by the FP assays (Figures 2.10 and 2.11). Thus, the observed binding is not due to the affinity tags on the protein targets but rather to specific domains of the targets themselves. Given that the highest degree of sequence similarity between hHSF1, hHSF2, and dHSF is in the DNA binding and trimerization domains (DBD-TD) [29] and that the previously selected dHSF aptamer was found to bind the DBD-TD of dHSF, [24] we predict these novel hHSF aptamers are likely to bind the HSF proteins in a similar fashion. Contrary to their functional similarity, these two aptamers did not show any similarity in secondary structure as predicted by mFold [30] (Figure 2.8). Although beyond the scope of the present work, the detailed mechanism of these and other potential aptamers' binding to HSF proteins as well as the consequences of binding await further study. However, successful selection of two distinct high-affinity aptamers, hHSF1-R5-1 and hHSF2-R5-2, targeting two closely related proteins, hHSF1 and hHSF2 respectively, in a single selection demonstrates that our microcolumn-based SELEX technology is capable of yielding high-affinity aptamers (K < 20 nM) in as little as five rounds of selection, whereas most conventional SELEX methods require typically 12 rounds of selection [7].

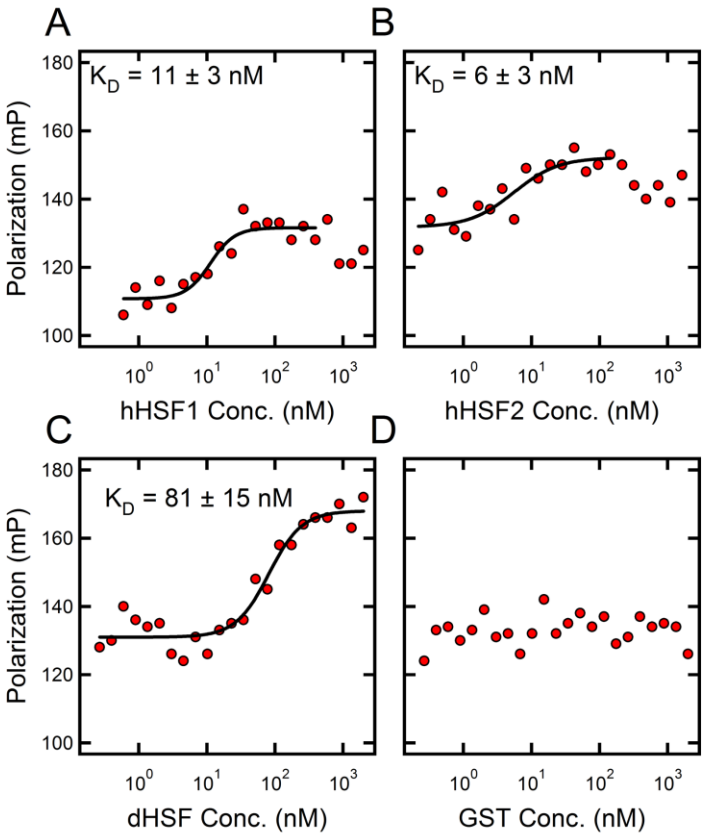


Figure 2.10. Fluorescence polarization (FP) assay results for binding of hHSF1-R5-1 aptamer to GST-hHSF1 (A), GST-hHSF2 (B), hexahistidine-dHSF (C), and GST-tag (D). The solid line in each panel (except D) is the best-fit of the Hill equation to the experimental data with the given K_D value.

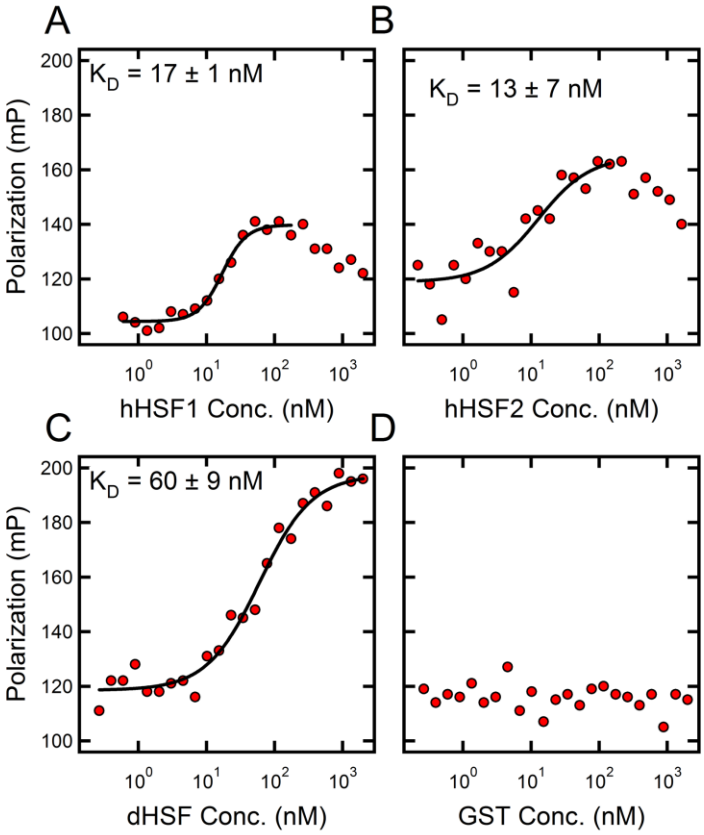


Figure 2.11. Fluorescence polarization (FP) assay results for binding of hHSF2-R5-2 aptamer to GST-hHSF1 (A), GST-hHSF2 (B), hexahistidine-dHSF (C), and GST-tag (D). The solid line in each panel (except D) is the best-fit of the Hill equation to the experimental data with the given K_D value.

CONCLUSIONS

We have developed a microcolumn-based technology for the efficient selection of nucleic acid aptamers for multiple targets. Our microcolumns have a number of advantages over other chromatography-based processes. First, they can be readily assembled in various configurations to accommodate multiple targets during the selection step; as a proof of principle, we used a single aliquot of starting RNA aptamer library to perform selections for seven different immobilized target proteins in separate microcolumns. Second, as either assembled or disassembled units, they do not require any specialized equipment to perform the selection step; we used a multirack syringe pump to control the solution flow rates during the selection step. Third, they require very small amounts of resin and target molecules. For this study, we focused on the selection of RNA aptamers; however, our approach would also work for DNA aptamers. The multiplex selection presented in this work is simpler and thus easier to perform than previous microfluidic techniques that use sol–gel chemistry [8]. Also, compared to filter-plate-based methods [9], our modular columns allow "counter-selections" to be done simultaneously with the selection step to enhance target specificity.

The overall performance of our devices was evaluated in two different sets of experiments. In the first, we looked at a single selection step and optimized its performance by monitoring the enrichment of GFPapt, a known RNA aptamer for GFP. The performance was strongly dependent on the amount of target immobilized onto the resin and the flow rates for both the loading and washing steps. In the second, we used the optimal conditions as part of a selection and amplification strategy to verify specific GFPapt partitioning over three protein preparations (UBLCP1, CHK2, and GFP), followed by a complete multiplex selection against hHSF1 and hHSF2. High-throughput sequencing analysis of the selected pools from multiple rounds

showed an enrichment of specific aptamer sequences. For hHSF1 and hHSF2, the sequences from round 5 with the highest multiplicity values also had high interround enrichments. Those sequences were target-specific and could be easily identified as being preferentially enriched after just three selection rounds. We tested a single candidate each to hHSF1 and hHSF2 and found both to be high-affinity aptamers to full-length hHSFs.

Although HSF proteins have been extensively studied and characterized, we are still limited by the available methods and lack approaches to perturb the activity of specific factors to tease apart molecular interactions. Aptamers can act as inhibitors that bind to a protein surface and disrupt specific interactions or functions. When expressed in vivo in a temporally and spatially controlled manner, these aptamers provide a way to rapidly disrupt targeted domains of proteins and efficiently assess their primary functions and mechanisms of actions. We previously demonstrated the utility of inhibitory RNA aptamers to study macromolecular interactions in vivo [24,25,31]. However, there were some limitations on the methodology used to select those aptamers, and we believe that our new method will significantly improve the SELEX efficiency by: (1) allowing the selection of aptamers for many targets, including different domains of a single protein, at the same time and (2) reducing the number of SELEX rounds to achieve the selection of high-affinity aptamers. With this multiplexed technology, we believe it will be possible to efficiently select aptamers that bind to the distinct domains of HSF and other proteins, which will be extremely valuable to study the interactions of these proteins.

Extensions of the microcolumn devices and approach developed in this work could be used for selection strategies with various combinatorial libraries, including but not limited to genomic sequences [32], mRNA display [33], and peptide nucleic acids [34]. Also, our multiplex approach would easily facilitate the discovery of

multivalent aptamers to distinct target binding sites [35] by performing a final selection by use of serially arranged microcolumns with different target subunits in each. Finally, our microcolumn devices could be used to discriminate for aptamers based on their on- or off-rate binding characteristics by enforcing certain restrictions on the flow rates used for the loading step and washing step, respectively [36].

REFERENCES

1. Ciesiolka J, Gorski J, Yarus M (1995) Selection of an Rna Domain That Binds Zn2+. Rna-a Publication of the Rna Society 1: 538-550.

2. Shangguan D, Li Y, Tang Z, Cao ZC, Chen HW, et al. (2006) Aptamers evolved from live cells as effective molecular probes for cancer study. Proc Natl Acad Sci U S A 103: 11838-11843.

3. Tombelli S, Mascini M (2009) Aptamers as molecular tools for bioanalytical methods. Curr Opin Mol Ther 11: 179-188.

4. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.

5. Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. Gene 82: 83-87.

6. Ellington AD, Szostak JW (1990) Invitro Selection of Rna Molecules That Bind Specific Ligands. Nature 346: 818-822.

7. Gopinath SC (2007) Methods developed for SELEX. Anal Bioanal Chem 387: 171-182.

8. Park SM, Ahn JY, Jo M, Lee DK, Lis JT, et al. (2009) Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. Lab Chip 9: 1206-1212.

9. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, et al. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20: 861-873.

10. Nieuwlandt D, Wecker M, Gold L (1995) In vitro selection of RNA ligands to substance P. Biochemistry 34: 5651-5659.

11. Liu JJ, Stormo GD (2005) Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. Nucleic Acids Research 33.

12. Cheng C, Dong J, Yao L, Chen A, Jia R, et al. (2008) Potent inhibition of human influenza H5N1 virus by oligonucleotides derived by SELEX. Biochem Biophys Res Commun 366: 670-674.

13. Shui B, Ozer A, Zipfel W, Sahu N, Singh A, et al. (2012) RNA aptamers that functionally interact with green fluorescent protein and its derivatives. Nucleic Acids Research 40: e39.

14. Holeman LA, Robinson SL, Szostak JW, Wilson C (1998) Isolation and characterization of fluorophore-binding RNA aptamers. Fold Des 3: 423-431.

15. Berg OG, von Hippel PH (1987) Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J Mol Biol 193: 723-750.

16. Cox JC, Rudolph P, Ellington AD (1998) Automated RNA selection. Biotechnol Prog 14: 845-850.

17. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.

18. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.

19. Pagano JM, Farley BM, McCoig LM, Ryder SP (2007) Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. J Biol Chem 282: 8883-8894.

20. Eulberg D, Buchner K, Maasch C, Klussmann S (2005) Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonist. Nucleic Acids Research 33: e45.

21. Nitsche A, Kurth A, Dunkhorst A, Panke O, Sielaff H, et al. (2007) One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. Bmc Biotechnology 7.

22. Leventis R, Silvius JR (2010) Quantitative experimental assessment of macromolecular crowding effects at membrane surfaces. Biophys J 99: 2125-2133.

23. Dai C, Whitesell L, Rogers AB, Lindquist S (2007) Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell 130: 1005-1018.

24. Zhao XC, Shi H, Sevilimedu A, Liachko N, Nelson HCM, et al. (2006) An RNA aptamer that interferes with the DNA binding of the HSF transcription activator. Nucleic Acids Research 34: 3755-3761.

25. Salamanca HH, Fuda N, Shi H, Lis JT (2011) An RNA aptamer perturbs heat shock transcription factor activity in Drosophila melanogaster. Nucleic Acids Research 39: 6729-6740.

26. Unpublished Work, Cornell University (2009).

27. Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, et al. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. Proc Natl Acad Sci U S A 107: 15373-15378.

28. Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. PLoS One 5: e9169.

29. Anckar J, Sistonen L (2007) Heat shock factor 1 as a coordinator of stress and developmental pathways. Adv Exp Med Biol 594: 78-88.

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

31. Shi H, Hoffman BE, Lis JT (1999) RNA aptamers as effective protein antagonists in a multicellular organism. Proc Natl Acad Sci U S A 96: 10033-10038.

32. Singer BS, Shtatland T, Brown D, Gold L (1997) Libraries for genomic SELEX. Nucleic Acids Research 25: 781-786.

33. Wilson DS, Keefe AD, Szostak JW (2001) The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A 98: 3750-3755.

34. Brudno Y, Birnbaum ME, Kleiner RE, Liu DR (2010) An in vitro translation, selection and amplification system for peptide nucleic acids. Nature Chemical Biology 6: 148-155.

35. Gong Q, Wang JP, Ahmad KM, Csordas AT, Zhou JH, et al. (2012) Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. Anal Chem 84: 5365-5371.

36. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.

CHAPTER 3

RAPID-SELEX FOR RNA APTAMERS[‡]

ABSTRACT

Aptamers are high-affinity ligands selected from DNA or RNA libraries via SELEX, a repetitive in vitro process of sequential selection and amplification steps. RNA SELEX is more complicated than DNA SELEX because of the additional transcription and reverse transcription steps. Here, we report a new selection scheme, RAPID-SELEX (RNA Aptamer Isolation via Dual-cycles SELEX), that simplifies this process by systematically skipping unnecessary amplification steps. Using affinity microcolumns, we were able to complete a multiplex selection for protein targets, CHK2 and UBLCP1, in a third of the time required for analogous selections using a conventional SELEX approach. High-throughput sequencing of the enriched pools from both RAPID and SELEX revealed many identical candidate aptamers from the starting pool of 5×10^{15} sequences. For CHK2, the same sequence was preferentially enriched in both selections as the top candidate and was found to bind to its respective target. These results demonstrate the efficiency and, most importantly, the robustness of our selection scheme. RAPID provides a generalized approach that can be used with any selection technology to accelerate the rate of aptamer discovery, without compromising selection performance.

---

‡The following sections are reprinted with permission from: Szeto, K.*, Latulippe, D.R.*, Ozer, A., Pagano, J.M., White, B.S., Shalloway, D., Lis, J.T., and Craighead, H.G. 2013. RAPID-SELEX for RNA Aptamers. PLoS ONE. 8(12) e82667, doi: 10.1371/journal.pone.0082667, with modifications to conform to the required format.

* KS and DRL are co-first authors.

INTRODUCTION

Aptamers are high-affinity ligands selected from large libraries of random oligonucleotides that can contain up to 10^{16} unique sequences. SELEX (Systematic Evolution of Ligands by EXponential enrichment) [1-3], an in vitro selection method, can isolate aptamers with high-affinity and specificity for a wide range of target molecules from DNA or RNA libraries [4-6]. This is achieved by iteratively selecting and amplifying target-bound sequences to preferentially enrich those sequences with the highest affinity to the target. Traditionally, after 10 to 15 iterations, one or several aptamers may be identified from the enriched pool, a process that may take months to complete. If an RNA aptamer is desired, this process takes even longer due to additional steps required for reverse transcription to amplifiable cDNA and subsequent transcription back to RNA. A disproportionate amount of time and effort is dedicated to amplifying RNA pools compared to the actual selection steps where aptamer enrichment takes place.

Recent work has focused on improving selection efficiency and enriching for aptamers with particular target-binding properties. This has resulted in modifications to the conventional SELEX strategy including the use of multiple targets to control specificity [7-9], changing the characteristics of the nucleic acid library [10-16], using different substrates for presentation of target molecules [1,17-20], and varying the separation technique [1,17,21,22]. Work has also been done to improve the throughput of aptamer discovery by utilizing high-throughput sequencing [17,23-26] or by performing parallel selections [19,27]. A number of automated selection strategies have also been introduced [28]. However, fully automated systems lack the quality

controls and evaluations that are applied when manual selections are performed [29]. Recently, we reported a multiplexed microcolumn technique that optimized selection parameters based on enrichment of a specific aptamer and demonstrated the ability to efficiently perform selections against multiple targets in parallel [30]. However, there is still a lack of thorough characterization and knowledge about the most efficient or effective methods and conditions for performing selections with emerging technologies. Improvements in this domain would not only increase the rate of aptamer selections, but have the potential to improve the rate and quality of downstream aptamer identification and refinement [30,31]

Despite many advances, only a few selection approaches diverge from the core methodology of traditional SELEX. To our knowledge, only one method breaks from the typical cycle of iterative and sequential selection and amplification steps; Non-SELEX [32] was shown to quickly generate DNA aptamers by repeated selections from an enriched library without any amplification steps. This methodology only takes about an hour to complete and is particularly useful for libraries that cannot be amplified. However, the capillary electrophoresis-based platform used for Non-SELEX requires tiny injection volumes (~150 nL) to achieve efficient separations and only a small fraction of the sequences recovered from a given selection cycle are re-injected for the subsequent cycle. This constraint significantly lowers the total number of sequence candidates that can be investigated, decreasing the complexity and diversity of the injected library by 5 or 6 orders of magnitude. Despite these restrictions, Non-SELEX was successfully used to identify DNA aptamers to h-RAS protein, bovine catalase and signal transduction proteins [32-34], which suggests that

in some cases aptamers may be much more abundant in random pools than previously thought. However, without the amplification steps utilized in traditional SELEX, this technique makes identifying aptamer candidates via population-based methods difficult. This limits the potential for using high-throughput sequencing, which has been used to characterize sequence distributions and their cycle-to-cycle dynamics, and has proven to be a powerful technique for identifying enriching aptamers with great sensitivity [17,23,25,26,30].

Here we propose a new scheme, RNA Aptamer Isolation via Dual-cycles SELEX (RAPID-SELEX or RAPID for short), which combines the efficiency of Non-SELEX with the robustness of conventional SELEX and provides a generalized approach for accelerating the rate of aptamer selections. RAPID significantly decreases the time required for RNA aptamer selections by systematically eliminating unnecessary amplification steps and performing amplifications only when higher numbers of certain sequences (referred to as the copy number) or higher pool concentrations are required. This results in a process that maximizes enrichment per unit time, rather than enrichment per cycle. For each additional selection cycle performed without amplification (Non-Amplification Cycle), the additional effort associated with RNA specific processes, such as reverse transcription and transcription is eliminated in addition to the typical PCR amplification of DNA templates. Furthermore, RAPID can be applied to any selection mode and used with any technology, including those that utilize whole cells and target cell surface proteins as in Cell-SELEX [18]. We demonstrate the improved efficiency of RAPID, by comparing and analyzing its sequence candidates to those generated from

conventional SELEX using our previously described, microcolumn-based platform [30] to the target proteins, CHK2 and UBLCP1. CHK2 and UBLCP1 are a kinase and a phosphatase, respectively, and were chosen because they were readily available and no aptamer selections had been previously performed against them. After completing six selection cycles, RAPID had enriched many of the same candidates, but in only a third of the time required for conventional SELEX.

MATERIAL AND METHODS

Protein preparation

As previously described [30], recombinant hexahistidine-tagged CHK2 and UBLCP1 proteins were expressed in BL21(DE3)-RIPL E. coli cells (Agilent Technologies). LB cultures supplemented with 100 µg/ml ampicillin were inoculated with starter LB culture derived from a single colony and grown at 37 °C until OD_{600} reached 0.6. Protein expression was induced with 0.2 mM IPTG at 18-22 °C for ~16 hours. After centrifugation, the bacterial pellet was collected and processed according to the manufacturer's instructions for Ni-NTA Superflow resin (Qiagen). SDS-PAGE was used to determine the purity and quality of the final protein product. The resulting proteins were dialyzed with 1×PBS with 5 mM 2-mercaptoethanol and 0.01% Triton X-100. The proteins were evaluated for purity (~90-95%) and were stored in small aliquots with 20% glycerol.

RNA library preparation

As previously described [30], a synthesized DNA library was purchased from GenScript. To increase the diversity of the initial library and to include higher order RNA structural classes, we chose to use a random region of 70 nucleotides (nt); this length averages about 4.5 structural features (vertexes) [35]. Including flanking constant regions, sequences in the library have 120 nts, as described by the scheme: 5'-AAGCTTCGTCAAGTCTGCAGTGAA-N70-GAATTCGTAGATGTGGATCCATTCCC-3'. This length is the practical limit for efficient commercial synthesis of DNA templates. The single-stranded DNA template library was converted to double-stranded DNA while introducing the T7 promoter using Klenow exo- (NEB) and the Lib-FOR oligonucleotide, 5'-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3'. The resulting library was later amplified in a 1 L PCR reaction using Taq DNA polymerase, Lib-FOR oligonucleotide, and the Lib-REV oligo, 5'-AAGCTTCGTCAAGTCTGCAGTGAA-3'. A single aliquot capturing the complexity of the entire library (5×10^{15} unique sequences) was transcribed with T7 RNA polymerase in an 88 mL reaction yielding 1200-fold amplification. An aliquot of this RNA library, corresponding to an average of 4 to 6 copies of each unique sequence, was used as the starting pool for each selection method.

Multiplex SELEX and RAPID

The protein immobilization was described previously [30]. Briefly, a new batch of resin was prepared for each protein target. Ni-NTA Superflow resin was

incubated in binding buffer (25 mM Tris-HCl, pH 8.0, 10 mM NaCl, 25 mM KCl, 5 mM $MgCl_2$) with each protein to the optimal final concentration of ~0.6 µg protein/µl of resin and then loaded into custom fabricated microcolumns [30]. For both SELEX and RAPID, three microcolumns were serially connected beginning with an Empty microcolumn, followed by UBLCP1 and ending with CHK2. Fresh aliquots of the RNA Library were prepared in 1 mL binding buffer by heat denaturing at 65˚C for 5 minutes, renaturing at 25˚C for 30 minutes and finally adding 200U of Superase-In RNase Inhibitor (Invitrogen). 10 µL samples were taken as 1% standards for subsequent quantitation by qPCR.

For the SELEX cycles, 1 mL of blocking buffer (binding buffer supplemented with 0.3 µg/µL yeast tRNA) was injected into the microcolumn assembly at a rate of 100 µL/min. The library was injected at the optimum rate of 1 µL/min using a multi-rack syringe pump (Harvard Apparatus) [30]. After binding the library, the microcolumns were reconfigured to run in parallel, and a 3 mL washing step was performed at the optimum rate of 3 mL/min with binding buffer containing 10 mM imidazole. Finally, the protein and bound sequences were collected from the microcolumns by flowing 400 µL of elution buffer (binding buffer supplemented with 50 mM EDTA) at 50 µL/min. By chelating the nickel ion (Ni^{+2}) from the resin with EDTA, protein-resin binding was disrupted allowing the recovery of all protein-RNA complexes and thus avoiding elution bias against potential Mg-independent binding aptamers. Each RNA sample was then phenol:chloroform and chloroform extracted, ethanol precipitated together with 1 µL of GlycoBlue (Ambion) and 40 µg of yeast tRNA (Invitrogen), and re-suspended in 20 µL of DEPC-treated water. These were

then reverse transcribed, PCR amplified, and transcribed into RNA (see below for details) for the next selection cycle. Five more SELEX cycles using the three microcolumns were completed in parallel, decreasing the washing flow rate by 10-fold at Cycles 3 and 6 to accommodate possible increases in the bulk affinity of the enriched pools. The input material was also decreased by 20-fold each cycle from Cycle 2 to 4 to decrease the time and reagents needed.

For the RAPID cycles, 1 mL of blocking buffer was injected into the serial microcolumn assembly at 100 µL/min. The library injections were performed at 10 µL/min to allow the completion of multiple selection cycles in one day. For the wash step, we used a 3 mL two-step wash at 1 mL/min for 1 minute, followed by 70 µL/min for 29 minutes. This combined the observed benefits of a brief, harsh wash for eliminating weakly bound or unbound molecules, with that of a longer wash for discriminating among more strongly bound molecules [30]. Elution buffer was then injected to recover bound sequences, which were then phenol:chloroform and chloroform extracted, ethanol precipitated, re-suspended in 1 mL binding buffer, and then used as the input pool for the next selection cycle. We took 1% standards/samples from each new pool and then the selection steps were repeated with all of the microcolumns in parallel. Following the completion of the elution step after the second cycle, each RNA sample was extracted, precipitated, and re-suspended in 20 µL of DEPC-treated water and processed for the next selection cycle. Two more RAPID “dual-cycles” (one Non-Amplification and one Amplification Cycle) were completed using the three microcolumns in parallel, decreasing the input material by 20-fold after each amplification cycle (Cycle 3 and 5).

The amplification and quantification of both the SELEX and RAPID pools were performed in the same way. All the resuspended samples and standards were reverse transcribed in 60 µL reactions with MMLV-RT and 30 pmol of Lib-REV primer. The cDNA samples were treated with RNaseH (Ambion) and a small amount analyzed on a LightCycler 480 qPCR instrument (Roche) to determine the amount of RNA that was recovered and to determine the optimal number of PCR cycles. 400 µL PCR reactions with 300 pmol of each primer were performed for each pool, followed by phenol:chloroform and chloroform extractions, and finally purified using DNA Clean & Concentrator (Zymo Research) spin columns. A small fraction (~1/4) of the purified PCR product was used to generate new RNA pools in 72 µL transcription reactions with T7 RNA polymerase. The template DNA was removed by DNaseI digestion and the resulting RNA pool was purified by phenol:chloroform and chloroform extractions and ethanol precipitation.

High-throughput sequencing and analysis

A detailed description has been reported [30]. Briefly, PCR products from each target pool for various selection rounds were PCR amplified using 6 nt barcoded primers with adapters for the HiSeq 2000 (Illumina) sequencing platform. The barcoded PCR products were PAGE-purified, phenol:chloroform and chloroform extracted, ethanol precipitated, and then re-suspended in 10 mM Tris-HCl pH 7.5 buffer. High-throughput sequencing was performed by the sequencing core facility at Life Sciences Core Laboratories Center, Cornell University. After removing ambiguous and poor scoring sequences the remaining sequences were separated into

pools based on the barcode sequences. Then sequences with 85% sequence identity were clustered together. This identity threshold is set to ensure that truly unique sequences with 85% identity (or higher) are unlikely to be present even within our large library size (2.5×10^{15}) due to the vast potential 70 nt random sequence space (4^{70} = ~1.4×10^{42}) and thus such detected sequences account for PCR and sequencing errors. The sequence with the highest number of reads, hereafter referred to as the sequence multiplicity, within each cluster was identified as the cluster's true sequence and used as the representative sequence for that cluster. The total multiplicity of a cluster was defined as the sum of multiplicities within the cluster. All the representative sequences in each pool were sorted based on their multiplicity to identify candidate aptamers for each protein target. The top 10,000 highest multiplicity sequences for each pool are available online. Sequence comparisons, histograms and scatterplots were performed and generated in MATLAB (Mathworks).

Candidate sequence purification

The DNA templates for candidate aptamers were PCR amplified from the final Cycle 6 pool using Phusion Polymerase (NEB), the Lib-REV oligonucleotide, and an aptamer-specific oligonucleotide that spans the forward constant region and approximately 30 nt of the candidate's unique, random region. The resulting PCR product was double-digested with BamHI and PstI, and ligated using low melt agarose "in-gel" ligation (EZ Clone Systems) into a similarly cut pGEM3Z-N70Apt plasmid. PGEM3Z-N70Apt plasmid was obtained by cloning a random full-length aptamer template from the N70 library together with T7 promoter into the pGEM3Z vector

(Promega) between NarI and HindIII sites. Three clones were sequenced to obtain a consensus for the full-length sequence of each candidate aptamer. The RNA aptamer was transcribed from the candidate's DNA templates, which were generated by PCR from the sequenced plasmid using the same primers.

Fluorescence EMSA and Polarization assays

The RNA samples were 3'-end labelled with fluorescein 5-thiosemicarbazide (Invitrogen) as described previously [36]. 50 µL binding reactions were prepared with 2 nM fluorescently-labelled RNA and decreasing amounts of protein (2000 to 0 nM) in binding buffer containing 0.01% IGEPAL CA630, 10 µg/ml yeast tRNA, and 3U of SUPERase•In RNase Inhibitor. Reactions were prepared in black 96-well half area microplates (Corning) and incubated at room temperature for 2 hours. The plates were scanned on a Synergy H1 microplate reader (BioTek) using the Ex: 485/20 Em: 528/20 filter set to determine the Fluorescence Polarization (FP). The polarization *P* is determined from the total parallel and perpendicular polarized fluorescence according to:

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}} \qquad (1)$$

For Fluorescence Electrophoretic Mobility Shift Assays (F-EMSA), the same samples used for the FP measurements were spiked with 6× loading dye and loaded into the wells of a refrigerated 5% agarose gel prepared with 0.5×TBE buffer. The gel was run for 90 minutes at 120 volts in refrigerated 0.5×TBE buffer. Images were acquired using the fluorescein scan settings on a Typhoon 9400 imager (GE Healthcare Life

Sciences) and the resulting bands were quantified with ImageJ. The dissociation constant, K_d, was determined by fitting the binding results, *Y*, from the FP and F-EMSA to the Hill equation:

$$Y = Y_0 + \frac{Y_{MAX} - Y_0}{1 + \left[\frac{K_d}{X}\right]^n} \qquad (2)$$

where Y_{MAX} is the maximum signal from binding, Y_0 is background, *n* is the Hill coefficient, and *X* is the protein concentration.

RESULTS

SELEX versus RAPID

Traditional SELEX is performed with a random library via iterative cycles of sequential steps (binding, partitioning, and amplification of target-bound sequences) until an aptamer emerges. To improve the efficiency of these selections, we developed and tested a hybrid selection scheme between SELEX and Non-SELEX that utilizes two cycles; one that includes amplifications and one which does not. For simplicity, we differentiate these two cycles as Amplification and Non-Amplification Cycles (Figure 3.1.A). By systematically eliminating certain amplification steps, RNA selections can be performed in much less time, and require less reagents and other costly materials. In addition, removing unnecessary amplification steps minimizes their potential biases [24,37] and also reduces large input libraries and pools to more convenient size scales when performing amplifications. Thus, rapid sequence convergence can be obtained via Non-Amplification Cycles, while diverse sequence

populations with high aptamer copy numbers are maintained through critical periodic Amplification Cycles.

To illustrate the validity of the RAPID method for RNA aptamer selections, we compared the simplest RAPID protocol (a single non-amplification cycle followed by an amplification cycle) to conventional SELEX (amplification at every cycle). Representative timelines for two cycles of RAPID and conventional SELEX conducted with the exact same selection conditions are shown in Figure 3.1.B. Completion of one cycle of conventional SELEX takes about 24 hours, over 80% of which is needed for the amplification step. In contrast, by adding one Non-Amplification Cycle, RAPID completes two selection cycles in nearly the same amount of time. For both methods, we define a selection "round" to necessarily include the amplification steps. In this way, a round of RAPID is comparable in time and effort to a round of SELEX; a round and a cycle are interchangeable terms in conventional SELEX.

To evaluate the advantage of using RAPID, we completed six selection cycles on the same set of targets using both the RAPID and conventional SELEX methods. As shown in Figure 3.1.C, SELEX took a total of 255 hours using the optimal parameters for aptamer enrichment on the microcolumns as determined in our previous work [30]. RAPID took only 84 hours to complete the three rounds with six selection cycles (Figure 3.1.C). However, different parameters were used to allow for the completion of two selection steps within one working day (i.e. a 10 hour time period). With this simple design, RAPID was straightforward to execute and took one third the time to complete as SELEX. If the same selection step parameters were used

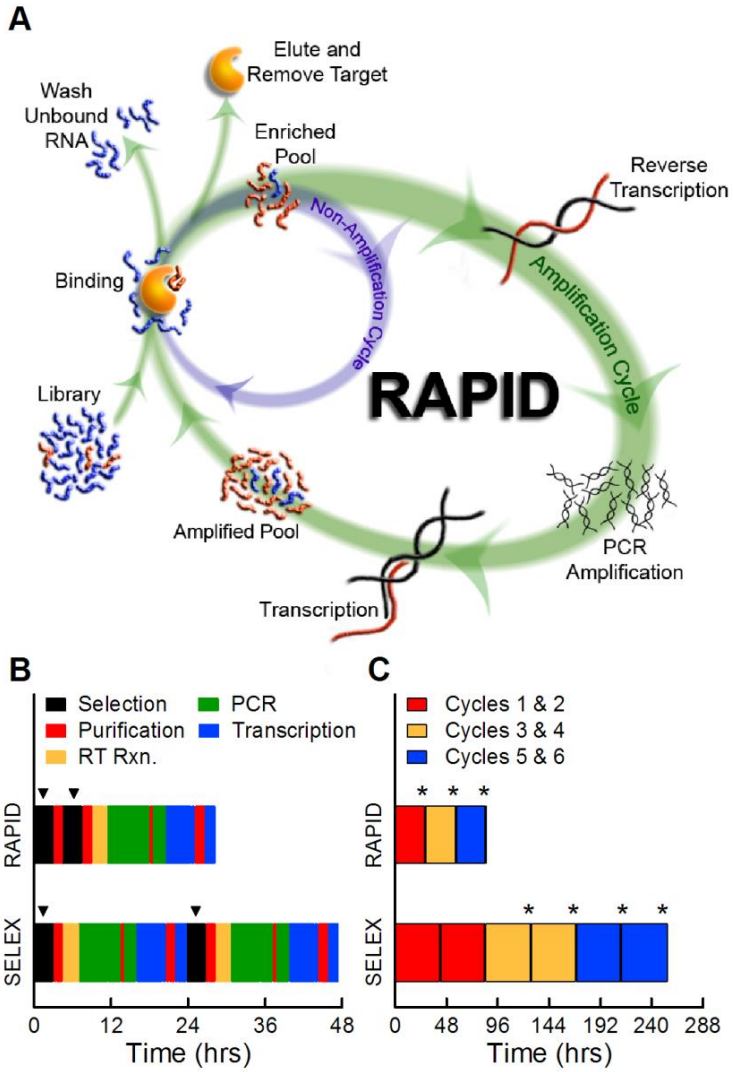


Figure 3.1. RNA Aptamer Isolation via Dual-cycles (RAPID). (A) Diagram of the RAPID process. The starting library or the enriched pool from the previous selection step can either go through the (inner) Non-Amplification Cycle and be used immediately in the next selection or go through the regular (outer) Amplification Cycle. (B) An example of processing times for SELEX and RAPID to complete two full selection cycles. Each selection is indicated with black blocks and arrowheads (▼) on top. (C) The total time required to complete six cycles of SELEX under optimal enrichment conditions, and six cycles of RAPID performed by alternating between Non-Amplification and Amplification Cycles; each colored block represents the total processing time between amplification steps. Asterisks (*) indicate the enriched and amplified pools that were analyzed via high-throughput sequencing.

for both processes, RAPID would have been completed in half the time needed for SELEX (Figure 3.1.B).

Ensemble binding of enriched aptamer pools

To monitor the progress of the selections, the recovery of bound RNA during each selection step was measured using quantitative PCR (qPCR). Figure 3.2.A shows the results for all six SELEX cycles to the Empty, UBLCP1 and CHK2 microcolumns. An increase in the fraction of bound RNA was observed from cycle to cycle for all three samples. The empty microcolumns generally bound an amount of RNA comparable to that bound to the microcolumns containing the two protein targets. This is because nearly all the recovered sequences in early selection cycles represent background and non-specific binding sequences. However, the two protein targets show higher recoveries than the Empty microcolumn, with the CHK2 target demonstrating the highest levels for the later cycles. Figure 3.2.B shows the results for all six cycles of RAPID to the same three targets. The recovery of the aptamer library with the RAPID method showed fluctuations from cycle to cycle that we believe are characteristic of the varying input concentrations since the total amount of material available following a Non-Amplification Cycle (1, 3, and 5) is lower compared to that following an Amplification cycle. This effect causes an increase in the recovery observed during the Amplification cycle. Despite these concentration induced fluctuations, CHK2 consistently showed the higher recovery of the two protein targets.

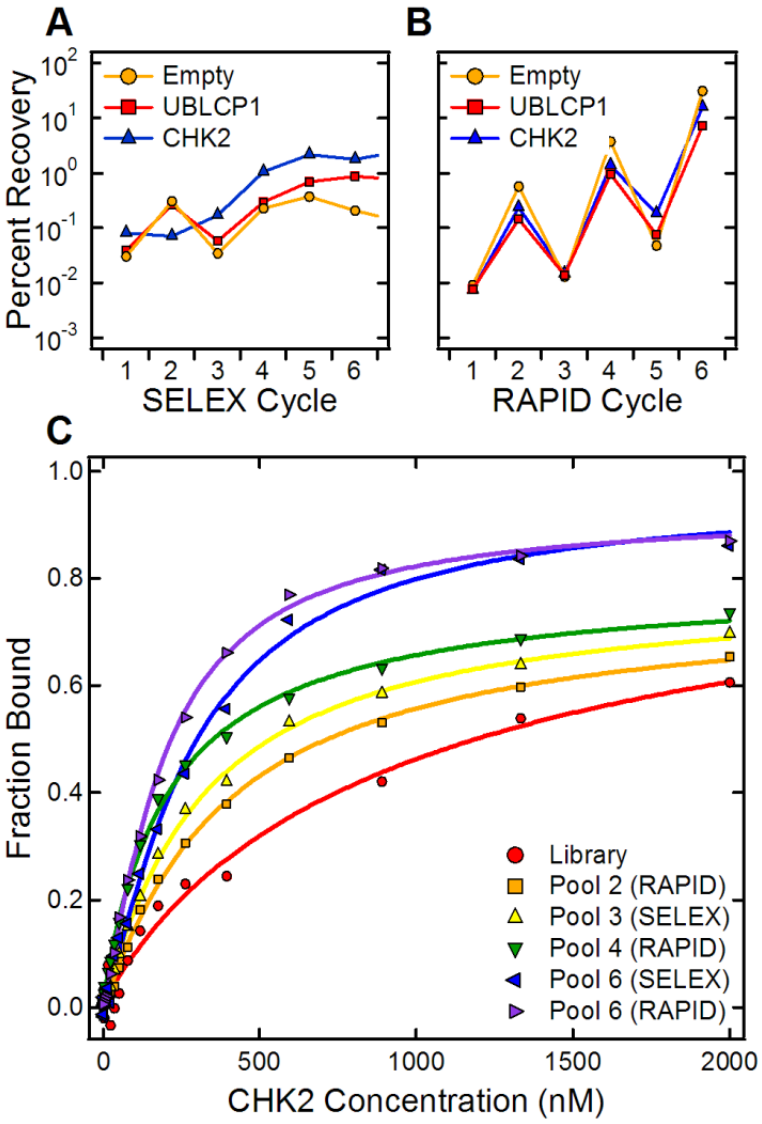


Figure 3.2. Binding of RNA after each selection cycle. (A) Percent RNA recovery for SELEX cycles for Empty (orange circles), UBLCP1 (red squares), and CHK2 (blue triangles) microcolumns. In this mode, there is a clear distinction between the protein-bound and the Empty microcolumns. (B) Percent RNA recovery for RAPID cycles for the same targets. In this mode, there are significant increases in the percent aptamer recoveries following selections with non-amplified pools at Cycles 2, 4, and 6, followed by a concentration induced drop with the amplified pools at Cycles 3 and 5. (C) Test of enriched pool binding to CHK2 protein preparation. F-EMSA shows the progression of bulk binding affinity increase for both SELEX and RAPID enriched pools with the RAPID Cycle 6 pool showing higher bulk binding than the SELEX Cycle 6 pool.

To evaluate improvements in target binding, Fluorescence Electrophoretic Mobility Shift Assays (F-EMSA) were performed with the initial random library and five enriched pools from the selection cycles for the CHK2 protein: RAPID cycle 2, SELEX cycle 3, RAPID cycle 4, SELEX cycle 6, and RAPID cycle 6. For each pool, the percent of input RNA that was bound at the highest protein concentration and the apparent ensemble dissociation constant, $K_{d\text{-}app}$, were calculated. The latter was determined by fitting the F-EMSA data to the Hill equation. The results shown in Figure 3.2.C indicate a general improvement in bulk affinity and an increased pool binding fraction at later cycles. The input library had a $K_{d\text{-}app}$ value greater than 1 μM, with 59% of input RNA bound. For SELEX, the Cycle 3 pool had a $K_{d\text{-}app}$ = 315±26 nM (69% bound) while the Cycle 6 pool had $K_{d\text{-}app}$ = 281±24 nM (86% bound). For RAPID, the Cycle 2, 4, and 6 pools had $K_{d\text{-}app}$ values of 390±34 nM (65% bound), 209±19 nM (72% bound), and 191±7 nM (87% bound), respectively. Across the cycles, the fraction of bound RNA increased monotonically from 59% for the starting library to 87% for the RAPID cycle 6 pool. In addition, the RAPID Cycle 6 pool showed a slightly higher bulk affinity for the protein than the SELEX Cycle 6 pool, which suggests that RAPID was enriching pools comparably to SELEX.

Population distributions from high-throughput sequencing analysis of selection pools

High-throughput sequencing was performed on selected pools to identify candidate aptamers and to compare the cycle-to-cycle enrichments of specific sequences from both the RAPID and conventional SELEX pools. As indicated in Figure 3.1.C, the four SELEX pools for Cycles 3, 4, 5, and 6 and all three of the

amplified RAPID pools were sequenced. Because the total number of sequencing reads for each pool varied between 5.6 and 9.4 million reads, the multiplicity of each sequence (number of times each sequence appeared) was normalized to 10^7 reads. We chose to analyze the sequences with the highest multiplicity (top 10,000) from each pool, because this was sufficient to cover 10-20% of the total sequence reads from the Cycle 6 pools. The top 10,000 sequences for each pool are plotted as a histogram to compare the population distributions for each of the RAPID and SELEX pools in Figure 3.3.A and 3.3.B, respectively. The histograms clearly show the convergence of the protein targets' sequences toward higher multiplicities at higher cycle numbers. As expected, there was minimal increase in multiplicity observed in the Empty columns which is consistent with the notion that RNA molecules bind randomly and non-specifically to the Empty column without enriching any specific RNA sequence. Overall, the two methods appear to be converging sequences at similar rates suggesting that RAPID's Non-Amplification cycles perform comparably to SELEX cycles. A quantitative comparison shows that the RAPID pools are actually more converged than the SELEX pools (Figure 3.4).

Multiplicity versus Cycle 4 to Cycle 6 enrichments

To further investigate and compare the evolving RNA pools obtained with RAPID and SELEX, the enrichments of individual sequences were calculated from the ratio of multiplicity values from two cycles [17]. The multiplicity values for the top 10,000 sequences in Cycle 6 were plotted versus their corresponding enrichment values from Cycle 4 for both selection methods (Figure 3.5). For both protein targets,

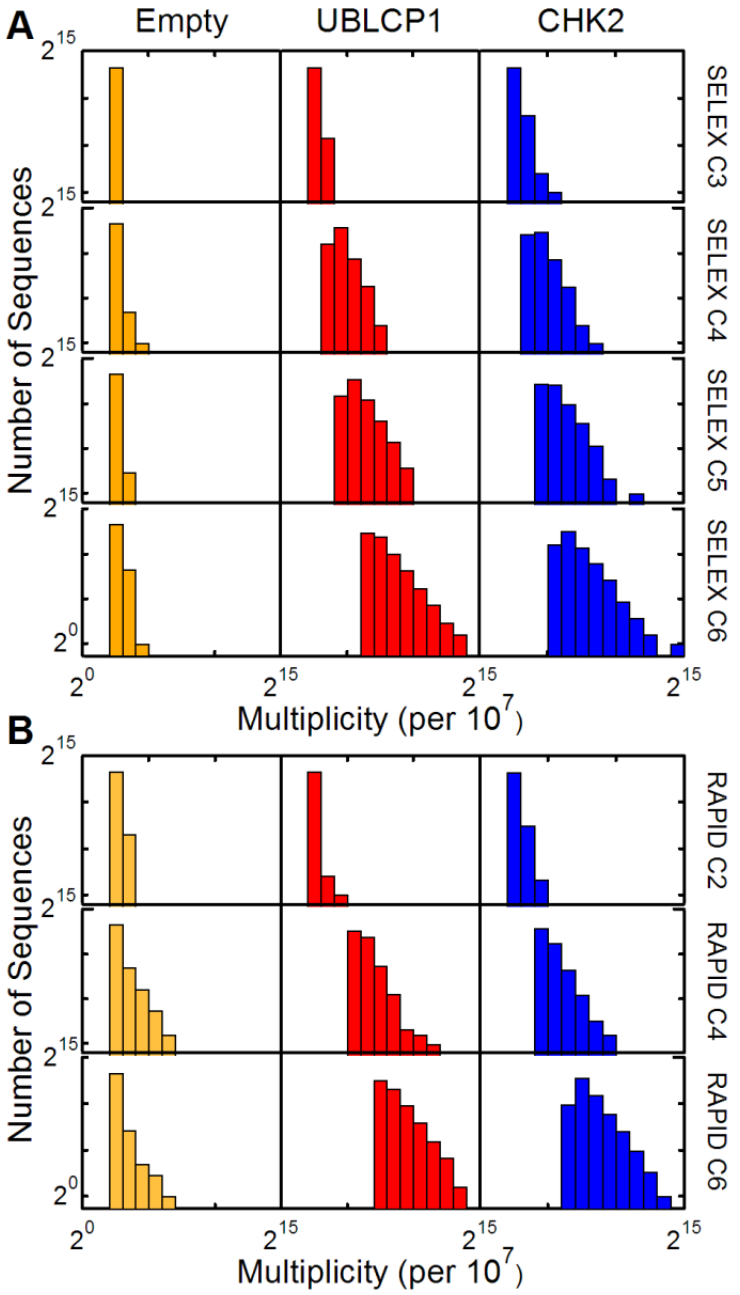


Figure 3.3. Sequence multiplicity distributions for various cycles of SELEX and RAPID. (A) Distributions of the top 10,000 Empty, UBLCP1 and CHK2 sequences for SELEX Cycles 3 to 6. (B) The same Sequence multiplicity distributions of RAPID Cycles 2, 4 and 6 for the same targets.

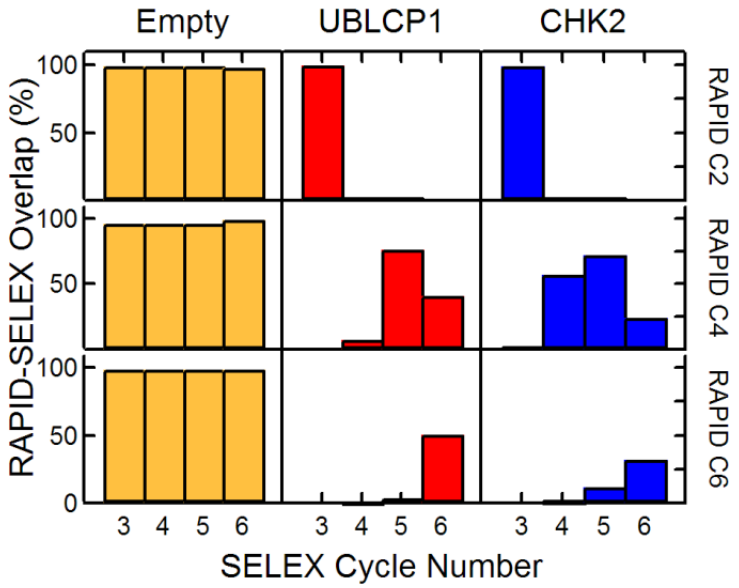


Figure 3.4. The similarity between RAPID and SELEX pool distributions. For each target, similarity between pools is determined by calculating the percent overlap of each RAPID cycle's distribution with each SELEX cycle's. The highest valued SELEX cycle against a given RAPID cycle is considered to be most similar to the given RAPID cycle. For both protein targets, the RAPID pools Cycle 2 and 4 distributions are most similar to the "later" SELEX Cycle 3 and 5 distributions, respectively. For the Empty columns, the overlap values are close to 100% between all of the pools confirming that there was negligible sequence convergence beyond the initial library within the Empty column's pools.

these two metrics were well correlated. However, the RAPID pools (Figure 3.5.D and 3.5.F) have higher multiplicities at equivalent enrichments than the SELEX pools (Figure 3.5.C and 3.5.E), and more of the top enriched sequences were identified in Cycle 4 of RAPID. In the RAPID pools, UBLCP1 and CHK2 had 6,565 and 5,063 sequences, respectively, in common between the Cycle 4 and 6 pools' top 10,000 sequences. For comparison, in the SELEX pools, UBLCP1 and CHK2 had 3,281 and 3,262 sequences, respectively, ranking in the top 10,000 of both pools. Thus, the RAPID pools have almost twice as many preserved sequences between cycles over SELEX, which is consistent with the improved convergence and enrichment data. In contrast, Figures 3.5.A and 3.5.B show that the Empty column had very few sequences

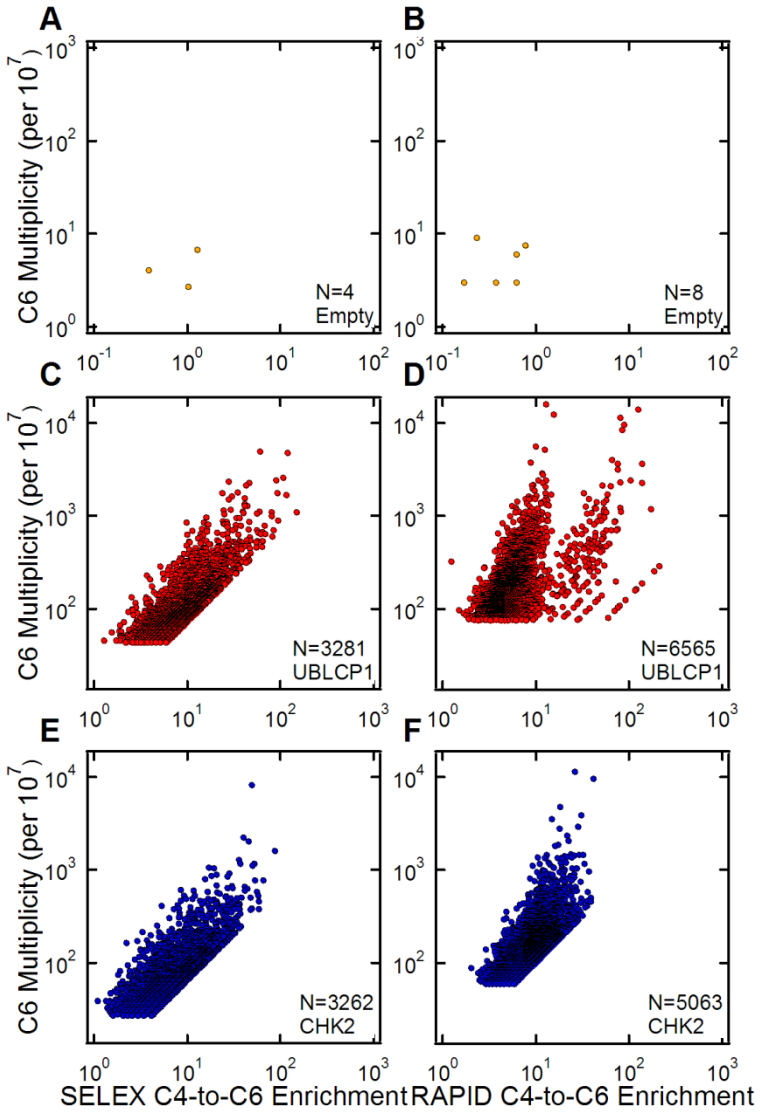


Figure 3.5. The relationship between sequence multiplicity and enrichment. (A and B) Scatter plots of sequences' multiplicity and enrichment within the top 10,000 highest multiplicity sequences from Cycle 6 of SELEX and RAPID for the Empty microcolumns. Multiplicity values have been normalized to counts per 10^7 and enrichment is calculated as the ratio of Cycle 6 multiplicities to Cycle 4 multiplicities for any sequence found in both pools. Some data points are obscured due to overlapping values. (C and D) Scatter plots of sequences' multiplicity and Cycle 4-to-Cycle 6 enrichment within the top 10,000 highest multiplicity sequences from Cycle 6 of UBLCP1 SELEX and RAPID. (E and F) Scatter plots of sequences' multiplicity and enrichment within the top 10,000 highest multiplicity sequences from Cycle 6 of CHK2 SELEX and RAPID. RAPID sequences show significantly higher multiplicities at lower enrichments than SELEX.

in both pools with only 4 in SELEX and 8 in RAPID. In addition, the majority of the Empty-column sequences had enrichment values less than one between the two cycles, which is expected if the binding and copy number for those sequences is random.

Independent RAPID and SELEX enrich identical sequences.

A closer examination of the sequencing results for the two Cycle 6 pools of each protein revealed identical sequences that had achieved very high multiplicities in both RAPID and SELEX. Among the top five candidates, UBLCP1's highest-ranked sequence in RAPID was ranked fifth in SELEX and its top-ranked sequence in SELEX was ranked third in RAPID (Figure 3.6.A). Furthermore, the top-ranked CHK2 sequence in RAPID was also the top ranked sequence in SELEX (Figure 3.6.B). This analysis was done using the entire random region of each candidate (i.e. not a short sequence motif), so each sequence represented the identical sequence that was selected from the 5×10^{15} random sequence library using RAPID and SELEX.

To extend this analysis, we searched for additional sequences common to each target's RAPID and SELEX Cycle 6 pools and found that many sequences among their top 10,000 were common and highly represented in both methods. Scatter plots relating the multiplicities of sequences represented in both pools are shown in Figures 3.6.C and 3.6.D. In total, we found 687 sequences that were common in both UBLCP1 pools, and 1317 sequences that were common in both CHK2 pools. Analysis for the Empty column yielded only a single common sequence with negligible multiplicities. It is difficult to prove that identical sequences identified in multiple selections are not the result of cross-contamination between simultaneous side-by-side selections;

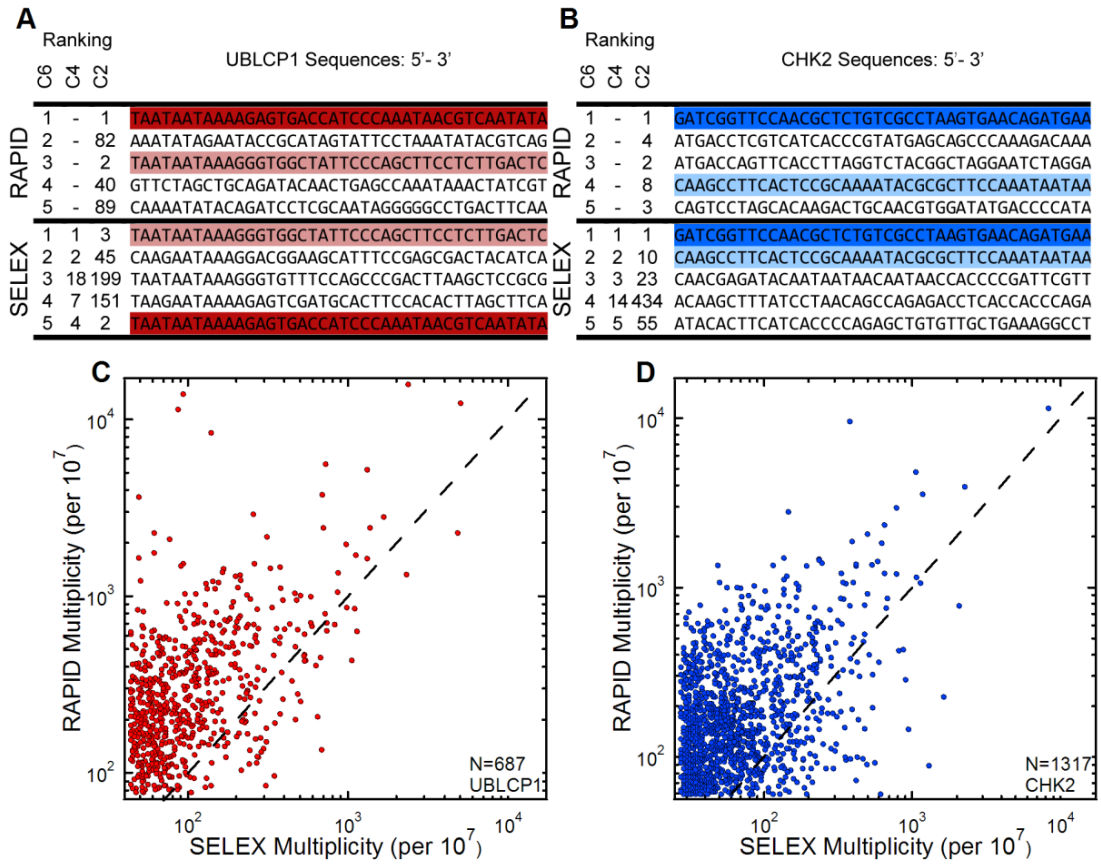


Figure 3.6. Relationship of the SELEX and RAPID selected sequences in Cycle 6 pools. (A and B) The first 40 random bases of the top 5 UBLCP1 and CHK2 sequences from Cycle 6 in RAPID (top) and SELEX (bottom). Identical sequences between both methods are highlighted with matching colors. The ranks of each sequence at earlier cycles (4, 5 and 6) are also shown. (C) A scatter plot of the 687 common sequences for UBLCP1 in SELEX and RAPID Cycle 6 pools; the dashed line represents a 1:1 correlation between multiplicities in the two pools. (D) The same analysis for CHK2 yielded 1317 common sequences. On average, RAPID pools were enriched above SELEX pools.

however, RAPID and SELEX were performed independently of each other at different times making contamination between methods unlikely. In addition, almost all of the common sequences were unique to each target (Figure 3.7) and most appeared more highly enriched in the RAPID Cycle 6 pools. On average, the RAPID selected sequences represented higher fractions of their pools having enriched approximately 3-fold more than from SELEX: UBLCP1 by a factor of 2.6 ⁎ 2.3 (1.1 – 6.0-fold) and

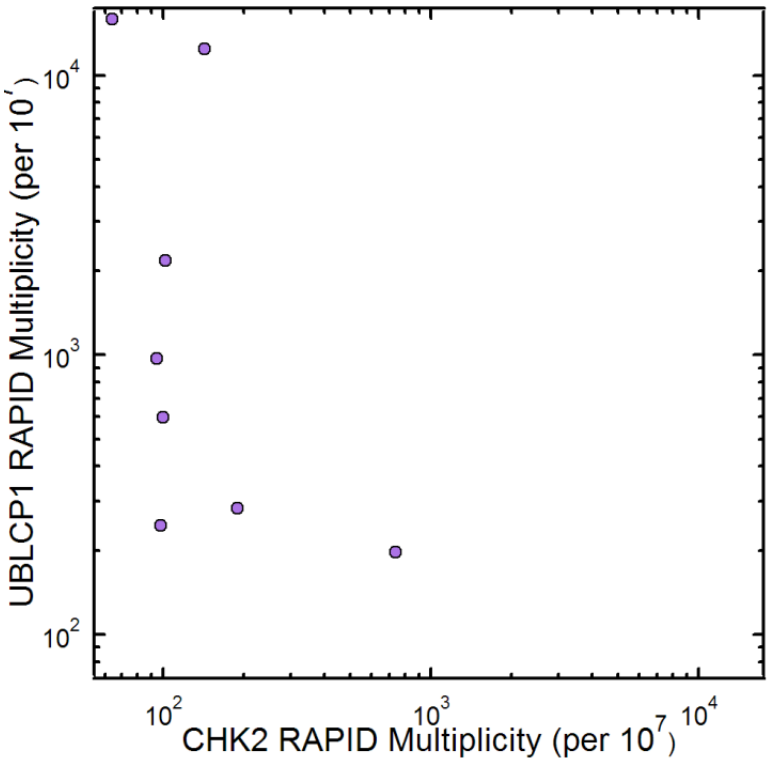


Figure 3.7. Sequences that are common to both UBLCP1 and CHK2 selected RAPID Cycle 6 pools. Of the 2004 sequences of interest (687 and 1317 sequences common between Cycle 6 of RAPID and SELEX pools for UBLCP1 and CHK2, respectively), only 8 of them were also common between the two target pools. This is likely due to a trace cross-contamination and strongly suggests that the unique sequences in each pool are target specific.

CHK2 by a factor of 2.8 ⁎ 2.2 (1.3 – 6.2-fold). These were determined by finding the geometric mean and standard deviation for the enrichments, thus the enrichments and their standard deviations are expressed as multiplicative factors.

Aptamer binding to CHK2 protein

The sequence for CHK2 identified as the top-ranked one in both selection methods, hereafter referred to as C6M1, was tested for its binding affinity to CHK2. After C6M1 was isolated from the Cycle 6 pools, it was labeled with fluorescein, and then evaluated via the Fluorescent Electrophoretic Mobility Shift Assay (F-

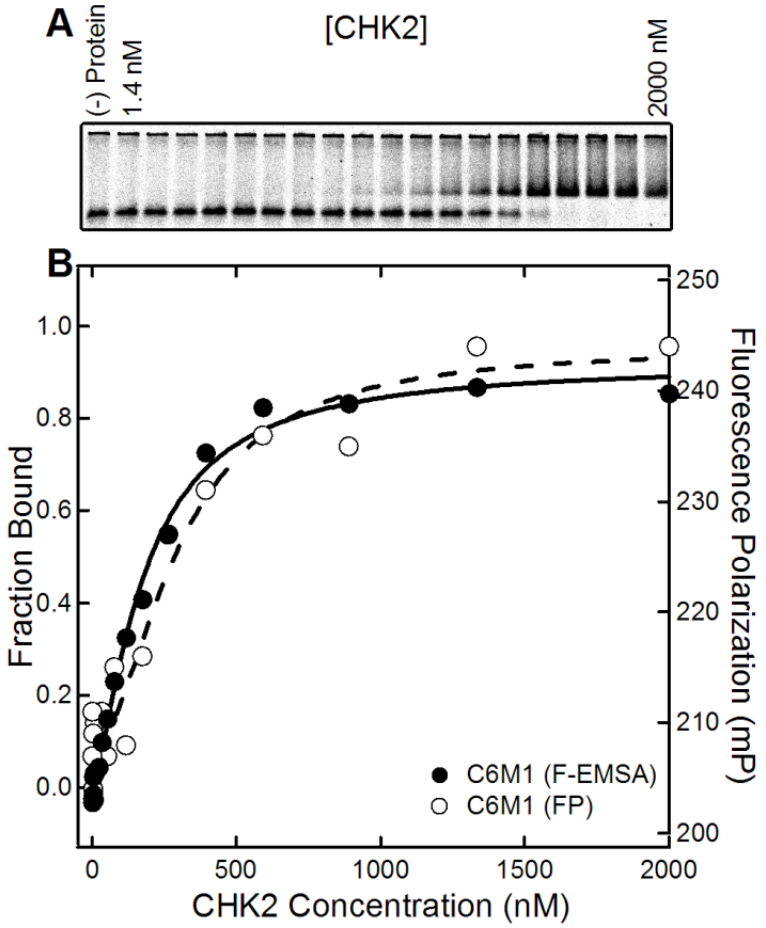


Figure 3.8. Binding test of the CHK2 protein prep's highest multiplicity Cycle 6 aptamer candidate C6M1. The sequence is given by the two flanking constant regions, and the random region: GATCGGTTCCAACGCTCTGTCGCCTAAGTGAACAGATGAAGAAAAAATAGCCCAA TAAGAGGCAACAATCT . (A) Gel image of F-EMSA for C6M1 aptamer incubated with no protein or the CHK2 protein prep ranging from 1.4 nM to 2000 nM, in 1.5-fold increments. (B) Binding curves for C6M1 using F-EMSA and FP. The left axis shows the calculated fraction bound from F-EMSA (solid line, black circles), while the right axis shows the fluorescence polarization from C6M1 (dotted line, white circles). The fitted K_d for the two curves are 180±13 nM and 299±53 nM, respectively

EMSA). Figure 3.8.A shows an image of the resulting gel shift assay. The fraction of bound RNA was evaluated from the gel image and plotted as the filled symbols in Figure 3.8.B. The solid line fit to the data was done using the Hill equation which yielded a K_d value of 180 ± 13 nM. In order to ensure that the observed binding was

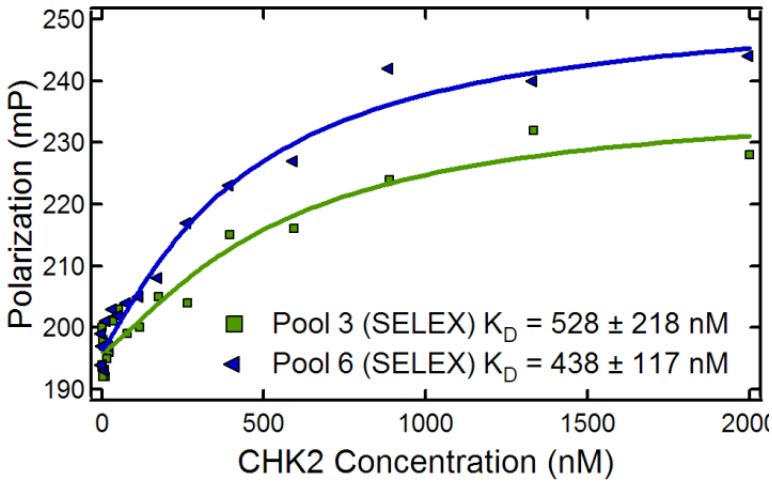


Figure 3.9. Fluorescent polarization binding assays of bulk SELEX pools to CHK2. The fitted Kd's for the Cycle 3 and Cycle 6 pools are higher than F-EMSA (Fig. 2). All of the tested pools and C6M1 have calculated dissociation constants 1.6-fold higher when measured from fluorescence polarization compared to F-EMSA.

not a gel artifact, a Fluorescence Polarization (FP) assay was also performed. The polarization results and curve fit are shown as the open symbols and dashed line in Figure 3.8.B. The calculated K_d is 299 ± 53 nM, which is 1.6-fold higher than determined with F-EMSA. This factor is consistent with other FP assays performed on some of the labeled bulk SELEX pools (Figure 3.9). Currently, we have not ruled out potential aptamer binding to a contaminant in our protein preparation. If this were the case, given the purity of our preparations, the binding affinity of C6M1 would be underestimated by at least an order of magnitude and thus the approximate K_d value would be less than 20 nM. However, for the purposes of this manuscript, the results and conclusions of this work remain the same in either case.

DISCUSSION

RAPID SELEX is capable of isolating aptamers in less time than conventional SELEX. Standard binding assays with the amplified pools clearly revealed cycle-to-cycle affinity enrichment for two protein targets, CHK2 and UBLCP1, using both RAPID and conventional SELEX. Further, higher affinities and total binding to CHK2 were observed for pools from later selection cycles. We found that the two Cycle 6 pools bound with comparable affinity, although the RAPID pool bound slightly better (~1.5-fold higher). Even though the RAPID selections were not performed with the optimal conditions used in SELEX, this suggests that the Non-Amplified RAPID pools did not suffer in performance compared to the SELEX pools, which would support the use of RAPID in many if not most selection strategies.

As with the binding affinities, we found that despite having half the amplification steps as SELEX, the RAPID pools had slightly more converged sequence distributions. This is in good agreement with the ordered binding curves mentioned above, which suggested that the RAPID pools should have slightly more converged distributions. This is in fact what we observed (Figure 3.3 and 3.5), and recalling our definition of a selection "round" that necessarily includes amplification steps, we found that one RAPID round was most similar to three SELEX rounds in terms of convergence (Figure 3.4). Similarly, two RAPID rounds yielded convergence similar to five SELEX rounds. This is particularly noteworthy since we found that our top candidate aptamers had acquired their high rankings after just two rounds of RAPID (four cycles).

Finally, we found that among the top 10,000 ranked Cycle 6 sequences from both selections, a large percentage (7% and 13%) were identical. This kind of reproducibility from different SELEX experiments has been addressed before; however, in this past study, sequencing was done at much less depth (less than 100 clones) and the identified aptamers generally contained short motifs which were determined to be highly represented in starting pools [38]. We found no sequence motifs in any of our pools and therefore restricted our analysis to the entire sequence of the ~70 nt random region. Independent enrichment of the identical rare sequences (~1 in 10^{15}) in both selection methods demonstrate the effectiveness and the robustness of our selection protocols. However, in further support of RAPID, we found that among those identical sequences, the great majority were more enriched an average of ~ 3-fold, in the RAPID Cycle 6 pools over the SELEX pools. As mentioned previously, the top aptamer candidates were actually resolved by Cycle 4 in both selections. This reflects the power of high-throughput sequencing for identifying enriching aptamers with great sensitivity many cycles before complete convergence. From these data, we chose to isolate our best candidate aptamer for CHK2, C6M1, and showed that the raw aptamer was indeed able to bind to its target. Further development and characterization of CHK2 and UBLCP1 specific aptamers is beyond the scope of this work and therefore not fully investigated. However, RAPID was able to generate the same results as SELEX in only one third the time.

In addition to specific protein binding results, we studied the impact that the empty microcolumns and downstream processing had on the selections. Interestingly, we noticed that the empty microcolumns generally bound a comparable amount of

RNA as the two protein targets (Figure 3.2). This is not surprising because aptamers are assumed to be rare in the starting library; nearly all the recovered sequences in an initial selection represent background and non-specific binding sequences. Despite this, there was negligible sequence convergence from cycle to cycle (Figure 3.3). The collective set of high-throughput sequencing results for the Empty microcolumns also suggest that there was negligible sequence bias in the starting library [17] as well as negligible contributions from the microcolumns and the enzymatic processes (PCR, transcription, etc.) to the overall sequence enrichment in the two protein target pools [24].

While we demonstrated RAPID using the simple pairing of one Non-Amplification Cycle followed by one Amplification Cycle, the efficiency of RAPID may be further improved. In general, more Non-Amplification Cycles can be performed between Amplification Cycles, though the number will be limited by practical considerations. Non-Amplification Cycles have the potential to significantly increase the efficiency of selections through the rapid accumulation of affinity enrichments in a short period of time. However, despite higher binding efficiencies, this process also depletes the population of high affinity sequences. Assuming (or requiring) a minimum binding probability, P_A, for a population of aptamers, the number of Non-Amplification cycles can be increased as long as an acceptable copy number of high affinity aptamers, N_{min}, is estimated to always be present before each cycle (N_{min} should be chosen such that $N_{min} \geq (P_A)^{-1}$ so that at least one copy of an aptamer is expected to remain after the last cycle). This can be expressed as:

$$N_{\min} \leq N_A \times (P_A)^{i-1} \qquad (3)$$

where N_A is the initial (or amplified pool's) copy number of the aptamer population and *i-1* is the maximum number of Non-Amplification Cycles, with the i^{th} cycle being an Amplification Cycle which must be done to replenish the pool's sequence populations. In addition, each Non-Amplification Cycle decreases the input material for the subsequent cycle which may result in increased binding fractions and reduced enrichment yields, diminishing the practicality of continued Non-Amplification Cycles. Using a simple measurement of total binding, the number of Non-Amplification Cycles can be increased as long as an acceptable enrichment, $E_{min,}$ of high affinity aptamers is estimated to have resulted after each cycle. This can be expressed as:

$$E_{\min} \leq \frac{P_A}{P_B(n,i)} \qquad (4)$$

where $P_B(n,i)$ is the background binding probability at the n^{th} total selection cycle with *i* cycles since the last amplification. If this expression ever proves false, amplification of the pool can be used to increase the concentration and selection stringency to improve future enrichments.

Together, the above two expressions place upper limits on the total number of Non-Amplification cycles that can be performed between Amplification Cycles, and maximizes the potential efficiency of RNA selections. Applying these expressions to our simple RAPID protocol required a minimum binding probability for aptamer candidates of about 40% (to ensure 1 copy survives the first round) which is typical of binding efficiencies demonstrated on our microcolumns [30]. Taking into account the amount of amplification and the measured background binding over the six cycles, our highest candidates should represent between 1 in 100 -1000 sequences. In fact our top

candidates are represented in the middle of this range. Altogether, our results make a compelling case for RAPID both in its efficiency, and its cycle-to-cycle performance.

Although we used our microcolumn-based processes to perform all selections, RAPID may be used in combination with any selection mode or technology to save time, reagents, and to rapidly converge selected pools. RAPID could be particularly useful for slow selections requiring many cycles, or when complete sequence convergence is needed so that conventional cloning methods can be used to identify candidates. Although the time-saving benefits would be less compared to RNA-based selections, RAPID can also be extended to DNA selections. We used high-throughput sequencing to quantify selected pools as described by histograms of converging multiplicities, and scatter plots of sequence enrichments and identical sequences derived from two independent selection methods. Similar detailed analyses could be used to gain higher confidence in aptamer candidates through replicate selections, or to make more quantitative evaluations of different selection schemes and technologies. In particular, with a standardized pool and target, these analyses could be used to objectively rank, compare, and optimize different selection techniques.

REFERENCES

1. Ellington AD, Szostak JW (1990) Invitro Selection of Rna Molecules That Bind Specific Ligands. Nature 346: 818-822.

2. Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. Gene 82: 83-87.

3. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.

4. Ciesiolka J, Gorski J, Yarus M (1995) Selection of an Rna Domain That Binds Zn2+. Rna-a Publication of the Rna Society 1: 538-550.

5. Nitsche A, Kurth A, Dunkhorst A, Panke O, Sielaff H, et al. (2007) One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. Bmc Biotechnology 7.

6. Paige JS, Wu KY, Jaffrey SR (2011) RNA Mimics of Green Fluorescent Protein. Science 333: 642-646.

7. Geiger A, Burgstaller P, vonderEltz H, Roeder A, Famulok M (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. Nucleic Acids Research 24: 1029-1036.

8. Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-Resolution Molecular Discrimination by Rna. Science 263: 1425-1429.

9. Gong Q, Wang JP, Ahmad KM, Csordas AT, Zhou JH, et al. (2012) Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. Anal Chem 84: 5365-5371.

10. Klussmann S, Nolte A, Bald R, Erdmann VA, Furste JP (1996) Mirror-image RNA that binds D-adenosine. Nature Biotechnology 14: 1112-1115.

11. Latham JA, Johnson R, Toole JJ (1994) The Application of a Modified Nucleotide in Aptamer Selection - Novel Thrombin Aptamers Containing 5-(1-Pentynyl)-2'-Deoxyuridine. Nucleic Acids Research 22: 2817-2822.

12. Jensen KB, Atkinson BL, Willis MC, Koch TH, Gold L (1995) Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. Proc Natl Acad Sci U S A 92: 12220-12224.

13. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.

14. Burmeister PE, Lewis SD, Silva RF, Preiss JR, Horwitz LR, et al. (2005) Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. Chem Biol 12: 25-33.

15. Green LS, Jellinek D, Bell C, Beebe LA, Feistner BD, et al. (1995) Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor. Chem Biol 2: 683-695.

16. Ruckman J, Green LS, Beeson J, Waugh S, Gillette WL, et al. (1998) 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. J Biol Chem 273: 20556-20567.

17. Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, et al. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. Proc Natl Acad Sci U S A 107: 15373-15378.

18. Daniels DA, Chen H, Hicke BJ, Swiderek KM, Gold L (2003) A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment. Proc Natl Acad Sci U S A 100: 15416-15421.

19. Park SM, Ahn JY, Jo M, Lee DK, Lis JT, et al. (2009) Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. Lab Chip 9: 1206-1212.

20. Peng L, Stephens BJ, Bonin K, Cubicciotti R, Guthold M (2007) A combined atomic force/fluorescence microscopy technique to select aptamers in a single cycle from a small pool of random oligonucleotides. Microscopy Research and Technique 70: 372-381.

21. Mendonsa SD, Bowser MT (2004) In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis. Anal Chem 76: 5387-5392.

22. Raddatz MSL, Dolf A, Endl E, Knolle P, Famulok M, et al. (2008) Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. Angewandte Chemie-International Edition 47: 5190-5193.

23. Schutze T, Wilhelm B, Greiner N, Braun H, Peter F, et al. (2011) Probing the SELEX process with next-generation sequencing. PLoS One 6: e29604.

24. Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. PLoS One 5: e9169.

25. Ditzler MA, Lange MJ, Bose D, Bottoms CA, Virkler KF, et al. (2013) High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. Nucleic Acids Research 41: 1873-1884.

26. Thiel WH, Bair T, Peek AS, Liu X, Dassie J, et al. (2012) Rapid identification of cell-specific, internalizing RNA aptamers with bioinformatics analyses of a cell-based aptamer selection. PLoS One 7: e43836.

27. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, et al. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20: 861-873.

28. Cox JC, Rudolph P, Ellington AD (1998) Automated RNA selection. Biotechnol Prog 14: 845-850.

29. Cox JC, Ellington AD (2001) Automated selection of anti-protein aptamers. Bioorg Med Chem 9: 2525-2531.

30. Latulippe DR, Szeto K, Ozer A, Duarte FM, Kelly CV, et al. (2013) Multiplexed microcolumn-based process for efficient selection of RNA aptamers. Anal Chem.

31. Ozer A, White BS, Lis JT, Shalloway D (2013) Density-dependent cooperative non-specific binding in solid-phase SELEX affinity selection. Nucleic Acids Research.

32. Berezovski M, Musheev M, Drabovich A, Krylov SN (2006) Non-SELEX selection of aptamers. J Am Chem Soc 128: 1410-1411.

33. Ashley J, Ji KL, Li SFY (2012) Selection of bovine catalase aptamers using non-SELEX. Electrophoresis 33: 2783-2789.

34. Tok J, Lai J, Leung T, Li SFY (2010) Selection of aptamers for signal transduction proteins by capillary electrophoresis. Electrophoresis 31: 2055-2062.

35. Gevertz J, Gan HH, Schilick T (2005) In vitro RNA random pools are not structurally diverse: A computational analysis. Rna-a Publication of the Rna Society 11: 853-863.

36. Pagano JM, Farley BM, McCoig LM, Ryder SP (2007) Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. J Biol Chem 282: 8883-8894.

37. Thiel WH, Bair T, Thiel KW, Dassie JP, Rockey WM, et al. (2011) Nucleotide Bias Observed with a Short SELEX RNA Aptamer Library. Nucleic Acid Therapeutics 21: 253-263.

38. Burke DH, Gold L (1997) RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. Nucleic Acids Research 25: 2020-2024.

CHAPTER 4

HIGH-THROUGHPUT BINDING CHARACTERIZATION OF RNA APTAMER SELECTIONS USING A MICROPLATE-BASED MULTIPLEX MICROCOLUMN DEVICE‡

ABSTRACT

We describe a versatile 96-well microplate-based device that utilizes affinity microcolumn chromatography to complement downstream plate-based processing in aptamer selections. This device is reconfigurable and is able to operate in serial and/or parallel mode with up to 96 microcolumns. We demonstrate the utility of this device by simultaneously performing characterizations of target binding using five RNA aptamers and a random library. This was accomplished through 96 total selection tests. Three sets of selections tested the effects of target concentration on aptamer binding compared to the random RNA library using aptamers to the proteins green fluorescent protein (GFP), human heat shock factor 1 (hHSF1, and negative elongation facto E (NELF-E). For all three targets, we found significant effects consistent with steric hindrance with optimum enrichments at predictable target concentrations. In a fourth selection set, we tested the partitioning efficiency and binding specificity of our three proteins' aptamers, as well as two suspected background binding sequences, to eight targets running serially. The targets included an empty

‡The following sections are reprinted with permission from: Szeto, K., Reinholt, S.J, Duarte, F.M., Pagano, J.M., Ozer, A., Yao, L., Lis, J.T., and Craighead, H.G. 2014. High-throughput Binding Characterization of RNA Aptamer Selections using a Microplate-based Multiplex Microcolumn Device. Analytical and Bioanalystical Chemistry. 406(11) pp2727-2732, doi: 10.1007/s00216-014-7661-7, with modifications to conform to the required format.

microcolumn, three affinity resins, three specific proteins, and a non-specific protein control. The aptamers showed significant enrichments only on their intended targets. Specifically, the hHSF1 and NELF-E aptamers enriched over 200-fold on their protein targets, and the GFP aptamer enriched 750-fold. By utilizing our device's plate-based format with other complementary plate-based systems for all downstream biochemical processes and analysis, high-throughput selections, characterizations, and optimization were performed to significantly reduce the time and cost for completing large-scale aptamer selections.

INTRODUCTION

Systematic Evolution of Ligands by EXponential enrichment (SELEX) is an *in vitro* selection method used to generate high affinity ligands for specific target compounds [1-3]. These selected molecules, called aptamers, are derived from large libraries of nucleic acids with random sequences through an iterative process of binding, partitioning, and amplification of sequences that bind to the target. This process enriches the initial random library for higher binding affinity sequences, and the cycle is repeated until the molecules in the enriched pools converge on the highest affinity sequence. Since this method was first introduced, aptamers have become valuable tools in biotechnology, diagnostics, and therapeutics [4].

There is interest in improving SELEX technology to obtain highly specific aptamers much more rapidly. However, despite their potential, many technologies are difficult to scale for multiplexed or parallel selections. For example, Park et al. and Ahn et al. used microfluidic sol-gel devices that could utilize up to five targets for

multiplexing [5,6], but currently, no large-scale microfluidic selections have been demonstrated. Large-scale parallel selections have been done with microplate technologies, which are of particular interest due to the availability of protocols and automated liquid handling devices [7,8]. However, in contrast to microfluidic devices that utilize flow and other dynamic behavior, most of these selections rely on traditional equilibrium solution binding [9] or interactions with target molecules that are bound or adsorbed to the plate surface [10,11].

Despite advances toward more sophisticated and automated SELEX, little has been done to characterize and optimize new or current technologies and recent binding studies show significant discrepancies with existing theory [12,13]. Therefore, empirical methods have been used to optimize selection conditions and aid the development of new models [14]. As new high-throughput technologies emerge, these studies will become even more important in order to obtain the most effective and robust selections under the available parameters.

To address these issues, we have developed a high-throughput device called Microplate-based Enrichment Device Used for the Selection of Aptamers (MEDUSA). This device is designed around a 96-well microplate format, which not only allows for high-throughput selections, but also complements existing plate-based methods and technologies for sample handling and has the potential for automation. MEDUSA is a substantial expansion of our previously reported modular and multiplexable microcolumns, which achieve non-equilibrium selections by utilizing dynamic flow rates shown to optimize the enrichment of aptamers [13]. We demonstrate the use of MEDUSA by performing 96 simultaneous selection tests to characterize the binding of

a number of RNA aptamers against various targets. In total, the characterization tests performed on MEDUSA shed light on the critical binding behaviors of specific and background binding aptamers that fundamentally limit the performance and sensitivity of solid-phase affinity selections.

MATERIAL AND METHODS

Preparation of recombinant protein targets

Recombinant proteins were expressed in BL21(DE3)-RIPL *E. coli* cells (Agilent Technologies) transformed with plasmids that encode for hexahistidine-tagged GFP, Drosophila NELF-E, and UBLCP1, or GST-tagged hHSF1 (Table 4.1). Two or four liter LB cultures supplemented with 100 μg/mL ampicillin were inoculated with starter LB culture derived from a single colony and grown at 37°C until the OD600 reached approximately 0.6. Protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After an additional incubation, bacteria were collected by centrifugation and the resulting pellet was processed according to the manufacturer's instructions for Ni-NTA Superflow (Qiagen) or Glutathione-agarose (Thermo Scientific) resins. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used to verify the purity and quality of the final protein product. Resulting protein preps were dialyzed against 1× PBS (supplemented with 5 mM 2-mercaptoethanol and 0.01% Triton X-100) and stored in small aliquots after addition of glycerol to a final concentration of 20%. NELF-E was prepared slightly differently [15].

Table 4.1. Properties of the target proteins

Protein	Molecular Weight (kDa)	Isoelectric Point	Affinity tag
GFP	27	5.5	Hexahistidine (N-terminus)
hHSF1	86	5.3	GST* (N-terminus)
NELF-E	36	8.9	Hexahistidine (N-terminus)
UBLCP1	37	6.1	Hexahistidine (N-terminus)

*GST tag ~ 30 kDa

Protein immobilization on affinity resins

Nickel-nitrilotriacetic acid (Ni-NTA) Superflow or glutathione-agarose (GSH) resins were extensively washed with binding buffer [10-mM N-2-hydroxyethylpiperazine-N′-ethanesulfonic acid −KOH pH 7.6, 125-mM NaCl, 25-mM KCl, 5-mM $MgCl_2$, and 0.02% Tween-20]. Hexahistidine- or GST-tagged proteins (see Table 4.1) were immobilized at the desired concentrations onto the washed resin in 10% slurry with binding buffer and incubated at 4°C with constant mixing for 1 h.

RNA library and aptamers

The random RNA library used in the experiments, hereafter referred to as N70 library, contains $\sim 5\times10^{15}$ sequences of 120-nucleotide (nt) RNA molecules and was prepared as described previously [13]. This library consists of a 70-nt random region flanked by two constant regions. Green fluorescent protein (GFP)-, human heat shock factor 1 (hHSF1)- and negative elongation factor E (NELF-E)-binding aptamers, GFPapt, HSFapt, and NELFapt, as well as the background binding sequences (BBSs), BBS1 and BBS2, were all derived from previous multiplex SELEX experiments [13,15-17]. HSFapt was previously identified as hHSF2-R5-2 using the N70 library and characterized elsewhere [13]. NELFapt was previously identified as Napt1 using

the N70 library [15]. The background binding sequences BBS1 and BBS2 were identified in several previous multiplex SELEX experiments using the N70 library for dozens of target proteins [13,15,16]. The GFP-binding RNA aptamer, GFPapt, used in this work was selected using a different library with a smaller random region and different constant regions; and was previously identified as AP3-1 [17].

The 84-nt GFP-binding RNA aptamer has the following sequence: 5'-AGCUUCUGGACUGCGAUGGGAGCACGAAACGUCGUGGCGCAAUUGGGUGGGGAAAGUCCUUAAAAGAGGGCCACCACAGAAGCU-3'. The forward and reverse oligos used for qPCR analyses were GFPapt-FOR: 5'-GCTTCTGGACTGCGATGGGAGCA-3' and GFPapt-REV: 5'-GCTTCTGTGGTGGCCCTCTTTTAAGGACT-3'.

The 117-nt hHSF1-binding RNA aptamer has the following sequence: 5'-GGGAAUGGAUCCACAUCUACGAAUUCAAUCAAGUCCCCAGACUCAGCAACACUGGACAGCGAUAUGCAGAUAACCAAGACCAAUUCACUCCAGUUCACUGCAGACUUGACGAAGCUU-3'. The two constant regions corresponding to the library design are denoted by underlines. The forward and reverse oligos used for qPCR analyses were HSFapt-FOR: 5'-AATCAAGTCCCCAGACTCAGCAACA-3' and HSFapt-REV: 5'-CTGGAGTGAATTGGTCTTGGTTATC-3'.

The 120-nt NELF-E-binding RNA aptamer has the following sequence: 5'-GGGAAUGGAUCCACAUCUACGAAUUCCCAACGACUGCCGAGCGAGAUUACGCUUGAGCGCCCCACUGAGGAUGCCCACGGGCGAUUGGGGCACGGCUUCACUGCAGACUUGACGAAGCUU-3' The two constant regions corresponding to the library design are denoted by underlines. The forward and reverse oligos used for

qPCR analyses were NELFapt-FOR: 5'-CCAACGACTGCCGAGCGAGATTAC-3' and NELFapt-REV: 5'-GCCGTGCCCCAATCGCCCGTG-3'.

BBS1 has the following sequence: 5'-GGGAAUGGAUCCACAUCUACGAAUUCCGCAGGGCUAGCCGCAUGCUCAGGCCUGGCGGGUAGGGAGUUAGGGUAGGGAGACCAGGAGAGCUGGCUUCACUGCAGACUUGACGAAGCUU-3'. The forward and reverse oligos used for qPCR analyses were BBS1-FOR: 5'-CGCAGGGCTAGCCGCATG-3' and BBS1-REV: 5'-GCCAGCTCTCCTGGTCTCC-3'.

BBS2 has the following sequence: 5'-GGGAAUGGAUCCACAUCUACGAAUUCCGAAGCUCGUGACGGUACCUCCUAAAAUGUCCAUGGGGAAGGGAGGGAAUGGGAAGGACAAUCGGACACCGUUCACUGCAGACUUGACGAAGCUU-3'. The forward and reverse oligos used for qPCR analyses were BBS2-FOR: 5'-CGAAGCTCGTGACGGTACC-3' and BBS2-REV: 5'-CGGTGTCCGATTGTCCTTC-3'.

The N70 library forward and reverse oligos used for qPCR analyses were Lib-FOR oligo: 5'-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGA-3' and Lib-REV oligo: 5'-AAGCTTCGTCAAGTCTGCAGTGAA-3'.

All of the oligos used in this work were obtained from Integrated DNA Technologies.

Preparation of protein- and background-binding aptamers

Sequence verified DNA templates for each one of the specific aptamers used in this study were transcribed using T7 RNA Polymerase. After transcription, the

samples were treated with DNase I (Ambion), PAGE-purified, phenol:chloroform and chloroform extracted, isopropanol precipitated, and then re-suspended in DEPC-treated H_2O.

RNA selections and quantification

For the sequence specificity-study with serially configured microcolumns, each triplicate of eight targets was exposed to 1 mL of a mixed RNA pool in binding buffer [4.75-nM N70 library, 50-pM GFPapt, 50-pM HSFapt, 50-pM NELFapt, 50-pM BBS1, 50-pM BBS2, and 10-µg/mL yeast tRNA (Invitrogen)]. Similarly for the protein concentration studies with parallel microcolumns, the mixed RNA pools consisted of 4.95-nM N70 library and 50-pM specific aptamer. The RNA pools were injected at a rate of 33 µL/min for 30 min with a 10 µL aliquot of each pool set aside and used as a standard for quantitative polymerase chain reaction (qPCR) analysis. All buffers and solutions were degassed prior to use and introduced into the microcolumns via programmable multichannel syringe pumps (Harvard Apparatus) with MEDUSA placed onto a 96-well format liquid waste reservoir.

After binding to the library, the serially configured microcolumns were reconfigured to run in parallel by removing the caps and silicone layers permitting the connectivity of microcolumns, and reassembling the device with the appropriate caps for a parallel configuration. Each of the 96 microcolumns was then washed with 3 mL of binding buffer at a rate of 300 µL/min to remove unbound RNA. Finally, MEDUSA was placed directly onto a 2-mL 96-well microplate, and the RNA/RNA-protein complexes were eluted from the individual microcolumns by flowing elution buffer

[binding buffer + 50 mM ethylenediaminetetraacetic acid (EDTA pH 8.0) for selections with Ni-NTA resin; binding buffer + 10 mM glutathione for selections with GSH resin; binding buffer + 10 mM maltose for selections with amylose-resin] at a rate of 50 μL/min for 12 min. The RNA elution samples and the input standards were phenol/chloroform-extracted and ethanol-precipitated together with 1 μL of GlycoBlue (Ambion) and 40 μg of yeast tRNA (Invitrogen), and the resulting pellet was resuspended in 20 μL of RNase-free water, and reverse transcribed with Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV-RT) in two 96-well microplates. The N70 library, HSFapt, NELFapt, BBS1, and BBS2 all contain the same 3' constant region and were reverse transcribed using Lib-REV primer complementary to the 3' constant region in the RNA. For the experiments containing GFPapt, 4 μL of the resuspended pools and the standards were reverse transcribed using the GFPapt-REV primer specific to GFPapt. A 10-μL volume of each of the cDNA products was used for quantitative PCR analysis using 384-well plates on a LightCycler 480 instrument (Roche) to determine the amount of RNA library and of each specific aptamer that was recovered from each microcolumn. Different sets of oligonucleotides (see above) were used to independently evaluate the amount of N70 library and specific aptamers in each pool.

Design and fabrication of MEDUSA

MEDUSA was modeled after a 96-well microplate. The 96 units of our device were based off of our previously reported modular and mutliplexable affinity microcolumns, which were shown to minimize reagent consumption while

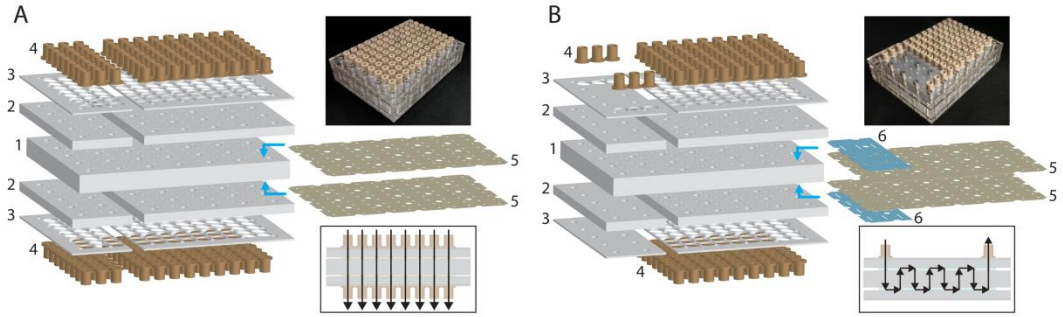


Figure 4.1. Diagram of the layers of MEDUSA in the order of assembly. (A) An exploded view of the customized device layers for configuring all 96 microcolumns to run in parallel. The flow path is shown in the lower boxed inset with no connections between microcolumns. The layers numbered 1 to 3 are the plastic layers: the middle layer (1) containing the microcolumns, the next outer two layers being the caps (2) and washers (3). The outermost layers (4) consist of inlet and outlet ports that are bonded to the final device. The two layers numbered (5) are silicone layers, which are bonded to the microcolumn layer (1) to hold porous frits against both sides of the microcolumns to retain affinity resin and to make liquid-tight seals across the entire device. A photograph of MEDUSA assembled in parallel is shown in the upper inset. (B) The customized device layers for configuring 24 of the microcolumns to run in series. The two additional silicone layers (6) shown in blue, as well as the smaller complementary plastic layers (2 and 3) on the left, are specifically programed to connect three sets of eight microcolumns within the device. The flow path is shown in the lower boxed inset with microcolumns connected in series via a serpentine route through eight microcolumns. MEDUSA assembled to run in series and parallel is shown in the upper inset

demonstrating significantly improved performances through optimizations of the selection parameters [13]. In order to allow for simple and versatile multiplexing and connectivity between microcolumns, our device was designed to be assembled in layers, with some of the layers "programed" for establishing connections within the device (see Figure 4.1). To fabricate the layers of MEDUSA, a two-dimensional CAD for each layer was designed and then cut using a CO_2 laser at 10.6 µm (Universal Laser Systems, VersaLaser). The speed, intensity, and density of laser pulses were optimized for each layer to obtain the highest quality and most reproducible cuts.

Each layer of MEDUSA was fabricated from either transparent biocompatible poly(methyl methacrylate) (PMMA) plastic or silicone. As seen in Figure 4.1.A (lower boxed inset), for parallelized microcolumns, there are 5 layers of plastic and 2 layers

of silicone as well as NanoPorts (IDEX Health and Science) for inputs and outputs on each side. The center most plastic layer (number “1” in Figure 4.1.A) is 1/2" thick and contains 96 microcolumns that each hold 10 µL of total volume. The next pair of layers (numbered “5” in Figure 4.1.A above and below the microcolumns) are 1/16” silicone layers for making a liquid tight seal across all 96 microcolumns. These layers contain 2 mm diameter holes for inserting porous polyethylene frits above and below each microcolumn to retain target-bound affinity resins, and have adhesive on one side for bonding to the microcolumn layer. The next pair of layers (numbered “2” in Figure 4.1.A) are 1/4" plastic capping layers which have small holes and NanoPorts (numbered “4”) bonded around them to allow solutions to flow in and out of the microcolumns. The outer most plastic layers (numbered “3”in Figure 4.1.A) are 1 mm thick and designed to simultaneously aid the alignment of the NanoPorts to the capping layers, as well as to bear and distribute forces from the assembly of all the layers by acting as a washer. All of the layers contain 35 evenly-spaced holes, with the middle microcolumn layer being threaded, for sealing the device together with screws (Figure 4.1.A and 4.1.B upper inset photographs). For serialized microcolumns (Figure 4.1.B, lower boxed inset), the design and assembly is similar. However, there are 2 additional layers of silicone (numbered “6” in Figure 4.1.B). These layers are fabricated in 1/32” silicone (no adhesive) and are programed to allow for the connectivity of microcolumns through small interconnecting channels.

RESULTS AND DISCUSSION

MEDUSA as an adaptable platform

MEDUSA was designed for high-throughput aptamer selections and characterizations of the SELEX process, and for versatility, allowing any combination of serial and parallel experiments. Due to the availability of plate-based processes and the potential for automation, we designed our device using the standard 96-well microplate layout, which easily couples with a typical 96-well plate for further post-selection sample processing. Furthermore, laser-cutting MEDUSA was ideal for rapid prototyping, requiring only 1 h to machine each device. For a universal device that does not necessitate customized plastic layers, a third layer of silicone could be used to similarly program the accessibility of all 96 possible input and output ports to the microcolumns. However, due to the inexpensive and rapid fabrication methods used, we decided instead to fabricate custom capping and washer layers that relay the same flow program by containing only the necessary input/output holes and NanoPorts. For the three sets of eight serialized microcolumns shown, this required only six holes/ports on the top layers and none on the bottom layers. This configuration also allowed for visual assessment of solutions flowing through the serialized microcolumns. The ease of fabrication for different programed parts, especially in thin silicone, allows for customized and versatile selections that can contain any number of parallel or serially connected microcolumns, as well as utilizing both configurations simultaneously. In cases where more than 96 microcolumns are desired, such as when 96 targets each require negative selections, additional microcolumn layers can be utilized in the assembly. As illustrated, our device was fabricated to accommodate

three sets of eight serialized devices, as well as 72 parallel selections (Figure 4.1). This combination was easily programed as described above. However, an even greater degree of versatility was achieved by dividing the capping, washer, and programed silicone layers into separately fabricated subsections that could be individually addressed and reconfigured without disrupting other microcolumns. This strategy also suggests the possibility of fabricating smaller versions of MEDUSA that contain the same general layout of a microplate, but occupy a smaller footprint by utilizing fewer microcolumns. This would allow users to handle smaller devices in less demanding applications, while benefiting from the standardized spacing and addressability of plate-based selections and sample processing.

Parallel selections reveal critical target concentration for aptamer enrichments

In our previous work, we found that GFP aptamer enrichments were limited by a critical GFP concentration that we attributed to steric hindrance [13]. Using MEDUSA, we decided to reproduce the GFP results with more data points, and to investigate the prevalence of this limiting effect by performing analogous studies with two additional proteins, hHSF1 and NELF-E, and their respective aptamers, HSFapt and NELFapt [13,15]. For each protein target, we chose to test eight concentration conditions starting at 10 μg/μL of resin with 2.5-fold dilutions down to 0.016 μg/μL in triplicate. The layout for all the samples on MEDUSA is illustrated in Figure 4.2 in the sections denoted II, III, and IV.

The binding results and enrichments for all three proteins are shown in Figure 4.3. The GFP microcolumns recovered a higher percentage of GFPapt and a lower

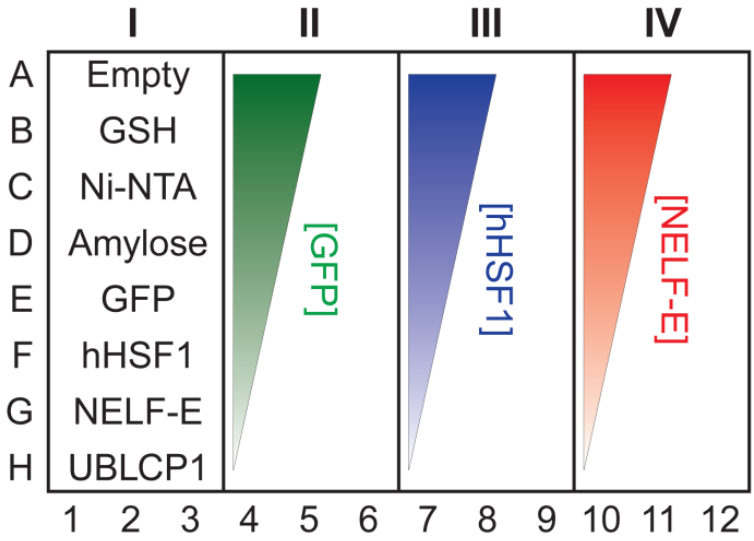


Figure 4.2. Layout for the 96 targets on MEDUSA according to its analogous microplate position given by the rows A-H, and the columns 1-12. In section I, the 8 indicated targets were connected in series from A to H to test the specificity and partitioning efficiency of various RNA aptamers. This was tested in triplicate in columns 1 to 3. Sections II, III, and IV tested the effects of target surface concentration on aptamer enrichments. The colored triangles indicate decreasing concentrations of each protein from 10 µg/µL (row A) to 0.016 µg/µL (row H) in 2.5-fold dilutions. Section II (green triangle) aimed to confirm previous enrichment behaviors shown with GFP. Sections III and IV tested the same concentrations of the proteins hHSF1 (blue triangle) and NELF-E (red triangle) to assess the prevalence of target surface concentration effects on binding due to steric hindrance or other effects in other aptamer selections

percentage of N70 library than those reported previously (Figure 4.3.A), due to lower flow rates that were used to increase HSFapt and NELFapt binding, since they have higher K_ds. This resulted in an expected increase in the enrichment of GFPapt over the N70 library; however, the characteristic shape and optimal concentration of 0.6 µg/µL for the enrichment curve are the same as previously reported (Figure 4.3.D). With hHSF1, the recovery of HSFapt followed a more typical sigmoidal shape, which saturated at increasing concentrations of hHSF1 (Fig. 2B). Similarly, the enrichment of HSFapt over the N70 library increased steadily and then saturated at higher concentrations (Figure 4.3.E). It is interesting, however, that HSFapt enrichment

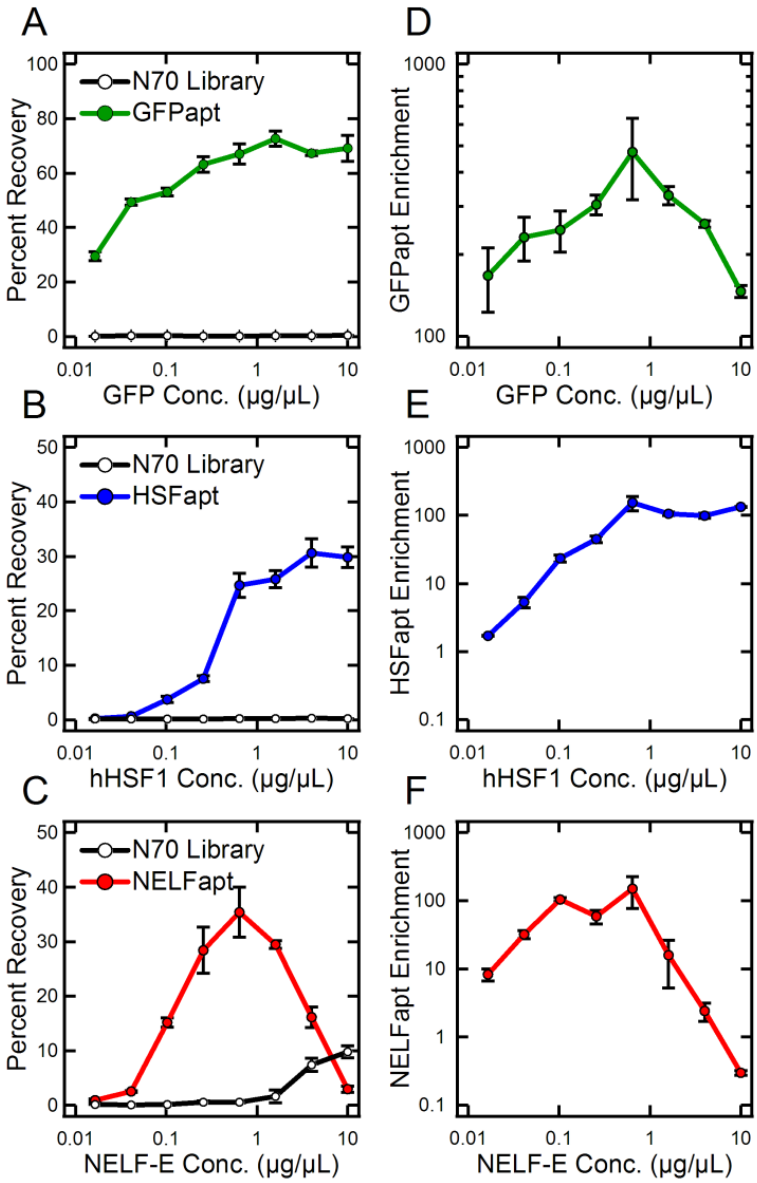


Figure 4.3 Recoveries and enrichments of specific RNA aptamers over the N70 library as a function of protein concentration. (A) The recovery of GFPapt and N70 library at various concentrations of GFP. Analogous data for the recovery of (B) HSFapt and N70 library from hHSF1, and (C) NELFapt and N70 library from NELF-E. (D-F) The calculated enrichments of the specific aptamers (GFPapt, HSFapt, NELFapt) over the random library. The error bars represent the standard deviation in recoveries or enrichments calculated for each condition performed in triplicate

plateaued at the optimal concentration for GFP. With NELF-E, there is a very clear NELFapt recovery optimum at this same concentration, with significant losses in recoveries at concentrations above 0.6 µg/µL (Figure 4.3.C). In addition, the recovery of the N70 library increased significantly above the optimum concentration for NELFapt, likely due to the fact that NELF-E contains an RNA Recognition Motif and can bind RNA non-specifically [15]. These two binding trends result in a drastic decrease in enrichment at higher concentrations of NELF-E, resulting in de-enrichment of NELFapt at the highest concentration of 10 µg/µL (Figure 4.3.F).

The three concentration studies with GFP, hHSF1, and NELF-E make a strong case for the general steric hindrance of target molecules that are over packed in solid-phase affinity selections. Although the GFP and hHSF1 aptamer recoveries do not show drastic decreases at high concentrations as with NELF-E, this binding behavior is affected by the selection flow rates and is clearly seen between our old and new GFP data. Importantly, the recoveries saturated well below 100%, which indicates the existence of some limiting effect. Most revealing is NELF-E, where the binding site for NELFapt appears to be particularly inaccessible at high concentrations, causing a significant loss in total aptamer binding. Furthermore, a simple calculation (assuming hard spheres for the resin) predicts that a critical surface density of proteins should occur between 0.1 and 1 µg/µL (depending on protein shape and size and the diameter of the resin beads). Since all three proteins are similar in size, it is not surprising that we observe the same critical concentration of 0.6 µg/µL, and the results suggest that the target concentration may be the most limiting parameter for enriching aptamers.

Multiplex serial selections show specificity of target and background binding sequences

Previously, we performed multiple partitions to input pools and libraries by connecting several microcolumns in multiplex selections [13]. In particular, we showed the highly specific and efficient partitioning of GFPapt to GFP over non-specific proteins and an empty microcolumn. This configuration is useful for multitasking DNA or RNA libraries on multiple unrelated selection targets, or to separate enriched pools for aptamers that bind to distinct sites on a complex target [18]. We decided to demonstrate similar multiplex selections using MEDUSA by extending this analysis to include several additional RNA aptamers and protein targets: GFP, hHSF1, NELF-E, and their respective aptamers. To thoroughly characterize the specific, non-specific, and background binding of each RNA aptamer, we also included a non-specific protein, UBLCP1, three commonly used affinity resins, GSH, Ni-NTA, and amylose, and empty microcolumns. Each of the four protein targets were immobilized onto their respective resins at 0.6 µg/µL. The eight targets were arranged in series for the multiplex selection and performed in triplicate to quantify the reproducibility of each aptamer's partitioning efficiency and specificity. The order of targets was as follows: empty, GSH, Ni-NTA, amylose, His-GFP, GST-hHSF1, His-NELF-E, His-UBLCP1, and is illustrated in Figure 4.1 in section I.

In addition to the random N70 RNA library and the aptamers to our three proteins, our test pool also included two suspected BBSs, BBS1 and BBS2. For all previous multiplex selections, we have performed high-throughput sequencing, which

Table 4.2. Frequencies of BBS1 and BBS2 in previous selections

Resin	BBS1 Dominant	BBS2 Dominant
Ni-NTA	1	7
GSH	3	4
Amylose	3	4
Empty (no Resin)	0	2

Summary of the number of times BBS1 or BBS2 has been identified in all previous selections. The numbers indicate the instances in which BBS1 was more highly enriched than BBS2 (or vice versa) on each target, grouped according to the resin on which each target was immobilized.

provided tremendous amounts of sequence data and sensitivity for early detection of aptamers [13,15,16]. However, comparison of the sequencing results for dozens of targets revealed several identical sequences that were frequently enriched, particularly in earlier cycles before target-binding aptamers began to dominate the pool. This was especially true for less aptagenic targets, where the two sequences, BBS1 and BBS2, were generally among the highest enriched candidates (see Table 4.2). From these data, we predicted that BBS1 would enrich on all targets by binding to the plastic device and the resins. This was also predicted for BBS2; however, we expected BBS2 to enrich more strongly than BBS1 on all targets, especially in microcolumns containing Ni-NTA (similar analyses have been used to identify sequences that bind specifically to Ni-NTA [19]).

The partitioning results for each RNA aptamer are shown in Figure 4.4 as enrichments over the random RNA library. Our specific aptamers each show striking enrichments only on their intended target. GFPapt enriched an average of 750-fold on GFP microcolumns, but only an average of 0.6-fold (de-enriched) on all other targets (Figure 4.4.A), which reflects its strong specificity for GFP. Similarly, HSFapt enriched an average of 232-fold on hHSF1 microcolumns, and only an average of 2-

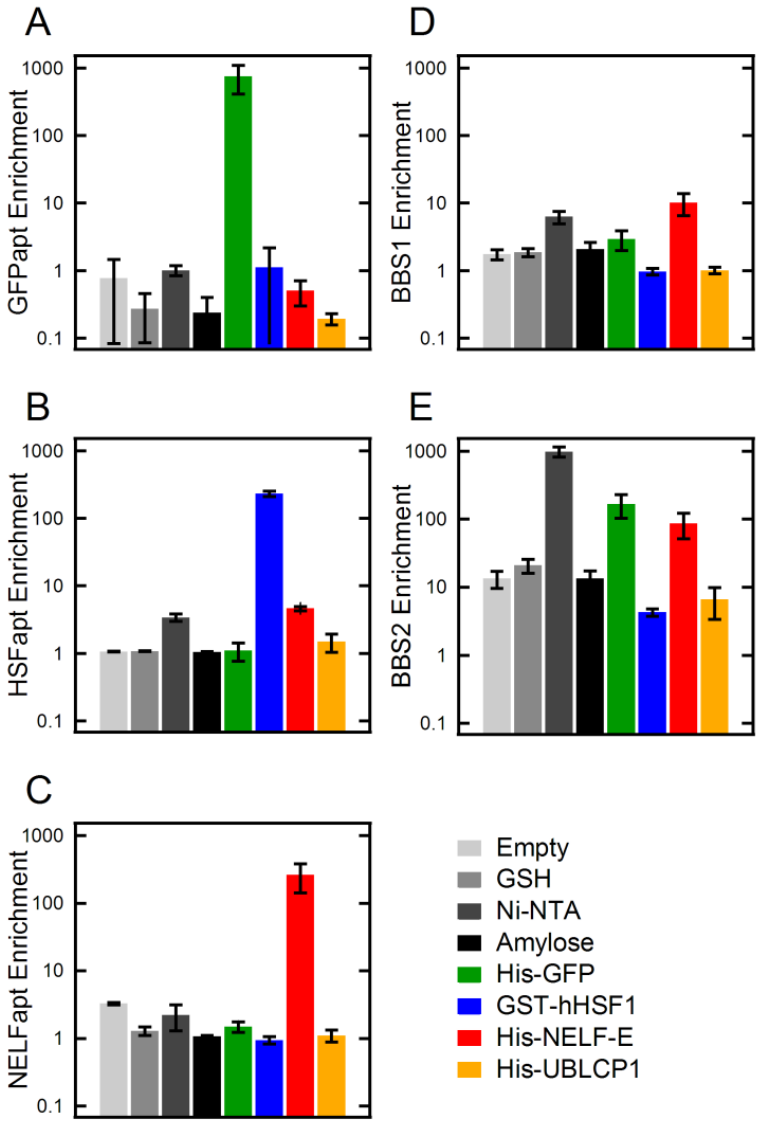


Figure 4.4. The enrichment of RNA aptamers over the N70 library on various targets connected in series. The enrichment of each protein-specific aptamer, GFPapt (A), HSFapt (B), and NELFapt (C), and non-specific aptamers, BBS1 aptamer (D), and BBS2 aptamer (E), on all 8 microcolumns. The error bars represent the standard deviation in enrichments calculated for each target performed in triplicate

fold on all other targets (Figure 4.4.B). NELFapt enriched an average of 262-fold on NELF-E, and only an average of 1.6-fold for all other targets (Figure 4.4.C).

For BBS1 and BBS2, we found good agreement with the qualitative analysis of all previous sequencing data [13,15,16]. BBS1 enriched on all targets as predicted; however, it enriched three times more in Ni-NTA-containing microcolumns (Ni-NTA, GFP, NELF-E, UBLCP1), with enrichments averaging 1.7 for non-Ni-NTA targets and 5.1 for the Ni-NTA targets (Figure 4.4.D). BBS2 also enriched as predicted, with enrichments higher than BBS1 on all targets (Figure 4.4.E). More specifically, BBS2 enriched an average of 13-fold on non-Ni-NTA targets and a surprising 311-fold on Ni-NTA targets suggesting that BBS2 has a specific affinity for Ni-NTA. In fact, for the first Ni-NTA target in the serial selection, blank Ni-NTA, enrichment averaged almost 1,000-fold. This is almost 80 times greater than non-Ni-NTA targets, and may reflect more accurately the specificity of BBS2 for Ni-NTA. In support of this hypothesis, we noticed that BBS2 was quickly depleted from the pool as it was injected across all the Ni-NTA-containing microcolumns, as seen by the monotonically decreasing enrichments of BBS2 to the Ni-NTA-containing targets from left to right.

Although negative selections are often used to separate sequences with specific affinities for sources of background binding, these are rarely completely effective at eliminating enrichment of non-specific RNAs. In contrast, we have found the repeated occurrence of BBSs in different SELEX experiments to be valuable indicators of the selection progress. Perhaps more interestingly, selections for and/or identification of

BBSs can be used to generate non-specific blocking reagents that are more effective than commonly used yeast tRNAs.

CONCLUSIONS

We believe that MEDUSA can be used to significantly increase the productivity of large-scale aptamer discovery efforts. By utilizing the versatility and programmability of MEDUSA, a larger configuration space of potential SELEX designs can be explored. Just as importantly, selections utilizing affinity chromatography can be thoroughly characterized. These kinds of data not only allow us to optimize future aptamer selections, but also clearly show that performances can be improved, while simultaneously consuming much less reagent, such as protein, making aptamer selections more accessible to targets that are difficult to purify or express in large quantities. Furthermore, such characterizations would not only improve future aptamer selections, but also aid in the development of more functional and applicable SELEX theories in solid-phase affinity selections.

REFERENCES

1. Ellington AD, Szostak JW (1990) Invitro Selection of Rna Molecules That Bind Specific Ligands. Nature 346: 818-822.

2. Joyce GF (1989) Amplification, mutation and selection of catalytic RNA. Gene 82: 83-87.

3. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.

4. Tombelli S, Mascini M (2009) Aptamers as molecular tools for bioanalytical methods. Curr Opin Mol Ther 11: 179-188.

5. Park SM, Ahn JY, Jo M, Lee DK, Lis JT, et al. (2009) Selection and elution of aptamers using nanoporous sol-gel arrays with integrated microheaters. Lab Chip 9: 1206-1212.

6. Ahn JY, Jo M, Dua P, Lee DK, Kim S (2011) A sol-gel-based microfluidics system enhances the efficiency of RNA aptamer selection. Oligonucleotides 21: 93-100.

7. Cox JC, Ellington AD (2001) Automated selection of anti-protein aptamers. Bioorg Med Chem 9: 2525-2531.

8. Cox JC, Rudolph P, Ellington AD (1998) Automated RNA selection. Biotechnol Prog 14: 845-850.

9. Eulberg D, Buchner K, Maasch C, Klussmann S (2005) Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonist. Nucleic Acids Research 33: e45.

10. Drolet DW, Jenison RD, Smith DE, Pratt D, Hicke BJ (1999) A high throughput platform for systematic evolution of ligands by exponential enrichment (SELEX). Comb Chem High Throughput Screen 2: 271-278.

11. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, et al. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20: 861-873.

12. Daniel C, Roupioz Y, Gasparutto D, Livache T, Buhot A (2013) Solution-Phase vs Surface-Phase Aptamer-Protein Affinity from a Label-Free Kinetic Biosensor. PLoS One 8: e75419.

13. Latulippe DR, Szeto K, Ozer A, Duarte FM, Kelly CV, et al. (2013) Multiplexed microcolumn-based process for efficient selection of RNA aptamers. Anal Chem 85: 3417-3424.

14. Ozer A, White BS, Lis JT, Shalloway D (2013) Density-dependent cooperative non-specific binding in solid-phase SELEX affinity selection. Nucleic Acids Research 41: 7167-7175.

15. Pagano JM, Kwak H, Waters CT, Sprouse RO, White BS, et al. (2014) Defining NELF-E RNA Binding in HIV-1 and Promoter-Proximal Pause Regions. PLoS Genetics 10: e1004090.

16. Szeto K, Latulippe DR, Ozer A, Pagano JM, White BS, et al. (2013) RAPID-SELEX for RNA Aptamers. PLoS One 8: e82667.

17. Shui B, Ozer A, Zipfel W, Sahu N, Singh A, et al. (2012) RNA aptamers that functionally interact with green fluorescent protein and its derivatives. Nucleic Acids Research 40: e39.

18. Gong Q, Wang JP, Ahmad KM, Csordas AT, Zhou JH, et al. (2012) Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. Anal Chem 84: 5365-5371.

19. Nastasijevic B, Becker NA, Wurster SE, Maher LJ, 3rd (2008) Sequence-specific binding of DNA and RNA to immobilized Nickel ions. Biochem Biophys Res Commun 366: 420-425.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

SUMMARY AND DISCUSSION OF RESULTS

High-throughput SELEX technologies

Aptamers have become powerful affinity reagents, demonstrating uses and functionalities far beyond antibodies. However, generating these high affinity molecules still requires a tremendous amount of time and effort. In addition, classical selections such as filter binding or affinity chromatography may consume, and potentially waste, large quantities of reagents. Finally, the complex and sophisticated selection schemes necessary to achieve aptamer qualities beyond tight binding (such as specificity) can be difficult to incorporate, and require significantly more time and effort to achieve. These difficulties only intensify when considering aptamer selections to multiple targets.

To address a number of these issues, we developed an affinity microcolumn-chromatography technology [1]. These microcolumns are orders of magnitude smaller in volume than traditional columns, which results in (at minimum) a proportional decrease in the consumption of target molecules. Under properly controlled conditions, the tiny volume of immobilized target molecules is sufficient to bind the majority of high affinity aptamers. The small volume also results in increased enrichments of aptamers, since high affinity sequences are quickly retained in small columns, where larger columns would predominantly recover background binding sequences. Due to their small size, the microcolumns can also be easily connected to one another. For this reason, we designed our microcolumns to be modular in nature, utilizing NanoPorts to facilitate standardized connections between multiple microcolumns and/or to fluid handling systems and other instrumentation. This serial

configuration of microcolumns allows for simple and simultaneous in-line negative SELEX to be performed against affinity resins, affinity tags or other sources of off-target binding. Similarly, this configuration allows for simultaneous counter SELEX to be performed isolating aptamers with high specificities between similar targets. Perhaps most importantly, these microcolumns help to achieve multiplexed parallelized selections.

With the catalogue of available aptamers growing, acceptance and use of aptamers has expanded dramatically. This has led to a demand for higher throughput selection technologies, which can accelerate and parallelize the selection processes. The microcolumns can easily accommodate multiple targets in parallelized selections, and by exploiting the non-equilibrium characteristics of our platform we can manipulate fluidic parameters, such as time and flow rates, to maximize the resulting aptamer enrichments. However, the microcolumns can also be used to parallelize the use of reagents, such as the random library. In most selections, the majority of the library is discarded with only a tiny fraction being bound and recovered. In general, it can be assumed that the subset of high affinity sequences are mutually exclusive among random targets. Therefore, in cases where dissimilar targets are being considered, this discarded library can be further partitioned. We demonstrated the ability to isolate aptamers to multiple targets through multiple partitions of the library, as well as through the continuous use of negative SELEX against the target's affinity tag (as well as the microcolumn). Multiplex SELEX was also demonstrated among targets bound to different kinds of resins simultaneously due to the variety of available affinity tags used for protein or target purifications.

At an even higher level of parallelization, aptamers can be simultaneously isolated to distinct domains of a large target complex. This can be achieved by an end-stage partitioning of an enriched aptamer pool [2]. This pool can be run across several

microcolumns, each containing a smaller subunit of the whole target, thus partitioning the mixed pool into subpools that each recognizes a specific region of the target. This is not only easier and less costly than performing individual selections to each domain separately, but also guarantees that the aptamers isolated to each subdomain properly recognize the whole complex, removing the possibility of isolating aptamers that bind to regions of the domains that are otherwise obscured in the complex.

To scale this microcolumn technology further, we designed a 96-well microplate formatted multiplex device called MEDUSA [3]. This format simplifies the handling of multiple targets by grouping them together into a single device. In addition, by constraining the layout of the device, MEDUSA couples directly to a microplate so the selection has the potential to be automated through liquid handling systems. Even more importantly, samples are eluted directly into a microplate and can be processed in a high-throughput and similarly automated manner through other microplate-based technologies. This device also retains the ability to perform selections in series and/or parallel, and the footprint of the device can be adjusted to accommodate smaller and less demanding numbers of selections. We demonstrated the use of MEDUSA utilizing all 96 available microcolumns to characterize the binding of various RNA aptamers in both serial and parallel formats.

Although the microcolumn platform is quite simple compared to other miniaturized SELEX technologies, our work shows that there is still a role for affinity chromatography in future high-throughput selections. In contrast to more complex and expensive technologies, the simplicity and ease of use of our microcolumns make them more accessible to researchers who want to perform aptamer selections, as the basic protocols, operating principles, and even fluidic equipment are already well understood and available in many labs. Not only are the microcolumns based on one of the most commonly used methods for aptamer selections, but the simple application

of controlled fluidics allows researchers to manipulate and tune a wider range of parameters. Ultimately, the multiplex nature of our technology can not only be used to highly parallelize the selection process, but to also exercise much more complex and sophisticated selection schemes.

The SELEX method and binding characterization

Traditionally, the SELEX method is treated as an equilibrium solution-phase binding problem. The associated theories on optimal SELEX strategies are quite complex, and generally emphasize the importance of target and ligand concentrations assuming some specific and background binding characteristics of the library or the enriched pool [4]. Some simple "rules of thumb" have been deduced from these theories suggesting "competitive" regimes for selections. The concentration of target molecules is often set relative to the existence of an aptamer with an assumed dissociation constant (K_d), and the library is set to be in certain excess relative to the target concentration (i.e. 10:1 or 100:1) [5]. However, outside of solution-phase SELEX, target molecules are bound to a solid support, and accurately describing their concentration can become difficult as local (surface) concentrations can be much larger than the average solution (volume) concentration. Properly interpreting target and ligand concentrations is made even more difficult when the target is confined to a small volume as in our microcolumns, and a much larger volume of ligand solution is flowed across it. Most importantly, the finite size and orientation of surface bound target molecules cannot be ignored.

In solid-phase selections, high target concentrations can allow adjacent target molecules to interact with each other. In all of our specific protein-binding aptamer tests, increasing concentrations of target was shown to eventually result in a

phenomenon consistent with steric hindrance, which lowered the accessibility and effective concentration of target proteins to the aptamers [1,3]. This resulted in an observed optimum target concentration that can be estimated easily through geometric arguments alone. In contrast, studies with background binding to peptide targets demonstrated significant cooperativities, where non-specific aptamers interacted strongly with multiple nearby target molecules, and significantly increased the amount of background binding (in some cases by over 2 orders of magnitude!) [6]. In general, there are additional sources of background binding other than the target molecules. This can include any surfaces and substrates that library molecules can come into contact with including tubes, filters, and in our case the affinity resins and the microcolumns themselves. Importantly, this form of background binding cannot be simply reduced by lowering the target concentration. In fact, lowering target concentrations too far ultimately increases the proportion of background binding over target binding aptamers.

Combining the effects of both steric hindrance and cooperativity at high concentrations of target molecules significantly complicates the aptamer enrichment process. Particularly, if either of these effects varies among high (or low) affinity binders, differential binding can make proper identification of the highest affinity aptamer difficult, as this determination may be dependent on the target concentration. Using common metrics such as the highest multiplicity sequence in a pool, or even the fastest enriching sequence between pools, may not successfully identify the aptamer with the highest 1:1 binding affinity as in solution-phase binding or surface-/solid-phase environments where target molecules are not tightly packed (i.e. multiple target molecule interactions are negligible). In fact, there may not be any agreement between the ranks of aptamer candidates based on their actual 1:1 binding affinity and their effective binding affinity in high target concentration environments.

Given the binding trends that have been observed from high to low target concentration regimes, the results support a local optimum target concentration that is determined primarily on the basis of background binding and the size of target (and aptamer) molecules, instead of on an assumed aptamer K_d or other idealized theoretical parameters. Using our microcolumns in conjunction with the observed optimum target concentration of ~ 0.6 µg/µL, we demonstrated improved aptamer enrichments and significantly reduced the time and cost to isolate aptamers. Furthermore, by miniaturizing the traditional column and using non-saturating optimum target concentrations, the reduction in the total consumed target molecules and other reagents have made aptamer selections accessible to a wider range of targets that are difficult to purify in large quantities. This is even more apparent as the number of selection cycles needed has been significantly reduced compared to classical methods.

Even when enrichments for specific aptamer binding have been maximized, it is important to note that background binding may not always be non-specific. To the contrary, concurrent selections for specific background binding aptamers and target binding aptamers are likely taking place [7]. In our own work, we have identified two noteworthy sequences that have appeared in at least 25 selections, which we believe were due to their preferential or specific background binding characteristics [3]. This was tested and verified in our MEDUSA experiments, which showed that these sequences enriched over the random sequence library on the affinity resins and the empty device. This highlights one unexpected advantage of multiplex SELEX, which has allowed us to perform many selections under identical or similar conditions and compare their sequence content to identify bias, processing artefacts, or reoccurring sequences indicating background binding to a common source.

Ultimately, lists of background binding sequences can be used to diagnose background binding in future selections, and be accounted/corrected for to aid in target-binding aptamer identifications, since the presence of these kinds of sequences are likely inevitable and impractical to completely eliminate. However, these sequences could be used to generate specific and more effective blocking reagents to minimize background depending on the materials being used in the selection. Furthermore, these sequences are likely unique to each random library and selection technology, in which case their unique sequence signatures can be used to indicate the status of individual selections whose target-binding aptamers must compete with background sources for sequence sampling. Rare and/or weakly binding aptamers (or the absence of an effective aptamer) would be sampled less frequently than abundant and/or high affinity aptamers compared to background binding sequences. These qualitative trends could be used to indicate poor selection performances, possibly due to the ineffective parameters and/or selection design being utilized, or due to a non-aptagenic selection target, which is often the case.

Throughout this work, we have encountered several factors that have significant implications not only on how to most effectively perform aptamer selections, but also how to adequately perform the subsequent sequence analysis in identifying true high affinity target-binding aptamers. It has become quite clear that despite improving the overall rate of aptamer selections and streamlining of the various processes, new (and even established) technologies pose a significantly more complex optimization problem for SELEX. As new technologies emerge and demands for faster and cheaper aptamer selections increase, a deeper understanding of how to adequately model and optimize their particular binding behavior will be necessary. These efforts have been and will continue to be greatly aided by new and more powerful sequencing and analysis techniques.

High-throughput sequencing and analysis

As more sophisticated SELEX technologies and methodologies emerge, more powerful analytical methods will become necessary. This is not only to facilitate more detailed sequence analysis, but also to identify aptamers with much greater sensitivity. This can be done by searching for and identifying consensus sequence motifs, or by identifying conserved secondary structures [8]. However, the simplest and arguably most important improvement to sensitivity can be achieved by identifying rare sequences via high-throughput sequencing.

With today's sequencing technologies it is now possible to acquire 10^8 sequencing reads from any given pool, which has allowed researchers to identify highly enriched sequences (high multiplicity) after many fewer selection cycles than would be necessary for the pool to fully converge to that sequence. This sensitivity has also opened up other modes of analysis, such as round-to-round enrichments and enrichment trajectories [9,10]. Enrichment analysis can overcome biases or contamination in selection pools by ranking sequences based on how rapidly they are converging rather than on their multiplicity. Some groups have monitored both enrichments and multiplicities of sequences over many rounds and revealed quite dynamic and unexpected phenomena [11]. With so many sequences available, the evolution of aptamers throughout various cycles and selection conditions can be used to model the binding behavior and help to generate better protocols and predictive methods for identifying the best binding aptamers through population-based metrics alone.

As selection schemes become more complex, high-throughput sequencing and comparative sequencing methods can be used to identify and/or correct for a number

of other sequencing phenomena. As mentioned previously, by comparing sequence content between targets, similarities can be identified, such as common background binding sequences. Sequences can also be divided into specific and non-specific binding candidates based on whether or not they are uniquely represented in any one pool. In fact, searching the sequence content among many targets, similarities may emerge revealing a deeper connection between aptamer sequence and structure to the characteristics of the selection target. Finally, the improved sequence sensitivity can be used to identify mutations of sequences, which, through appropriate analysis, may prove to be better binders than their parent sequence, or simply an inconsequential effect from PCR or sequencing errors. We have developed clustering algorithms to reduce pool complexities to more representative levels by grouping together nearly identical sequences which we believe are (more often than not) artefacts from imperfect processing of the pools [1].

By using new high-throughput sequencing and analysis tools with a standardized target and library, researchers can finally begin to objectively rank and characterize the effectiveness of the wide array of different SELEX technologies and methods, which are currently still evaluated primarily on the basis of success or failure to isolate aptamers. Many researchers cite their technology's ability to isolate higher affinity aptamers than previous attempts; however, this can be due to slight changes in selection conditions rather than the technology itself. Even more importantly, sequence and affinity differences may be the consequence of simple sampling variations due to differing source libraries.

The sampled sequence space of about 10^{15} different sequences for libraries larger than about 25 randomized bases can be safely assumed to be mutually exclusive for independent but identically synthesized libraries. This is especially true for larger libraries such as our N70 library which has more than 10^{42} possible sequences.

Different, but identically synthesized pools may contain superior sequence candidates for different targets, resulting in a randomly acquired advantage, that may be unfairly credited to the selection method or technology. In addition, sequence libraries can be synthesized from a variety of different natural and modified nucleic acids, as well as vary significantly in length from each other; and it's currently unclear if/how technologies differing in these ways can be fairly compared. Even when all the selection parameters are identical, the "winningest" aptamer sequence can vary between selections simply because of the stochastics of binding and recovery of sequences in the first cycle or two when copy numbers are very low. In addition, losses and amplification errors can be incurred during post-selection processing of the enriched pools. These sampling problems make comparisons of selections unfair when based purely on success and the characteristics of a single aptamer.

With the resources available today, and given the wide array of different selection technologies, rigorous and unbiased methods of comparison and characterization are needed. A fair and quantitative comparison could be achieved by standardizing the target molecule and the library, and comparing thousands or millions of enriched sequences generated using different technologies, methods or selection parameters. By identifying identical sequences and characterizing their affinities, scoring metrics could be designed that do not rely solely on success as defined by a single sequence, but rather on the populations of unique and identical sequences compared to previous methods. We used this format as well as the above sequencing and analysis tools to test and validate our RAPID-SELEX method, which proved to be much more efficient compared to traditional SELEX [10].

The use of high-throughput sequencing represents a tremendous advancement in aptamer selection technologies. Its ability to detect enriching sequences at such low levels has accelerated and even enabled otherwise impossible identification of high

affinity aptamers. Furthermore, the improved sophistication in sequence analysis has led to many new and powerful techniques for identifying and optimizing aptamer sequences and structures. However, these methods are still in their infancy, and have yet to be used in rigorous characterizations of SELEX technologies or methods.

Time as an explicit variable of SELEX

Since its inception, many innovations around the SELEX method have focused on reducing the time needed to isolate high affinity aptamers. More accurately, new methods and technologies advertise their potential to reduce the number of SELEX cycles needed to converge an aptamer. This paradigm is consistent with the numerous equilibrium solution-phase SELEX theories, which aim to describe the most optimal library/pool and target concentrations that converge an aptamer in the least number of cycles. Not surprisingly though, time is not considered explicitly in these theories, and in laboratory selections the times used to complete various steps are often poorly, if ever, justified. Although citing the number of cycles does illustrate to some extent the amount of time as well as the amount of reagents consumed, the true length of a "cycle" can vary widely among different methods and technologies obscuring their actual advantages or even potential disadvantages.

There are three scales of time that are important in describing each technology's potential to "save time". These include the times allocated for each step in the selection cycle, the total cyle time (the sum of the individual step times), and the total selection time (the sum of all cycles). The individual selection step times are probably the most poorly characterized of all three time scales. In many cases, these times seem to be chosen arbitrarily, or based on unproven assumptions. Perhaps the most important time-dependent steps are the binding and washing steps. The time for these steps is especially important in non-equilibrium SELEX formats, such as our

microcolumns, and can be used to manipulate the binding kinetics within the device. Intuitively, one would expect that short binding times would discriminate sequences based on their on-rate (k_{on}), which should (on average) preferentially bind higher affinity aptamers over lower affinity aptamers. Similarly, long wash times would be expected to discriminate sequences based on their off-rate (k_{off}), which should (on average) preferentially eliminate low affinity aptamers over high affinity aptamers.

The example above illustrates the possibility to utilize a fluidic platform to manipulate kinetic parameters, and highlights the potential advantage of using technologies such as our microcolumns to improve aptamer enrichments. However, when these parameters were tested in model selections with fixed volumes (i.e. fixed concentrations), the results were surprising [1]. To control the binding time, we varied the flow rate of a 1 mL model library, which in turn adjusted the exposure time of individual sequences to target molecules in the microcolumns. Although the aptamer behaved as expected in the microcolumns by binding less with shorter binding times, we found that the random library bound more non-specifically with short binding times. This is in complete disagreement with simple kinetic binding theory, and the result was an enrichment trend that varied inversely with time compared to predictions. Similarly, the washing time was controlled by varying the flow rate of a 3 mL washing buffer. As before, the aptamer behaved as expected in the microcolumns binding less with longer washing times, but we found that the random library bound more non-specifically with longer washing times. This too is in complete disagreement with kinetic theory and resulted in a similarly inverse enrichment trend compared to predictions. Surprisingly, the observed trends are ideal for selections as they suggest that the conditions that most efficiently bind and retain high affinity aptamers are also the same conditions that minimize background binding. This means that the most conservative conditions for aptamer binding result in the highest

enrichments, eliminating design considerations that would require deliberately losing potential aptamers for the sake of improved enrichments.

These results not only demonstrate the potential to utilize fluidic parameters to optimize aptamer selections, but also reveal the importance of characterizing selection performances for the relevant and tunable parameters. As our results show, high affinity aptamers are well modeled by kinetic theory; however, background and non-specific binding (as discussed in the previous sections) is much more complex and seems to be much more sensitive to the flow rate than to the step time. This is likely due to flow characteristics near surfaces such as shear, high pressures, and other more complex phenomena (such as the compressibility of the affinity resin) that may preferentially affect low affinity binders. These complex phenomena are beyond the capacity of basic kinetics, and its failure to qualitatively predict the observed enrichment trends points to the dangers in using simple arguments and assumptions, like the ones illustrated above, to choose the temporal conditions for aptamer selections. In our case, the assumed "optimal" conditions within the tested range would have resulted in little to no aptamer enrichments. These kinds of operational and design mistakes likely occur unnoticed in many selections, requiring significantly more selection cycles or resulting in failures that are incorrectly attributed to the difficulties of the particular selection target.

The selection time is the most commonly reported metric and, as mentioned, is usually given in number of cycles. This number can be easily reduced depending on the methods used for sequencing aptamer candidates. If high-throughput sequencing is used, aptamers can theoretically be identified with sensitivities down to 1 in 10^8; and with pools that start with unique sequences at about 1 in 10^{15}, high-throughput sequencing can effectively cut the number of rounds needed in half. Although the information content from high-throughput sequencing is immense and can often make

a critical difference in identifying aptamers, it also is quite costly and can require significant time to complete and fully analyze. This time is rarely if ever considered when reporting a successful aptamer candidate relative to less powerful but simpler methods, and it is conceivable that in some cases, the application of high-throughput sequencing saves neither time nor costs in identifying candidates despite eliminating many SELEX cycles. In fact, with today's current methods, SELEX technologies that accelerate aptamer selections and high-throughput sequencing analysis are at odds with each other. On one hand, SELEX technologies aim to converge to the highest affinity aptamer as efficiently as possible, which presumably has the absolute performance limit of binding and recovering only a single sequence of the highest affinity in a single cycle. However, from a sequencing standpoint a single recovered sequence makes it impossible to distinguish between a successful and an unsuccessful selection, and all population-based metrics become meaningless for discriminating aptamers from non-aptamers. In contrast, high-throughput sequencing methods require a rich diversity of sequences that can be used to determine one or more sequence motifs or secondary structural elements that are critical to an aptamer. This information can even be used to generate a higher performing consensus-based aptamer that may not be fully represented in any one sequence. However, the ideal diversity and analytical power of these tools scales proportionally with the capacity of the sequencing method used, which limits the desired efficiency of the SELEX technology being utilized. As more efficient methods for converging aptamers are developed, the potential richness and value from high-throughput sequencing and analysis may be diminished, requiring new approaches that maximize the utility of both.

The potential for methods, such as high-throughput sequencing, to simply reduce time and costs in selections is largely dependent on the number of selection

cycles that are avoided and the time required to complete each cycle. In some cases, selection cycles can be done daily or even more frequently [12]. This diminishes the value of such costly analysis for simple selections, especially those that are well executed and free from bias and other potentially constraining phenomena. In contrast, in one of our multiplex SELEX experiments using our microcolumns under optimal temporal (and target concentration) conditions, we required an average of 42.5 hours of work for each round, which was typically achieved over the course of one week [1]. After 5 rounds, we performed high-throughput sequencing, that revealed that not only had we almost completely converged our pools onto the candidate aptamers, but also that we were able to confidently identify these sequences as candidate aptamers after only 3 rounds. This suggests that we could have saved 1 or 2 weeks' worth of time and reagents, cutting our SELEX efforts nearly in half. In other selections, sequences were far from converged even after 6 rounds of SELEX, and aptamers candidates were only identified because of the sensitivity of high-throughput sequencing [10].

Although technologies like our microcolumns and high-throughput sequencing have been shown to drastically decrease the number of cycles of SELEX needed to confidently identify aptamer candidates, little work has been done to address the efficiency of SELEX as a whole. Despite more than two decades of progress, only a few studies have been done to globally optimize SELEX. We first introduced our microcolumns by utilizing parameters that maximize enrichments at each selection step. However, no discussion was made about the potential net gain or loss in efficiency that could have resulted from the optimal selection steps, such as the 1 µL/min binding flow rate that required almost 17 hours to complete [1]. In general, the appropriateness of these conditions depends also on the total time to complete the selection cycle (and potentially on reagent consumption), since the total enrichment of aptamers can accumulate faster despite using less efficient parameters, provided that

they allow for much more rapid completion of each cycle of SELEX. In addition, only one previous study had diverged from the traditional SELEX method of iterative binding, partitioning, and amplifying enriched pools. This Non-SELEX method was shown to effectively isolate aptamers to targets despite several cycles of binding and partitioning without any amplification [12]. This resulted in tremendous savings in time and reagents, since amplifications are often the rate-limiting steps, and also brought into question the effectiveness of the traditional SELEX method, especially in later selection cycles.

In search of the global SELEX optimum, as well as a more efficient SELEX method, we developed a new selection method called RAPID-SELEX, and compared its performance to traditional SELEX [10]. For traditional SELEX, six cycles of selections were performed under optimal conditions for each parameter (~17 hour binding step, and a "tuned" wash), with amplifications performed after every cycle. However, for RAPID, only every other selection had its pool amplified allowing two selection cycles to be performed quickly back-to-back in each full RAPID-SELEX "round" (here, a round is defined to include amplifications). In addition, the binding steps were shortened to less than 2 hours to speed up the selection process, and the wash steps were set to 30 minutes and performed at two different flow rates to mitigate binding in the high and low affinity regimes. Together, these changes significantly simplified and accelerated the selection steps (especially for RNA pools), such that RAPID was completed in one third the time as SELEX. However, despite using less optimal selection parameters, high-throughput sequencing revealed that the RAPID selections had enriched thousands of aptamers identical to those from conventional SELEX. In fact, these sequences were enriched approximately 3-fold more in RAPID, and detailed analysis of the various pools clearly showed that RAPID had consistently outperformed SELEX. This not only revealed the feasibility of

eliminating amplification steps to save time and reagents, but also brought into question the necessity for amplifying pools to maintain "competitiveness" at every cycle. These results suggest that the pool concentration may be another poorly understood quantity in solid-phase affinity selections. Ultimately, a more general method like RAPID-SELEX can allow for more efficient selections than traditional SELEX by eliminating amplifications, unless sequence copy numbers are too low or higher pool concentrations are needed to increase enrichments.

Although the primary goal of most new SELEX techniques is to reduce the time to generate aptamers, time often is not considered directly. The majority of SELEX technologies aim to simply reduce the number of selection cycles, with a reduction in time being an assumed and natural consequence. However, this can be a misguided approach to optimizing the SELEX process. In every experiment we have performed to date, time has played an unexpectedly important role in selection efficiency as oftentimes our intuition about how to manipulate time for the most efficient processing proves to be incorrect. Most importantly, with new technologies and a more explicit treatment of time, the SELEX optimization problem has become much more complex. Unfortunately, little development of new theoretical methods, empirical characterization, or modeling has been done to explain and accommodate the increasing complexity of the aptamer enrichment process within ever more sophisticated technologies. Our work represents only a small step in what we hope will be a new and continued effort to characterize selection technologies, and to use this in conjunction with methods, such as RAPID and high-throughput sequencing, to develop a more global approach to optimization.

FUTURE DIRECTIONS

Demonstrating our full capabilities

Although our microcolumns are ideal for multiplexing due to their small and modular nature, we have yet to demonstrate many of the advantages of using the microcolumns. Multiplex selections have been used to generate aptamers to multiple targets, but the use of negative and counter selections to yield a highly specific aptamer between similar targets has not yet been demonstrated. This would involve a careful selection scheme between targets, which *requires* these multi-microcolumn methods for success. Similarly, our columns could be used to isolate unique aptamers to multiple domains of a large target complex, that would involve initial selections to the complex, and end with multiplex selections with its individual domains. Experiments utilizing this configuration have yet to be demonstrated, but methods such as this could become standard for generating multiple non-interfering aptamers to a single target. Finally, using the microcolumns, selections have been performed using RAPID, which significantly reduced the time and effort needed to enrich aptamers. However, we have yet to utilize the quantitative selection scheme that would maximize the use of RAPID. This would not only reveal the full capacity for RAPID to reduce selection cost and time, but could also be used to further optimize selection protocols.

Possible modifications to our microcolumn processes

To date, our microcolumns and their associated processes have been used successfully to select for novel aptamers, and MEDUSA has demonstrated tremendous promise for increased throughput and potential automation for large scale selections. However, currently no novel large-scale multiplex or parallel selections have been demonstrated using MEDUSA, as selections of this magnitude are generally

unprecedented in most laboratory settings where the diversity of potential targets needed to fully utilize its capabilities are not available. Specific applications for using this device will need to be considered given the current environment of simple small-scale or single target selections. In addition, the full compatibility of MEDUSA's microplate layout has not been fully exploited or demonstrated. Although all test selections were performed and quantified using plate-based processes, these were still executed manually. Eventually, selections with MEDUSA should be automated to make use of the standardized fluidic manipulation and plate-based handling available with robotic workstations. This would make large-scale selections easier and more efficient for researchers to perform.

One of the advantages of the microcolumns is their capacity to be multiplexed through reconfigurable connections that facilitate serial/parallel fluid networks. However, this has always been performed manually and is likely an undesirable intervention for future and more automated versions of the microcolumns, especially MEDUSA. Although simple to execute, the multiple programed layers within the device that allow for the microcolumns' connectivity rely on an array of screws to form tight liquid seals, which must be removed when changing configurations. This process can be a nuisance if not tedious as fluidic connections need to be disconnected and reconnected as well. More elegant solutions for alternating between modes may need to be explored. Using similar programming strategies, simple valving layers could be introduced into the device which would act as basic switches between serial and parallel modes [13]. However, generalizing this onto a device would be difficult and may require customized devices, limiting its practicality.

Aside from improving the operation of the microcolumns, there are a number of additional features that have not yet been explored for the microcolumns. For example, Cell-SELEX is often considered a separate entity in aptamer selections, as

most selection methods for aptamers cannot be simply extended to cells. Instead, specific protocols have been designed to accommodate the complications of handling living cells. Initial attempts to form packed columns of cells within the microcolumns were unsuccessful as cells would either flow through the frits or form a dense plug that solutions could not flow through. Some work was done to immobilize cells onto affinity resins in order to improve flow through the column, but the immobilization proved highly inefficient and required chemical modifications to the cells surface. Future work could look into culturing cells onto microbeads as a means of immobilizing a high density of cells onto a solid column matrix. This would provide a more general mechanism of immobilization than chemical modifications, and would open up a potentially new avenue for cell selections in microcolumns that has not currently been explored.

Finally, only flow and time have been used as a means to manipulate the binding characteristics within the microcolumns. However, there exist a number of ways by which additional thermodynamic thresholds can be applied to selections. For example, sequence binding can be purposely biased through competition by forcing candidate aptamers to dissociate from other complexes [14,15]. This would only allow aptamers to be recovered that have affinities high enough to overcome the barrier of association with anchoring molecules, such as biotin-streptavidin or dsDNA to ssDNA. Sequence dissociation from targets can be purposefully biased through long washes or in the presence of competitor molecules [16]. This can also be achieved by manipulating the effective binding affinity of aptamers, which can be done by eluting bound sequences in gradients of increasing temperature or salts [17,18]. This has the effect of reducing the binding affinity causing more weakly bound sequences to dissociate earlier in the gradient. The highest affinity sequences could be recovered at the end of the gradient, either through sufficient reduction in its binding affinity, or

through an unbiased chemical elution, which allows recovery of the target along with bound sequences.

Further elucidating the SELEX process

Throughout this work, much effort has been dedicated toward optimizing and characterizing as many of the selection parameters as possible. In particular, we have found a crucial dependence of target concentration with target (or aptamer) size. However, currently, only similarly sized molecules have been tested, and a stronger case could be made by revealing similar behavior with differently sized targets and/or aptamers. Also, although it was briefly investigated in developing RAPID-SELEX, the effect of library concentration in aptamer enrichments has not been rigorously explored in experiments. Theory suggests that optimal target concentrations should vary with the library concentration and with the evolving bulk binding affinity of the pool; however (as mentioned above), our results suggest that the optimum target concentration is geometrically constrained and dominated by molecule size rather than affinity. It is currently unclear how these new factors will ultimately affect the binding dependence on the library concentration, but current results using RAPID have been surprising given the less "competitive" nature of their selections compared to traditionally executed SELEX.

In addition to concentration, the effect of library length on aptamer frequency or binding affinity to targets of various sizes has not been investigated experimentally [19]. It is conceivable that in general, small targets are more likely to have short, high affinity aptamers compared to larger targets. This, of course, is dependent on the type of target being selected against, such as DNA or RNA binding proteins that may associate strongly with a specific short sequence. However, aside from the added *potential* complexity gained by using large libraries, ultimately the number of

sequences, and thus aptamer structures, that can be probed is limited to the same number for all libraries larger than ~25 nt. Assuming there exists a highest affinity aptamer configuration of some unique sequence and length, it is an obvious statement to say that longer or shorter libraries will generate poorer affinity aptamers due to insufficient nucleotides for base pairing and other interactions, or from excess nucleotides forming structures that interfere with or abolish the ideal aptamer configuration. For example, many researchers find the need to minimize aptamers by removing excess and unneeded nucleotides through trial and error. A systematic study has yet to be done to help generate rules in choosing the ideal library length given a target's general characteristics, such as its size. This would not only help researchers to choose library lengths that might maximize the likelihood of isolating an aptamer, but also may reduce the work needed to optimize or minimize the resultant aptamer sequence.

Finally, much of the available theory on optimal SELEX overly simplifies the selection process given the array of complex and sophisticated technologies available. However, perhaps their greatest weakness lies in the number of assumptions that are necessary in order to rigorously apply the theories, as well as to evaluate their accuracy. With the invention of techniques such as HiTS-FLIP and HiTS-RAP, which couple high-throughput sequencing of pools to high-throughput affinity measurements, not only can the guess work in identifying an aptamer be completely eliminated, but detailed snapshots of the affinity profile for various enriched pools and the random library can be obtained [20,21]. This would eliminate assumptions that are made about the affinity distribution of the starting pools, and allow detailed modeling of sequence evolutions between cycles. Applying these techniques could also help address existing analytical questions about the effectiveness of the more conventional methods that we have used for identifying aptamers, such as clustering, multiplicity or

enrichments. Furthermore, coupling this with analysis of predicted secondary structures may prove quite powerful compared to sequence (or consensus sequence) populations alone.

OUTLOOK

Selection technologies have come a long way in the last two decades. However, the complex processes and lack of detailed understanding of new methods has hindered progress in the field toward the development of more robust and accessible technologies. Although new methods are constantly emerging, the field has still not matured into a carefully controlled and detailed science. Despite such little work that has been done to address these gaps, this presents a tremendous opportunity for engineers and physicist to apply their rigorous and systematic methods in order to decipher the inner workings of the selection process within ever more complex technologies. We have only begun to scratch the surface in applying these methods to a simple affinity chromatography-base technology. Similar studies could not only improve the performance of other technologies, but also help develop more integrated and automated technologies. These would be more complex and more difficult to control than other technologies, but would make selections easier and less costly to perform, assuming our understanding of the processes and performance characteristics was sufficiently detailed. Ultimately, characterization of aptamer selections is only just beginning, and will enable a revolution in future aptamer selections.

REFERENCES

1. Latulippe DR, Szeto K, Ozer A, Duarte FM, Kelly CV, et al. (2013) Multiplexed microcolumn-based process for efficient selection of RNA aptamers. Anal Chem 85: 3417-3424.

2. Gong Q, Wang JP, Ahmad KM, Csordas AT, Zhou JH, et al. (2012) Selection Strategy to Generate Aptamer Pairs that Bind to Distinct Sites on Protein Targets. Anal Chem 84: 5365-5371.

3. Szeto K, Reinholt SJ, Duarte FM, Pagano JM, Ozer A, et al. (2014) High-throughput binding characterization of RNA aptamer selections using a microplate-based multiplex microcolumn device. Anal Bioanal Chem.

4. Irvine D, Tuerk C, Gold L (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. J Mol Biol 222: 739-761.

5. Aita T, Nishigaki K, Husimi Y (2012) Theoretical consideration of selective enrichment in in vitro selection: Optimal concentration of target molecules. Math Biosci 240: 201-211.

6. Ozer A, White BS, Lis JT, Shalloway D (2013) Density-dependent cooperative non-specific binding in solid-phase SELEX affinity selection. Nucleic Acids Research 41: 7167-7175.

7. Nastasijevic B, Becker NA, Wurster SE, Maher LJ, 3rd (2008) Sequence-specific binding of DNA and RNA to immobilized Nickel ions. Biochem Biophys Res Commun 366: 420-425.

8. Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. PLoS One 5: e9169.

9. Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, et al. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. Proc Natl Acad Sci U S A 107: 15373-15378.

10. Szeto K, Latulippe DR, Ozer A, Pagano JM, White BS, et al. (2013) RAPID-SELEX for RNA Aptamers. PLoS One 8: e82667.

11. Schutze T, Wilhelm B, Greiner N, Braun H, Peter F, et al. (2011) Probing the SELEX process with next-generation sequencing. PLoS One 6: e29604.

12. Berezovski M, Musheev M, Drabovich A, Krylov SN (2006) Non-SELEX selection of aptamers. J Am Chem Soc 128: 1410-1411.

13. Lee S, Kang J, Ren S, Laurell T, Kim S, et al. (2013) A cross-contamination-free SELEX platform for a multi-target selection strategy. Biochip Journal 7: 38-45.

14. Miyachi Y, Shimizu N, Ogino C, Kondo A (2010) Selection of DNA aptamers using atomic force microscopy. Nucleic Acids Research 38.

15. Fan M, McBurnett SR, Andrews CJ, Allman AM, Bruno JG, et al. (2008) Aptamer selection express: a novel method for rapid single-step selection and sensing of aptamers. J Biomol Tech 19: 311-319.

16. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. (2010) Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS One 5: e15004.

17. Calik P, Balci O, Ozdamar TH (2010) Human growth hormone-specific aptamer identification using improved oligonucleotide ligand evolution method. Protein Expr Purif 69: 21-28.

18. Arnold S, Pampalakis G, Kantiotou K, Silva D, Cortez C, et al. (2012) One round of SELEX for the generation of DNA aptamers directed against KLK6. Biol Chem 393: 343-353.

19. Gevertz J, Gan HH, Schlick T (2005) In vitro RNA random pools are not structurally diverse: a computational analysis. Rna-a Publication of the Rna Society 11: 853-863.

20. Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, et al. (Submitted) Comprehensive analysis of RNA-protein interactions by high throughput seqencing-RNA affinity profiling. Nature Methods.

21. Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, et al. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. Nature Biotechnology 29: 659-664.