Partially Robotic Selection of Aptamers to Red Blood Cell Protein Glycophorin A

Evan Bushnik

Submitted to the Faculty of Graduate and Postdoctoral Studies in Partial Fulfillment of the Requirements for the M.Sc. Degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Acknowledgements

I would like to thank my supervisor Maxim Berezovski for all of his support, both financially and emotionally as I completed this work. His ideas and insights were always helpful as I moved through the various stages of the project. I would also like to thank all of the members of the Berezovski laboratory, as every one of them has helped me in some way through the following years. To Shahrokh Ghobadloo for teaching me the basics of aptamer selection and flow cytometry and always having the time to listen to my concerns. To Thao Nguyen for the immense emotional and scientific support throughout my degree. I would like to thank Christopher Clouthier, for his expertise on the robotic platform, the Hamilton Robotics Biolab Star, making this project possible, for his contributions in both terms of methods and ideas as the project continued such as his work on the development of the ePCR method, and for being an amazing friend. To Pavel Milman for his many hours in developing aptamers, which served as a stepping stone for my research.

Finally, I would like to thank my family, for their constant support throughout the process

Abstract

Aptamers are small DNA ligands that have been manually selected to strongly and specifically bind a target of interest. These molecules may prove superior to modern antibodies in a number of ways including price and reproducibility. One of the major advantages of using aptamers as opposed to antibodies is the relative speed of development. This, coupled with the repetitive nature of aptamer selection, means that the entire process is a possible target for automation. In the following experiments, a ssDNA aptamer is developed against the human red blood cell protein glycophorin A, partially through the novel use of a robotized benchtop. The process also utilizes an adapted protocol for emulsion PCR to further increase the efficiency of the selection process. After 11 rounds of selection, the DNA pools were sequenced leading to the generation of 14 potential aptamers. These aptamers were tested with the isolated protein and with human red blood cells resulting in several of the aptamers being deemed potential binders. Further work with these identified sequences could result in aptamers that can be reliably used to tag and delicately separate red blood cells from other cells of interest within blood, such as stem cells. The novel approaches to selection used in this work may also lead to quicker and more efficient generation of future aptamers.

Table of Contents

List of Tables

List of Figures

Abbreviations

Cy5	Cyanine 5
ddNTP	Dideoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DREME	Discriminative regular expression motif elicitation
EDTA	Ethylenediaminetetraacetic Acid
ePCR	Emulsion Polymerase Chain Reaction
FACS	Fluorescence-activated Cell Sorting
FITC	Fluorescein isothiocyanate
G-Quadruplex	Guanine quadruplex
MACS	Magnetic-activated Cell Sorting
NGS	Next Generation Sequencing
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
RBC	Red Blood Cell
RNA	Ribonucleic Acid
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
ssDNA	Single Strand Deoxyribonucleic Acid
TBE	Tris-borate-ethylenediaminetetraacetic Acid

Chapter 1: Introduction

The term “aptamer” refers to a DNA or RNA molecule which has been evolved through a rigorous selection from a random pool to specifically recognize and bind to a target of interest. (1,2) These aptamers are generally from 60-120 nucleotides long, classifying them as short molecules. (1,2) The key to their binding is the ability to fold into stabilized 3D structures which can effectively be used to label and detect a target, and even potentially deliver and separate targets. (12,13,14)

1.1 SELEX – The Aptamer Selection Process

Aptamers are created through a stringent, iterative selection process known as the Systematic Evolution of Ligands through Exponential Amplification, often shortened to SELEX. Two major research groups were responsible for the discovery of this process. The first was published by the Gold group who used SELEX, a term they coined, to develop an RNA ligand to the bacteriophage T4 DNA ligase. It should be noted that the term aptamer was not yet used. They defined SELEX similarly to how it is still used today, with repeated steps of a) Incubation with Target, b) removal of non-binding nucleotides, c) the amplification of nucleotides which bind effectively. (2) It is also key to realize, that while the selected nucleotide was an effective binder, the starting random pool was much less diverse than what is typically used in modern SELEX, as the authors only randomized an 8 nucleotide sequence of a hairpin loop structure with a previously known target. There were only 65,536 sequences in the starting pool, which is many magnitudes fewer than most later aptamer selections. (3) Later in the same year, another group, the Szostak lab, performed another selection, this time finally using

the term aptamer which they had created. This selection was closer to current SELEX procedures with a starting pool of around 10^{13} sequences. (1)

Modern SELEX usually utilizes a starting pool of around $10^{15} – 10^{20}$ molecules. The actual SELEX method can vary, particularly in the method of target presentation, with methods using whole cells (4,5) , membrane affinity columns (6), and targets immobilized on beads (7,8,9), among others. There are pros and cons for each method. For the purpose of this work, we will focus on bead immobilized targets, the method used in this research.

1.1.1 Magnetic Bead-Immobilized SELEX

Bead immobilized targets present the most controlled environment, with the lowest amount of possible side targets. This is because that apart from the bead, which negative selection is conducted against, there are no possible binding targets apart from the target of interest. This method, however, lacks physiological context, and therefore runs the risk of selected aptamers losing their binding ability when presented with in-vitro conditions. (10,11) The bead method generally follows the process of attaching the target of interest to magnetic beads. Following this, the beads are incubated with the DNA or RNA pool. The beads are washed to remove non or weak binding molecules. Heat, or another process that interferes with the aptamer structure, is introduced causing the binding molecules to detach. The beads are removed from the binding pool using a magnet. The binding pool can be amplified through PCR and reintroduced to the target. (7,8,9) This process is commonly repeated for anywhere from 6-15 times (each repeat being defined as a round of SELEX), although shorter methods requiring fewer

rounds are currently being developed. (15). Figure 1.1 shows an example of a simple magnetic bead SELEX procedure.

1.1.2 Negative Selection

There is a flaw in the basic SELEX procedure, in that it selects for molecules that bind strongly, but not necessarily molecules that bind specifically. That is to say, while the strongest binders may be chosen through the SELEX process, they may not only bind to the target of interest, but merely be attracted to anything available for binding. To counter this phenomenon, most modern SELEX procedures involve steps of negative selection, originally referred to as subtractive selection. (44) In negative selection, the binding pool of molecules is incubated with an undesirable target instead of the target of interest. The molecules that are washed away, and therefore not good binders, are saved for the next round of selection. Conducting a round of negative selection per every few rounds of regular selection is an effective way to ensure that the evolving pool continues to bind only to the desired target. (45)

1.1.3 Issues with the Current SELEX Process

One issue identified in the SELEX process is an incompatibility of basic PCR with an aptamer pool (34). In the field of aptamer amplification, it has been found that within five cycles of SELEX, by products and artifacts begin forming and within another five cycles they are even more prevalent. (35) In Chapter 3 I will demonstrate that these artifacts can cause issues even sooner within the selection process. The main

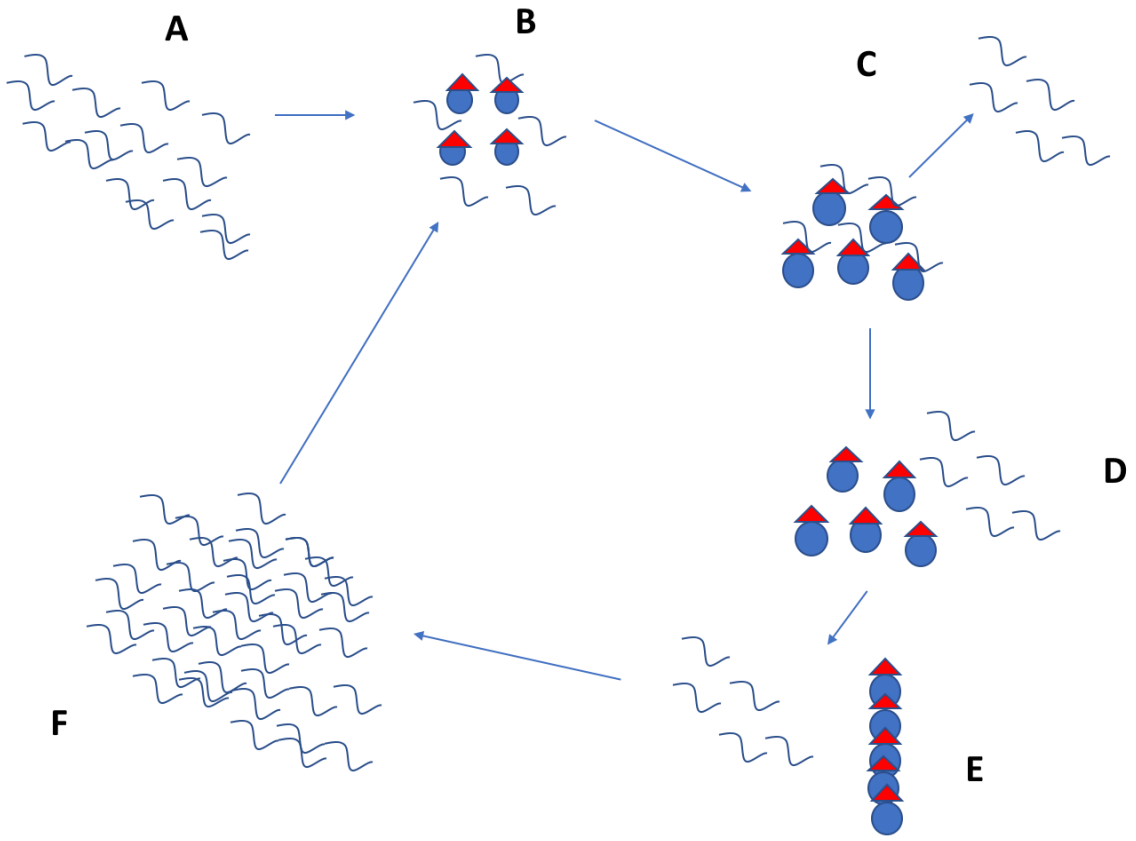


Figure 1.1: An example of a simple bead SELEX procedure. The First Round of SELEX starts with (A) a library of random DNA, usually in the 10^{15} – 10^{20} molecule range. In (B) the random library is incubated with beads that have been attached to the target molecule of interest. The incubation time varies, but usually decreases with each round of SELEX. In (C) the beads are washed repeatedly and stringently to remove any DNA which does not bind strongly or at all to the target of interest. In (D) the DNA which has withstood washing is removed from the target through the introduction of heat which unfolds the DNA. In (E) the beads are removed from the solution by the introduction of a strong magnet which pulls them to the side of the container. Finally in (F) the DNA that was bound to the target is amplified exponentially by PCR. This amplified material can be reintroduced to the target on beads (B) in the next round of SELEX. This process is repeated 6-20 rounds depending on the specific selection.

cause behind the buildup of PCR artifacts is that aptamers are by definition a DNA strand with complex G and C nucleotide rich folding structure (12.13.14) In contrast, PCR most effectively and preferentially replicates DNA with a straight, A and T rich shape. (35) Due to this, if after a round of selection the actual aptamers comprise of most of the pool, after PCR amplification it is likely that the background artifacts will be overly represented. This process serves to effectively reverse the selection round and can often stop the selection process. This occurs when the aptamers are represented so poorly after PCR, that the pool then fails to bind to the target in the next round. (34) To get around this issue, a new way of doing PCR must be developed.

1.2 Emulsion PCR – A New Way of Amplifying SELEX Product

Emulsion PCR is an alternative method of PCR where the end result more accurately reflects the contents of the starting pool. (34,37) While the basic principles of PCR are the same, the major difference is that every DNA strand that is to be amplified is separately encased in an oil bubble. Because each DNA strand is encased in its own environment separate of other DNA strands, the innate bias of PCR described before cannot affect which strands are amplified. Due to the goal of having no more than one strand of DNA per oil bubble, it was found that ePCR works best with a low template concentration. (34,37) This means that more cycles of PCR must be conducted to receive the same final product. This was found to not be a problem, as using ePCR, more cycles could be conducted with less by-product. Classic PCR was found to have by-product building up at 14 cycles of PCR, while ePCR was able to run for 40 cycles with far less by-product. (34) It is possible that selections with fewer rounds are possible using ePCR, as classic SELEX struggles with reversal in selection progress when PCR

is run. With these positive effects and insignificant negatives, it can be concluded that emulsion PCR stands to improve the aptamer process. Figure 1.2 is a representation of the ssDNA captured in emulsions during emulsion PCR.

1.3 Partial Automation of Aptamer Selection

A defining characteristic of SELEX is that it involves the repetition of the same processes. (36) It is then possible that after optimization is complete, aptamer selection is not a job that is restricted to research scientists. It has therefore been suggested by ours and other research groups that the SELEX process could be performed by a machine. While a chip-based aptamer selection process has been experimented with, (36) our lab strives to automate the process in a different method. The idea is to run the bulk of SELEX robotically, similar to how it is run now, but through an automated lab bench setup. The emulsion PCR steps would still need to be manually conducted, due to the danger of the emulsion breaking ethanol with the lack of ventilation on the robot. Additionally, there would still need to be a manual gel analysis and by-product removal at the end of each round before next incubation of product with the target. This is because the available robotic setup does not have gel running capabilities. As ePCR based SELEX

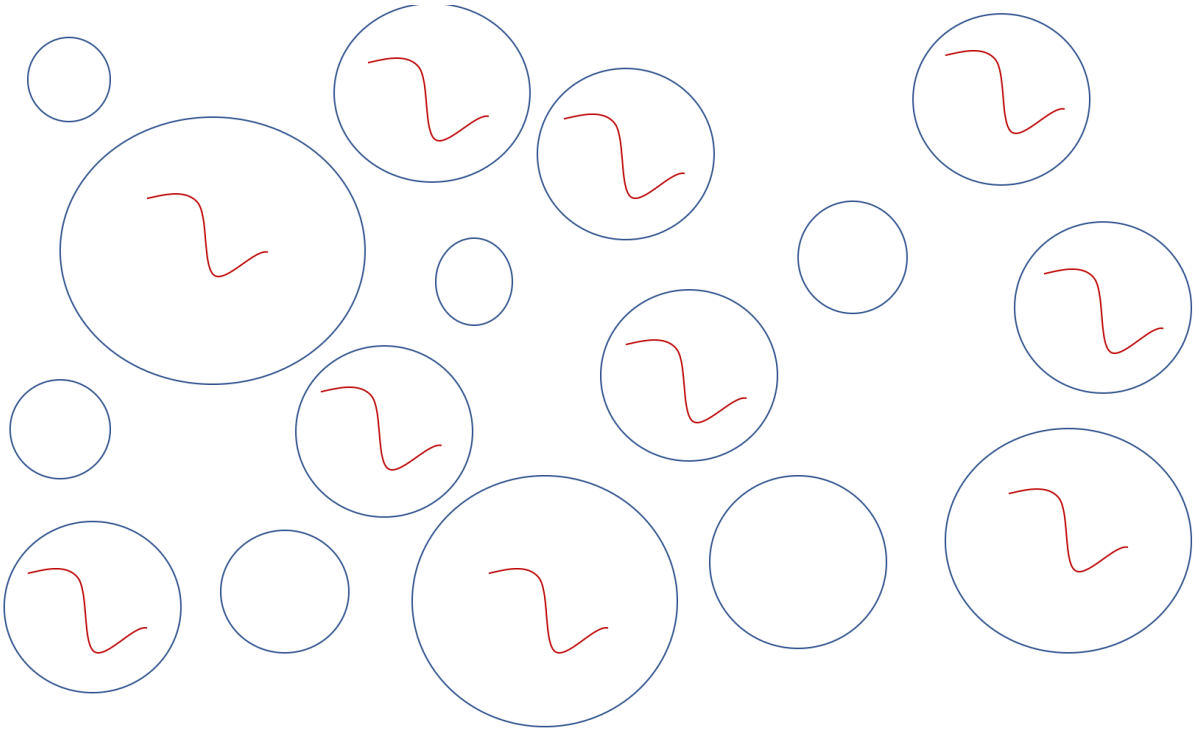


Figure 1.2: A typical view of the ssDNA after it has been captured in micelles. Every red strand represents a single strand of ssDNA from the pool to be amplified. The phase inside the droplets is the aqueous phase which is pre-mixed with all of the necessary PCR reagents. The majority phase is mineral oil. Every micelle does not contain ssDNA, but the important factor is that each emulsion contains no more than one or two pieces of ssDNA. This creates a locked environment for PCR of every individual strand, so that PCR bias is not allowed to influence the end pool.

results in less by-product, (34,37) it may also be possible to run multiple rounds without manually purifying the result of by-product in the future.

1.4 Sequencing of SELEX Product

After the SELEX procedure, the end result is a pool of sequences that are highly specific to the desired target, having lasted multiple rounds of selection. It is important to identify the specific sequences to be able to reproduce any discovered aptamers. One cannot sequence a random strand of DNA from the pool and assume this will be an efficient aptamer. Aptamer pools are greatly heterogeneous, even after a selection. (33) In addition, the most common sequences in the pool may not be the best aptamers, which is due to the difficulty in replicating DNA with high GC content. (34) It is therefore necessary to use a method that will give the most whole and accurate picture of the entire final DNA pool.

1.4.1 Sangar Sequencing and Molecular Cloning of Aptamers

Early aptamer work by the Szostak lab utilized molecular cloning and Sanger sequencing. (1) Both of these methods, while the best available options in the 1990s, are labor intensive and low throughput. (14,38,39)

Bacterial transformation is a process made standard by Cohen et al in the early 1970s (29). In this process, DNA of interest is introduced into a circularized vector, within a reporter gene such as LacZ. These vectors are up taken into bacterial cells, which replicate themselves with the DNA desired. When the bacteria is plated on antibiotic fortified agar, the presence of the reporter gene allows bacteria that have taken in the DNA to be identified. In this example, LacZ would cause the bacteria to produce a blue pigment, however, if the foreign, synthetic DNA sequence is incorporated into said Lac

Z gene, it will not function, resulting in a white colony. (30) In a span of approximately 12 hours it can convert nanogram quantities of DNA into microgram amounts after purification. This method of replication is still used to this day, as amplifying the DNA within a bacterial cell will take advantage of the bacteria's natural proof reading measures and repair exonucleases. This is in contrast to PCR, which generally uses the error prone Taq polymerase and lacks any repair mechanism for improperly incorporated nucleotides. (31)

Sangar sequencing involves the amplification of DNA with a master mix that has been spiked with chain terminating ddNTPs. These are essentially the same as naturally occurring dNTPs, with the exception of a missing 3' OH group. This results in a complete stop to any further extension of the complementary strand. (32) Due to the frequent and completely random incorporation of these ddNTPs, the resulting amplified DNA pool is mixed with every possible combination of premature termination. Gel electrophoresis is used to order the DNA by size and (therefore termination order) and by differentially labeling the four different ddNTPs, the exact sequence of the DNA can be determined. (28)

While these two methods are cheap and effective, these methods can be used to capture a randomly selected portion of a massive DNA population. This portion is then assumed to represent the most abundant sequences. (1) The obvious problem with this, is that the best aptamer may not be the most abundant, particularly with the difficulty in amplifying DNA with high levels of GC content. Even still, the random nature of selection does not ensure an accurate view of the aptamer population. The solution to this is to move on to different technologies.

1.4.2 Next Generation Sequencing of Aptamers

In recent years, aptamer research has relied on high throughput sequencing, otherwise known as next generation sequencing. This technology allows for the identification and sequencing of every (or close to every) sequence in the pool, allowing for more accurate and thorough analysis. (16) In addition, this method can be conducted at a fraction of the time and cost of conventional Sanger sequencing. While there are several platforms, the one currently owned by Illumina is the method utilized in these experiments. The process begins with library preparation in which the DNA of interest is fragmented, and specialized adapters are ligated to both ends of each strand. Next the library is loaded onto a flow cell and each fragment is hybridized to the flow cell surface. Through bridge amplification, the bound fragments are amplified into clusters of identical sequences, resulting in clusters for each individual sequence in the pool. Next in the sequencing step, fluorescently labelled nucleotides, each nucleotide labelled with a unique fluorescent dye, and other sequencing reagents are added and the first base on the sequence is incorporated and imaged. The cycle is repeated for each nucleotide in the pool, by cleaving the dye and terminating groups and washing these groups away. The next dyed nucleotide is incorporated into the sample, and the cycle is continued for each nucleotide in the DNA strand. This allows for an exact sequence to be recorded for every cluster, with each cluster corresponding to one DNA sequence in the original pool. (27) Overall this process is said to outperform Sanger sequencing by a factor of 100 to 1000. (28) Figure 1.3 is an overview of the Illumina sequencing method.

1.4.3 Analysis of Sequence Results

While one might assume that any sequenced results would be viable aptamers, it is important to note that PCR artifacts may still represent a large portion of the population. To identify the sequences that are desired a number of factors are analyzed including copy number, enrichment fold, and analyzing repeating motifs. (17,18,19) Copy Number is a logical factor to assess, being that the more a sequence repeats in a binding pool, the more likely it is to be a factor in the binding. (18) Similarly, enrichment fold allows one to analyze how much more a sequence appears than would be through random chance. (19). Finally, by analyzing repeating motifs in the binding pool, one can determine which motifs appear to be common in sequences that bind to the target. This allows the analyzer to gauge whether or not a sequence is a binding sequence based on the appearance of certain clusters of nucleotides. (17). In addition to these methods, basic screening such as looking for repeated strings of primers and more complex analysis such as analyzing the shape of the predicted structure of the sequence can be used. There are also helpful programs for analyzing the number of sequences that NGS produces. Alignment software such as Clustal Omega can allow one to compare homology of sequences by aligning them based on similar sequences. DNA and RNA folding software, such as that available on IDT called OligoAnalyzer 3.1 (49), allow a researcher to input a sequence and analyze the predicted structures. Specifically, for aptamers, a program called QGRS Mapper (48) is available which

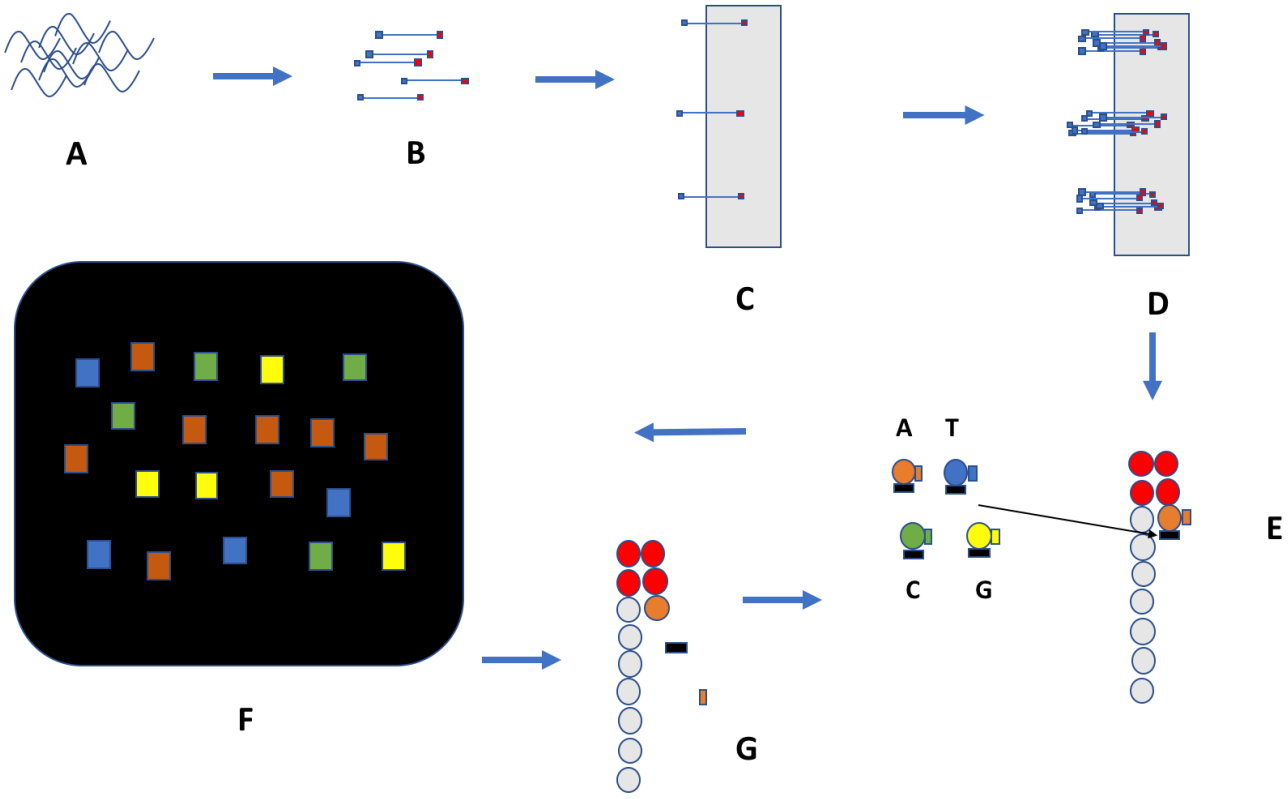


Figure 1.3: A general overview of the Illumina sequencing method. The process starts with (A) a single strand DNA pool for which the sequence are desired. In (B) specialized linkers are added to both ends of every strand in the pool. This allows for (C) where one end of the DNA is hybridized to the flow cell wall, resulting in DNA strands standing upright on the flow cell surface. In (D) every single DNA strand is replicated into a cluster of DNA strands, in a process called bridge amplification. In (E) we can see a zoomed in view of one single strand DNA. Fluorescently labelled nucleotides with a termination sequence which stops further nucleotide addition are added to the DNA. This allows for a single fluorescent nucleotide corresponding to the first nucleotide in the ssDNA to attach. In (F) the flow cells are fluorescently imaged, with each cluster fluorescing in the colour of the last attached nucleotide. The flow cells are washed in (G) resulting in the fluorescent labels and termination sequences detaching. The DNA again goes to (E) where the next fluorescent nucleotide is added. This process is repeated until the entire DNA sequence is recorded.

analyses a sequence and predicts the number of G-quadruplexes within the 3D structure.

1.4.4 Structural Indication of Useful Aptamers – G quadruplex

Another method of sorting out potentially good aptamers from sequencing results is to look for certain structures formed by the aptamers that are common among sequences that bind strongly or have other beneficial qualities. An example of such a structure is a G-quadruplex. A G-quadruplex is a helix structure formed by multiple tetrads of guanine (G) nucleotides. (41) The overall structure is stabilized by the presence of a cation in the centre column of the molecule. G-quadruplex containing aptamers have been studied and shown to contain several advantages over aptamers not containing these structures. These include an enhanced cellular uptake, higher chemical and thermodynamic stability, and such aptamers are non-immunogenic. (40) It has therefore been suggested that G-quadruplex aptamers are more suitable than standard aptamers for biochemistry and medicinal uses. Figure 1.4 shows the structure of a typical G-quadruplex.

1.5 Stem Cells and the Desire to Isolate Them

Recently emerging research and therapy using stem cells have led to a requirement of a high number of such cells with a high purity. Currently the most powerful methods are affinity-based methods such as FACS (Fluorescence-Activated Cell Sorting) and MACS (Magnetic-Activated Cell Sorting). (20) These methods use antibodies which are specific to stem cells, with a tag that can be recognized and picked out by a machine. In FACS, antibodies are tagged with fluorescent markers, and sorted based on light-scattering. This method is 95% effective, however, it is extremely expensive and

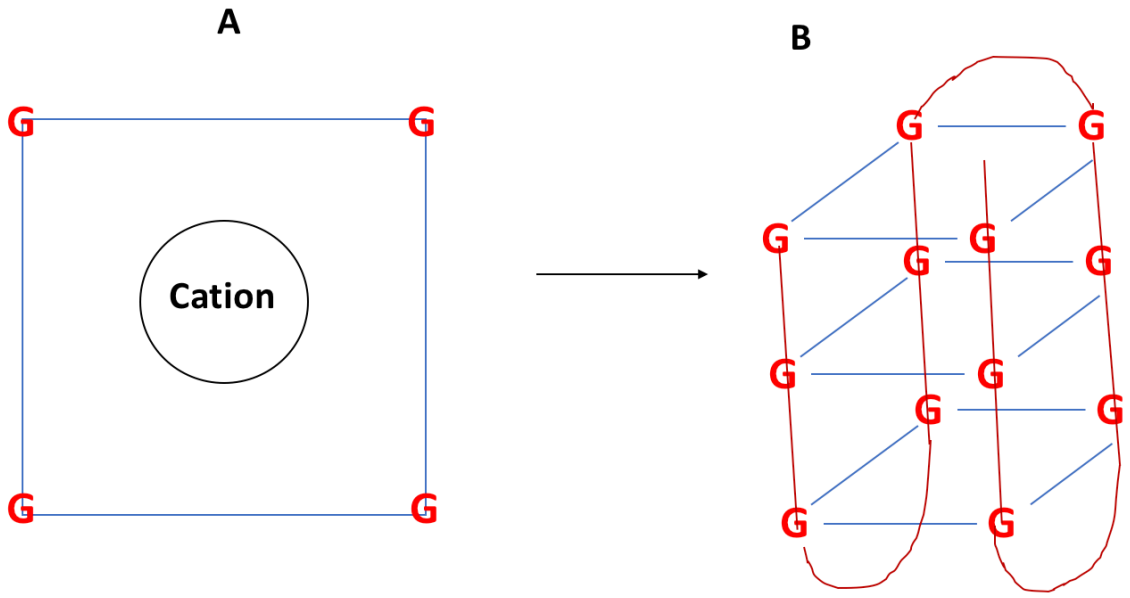


Figure 1.4: The structure of a typical G-quadruplex. In (A) is pictured a typical G-tetrad structure. Multiple G-tetrads come together to form a G-quadruplex as shown in (B). (58)

relatively slow (10^7 cells per hour) (21) Alternatively, MACS separates using antibodies attached to magnetically coated beads and a magnetic sorter to separate the stem cells. This process is faster than FACS at 10^{11} cells per hour, however, achieves only 75% purity. (22) Of importance to note, there is concern over how to remove the magnetic beads after sorting, as the stem cells are delicate to basic methods usually used to remove target from the bead, such as heat. (22) As both of these methods have major flaws, a future alternative is required if the demand for stem cells is to possibly be met. (20)

1.5.1 Potential Use of Aptamers to Separate Stem Cells

SELEX generated aptamers are a possible alternative to the previously established FACS and MACS methods. (20) The first major advantage of using aptamers in this circumstance is that the structure of the target does not need to be known as the aptamer is evolved through direct incubation. This is useful as the surface markers of many stem cells are not well mapped (23). In addition, aptamers are generally much cheaper to produce than antibodies, are much longer lasting, have higher reproducibility, and can attach to non-antigenic targets. (24) There are still problems with this method as the SELEX method so far has proven insufficient in separating stem cells in high purity, and the problem of removing the magnetic beads is not addressed. A potential work-around to the current problems with aptamer-based stem cell separation may exist. In theory, this would be to use the aptamers not to separate the stem cells directly, but to remove the major impurities in stem cell separation. This could be used in combination with MACS methods to improve purity or by itself. This would

also solve the problem with removing the magnetic beads, as the stem cells would not be directly attached to said beads.

1.6 The Human Red Blood Cell as an Aptamer Target

There are multiple targets one could create aptamers against for purifying stem cell samples, but the most abundant cell in blood is the RBCs (Red Blood Cells), so this is a large potential impurity. (25) Red Blood Cells when attached to a bead-bound aptamer, could be easily removed from a sample without the use of force which could damage or alter the stem cell by using a strong magnet. Whole cell SELEX, however, is inefficient due to the inability to control which target the aptamers bind. (43) When multiple aptamers are selected after sequencing, they will most often be aptamers to different targets and be difficult to compare. Additionally mass spectrometry must be conducted after an aptamer is selected to determine which target on the cell bound the aptamer. It is much simpler to pre-determine a target on the surface of the RBC, and select an aptamer to that target.

1.6.1 Human Glycophorin A as an Aptamer Target

The protein Glycophorin A, also referred to as GYPA, is an abundant protein on the surface of the red blood cell with over 1 million GYPA molecules on the surface of one RBC. (42) This protein along with Glycophorin B, bears the antigens of the Ss and MN blood type groups. As theoretically only one aptamer would need to attach to the RBC to identify and remove it from a blood sample, and each RBC having over 1 million of these surface proteins, it would be unlikely any RBCs would avoid aptamer tracking. Overall, these properties make it a potential target for the development of a RBC aptamer. As long as a blank GYPA protein is used, without the various antigens

attached, an aptamer to GYPA should bind to RBCs of any blood type, with high efficiency.

In addition, this protein is critical for identification of the RBC by the malaria parasite *Plasmodium falciparum*. (26) This secondary function results in more possible usages for an aptamer to this target even beyond cell separation, as a cell with the Glycophorin A proteins blocked with aptamers may not be recognized by the malaria parasite. Figure 1.5 is a PyMOL rendering of the Human GYPA molecule.

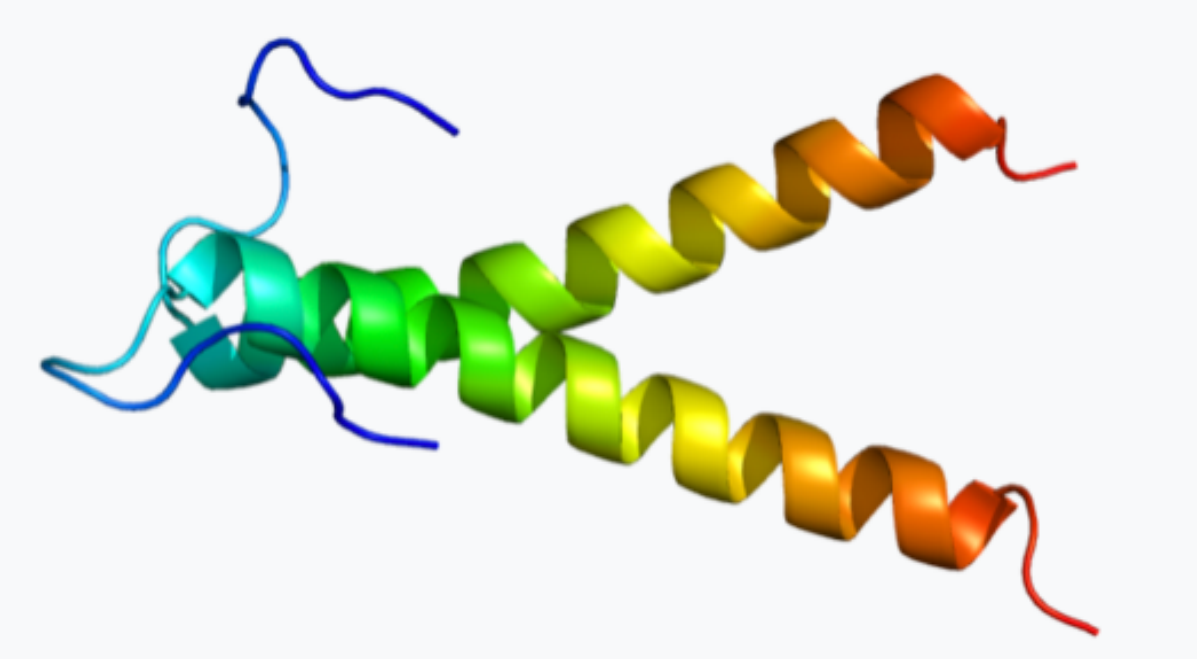


Figure 1.5: PyMOL ribbon diagram model of the Human GYPA molecule. The Blue regions represent the portion of the protein on the surface of the cell, where the antigens attach. The green and yellow regions are the transmembrane portions of the protein. The orange and red regions are the portions of the protein inside the cell. (59)

1.7 Thesis Overview

This thesis is divided into several chapters. Chapter 2 overviews all of the materials and methods used in the research contained in the subsequent chapters. Chapter 3 details the isolation of the Glycophorin A protein, adaptation and creation of an ePCR protocol, and the partially robotic selection of the DNA strands bound to the target. Chapter 4 presents the next generation sequencing results, and the predicted shapes and motifs of the best potential aptamers. Chapter 5 provides tests of the chosen sequences against the protein attached to beads, and some initial results of the sequences against whole RBCs. Finally, there is a section containing general conclusions as well as future work to be done on this project.

Chapter 2: Materials and Methods

2.1 Aptamer Selection

2.1.1 Protein and Bead Preparation

Lyophilized His-tagged Glycophorin A was purchased from Sino Biological. Pure Glycophorin A was added to PBS for a final concentration 250ug/mL. Dynabeads™ His-Tag Isolation and Pulldown magnetic beads were purchased from Thermo Fischer Scientific. Initial bead concentration is approximately $1.44x10^9$ beads per uL with a bead radius of 1um. $10x10^9$ beads were extracted from the initial vial and placed in a 1.5mL tube. The beads were washed on a magnet twice with PBS + 0.01% triton. The addition of the triton to the PBS to prevent bead clumping. The beads were resuspended in PBS + Calcium and Magnesium with a final concentration of $5x10^8$ beads/uL.

2.1.2 Attaching Glycophorin A to Magnetic Beads

Initially it was unsure if 5ug,10ug or 20ug of the selected protein should be added to $35x10^8$ beads for full saturation. Tests proved full saturation was achieved after 10ug was used, so this was the value used for all future bead preparations. 10ug of Glycophorin A was added to PBS with $3.5x10^9$ beads. The solution was incubated in an Eppendorf Thermomixer incubator and shaker at 800rpm at room temperature. Following incubation, the beads were washed four times with PBS + 0.01% triton on a magnet to remove any unbound protein. The protein and bead solution was resuspended in 200uL of PBS + 0.01% triton. The final concentration was $1.75x10^7$

beads/uL. Assuming full saturation, there would be 12.5 pmole of Glycophorin A per uL of bead and protein solution.

2.1.3 Antibodies to Glycophorin A on Beads to Test for Protein Saturation

To test for protein presence, FITC labelled anti-human Glycophorin A antibody clone HI264 was ordered from Biolegend. The initial concentration was 0.5 mg/mL. Dilutions of 1:50 (0.01mg/mL) 1:100 (0.005mg/mL), 1:200 (0.0025mg/mL), and 1:500 (0.001mg/mL) were prepared by mixing with PBS + Calcium and Magnesium. 5uL of each dilution was added to $8.75x10^7$ beads and incubated for 15 minutes. The solutions were washed once with PBS and analyzed by a Gallios flow cytometer (Beckman Coulter) to read FITC fluorescence for 20,000 events at excitation of 490nm.

It was prudent to determine how much protein needed to be added to saturate the beads. For each of three initial protein values (5ug, 10ug and 20ug), $8.75x10^7$ beads were added to 0.015ug of the selected antibody. The solution was incubated for 30 minutes at room temperature, washed once with PBS + Calcium and Magnesium. The concentration was set to $1.75x10^5$ beads/uL using PBS + Calcium and Magnesium. The solutions were analyzed by a Gallios flow cytometer by Beckman Coulter to read FITC fluorescence for each of the three protein concentrations where 20,000 events were read at excitation of 490nm.

2.1.4 ssDNA Library Composition

For the purpose of SELEX, a random ssDNA library was purchased from Integrated DNA Technologies with a length of 80 nucleotides and a complexity of 10^{15} sequences. The library consists of a forward and reverse primer region of 20 nucleotides each, with

a 40 nucleotide random sequence in the centre. The sequences were also tagged with Cy5 fluorescence. The sequence for the library was:

5'-/5Cy5/CTC CTC TGA CTG TAA CCA CG(N1)[(N1) x 39] GCA TAG GTA GTC CAG AAG GC-3'

2.1.5 Selection Round Zero

1uM of ssDNA library was heated to 95 degrees Celsius for 5 minutes to unfold any secondary DNA structure using an Eppendorf Thermomixer incubator and shaker. The ssDNA library was snap cooled to 4 degrees by placing the samples in a freezer. 1.44x 10^{10} beads without protein were incubated with the 1uM library for 30 minutes at room temperature. A strong magnet was introduced, and the beads were removed. The remaining library which did not bind to the plain beads were carried forward as the starting library for round 1.

2.1.6 Positive Selection – Robotic Portion

The selection pool comes from the PCR product of the previous round for every round except for round 1, where the selection pool was the final pool of the Selection Round Zero. 50uL of the selection pool was loaded onto the deck of the Hamilton Robotics Biolab Star. It should be noted that after each round there is approximately 200uL of generated ssDNA product, but only 50uL is used for the next round. The remaining sample is frozen at -80 degrees Celsius as a physical record of each round. The sample was loaded into a deep well 96 well plate. It was first heated to 95 degrees Celsius for 5 minutes, and then immediately snap cooled to 4 degrees Celsius to ensure the pool is in an open configuration. $1.75x10^{8}$ beads with Glycophorin A protein in PBS+ Calcium and Magnesium were pipetted by the robot into the ssDNA pool. This number of beads

ensures an approximate protein amount of 150 pmoles. The solution was incubated at room temperature for 45 minutes while shaking at 800 rpm. After 45 minutes a strong magnet was introduced, and the beads were removed from solution. The beads were washed in PBS + Calcium and Magnesium for 5 minutes at 800 rpm. This wash was repeated 2 times. For every two successive round, an additional wash repetition was added. The beads were resuspended in PBS+ Calcium and Magnesium and heated to 95 degrees Celsius to remove the bound ssDNA from the beads. The beads were finally removed from solution using a strong magnet. The final solution of ssDNA was transferred to a 1.5 mL Eppendorf tube and ejected from the Hamilton Robotics Biolab Star Robotic setup for manual processing.

2.1.7 Positive Selection – Emulsion PCR

The pool of ssDNA generated from the robotic portion of the SELEX process was amplified by emulsion PCR to reach concentrations high enough for the next round of SELEX. The volumes used in the following explanation are extremely important. Attempts were made to use different volumes with the same concentration ratios, and these results all failed. Because of this sensitivity, I have stated exact volumes used instead of concentrations.

First 10% stocks of Span 80, Tween 80, and Triton x-100 were prepared using mineral oil as a base. The oil used for the emulsion creation was prepared using 50.5 % Mineral Oil, 45% of 10% Span 80, 4% of 10% Tween 80, and 0.5% of 10% Triton x-100. 200uL of this oil in a 1.5mL Eppendorf tube was used per PCR reaction tube, with a total of 4 of these tubes prepared for one round.

10uL of DNA template (the result of the robotic selection) was added to 90uL of ice cold Phire Green Polymerase Master Mix. The use of Phire Green Polymerase is important, as the green loading dye is critical in distinguishing the two layers (oil and aqueous) later in the process. The Master Mix was created as follows: 7% DMSO, 1x Phire Green Buffer, 200uM dNTP, 0.4uM of Cy-5 labelled forward primer (5'-/5Cy5/CTC CTC TGA CTG TAA CCA CG-3') , 0.4uM reverse primer (5'-CG TAT CCA TCA GGT CTT CGG-3'), and 0.02U/uL of Hot Start Phire II DNA Polymerase (Thermo Scientific). It was important that this 100uL sample is kept on ice at all times when not being pipetted.

Of the 100uL of mixed template and PCR master mix, 10uL of this solution was added to the 200uL of oil in 1.5mL Eppendorf tube previously prepared. It was vortexed and another 10uL of the master mix solution was added every ten seconds while vortexing constantly. Vortexing was continued at maximum speed for an additional 5 minutes after all of the sample has been loaded into the oil.

Each of the 4 oil tubes created were separated into 6x50uL PCR tubes. Each tube was loaded into a thermocycler and run on the following optimized program: initial denaturing for 30 seconds at 98°C, followed by 40 cycles of denaturing at 98 degrees Celsius for 10 seconds + annealing for 15 seconds at 58 degrees Celsius + extending for 5 seconds at 72°C, final extension at 72°C for 20 seconds, and ending with cooling to 4°C. Each of the four sets of six tubes were recombined into four separate 1.5 mL Eppendorf tubes.

Each of the four tubes were centrifuged at 9000g for five minutes. There were two distinct layers. As much of the top oil layer as could be removed without disturbing the bottom layer was removed and discarded. 500uL of ether was added to each tube and

briefly vortexed, resulting in all of the emulsion breaking. The aqueous layer was easily distinguishable due to the Phire Green buffer. As much of the oil and ether was removed without disturbing the aqueous layer as was possible. The samples were placed in a fume hood incubating at 50 degrees Celsius using an Eppendorf Thermomixer incubator and shaker. Drying continued for approximately 30 minutes until all of the ether layer had evaporated. The remaining 100uL of sample was the double stranded PCR product and was highly concentrated. (approximately 200nM)

2.1.8 Regular Symmetric PCR – Used for Comparison

Regular PCR was only conducted in early attempts at SELEX but results are shown to compare to the ePCR.

5uL of DNA template was added to 45uL of Phire Green Polymerase Master Mix. The Master Mix was created as follows: 7% DMSO, 1x Phire Green Buffer, 200uM dNTP, 0.4uM of Cy-5 labelled forward primer (5'-/5Cy5/CTC CTC TGA CTG TAA CCA CG-3') , 0.4uM reverse primer (5'-CG TAT CCA TCA GGT CTT CGG-3'), and 0.02U/uL of Hot Start Phire II DNA Polymerase (Thermo Scientific). Five reactions were conducted at once. Each tube was loaded into a thermocycler and run on the following optimized program: initial denaturing for 30 seconds at 98°C, followed by 40 cycles of denaturing at 98°C for 10 seconds + annealing for 15 seconds at 58°C + extending for 5 seconds at 72°C, final extension at 72°C for 20 seconds, and ending with cooling to 4°C. The end result was the concentrated double strand PCR product.

2.1.9 Positive Selection - Digestion of PCR Product

As symmetric PCR was used, the end product is double stranded DNA. As the aptamers only have their 3D shape and function while single stranded, the double

stranded product had to be digested. 1U of lambda exonuclease enzyme (Thermo Scientific) in 1X lambda exonuclease buffer (Thermo Scientific) was added per 50uL of undigested PCR product. This mixture was incubated for 2 hours at 37 degrees Celsius using an Eppendorf Thermomixer incubator and shaker. At the end of 2 hours, the solution was heated to 80 degrees Celsius for 5 minutes to denature the enzyme.

2.1.10 Positive Selection – Gel Clean Up

Every round a manual clean-up of the digested product to remove impurities was conducted. The entire digested product was mixed with 1x BlueJuice gel loading dye (Thermo Scientific) and loaded onto a 2% agarose gel (Ultrapure Agarose by Invitrogen) with 0.5X TBE as the base and containing 1X GelRed (Biotium). The original starting library (at a concentration of approximately 100nM) was loaded into the gel in a separate lane to have an easy reference for the appropriate location of the desired product. The gel was run at 150V for 45 minutes. The gel was visualized for Cy-5 fluorescence by a Fluorochem Q gel imager (Alpha Innotech). After images were taken and filed, the band corresponding with 80 nucleotides (matching the 100nM library) was excised using a scalpel. The gel pieces containing the ssDNA of interest were chopped up into tiny bits. The gel pieces were weighed using an analytical balance and for every 100 mg of the gel, 200 uL of NTC binding buffer (Clontech) was added and the gel-buffer sample was incubated at 50 degrees Celsius using a hot water bath for 30 minutes or until the gel pieces were completely dissolved. The ssDNA was removed from the liquid agarose by using Nucleospin Column Clean-Up and Gel Extraction Kit (Clontech) as per the manufacturer's protocol. It should be noted that as the ssDNA of interest had a unique 3D structure it often clung to the clean-up column instead of

eluting in the final step. To ensure maximum elution, 50µL of PBS with calcium and magnesium was added to the column and incubated at 50°C in a hot water bath for 10 minutes, and centrifuged at 11000g for 1 minute. This was repeated 4 times, resulting in approximately 200uL of PBS with concentrated ssDNA.

2.1.11 Negative Selection

For every two rounds of SELEX, a round of negative selection was conducted using plain beads without protein as a target. This negative selection was conducted after round 2, 4, 6, 8, and 10. The ssDNA pool acquired from the gel clean-up was heated to 95 degrees Celsius for 5 minutes to unfold the DNA structure using an Eppendorf Thermomixer incubator and shaker. The ssDNA library was snap cooled to 4 degrees Celsius by placing the samples in a freezer. 1.44×10^{10} beads without protein were then incubated with the 1uM library for 30 minutes at room temperature. A strong magnet was introduced, and the beads were removed. The remaining library which did not bind to the plain beads was carried forward to the next round of SELEX.

2.2 Flow Cytometry Analysis

2.2.1 Preparing Aptamer Pools for Flow Cytometry Analysis

To check the aptamer pools under flow cytometry, a portion of every round (1-11) was amplified using emulsion PCR. The pools were amplified, digested, and purified using the technique previously stated. It should be noted that the samples amplified were already prepared for use in the subsequent round of SELEX, so this step is an additional amplification to maximize the concentration. The ePCR product from each round was loaded into a 96 well plate and analyzed by a plate reader (Biotek Cytation 3). The cy5 fluorescence was measured and compared to a previously prepared

standard curve to calculate concentration. The samples from all 11 rounds were normalized to 200nM for flow cytometry analysis.

2.2.2 Flow Cytometry Analysis of Selection Pools

A Gallios flow cytometer (Beckman Coulter) was used to analyze the binding of the pools of round 6 onwards to protein attached beads. Data for 20,000 events were recorded for all 6 rounds, starting library, plain beads, and a sample with FITC labelled anti-human Glycophorin A antibody to verify protein presence. All samples had first been incubated with $1.75x10^8$ beads for 45 minutes, and washed 3 times with PBS plus Calcium and Magnesium on a magnet. All samples were read using a 633nm laser excitation for cy5 fluorescence, except for the antibody sample which was excited at 490nm for FITC. The data was analyzed using the Kaluza Gallios program (Beckman Coulter) with median intensity values of each sample used for direct comparison of cy5 fluorescence intensity.

2.3 Barcoding Samples and Sequencing

To sequence the selected rounds all at once, the individual pools had to be barcoded. A sample of each round was amplified by ePCR, digested and purified as per the protocols listed previously. The only difference in protocol was the forward primer used when creating the PCR master mix. The barcoding forward primers used for each round were as follows in Table 2.1.

Table 2.1: Sequenced SELEX round pools and associated barcoding forward primers used.

Round of Selection	Barcoding Primer Used
Round 6	FS20: 5'-GTACTGACCTCCTCTGACTGTAACCACG-3'
Round 7	FS21: 5'-CTGACATACTCCTCTGACTGTAACCACG-3'
Round 8	FS22: 5'-ACTGATCACTCCTCTGACTGTAACCACG-3'
Round 11	FS23: 5'-CAGTACGTCTCCTCTGACTGTAACCACG-3'

A total amount of 300ng was needed and so each round was diluted to an appropriate amount. Each round was combined together and with other samples and sequenced in a single lane of Illumina Mi-Seq by StemCore Laboratories at the Ottawa Hospital Research Institute.

2.3.1 Analysis of NGS Data

The sequencing data was received by email as a compiled fastaq file. The files were converted to fasta format and analyzed using the online database called Galaxy. (46) Data was separated by barcoding primer and collapsed to amalgamate sequences. Using the program DREME motif analyzer (47) of the MEME NCBR database, motif analysis was performed. The structure of potential aptamers was analyzed using IDT's Oligo Analyzer Version 3.1. (49) Each potential aptamer was also checked for G quadruplex structures using the QGRS Mapper. (48)

2.4 Flow Cytometry Analysis of ssDNA Clones against Beads

Fourteen of the most promising sequences were ordered to analyze under flow cytometry. To save money, the sequences were not labelled fluorescently. Instead, a fluorescent label with the sequence CGTGGTTACAGTCAGAGGAG-Cy5 was ordered. This sequence was complementary to the forward primer on the aptamers, and as the primer region was unlikely to contribute to binding, it could be used to tag the aptamer with fluorescence. 100nM of each sequence was separately added to 100nM of the complementary fluorescent sequence, heated to 95 degrees, and snap cooled to 4 degrees. It was assumed that all sequences were fluorescently labelled. In addition a 0.01mg/mL, FITC labelled anti-human Glycophorin A antibody was prepared to verify presence of protein. All 14 potential aptamers at a concentration of 100nM, the antibody at 0.01mg/mL, and starting library at 100nM were incubated with $1.75x10^8$ beads for 45 minutes, and washed 3 times with PBS plus Calcium and Magnesium on a magnet. The samples were analyzed.

2.5 Analysis Using Whole RBCs

2.5.1 Preparation and Counting of Red Blood Cells

Whole Blood was purchased from BioreclamationIVT. Red Blood Cells were separated from whole blood through centrifugation at 300G for five minutes after adding one volume of PBS and collecting the bottom layer. This was repeated 3 times to ensure minimum contamination with other cell types. A 1/10 ratio of RBCs and Muse Cell Count and viability Mix (Millipore) was prepared. The sample was loaded into the Countess II Automated Cell Counter (Thermo Fischer) and cell count was recorded. It was determined that there were 10,000 cells/uL of RBCs. The RBCs were mixed with 1mg of

tRNA per 50mL of cells to block the positively charged sites which might interfere with proper aptamer binding tests.

2.5.2 Flow Cytometry Analysis of ssDNA Clones against RBCs

All 14 potential aptamers at 100nM, the antibody at 0.01mg/mL, and starting library at 100nM were incubated with 150,000 RBC cells for 45 mins. The cells were washed once with PBS with Calcium and Magnesium and analyzed by Gallios Flow Cytometer (Beckman Coulter) for 20,000 events. After this, to each tube containing aptamers 0.01mg/mL antibody was added and incubated for 15 minutes. The samples were re-analyzed by the flow cytometer for 20,000 events. This was to check for competitive binding between the antibody and aptamer. This process was repeated without the antibody test, but with a ssDNA concentration of 25nM, and 500nM.

Chapter 3: Specialized Aptamer Selection

3.1 Abstract

Aptamer research is an ongoing field, yet despite several strong advantages of aptamers over antibodies, aptamers have yet to attain a strong commercial appearance. It is therefore necessary to improve upon the current aptamer production process so as to produce better aptamers in a faster and more efficient way. Through the use of an adapted emulsion PCR protocol and through partial automation, we sought to achieve this goal in a 11 round aptamer selection against Human Glycophorin A. The product of the emulsion PCR steps was cleaner than the product of the same selection using conventional PCR, allowing for a more efficient selection. It was found that even in a raw pool, the ssDNA from rounds 6,7, and 8 bound more to the protein of interest than the starting random library.

3.2 Background

Despite being produced for approximately 30 years (2) aptamers have yet to dominate over antibodies in the scientific community with only one aptamer drug commercially available. This aptamer drug is pegaptanib (Macugen) for the treatment of age-related macular degeneration. (50) It is therefore apparent that there needs to be changes to some of the basic aptamer production methods to allow for the better production of more powerful aptamers. One of the weakest points in the aptamer production process is during PCR amplification, both due to the buildup of harmful byproduct and the natural tendency for PCR bias to underrepresent DNA with high GC content. (34,37) This was problematic during initial attempts at developing an aptamer to the current target (Glycophorin A) by our team. We have therefore adapted and altered a protocol

from the works of Shao et al (34) to use emulsion PCR to improve the amplification of our ssDNA pools from round to round. It was also noted while analyzing previous aptamer work by our lab that one possible reason for poor aptamer binding is that the stringency of the washing step of SELEX was not high enough to dislodge all ssDNA that does not bind effectively. We introduced more rigorous washing steps after each round so adherent binders could be recovered. Furthermore due to the repetitive nature of aptamer selection it has been suggested that aptamers do not necessarily need to be entirely manually selected. (36) For the purpose of moving towards the goal of complete automation, we started with automating the simpler steps in the selection process, with the more complex portions still completed manually.

In this chapter we shall overview the 11 round selection and verification of an aptamer against Human Glycophorin A, focusing on the changes to standard operating procedure in aptamer selection, namely the use of ePCR. The chapter will present the results of the selection, with flow cytometry data analyzing the binding of the raw pools when compared to the starting library.

3.3 Results

3.3.1 Verification of Glycophorin A Attachment to Beads

The first test conducted was to verify the presence of Glycophorin A on the beads after preparation as per manufacturer recommendations of 10ug of Glycophorin A per $3.5x10^9$ beads. To do this, an antibody to Glycophorin A, FITC labelled anti-human Glycophorin A antibody clone HI264, was purchased. Antibody concentrations of 1:50 (0.01mg/mL) 1:100 (0.005mg/mL), 1:200 (0.0025mg/mL), and 1:500 (0.001mg/mL) were added to the protein and bead mixture ($8.75x10^7$ beads) and incubated for 15

minutes. The results were analyzed by flow cytometry. Figure 3.1 shows the results of this experiment.

After the confirmation of a presence of Glycophorin A on the beads, it had to be determined how much glycophorin A needed to be incubated with the beads to reach bead saturation. 5ug, 10ug, and 20ug of Glycophorin A was incubated with the plain beads as per manufacturers protocol (see Chapter 2). A 1:50 dilution (0.01mg/mL) of antibody was mixed with each sample and the fluorescence was analyzed by flow cytometry after being washed to remove unbound antibody. Figure 3.2 shows the resultant protein values.

3.3.2 Emulsion PCR vs Conventional PCR

Initially selection of an aptamer for Glycophorin A was conducted using conventional PCR after every round of selection. This selection halted after round 3. This selection was repeated starting from round 0 with the exception of emulsion PCR being used where conventional PCR was used in the previous selection. This allows for a comparison of the effect of PCR method upon the selection quality. In Figure 3.3 the Round 3 gels of both selections are shown side-by-side. These gels were pictured directly after amplification and digestion as per the SELEX protocol described in Chapter 2.

3.3.3 Verification of Protein Binding from Raw Pools

After 11 rounds of SELEX, the raw pools were to be analyzed by flow cytometry to see if the selection was successful. It was deemed unlikely that the first five rounds would have strong binding , and therefore these pools were not analyzed. The remaining 6 pools were amplified to 200nM and incubated with the protein beads as per the

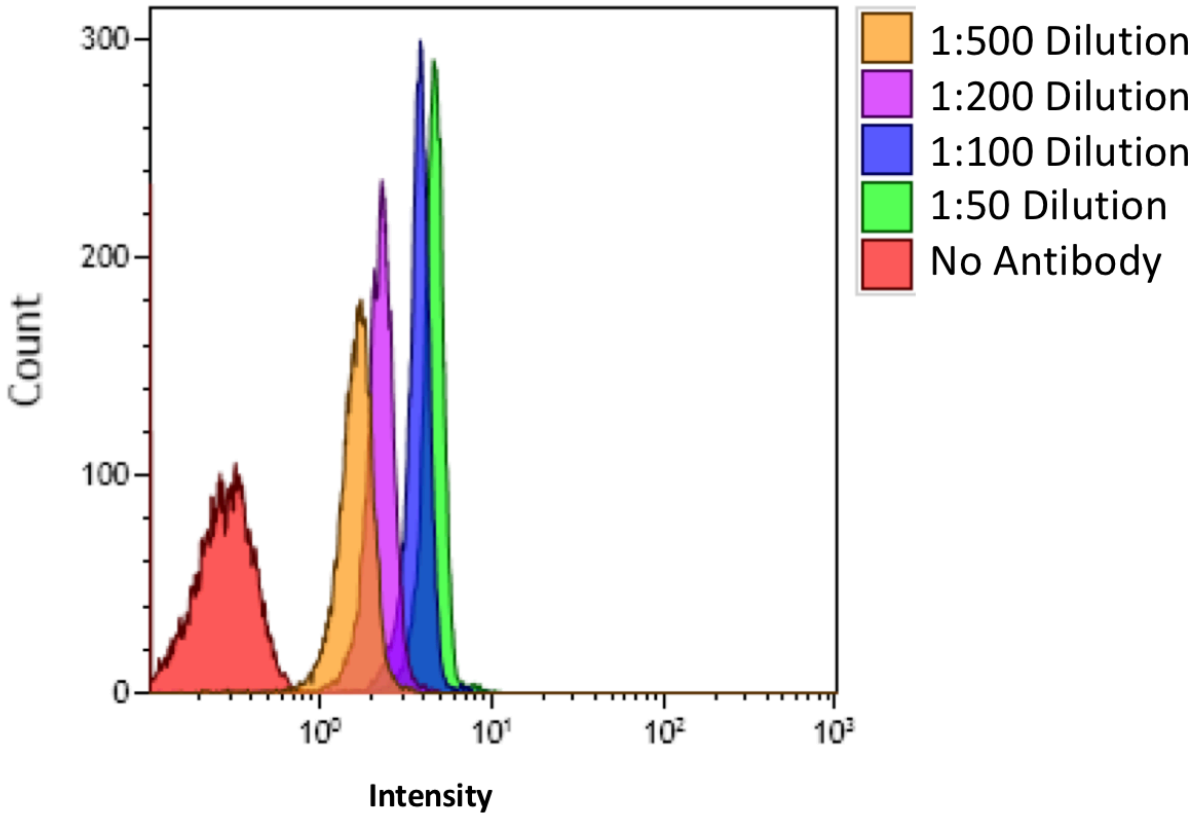


Figure 3.1: Verification of proper binding of Glycophorin A protein to beads. FITC labelled anti-human Glycophorin A antibody clone HI264 was prepared and concentrations of 1:50 (0.01mg/mL) 1:100 (0.005mg/mL), 1:200 (0.0025mg/mL), and 1:500 (0.001mg/mL) were added to magnetic beads after 10ug of Glycophorin had been incubated with $3.5x10^9$ beads. The red peak corresponds to the fluorescence of the protein beads without any antibody.

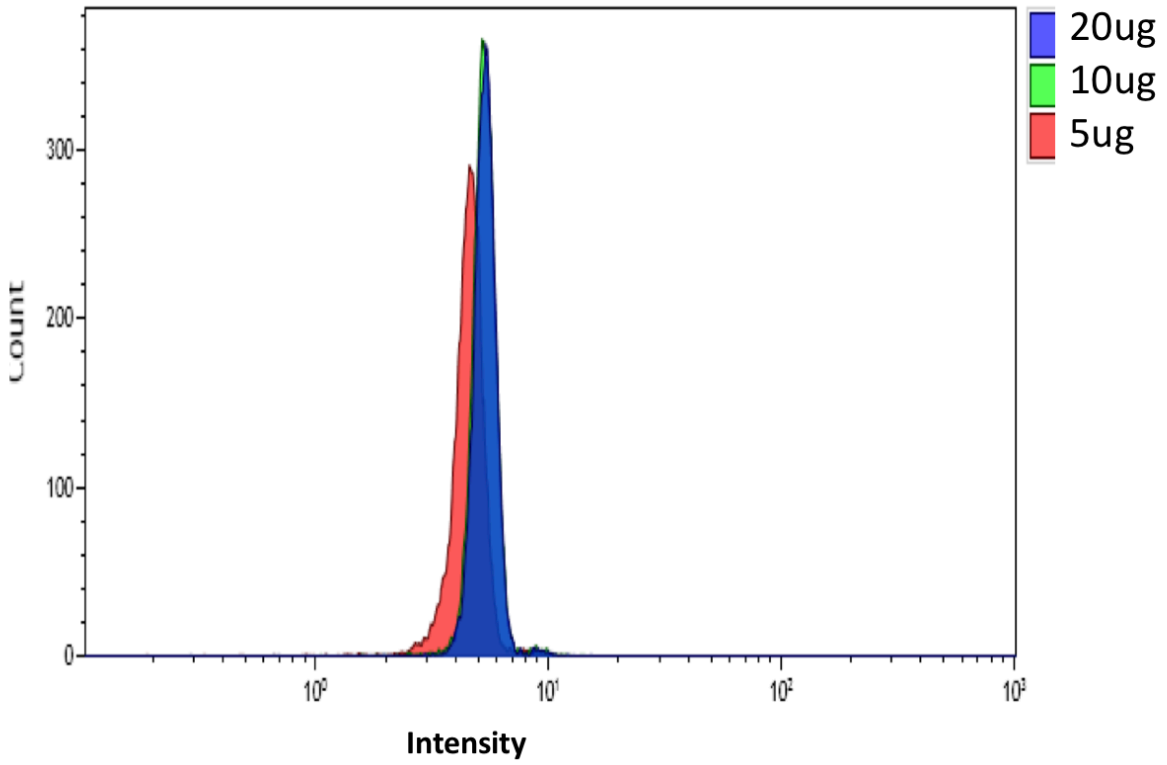
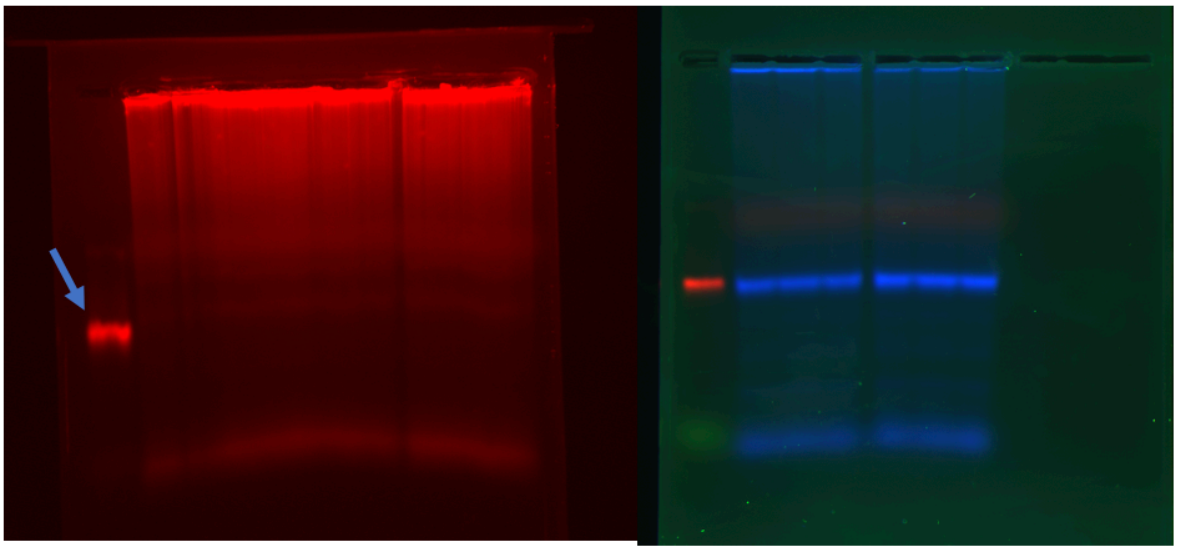


Figure 3.2: Determination of amount of protein per $35x10^8$ beads to fully saturate beads with protein. Three different protein amounts were selected and 0.01mg/mL of FITC labelled anti-human Glycophorin A antibody clone HI264 were added. After a 15 minute incubation these samples were all analyzed by flow cytometry.



A B

Figure 3.3: Effect of emulsion PCR on quantity of high molecular weight by-product versus ssDNA of desired length. (A) is a gel of the third round SELEX product after conventional PCR amplification. The small red band highlighted by the blue arrow is 100nM 80 base pair DNA library. The large red smear is the combined PCR product.
(B) is a gel of the third round of SELEX product after emulsion PCR amplification. The small red band is 100nM 80 base pair DNA library. The large blue bands indicate the combined PCR product.

Methods chapter. The beads from each round were fed through the flow cytometer and the Cy5 fluorescence was analyzed. Figure 3.4 and Figure 3.5 show the amount of DNA bound to the beads.

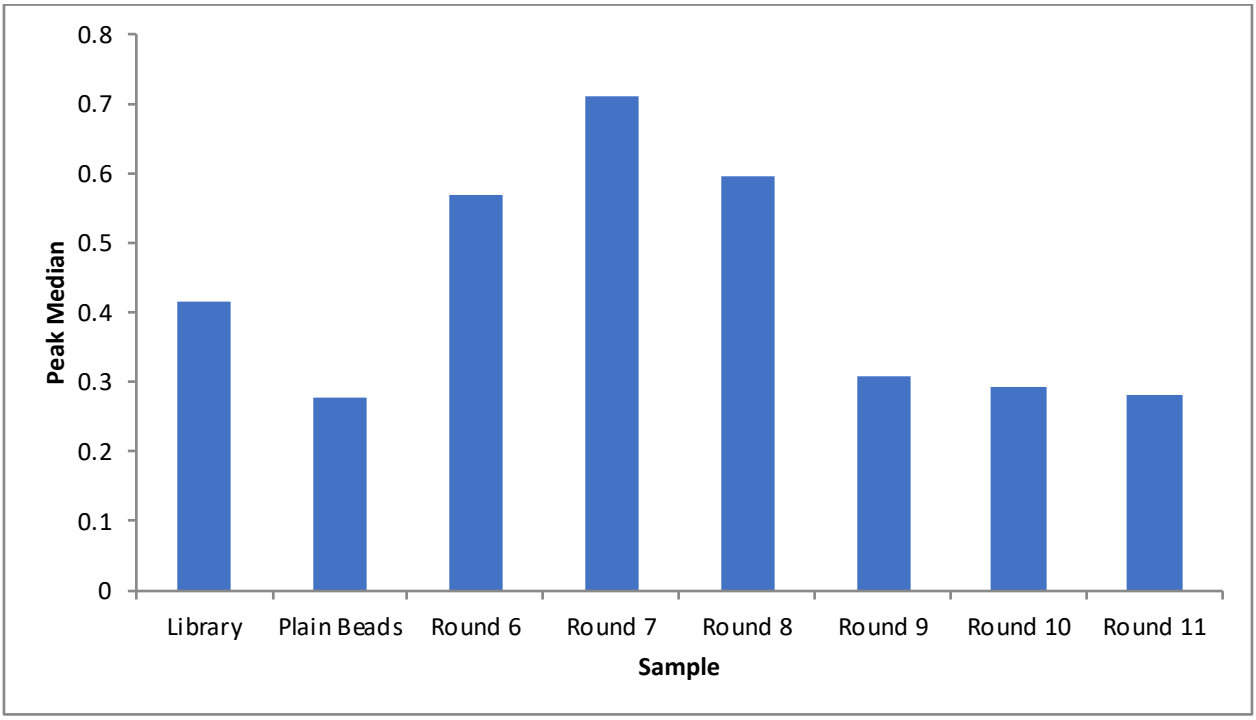


Figure 3.4: Graphed peak median of post-selection raw pool flow cytometry analysis. The raw selection pools from each round were incubated with the Glycophorin A beads. These beads were then compared in Cy5 fluorescent intensity to beads incubated with the starting library and plain Glycophorin A beads. The median of every peak corresponding to the various round pools, the library, and the plain beads with Glycophorin A were taken and graphed. The sample number was 1 for each bar.

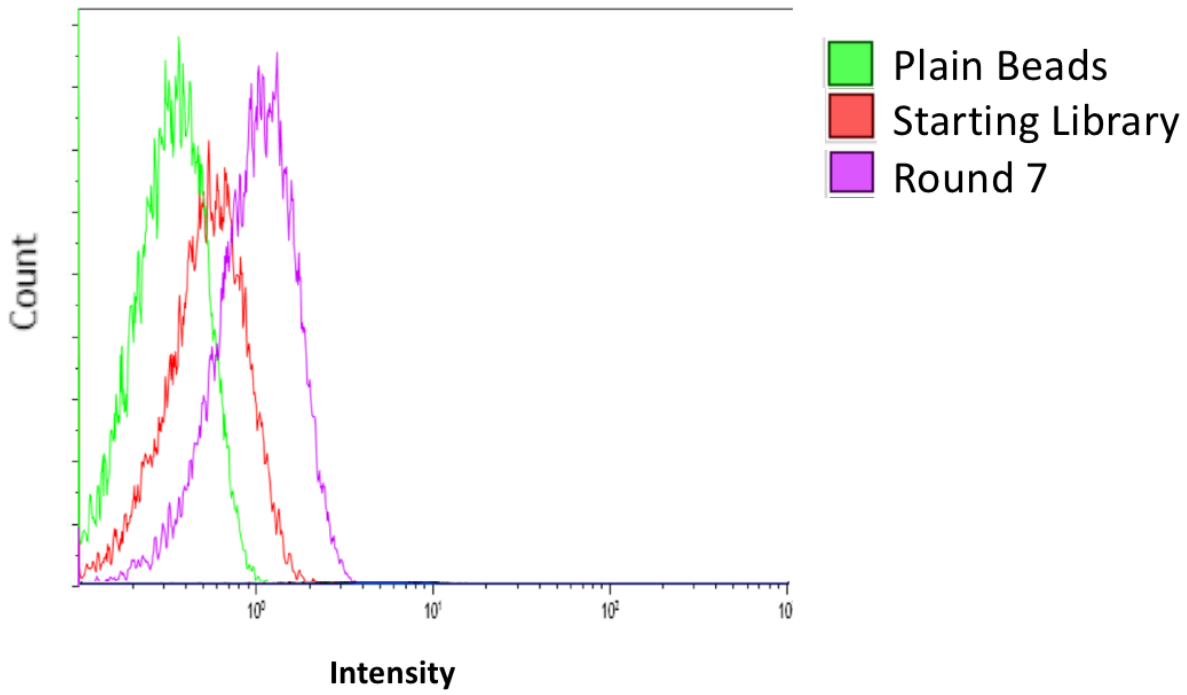


Figure 3.5: Raw histogram data comparing fluorescent intensity of the strongest binding round, the plain Glycophorin A beads, and the starting library. The raw selection pools from each round were incubated with the Glycophorin A beads. These beads were compared in Cy5 fluorescent intensity to beads incubated with the starting library and plain Glycophorin A beads. Shown is the Round 7 results superimposed onto those two previously mentioned controls. Round 7 showed the strongest fluorescence amongst the rounds tested.

3.4 Discussion

3.4.1 Verification of Glycophorin A Attachment to Beads

Figure 3.1 suggests that some Glycophorin A protein may have bound to the magnetic beads. The red peak corresponding to no antibody would be the expected peak if no protein was present. Even in the lowest dilution however, the FITC fluorescence is higher than the no antibody peak. This confirms that the beads bound some Glycophorin A.

Varying amounts of protein were added to a fixed number of beads to determine how much Glycophorin A is necessary to reach saturation of the beads. In Figure 3.2 the three peaks overlap, with only the 5ug peak being slightly lesser than the other two. This suggests that 5ug of Glycophorin A was insufficient to saturate the beads. The manufacturer's suggestion of 10ug of protein per 35x108 was sufficient.

3.4.2 Emulsion PCR vs Conventional PCR

In Figure 3.3, after three rounds of selection using conventional PCR, the gel of the eluted product displays high molecular weight by-product. A faint band was visible around the same size as the starting library, however, most of the fluorescence comes from product so heavy it barely leaves the initial well. This high molecular weight by-product has been previously encountered by aptamer researchers. One group sequenced the by-product to discover what was causing it. They suggested that it may be a result of a primer region of one strand attaching to a complementary strand in the random portion of another, leading to ever lengthening strands with every cycle. Despite their best efforts, they could not reduce this by-product without changing primers entirely. (51) The results found here suggest that the conditions created by emulsion

PCR reduce the formation of these by-products. During emulsion PCR, each strand of template DNA is trapped in its own emulsion or “bubble”. This means that as PCR continues, different strands in the library cannot bind to each other and create the DNA concatamers that hamper conventional PCR methods. In additions to the beneficial results shown by these gels, there are additional benefits to ePCR. Conventional PCR favours the production of AT rich, pin shaped DNA as opposed to GC rich, structure heavy DNA. This is a major negative in aptamer research where aptamers rely on 3D structure for their binding. (1,2) The non-biased approach to amplification in ePCR can only benefit the amplification of potential aptamers. This benefit cannot be visualized in a gel, as all 80 base pair DNA will appear the same in this type of gel, however it should be later seen in the form of improved aptamer pool binding at lower round numbers than most previous SELEX procedures.

3.4.3 Verification of Protein Binding from Raw Pools

After a selection is complete one may pick the highest round and send this pool for sequencing. (examples: 1,2,4) Instead, I first analyzed each of the raw pools through flow cytometry. This experiment was only conducted once for a number of reasons. First, there was a very limited amount of sample from each round, and this sample was in low concentration. Each sample had to be amplified for flow cytometry analysis to be able to detect the fluorescence. If there was to be enough sample to amplify for sequencing without amplifying multiple times over and possibly introducing more error, only a small amount of each of the saved pools from each round could be taken. In an effort to gather more data, our lab wanted to analyze the raw pools, and see if, as expected, the latest round had the most fluorescent binding. This experiment could

provide an additional level of scrutiny. This experiment could have been replicated thrice as was done with later flow cytometry work in this thesis. The lack of repeats can be listed as a source of error in the conclusions made from these figures.

Figure 3.4 and Figure 3.5 show the results of these experiments. Figure 3.4 shows the summary of all the data. The reasons behind also including Figure 3.5 was to show an example of a raw histogram, to give context to the bar graph shown in Figure 3.4. The round with the highest difference from the starting library is shown as a reference for all other rounds in comparison. Analyzing Figure 3.4, an interesting pattern appears. When viewing the raw pool binding, the rounds form a histogram shape when the median is presented in bar graph form. This suggest that up to round 7, the aptamer rounds continued to improve in terms of binding, but from round 8 onwards, the binding affinity decayed. This phenomenon suggests that after a certain point, the selection is actually hindering the binding affinity of the overall pool. This phenomenon is counter to what was expected to be found in aptamer selection, as most selection rely on over selection to narrow down the pool as much as possible to facilitate finding binders in the sequencing results. (52) A possible explanation for why this occurred in the selection is the method of washing. I started with an aggressive wash to ensure the only binding survivors were strong binders. This wash was increased every successive round. It is possible that after round 7, the wash became so aggressive that it removed even properly binding aptamers, and the pools from round 8 onwards became more composed of trace artifacts and contamination. Another possible explanation is that the selection finished early due to the use of emulsion PCR and strict washing. Perhaps after 7 rounds, almost all of the surviving sequences were strong binders. Further

rounds of selection may not have narrowed the pool, but still provided an opportunity for contamination and introduction of PCR artifacts. The second explanation is supported by previous studies which use ePCR in aptamer selection, where they report a lower number of rounds of selection are necessary to isolate good binders. (34,37). This explanation may also be supported by other forms of aptamer selection that use a more stringent method of washing or separating binding aptamers. Using more stringent methods, there are selections which are completed in much lower round numbers than traditionally accepted, as low as 5-8. (53,54,55) Because of these combined reasons, I favour explanation 2.

3.4.4 Selection of Pools for NGS

Moving forward with the next step of generating a useful aptamer, pools must be sequenced. Based on the results above, it was decided to send the raw pool with the best binding. As the difference in binding between rounds 6,7, and 8 may be negligible due to the error involved, all three were sequenced. In addition, the round 11 pool was analyzed as well because there was little binding. This serves as a negative control. Sequences present in high amounts in the round 11 raw pool may not be strong binders for the target.

3.5 Conclusion

The selection portion of this project was a success. The Emulsion PCR method appeared helpful, with ePCR samples showing less contamination than those using traditional PCR. Analysis of the raw pools indicated that the post-selection pools have stronger binding than the starting library. The strongest binders were the DNA sequences in the pool of round 7.

Chapter 4: NGS Results Analysis

4.1 Abstract

Aptamer selection is the first step in the development of a functional aptamer. The raw pools need to be sequenced to discover the exact identity of any binding molecules. Through the next generation sequencing of the raw pools for rounds 6, 7, 8 and 11, I sought to identify possible aptamers for the protein Glycophorin A. Through various methods such as analyzing sequence predicted structure and sequence rarity, 14 aptamers were isolated and synthesized for further testing.

4.2 Background

No matter how successful the initial selection, the raw pools generated by SELEX lack value by themselves. The aptamer pools are still heterogeneous, even after a long selection, and it would be difficult to replicate said pools without knowing what sequences comprise them. (33) To deconvolute the raw pools generated, next generation sequencing can be used to gain a more accurate picture of the percent composition of each sequence. (16)

After the sequencing results are attained, it would be arduous to test the binding of all of the generated sequences as thousands of different molecules remain. One can infer which of the sequences to synthesize. Here U use multiple methods of sequence analysis including analyzing the predicted structure using OligoAnalyzer 3.1 (49), searching for the presence of G quadruplex structures using the QGRS Mapper (48), and analyzing the relative rarity of each sequence within the pools.

This chapter we reviews the next generation sequencing results, and the subsequent analysis of the generated data. This includes the selection of potential aptamers to be tested for binding in the next chapter.

4.3 Results

4.3.1 Most Abundant Sequences

The most important and easiest to collect data on each of the pools recovered would be the relative abundance of sequences within the pools. The raw sequences were collapsed using Galaxy (56) allowing the ability to see the relative number of repeats for each sequence. Tables 4.1- 4.4 show the top five sequences in each of the four sequenced pools.

As each pool is analyzed it is noticeable that the number of repeats of the most abundant sequences becomes higher. The only sequence that is not the proper length is the third most abundant sequence in round 6, which appears to be a fragment. Every other abundant sequence is 80 nucleotides discounting the barcoding region. Furthermore, when accounting for the slight variation in sequences before the forward primer due to the barcoding (see Chapter 2), several sequences appear in multiple pools. The first sequence in rounds 6, 7, and 8 are the same sequence, but this sequence is not present in round 11. The 4th sequence in rounds 6, 8, 11 and the 2nd sequence in round 7 are the same sequence. The 2nd sequence in rounds 6 and 8, and the third sequence in round 7 are the same sequence, however, this sequence is not present in round 11.

Table 4.1: Most Abundant Sequences in Raw Pool Round 6

Sequence Abundance	Sequence	Number of Repeats
1	GTACTGACCTCCTCTGACTGTAACCACGTCGCGGGT AGGGGGAGGGCCGAGGAGGCTGTAGGTGGGTGGC ATAGGTAGTCCAGAAGCCA	561
2	GTACTGACCTCCTCTGACTGTAACCACGTGGAGGGG GCAAGTAGGTTAGGGGGGAGGAGCAGGCAGGGGCA TAGGTAGTCCAGAAGCCA	439
3	GTACTGACCTCCTCTGACTGTAACCACGGCATAGGTA GTCCAGAAGCCA	423
4	GTACTGACCTCCTCTGACTGTAACCACGTCTGCAGGG TAGGGGCGAGGGTAGGTAGGGGGGAGCGTGGGCAT AGGTAGTCCAGAAGCCA	380
5	GTACTGACCTCCTCTGACTGTAACCACGTGAGGGAGG GCGAGGGAGCGGCGCATGGGAAAGGGGTGGAGCATA GGTAGTCCAGAAGCCA	319

Table 4.2: Most Abundant Sequences in Raw Pool Round 7

Sequence Abundance	Sequence	Number of Repeats
1	CTGACATACTCCTCTGACTGTAACCACGTCGC GGGTAGGGGGAGGGCCGAGGAGGCTGTAGG TGGGTGGCATAGGTAGTCCAGAAGCCA	1123
2	CTGACATACTCCTCTGACTGTAACCACGTCTG CAGGGTAGGGGCGAGGGTAGGTAGGGGGGA GCGTGGGCATAGGTAGTCCAGAAGCCA	959
3	CTGACATACTCCTCTGACTGTAACCACGTGGA GGGGGCAAGTAGGTTAGGGGGGAGGAGCAG GCAGGGGCATAGGTAGTCCAGAAGCCA	893
4	CTGACATACTCCTCTGACTGTAACCACGTGGT AAGGGTAGAGGCTCGGAGGGGGGCCTTAGG GGGGTGGCATAGGTAGTCCAGAAGCCA	659
5	CTGACATACTCCTCTGACTGTAACCACGTGAG GGAGGGCGAGGGAGCGGCGCATGGGAAAGG GGTGGAGCATAGGTAGTCCAGAAGCCA	596

Table 4.3: Most Abundant Sequences in Raw Pool Round 8

Sequence Abundance	Sequence	Number of Repeats
1	ACTGATCACTCCTCTGACTGTAACCACGTCGCG GGTAGGGGGAGGGCCGAGGAGGCTGTAGGTG GGTGGCATAGGTAGTCCAGAAGCCA	3465
2	ACTGATCACTCCTCTGACTGTAACCACGTGGAG GGGGCAAGTAGGTTAGGGGGGAGGAGCAGGC AGGGGCATAGGTAGTCCAGAAGCCA	1218
3	ACTGATCACTCCTCTGACTGTAACCACGGTGAG AGGGGGGCAGGCCAGGGGGGGGCAGAGTGGA GGATGCATAGGTAGTCCAGAAGCCA	1196
4	ACTGATCACTCCTCTGACTGTAACCACGTCTGCA GGGTAGGGGCGAGGGTAGGTAGGGGGGAGCGT GGGCATAGGTAGTCCAGAAGCCA	1185
5	ACTGATCACTCCTCTGACTGTAACCACGGCGGGA GGCGAAGGGGGGGAGAGGAGGTAGGCGGGGCG ATGCATAGGTAGTCCAGAAGCCA	1024

Table 4.4: Most Abundant Sequences in Raw Pool Round 11

Sequence Abundance	Sequence	Number of Repeats
1	CAGTACGTCTCCTCTGACTGTAACCACGACAG GGGGCGGTAGGGTGTGGTCGAGGGGGGTGC TAGGGTGCATAGGTAGTCCAGAAGCCA	5150
2	CAGTACGTCTCCTCTGACTGTAACCACGGTGG GCAGGGGGGCGAGAGGGCTGAGGTGGTGCTT GGTGGGCATAGGTAGTCCAGAAGCCA	2703
3	CAGTACGTCTCCTCTGACTGTAACCACGGCAG GGGGCGGTAGGGTGTGGTCGAGGGGGGTGCT AGGGTGCATAGGTAGTCCAGAAGCCA	2572
4	CAGTACGTCTCCTCTGACTGTAACCACGTCTGC AGGGTAGGGGCGAGGGTAGGTAGGGGGGAGC GTGGGCATAGGTAGTCCAGAAGCCA	1905
5	CAGTACGTCTCCTCTGACTGTAACCACGCCATA TGCAAGTGGAGAACCGCTTTCAGAACATCGGAA GTGCATAGGTAGTCCAGAAGCCA	1637

4.3.2 Identification of 14 Possible Aptamer Sequences

After comparing and structure analysis, fourteen potential aptamers were identified. These sequences were chosen through a combination of the abundance data, such as that analyzed above, and the subsequent structure and G-quadruplex presence data which will be shown in the following sections. Table 4.5 shows the 14 sequences which were selected for further analysis and the given names for each aptamer. Going forward, each sequence will be referred to by this name for simplicity.

Table 4.5: Selected Potential Aptamers

Potential Aptamer	Sequence
Aptamer 1	CTCCTCTGACTGTAACCACGTCGCGGGTAGGGGGAGGGCCGAGG AGGCTGTAGGTGGGTGGCATAGGTAGTCCAGAAGCC
Aptamer 2	CTCCTCTGACTGTAACCACGTCTGCAGGGTAGGGGCGAGGGTAG GTAGGGGGGAGCGTGGGCATAGGTAGTCCAGAAGCC
Aptamer 3	CTCCTCTGACTGTAACCACGTGGAGGGGGCAAGTAGGTTAGGGG GGAGGAGCAGGCAGGGGCATAGGTAGTCCAGAAGCC
Aptamer 4	CTCCTCTGACTGTAACCACGTGGTAAGGGTAGAGGCTCGGAGGG GGGCCTTAGGGGGGTGGCATAGGTAGTCCAGAAGCC
Aptamer 5	CTCCTCTGACTGTAACCACGTGAGGGAGGGCGAGGGAGCGGCGC ATGGGAAAGGGGTGGAGCATAGGTAGTCCAGAAGCC
Aptamer 6	CTCCTCTGACTGTAACCACGCGGAGGGTACCGGAGCGGAGAGGG GAGTGGGAGGGGAGAGGCATAGGTAGTCCAGAAGCC
Aptamer 7	CTCCTCTGACTGTAACCACGGCGGGAGGCGAAGGGGGGGAGAGG AGGTAGGCGGGGCGATGCATAGGTAGTCCAGAAGCC
Aptamer 8	CTCCTCTGACTGTAACCACGTGAAGGAGGGAGAGGGAGGTGGTG CAGGGGGTCAGGGGTAGCATAGGTAGTCCAGAAGCCA
Aptamer 9	CTCCTCTGACTGTAACCACGCGGGGGGAGGAGGGAGAGGGGTG CGAGGTCAATGGGGGAAGCATAGGTAGTCCAGAAGCC
Aptamer 10	CTCCTCTGACTGTAACCACGCCAGGGGGGAAGGGAGTGGAAGGG GGGTTGCTGGGAGTTGGCATAGGTAGTCCAGAAGCC
Aptamer 11	CTCCTCTGACTGTAACCACGATGGGTGAAGGGTAGAGAGGGGGG TCGAGGGAGCCGGGGTGCATAGGTAGTCCAGAAGCC
Aptamer 12	CTCCTCTGACTGTAACCACGAGGGGGCAAGGGAGGTGGAGGGGG ATGGAAATGGGTGTGGGCATAGGTAGTCCAGAAGCC
Aptamer 13	CTCCTCTGACTGTAACCACGTGGCCAGGGAGGGAGGAGGGCAAG GGGGGCTGAACGTGCGGCATAGGTAGTCCAGAAGCC
Aptamer 14	CTCCTCTGACTGTAACCACGTGCGGTGAGGGAGAGGGGGGCGAG GTCGGGCTACGCGAGGGCATAGGTAGTCCAGAAGCC

4.3.3 G Quadruplex Content

One of the ways in which potential aptamers were identified was through looking for and identifying G Quadruplex structures within the sequence. This was conducted using the program QGRS Mapper (48). Table 4.6 shows each of the 14 identified potential aptamers and the number of associated G Quadruplex structures. Aptamers 1, 3, 9, and 12 were identified to have two G quadruplexes within the potential structure. All other identified aptamers have one.

Table 4.6: Potential G Quadruplex Structures Within Potential Aptamers

Potential Aptamer	Number of G Quadruplexes
Aptamer 1	2
Aptamer 2	1
Aptamer 3	2
Aptamer 4	1
Aptamer 5	1
Aptamer 6	1
Aptamer 7	1
Aptamer 8	1
Aptamer 9	2
Aptamer 10	1
Aptamer 11	1
Aptamer 12	2
Aptamer 13	1
Aptamer 14	1

4.3.4 Aptamer Structures

To gauge the likelihood of each sequence being a useful aptamer, the 2D structure was analyzed. A strongly binding aptamer relies heavily on a complex structure. In addition, similarities between the different structures may indicate common motifs. Each of the 14 aptamers was entered into Oligo Analyzer 3.1 and the most likely DNA structure was noted . Figure 4.1, Figure 4.2, and Figure 4.3 show these structures.

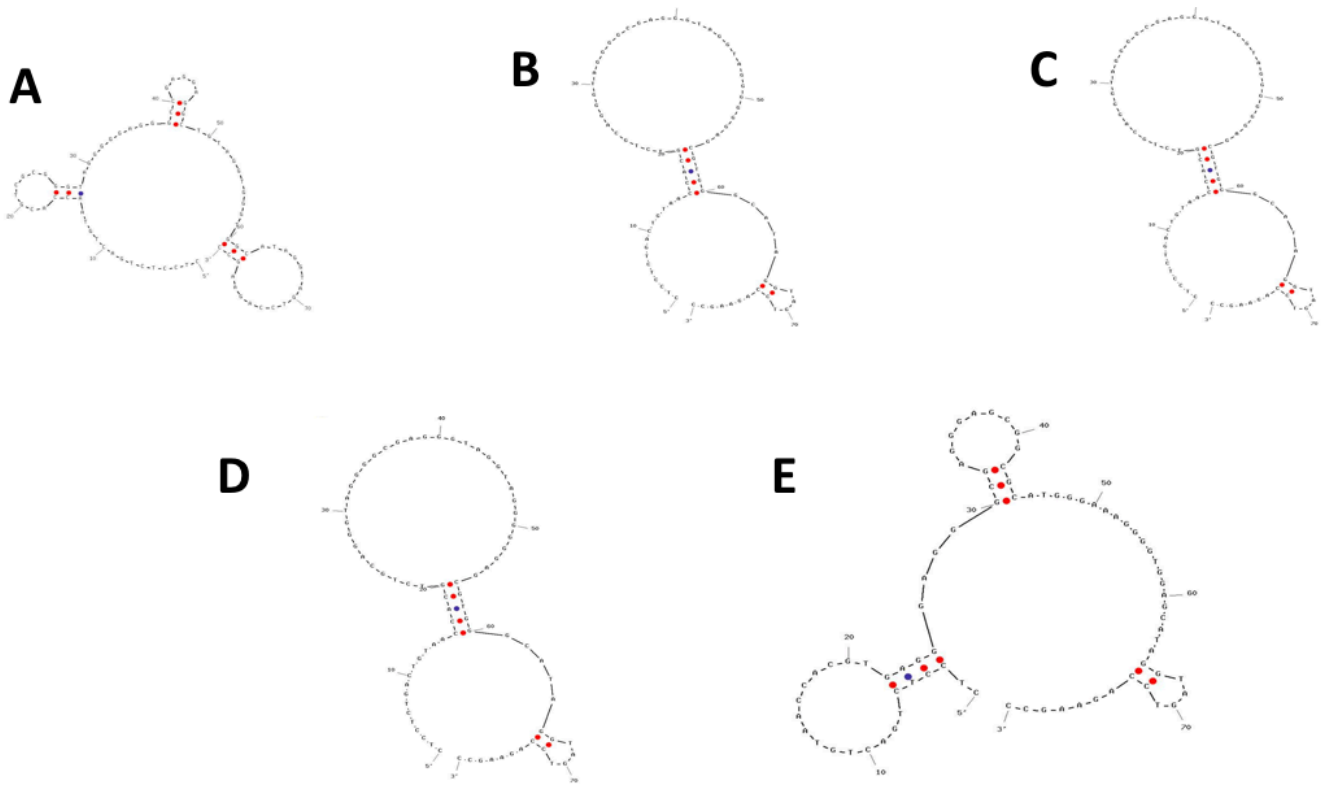


Figure 4.1: Predicted Structure of Selected Aptamers 1 through 5. The sequences of the selected possible aptamers were entered into the program Oligo Analyzer 3.1. The Number 1 predicted structure was taken for each aptamer. A shows the predicted structure for Aptamer 1. B shows the predicted structure for Aptamer 2. C shows the predicted structure for Aptamer 3. D shows the predicted structure for Aptamer 4. E shows the predicted structure for Aptamer 5.

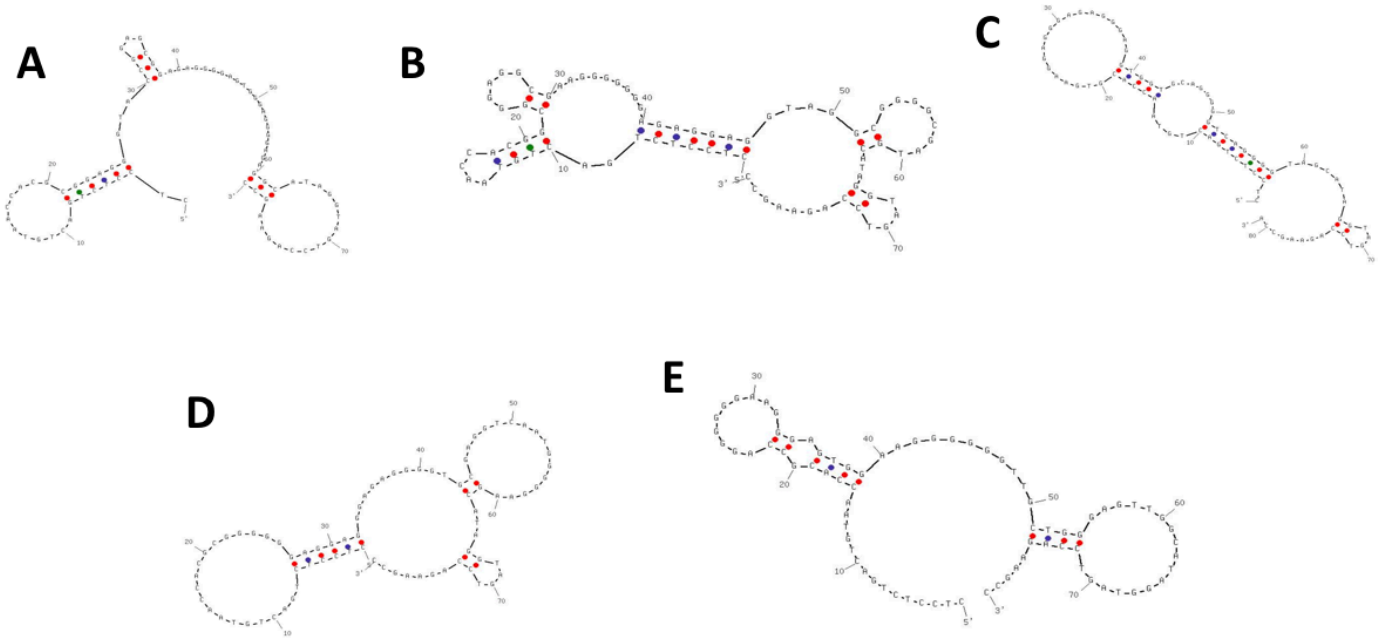


Figure 4.2: Predicted Structure of Selected Aptamers 6 through 10. The sequences of the selected possible aptamers were entered into the program Oligo Analyzer 3.1. The Number 1 predicted structure was taken for each aptamer. A shows the predicted structure for Aptamer 6. B shows the predicted structure for Aptamer 7. C shows the predicted structure for Aptamer 8. D shows the predicted structure for Aptamer 9. E shows the predicted structure for Aptamer 10.

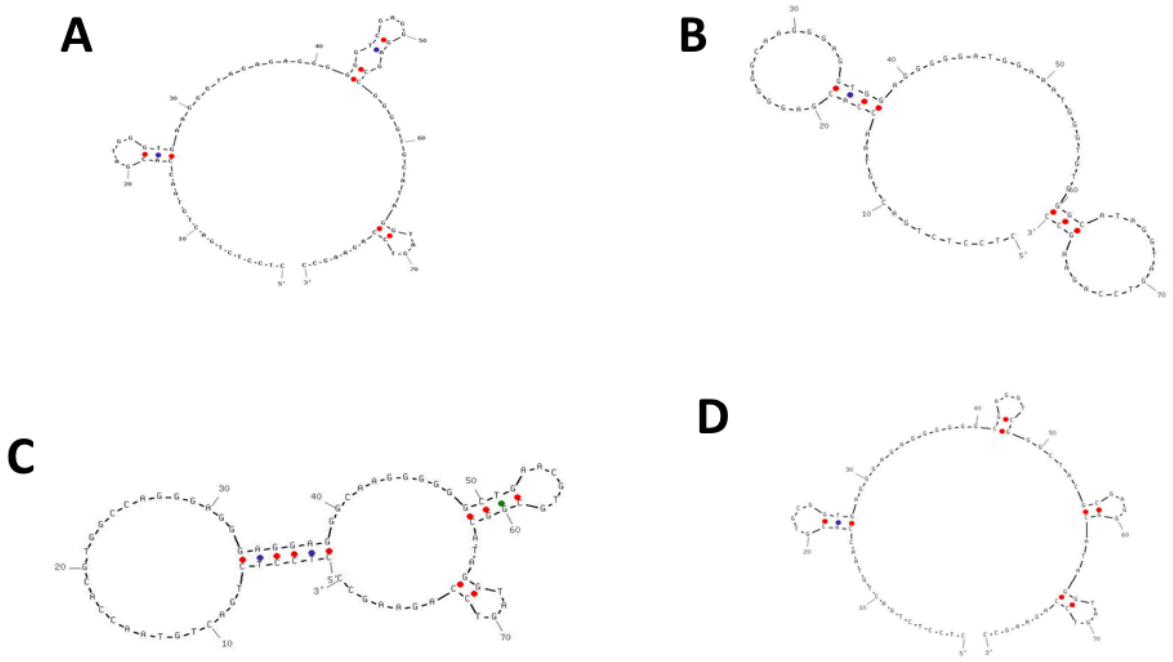


Figure 4.3: Predicted Structure of Selected Aptamers 11 through 14. The sequences of the selected possible aptamers were entered into the program Oligo Analyzer 3.1. The Number 1 predicted structure was taken for each aptamer. A shows the predicted structure for Aptamer 11. B shows the predicted structure for Aptamer 12. C shows the predicted structure for Aptamer 13. D shows the predicted structure for Aptamer 14.

4.3.5 DREME Motif Analysis

It is possible that through the selection to a common target, the aptamer pool evolved toward common motifs. To determine this, the top 75 sequences from each of the pools of round 6, 7, and 8 were entered into the program DREME Motif Analyzer (47) and processed. Interestingly the only discovered common motifs were in the common primers at the beginning and end of the sequences. Nothing within the random 40 nucleotide centre was determined to form a pattern among sequences.

4.4 Discussion

4.4.1 Most Abundant Sequences

The first observation of note is that the number of repeats of the most abundant sequences continues to become larger as the pool evolves via selection. As the selection continued, the sequences may have began to become less diverse. (1,2) It must be remembered, however, that the ultimate pool (round 11) had no binding to the target of interest. It can therefore be assumed that the latest pool, while less diverse, was favouring the selection of sequences that amplified easily as opposed to sequences that bound strongly. This suggests that the emulsion PCR protocol failed to entirely eliminate the inherent PCR bias to amplify sequences plentiful in A and T nucleotides as opposed to structure rich G and C plentiful molecules.

The immediate most important observation is that the most abundant sequences in all three binding rounds (rounds 6,7, and 8) are identical, and that this sequence is not present in the top 5 sequences of the non-binding round. Further analysis identifies this sequence in the round 11 pool, but at a position of 11^{th} most plentiful sequence. These considerartions make this sequence a possible promising aptamer based on probability.

4.4.2 G Quadruplex Content

All of the selected aptamers had at least one G quadruplex structure, with four of the sequences containing 2. Considering the relative small size of the aptamers, being only 80 nucleotides long, the observation that some of the sequences contained two of these structures is surprising. It has been shown that aptamers containing at least one G quadruplex structure have higher chemical and thermodynamic stability as well as an improved uptake by the body's cells. (40) Given that this aptamer is to be used with human red blood cells, these properties are desirable. Therefore all sequences within high abundance (within the first 5-10 sequences) in round pools 6,7, and 8 which contained a G quadruplex were screened for binding.

4.4.3 Aptamer Structures

Figures 4.1- 4.3 show the most stable structures for the 14 selected aptamers. Some interesting common structures appear. Aptamers 1, 5, and 6 share a similar structure. They all comprise of a large circular shape with three small circular structures branching off in similar locations. Aptamers 2, 3, and 4 are also similar. They comprise two circular structures attached by a thin straight region. There are similarities amongst the other structures, however, none of the others show the high similarity of those listed above. These similarities are interesting because they show that while the sequences themselves may not show common motifs, the predicted structure, which may be the determining factor of binding, has converged. The sequences with similar structure can be tested for similar binding affinity.

4.4.4 DREME Motif Analyzer

It was found that there were no common motifs among the top 75 sequences in each of the three binding pools. Based on this, perhaps there was no binding convergence of the pools. It must be noted that convergence of sequence is not a defining characteristic of binding, however it can be a method of identificiation.

4.5 Conclusion

Using various methods of screening the raw sequencing data, 14 potential aptamers were identified. This includes Aptamer 1, which was the sequence with the most repeats in all three binding pools. All 14 aptamers contained at least one G quadruplex, which bodes well for the quality of the final aptamer product. Furthermore, when analyzing structure, many of the aptamers have similar structures or have structures with some similarity to others. It remains to be tested if the binding of each of these identified aptamers against the protein Glycophorin A, and against whole red blood cells.

Chapter 5: Sequence Analysis

5.1 Abstract

While analysis of structural and frequency data of sequences is useful, a better way to determine if an aptamer is a strong binder is to directly test it with the target of interest. Through flow cytometry analysis, all 14 aptamers selected in Chapter 4 were tested for binding in triplicates against the Glycophorin A protein. After this, all 14 aptamers were tested in a similar fashion against whole human red blood cells (RBCs) to determine if the binding to the protein carried over to the whole cells. It was determined that aptamers 1, 5, and 6 had the strongest binding to the protein. Further testing needs to be done to confirm strongest binders to the RBCs. It was also determined that the aptamers did not share binding with the antibodies to the same protein.

5.2 Background

There are multiple methods to determine binding depending on the role of the aptamer, including fluorescent imaging after incubation with the target, and flow cytometry analysis (57). Here, flow cytometry provides a means of directly comparing the binding of multiple aptamers and comparing directly with the starting library. To do this, each ordered aptamer was attached to a cy5 fluorescent sequence which was complementary to the forward primer of all the aptamers. It was reasoned that the primer region, being non-unique, was unlikely to contribute to the binding of the aptamer. The fluorescence at 633nm was analyzed (the fluorescent wavelength of cy5), and the median of all generated peaks was compared.

In this chapter, flow cytometry analysis is first used to analyze the 14 aptamers when incubated with the protein glycophorin A. Following this I analyzed some initial

experiments of the 14 aptamers when incubated with whole red blood cells. Finally, when all 14 aptamers were mixed with whole red blood cells, the antibody FITC labelled anti-human Glycophorin A antibody clone HI264 was introduced to see if any of the. Aptamers share binding locations with the antibody.

5.3 Results

5.3.1 Aptamer Binding with Glycophorin A Beads

Each selected aptamer (See Chapter 4) and the starting library were incubated at a concentration of 100nM with $1.75x10^8$ beads with glycophorin A for 45 minutes. In addition, FITC labelled anti-human Glycophorin A antibody clone HI264 was also incubated at a concentration of 0.01mg/mL to verify the presence of the protein. The FITC fluorescence of the sample incubated with the antibody was high, verifying tpresence of glycophorin A. Figures 5.1- 5.3 show the median fluorescent intensity of each of the samples. Note that due to the limit of how many samples the Gallios flow cytometer could run at one time, the experiment was split into three. Unique controls had to be run for all three experiments, as flow cytometry results are relative, and can only be compared in relation to a control.

It is apparent that all 14 aptamers showed higher binding than the starting library. Aptamers 1, 5, 6 showed the highest binding. Aptamers 11 through 14 showed very weak binding.

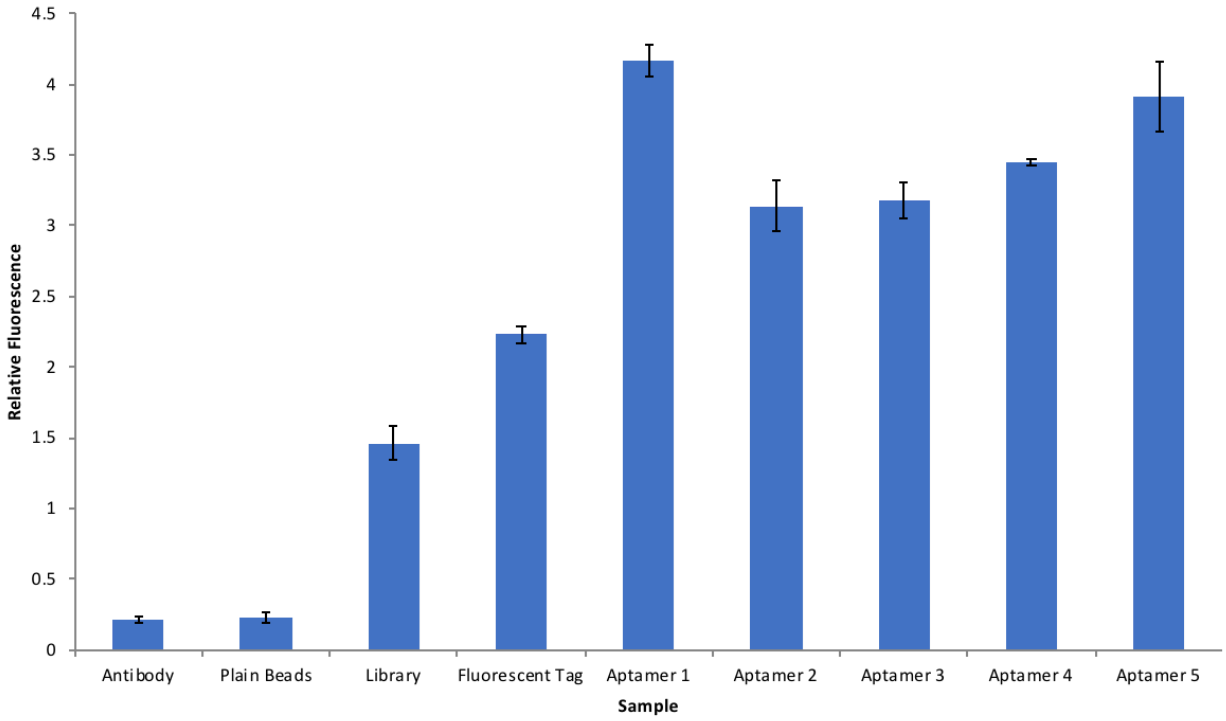


Figure 5.1: Median fluorescence of aptamers 1 through 5 when incubated with glycophorin A coated beads. All samples were incubated with glycophorin A coated beads for 45 minutes, then washed with PBS three separate times. Aptamers 1-5, starting library, and fluorescent tag were at a concentration of 100nM. The sample size for all samples is three. The error bars represent the standard deviation between the three samples.

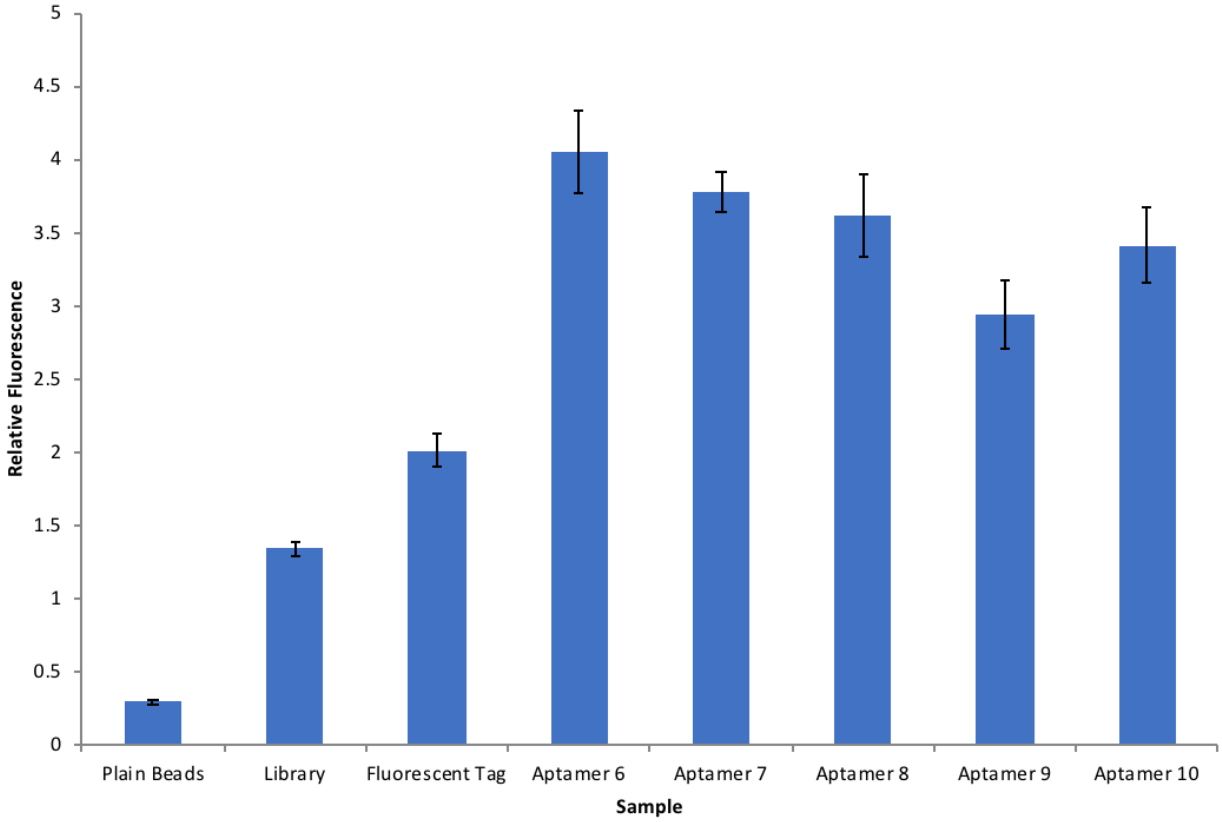


Figure 5.2: Median fluorescence of aptamers 6 through 10 when incubated with glycophorin A coated beads. All samples were incubated with glycophorin A coated beads for 45 minutes, and washed with PBS three separate times. Aptamers 6-10, starting library, and fluorescent tag were at a concentration of 100nM. The sample size is three. The error bars represent the standard deviation among the three samples.

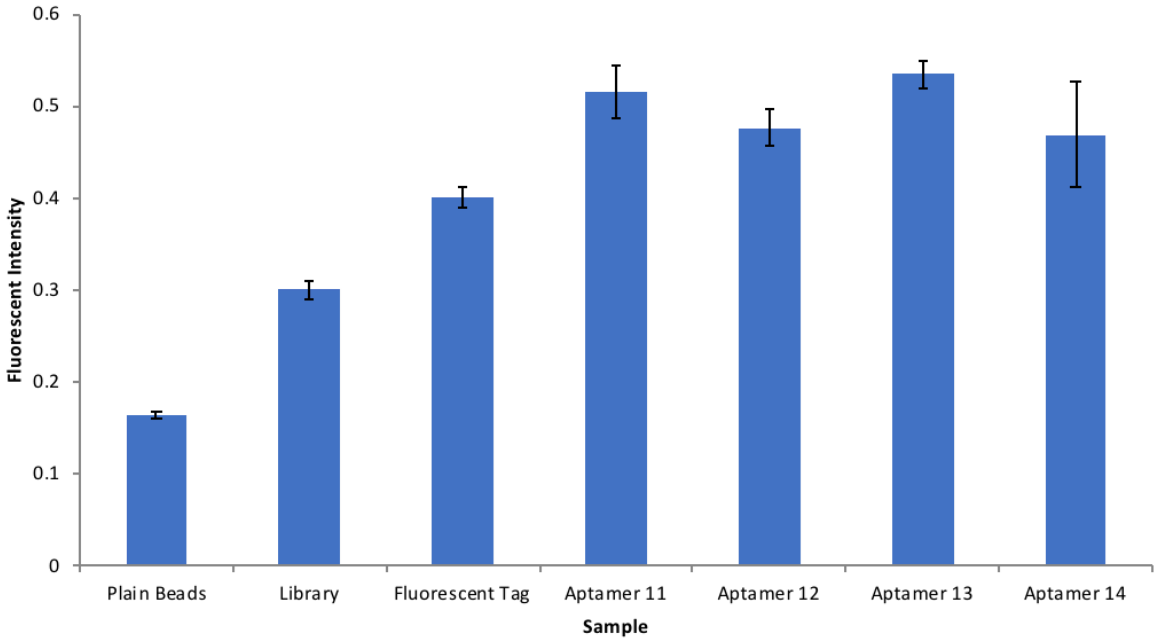


Figure 5.3: Median fluorescence of aptamers 11 through 14 when incubated with glycophorin A coated beads. All samples were incubated with glycophorin A coated beads for 45 minutes, and washed with PBS 3 separate times. Aptamers 11-14, starting library, and fluorescent tag were at a concentration of 100nM. The sample size is three. The error bars represent the standard deviation among the three samples.

5.3.2 Aptamer Binding with Whole Red Blood Cells

After measuring binding to Glycophorin A binding can be tested with whole red blood cells. Each aptamer, along with the fluorescent tag and the starting library were incubated with 150,000 RBCs for 45 minutes at concentrations of 25nM, 100nM, and 500nM. FITC labelled anti-human glycophorin A antibody clone HI264 was also incubated at a concentration of 0.01mg/mL to verify that Glycophorin A can be detected on the surface of whole red blood cells. The FITC fluorescence of the sample incubated with the antibody was high, verifying that the Glycophorin A on the surface of the RBCs could easily be detected. Figure 5.4 shows the fluorescent intensity of the samples at 25nM. Figure 5.5 shows the fluorescent intensity at 100nM. Figure 5.6 shows the fluorescent intensity at 500nM. The graph of the 25nM fluorescence shows very little difference between any of the aptamers, but a consistent improvement in fluorescence over the starting library. The graphs of the 100nM fluorescence shows aptamers 1 and 8 having the strongest fluorescence, but all aptamers were consistently higher than the library. The graph of the 500nM fluorescence shows aptamers 1, 7, 13, and 14 having the highest fluorescence. Some aptamers such as 8, 9, and 12 are around the same fluorescence as the library.

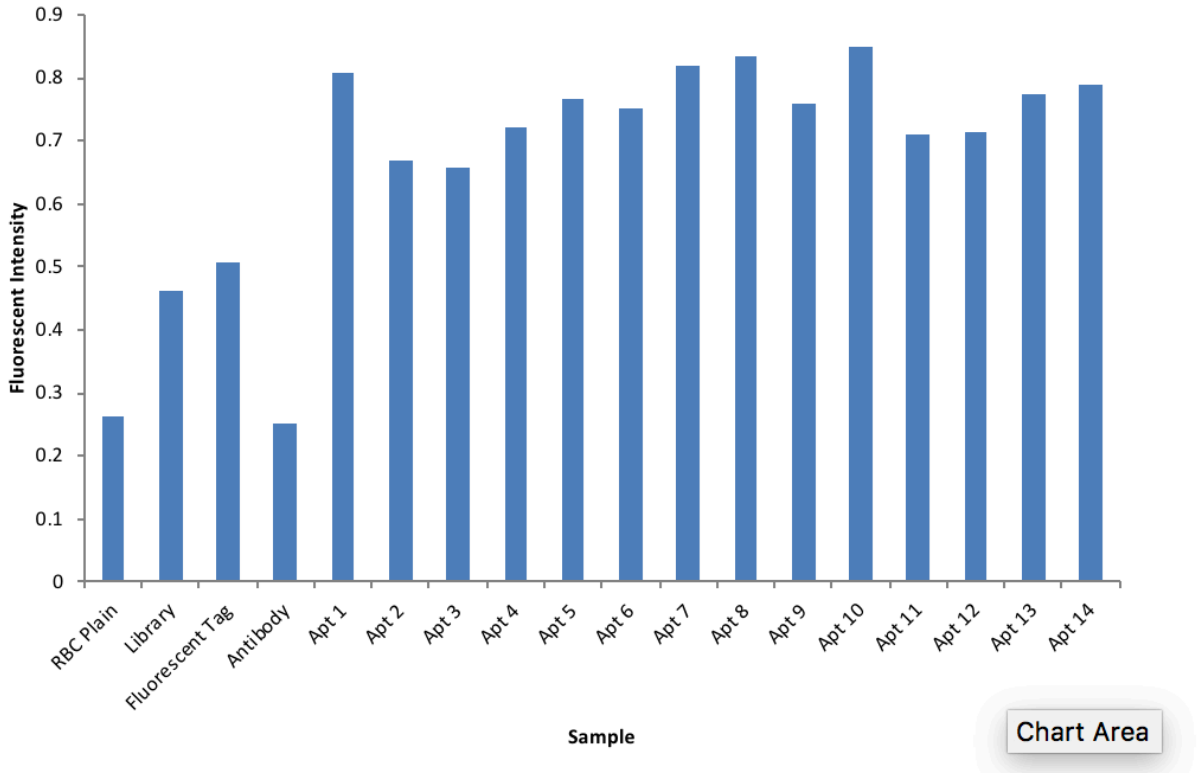


Figure 5.4: Median fluorescence of all selected aptamers with whole red blood cells at a concentration of 25nM. All samples were incubated with 150,000 RBCs for 45 minutes, washed with PBS three separate times. All aptamers, starting library, and fluorescent tag were at a concentration of 25nM. The sample size one.

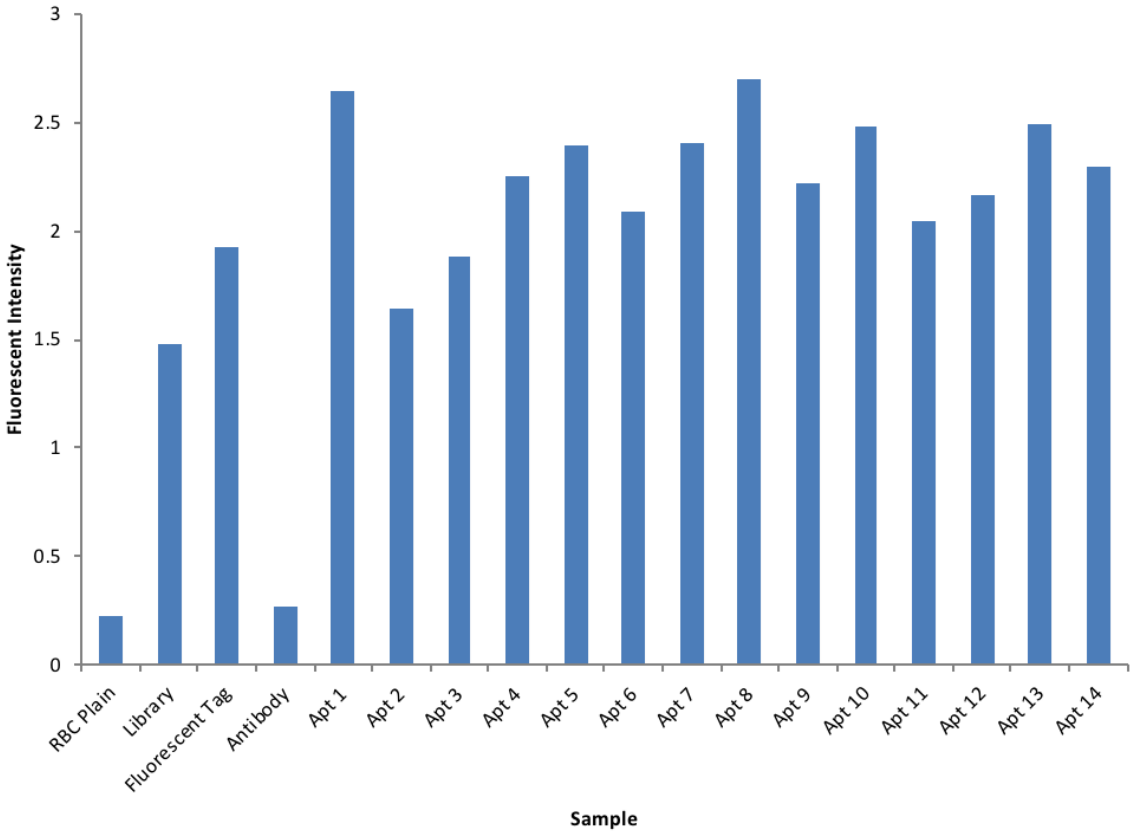


Figure 5.5: Median fluorescence of all selected aptamers with whole red blood cells at a concentration of 100nM. All samples were incubated with 150,000 RBCs for 45 minutes, washed with PBS three separate times. All aptamers, starting library, and fluorescent tag were at a concentration of 100nM. The sample size is one.

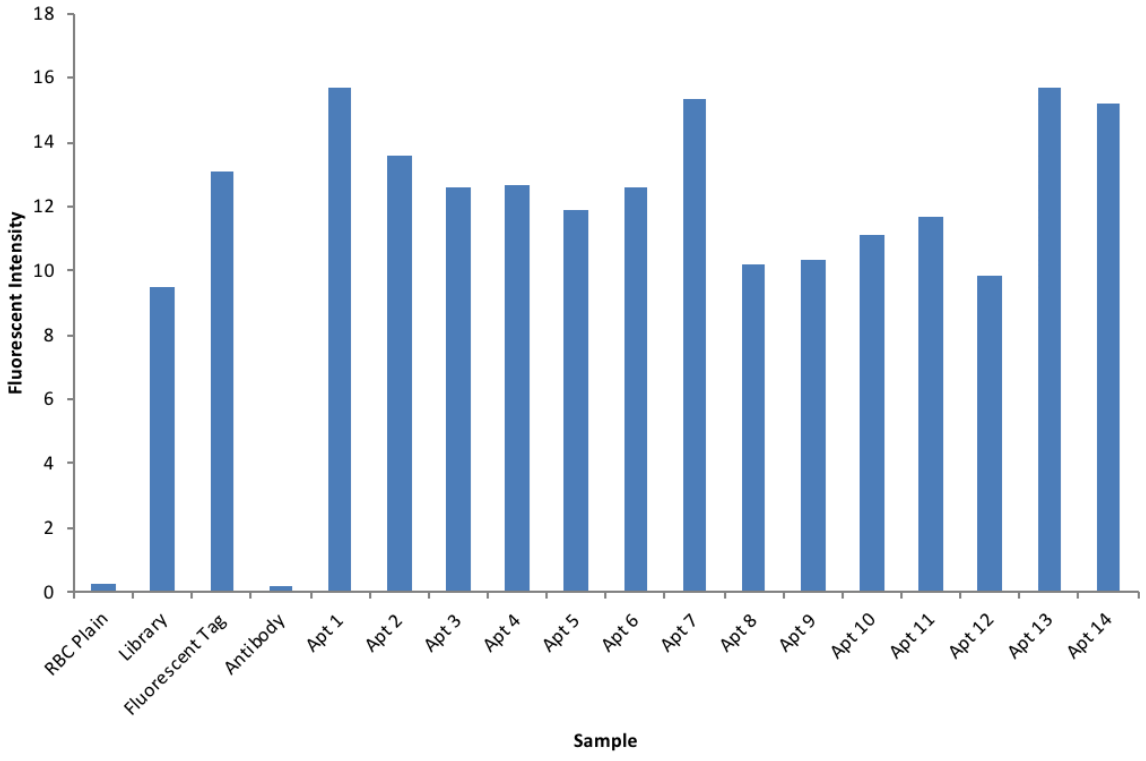


Figure 5.6: Median fluorescence of all selected aptamers with whole red blood cells at a concentration of 500nM. All samples were incubated with 150,000 RBCs for 45 minutes, washed with PBS 3 separate times. All aptamers, starting library, and fluorescent tag were at a concentration of 500nM. The sample size is one.

5.3.3 RBC Displacement Experiment

In an effort to see if the aptamers and antibody were binding to the same target on the RBC, all of the samples at 100nM were spiked with the FITC labelled anti-human glycophorin A antibody clone HI264 at a concentration of 0.01mg/mL after the initial experiment shown in Figure 5.5. They were incubated for 15 minutes, then washed thrice. Figure 5.7 shows the Cy5 fluorescence before and after this spike of antibody. Figure 5.8 shows the FITC fluorescence before and after this antibody spike. In Figure 5.7, the cy5 fluorescence of all aptamers only decreases by a small amount for all samples after the introduction of antibody. In Figure 5.8, the FITC fluorescence for all samples raises after the introduction of the antibody.

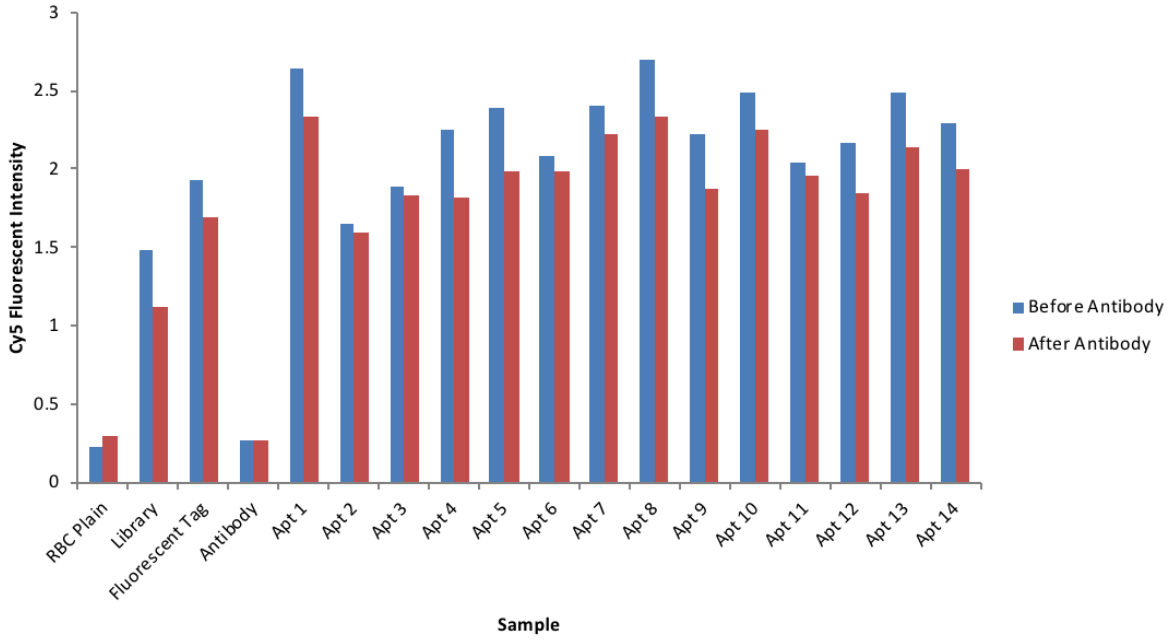


Figure 5.7: Median Cy5 fluorescence of all selected aptamers before and after the introduction of glycophorin A antibody. All samples were incubated with 150,000 RBCs for 45 minutes, then washed with PBS 3 separate times. All aptamers, starting library, and fluorescent tag are all at a concentration of 100nM. FITC labelled anti-human glycophorin A antibody clone HI264 at a concentration of 0.01mg/mL was added to the samples and incubated for 15 minutes. The samples were washed thrice more. The sample size for all samples is one.

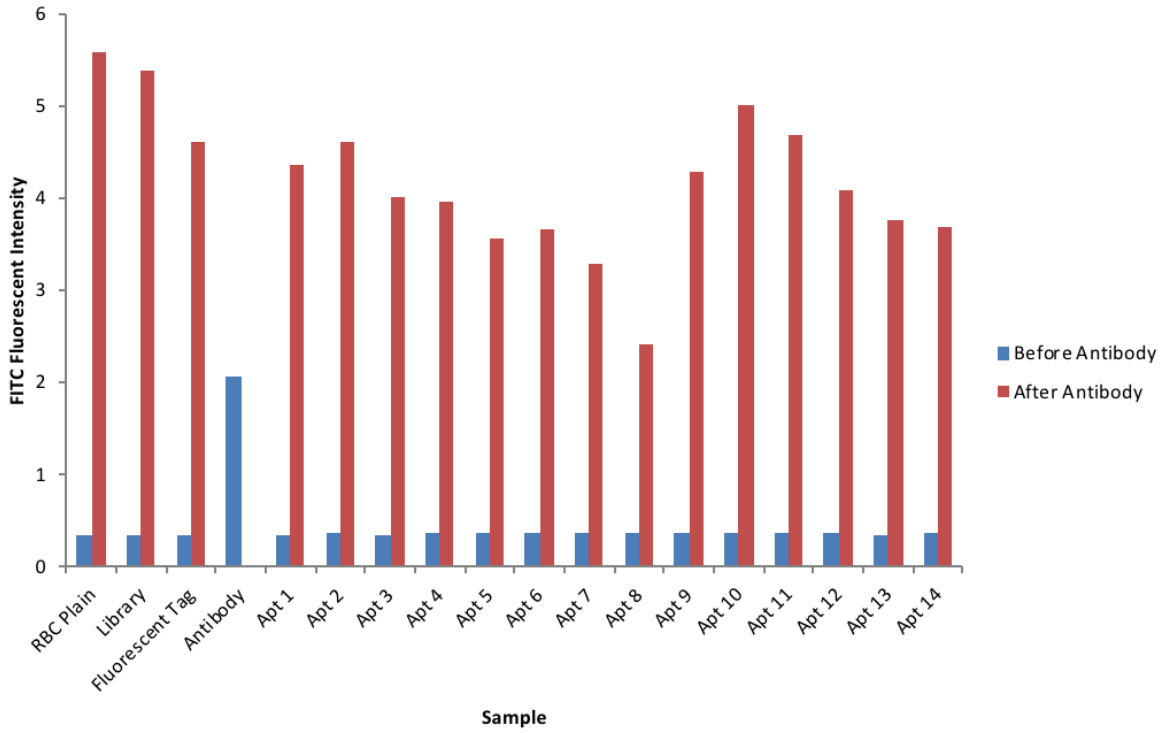


Figure 5.8: Median FITC fluorescence of all selected aptamers before and after the introduction of glycophorin A antibody. All samples were incubated with 150,000 RBCs for 45 minutes, then washed with PBS 3 separate times. All aptamers, starting library, and fluorescent tag are all at a concentration of 100nM. FITC labelled anti-human glycophorin A antibody clone HI264 at a concentration of 0.01mg/mL was added to the samples and incubated for 15 minutes. The samples were washed thrice more. The sample size for all samples is one.

5.4 Discussion

5.4.1 Aptamer Binding with Glycophorin A Beads

It is apparent that the strongest binding aptamers are Aptamer 1, Aptamer 5, and Aptamer 6. All three of these aptamers share a general predicted structure. This suggests that this structure may be effective at binding to the protein of interest. Furthermore, the later aptamers (11 through 14) had the most unique structure and also had the weakest binding. This further adds to the suggestion that the common structures of the first selected aptamers are the most effective at binding. The fluorescent tag was also tested with the starting library and aptamers as a control. The fluorescence is consistently stronger than that of the library but lesser than that of the aptamers. This is likely stronger than the library because the tag is small and sticky, therefore making it harder to wash away with only three washes. Simultaneously this indicates that this is NOT responsible for all of the fluorescence, however, as the fluorescence is still not as strong as the actual aptamers. The difference between the three trials conducted for each of the samples was very small. This is as expected and shows that the fluorescence generated by the aptamers is consistent between experiments. Overall, it would seem that nearly all 14 aptamers could be classified as effective aptamers for Glycophorin A, however, aptamers 1, 5 and 6 could be developed as the most effective binders.

5.4.2 Aptamer Binding with Whole Red Blood Cells

Three different concentrations of aptamers were tested against the red blood cells, as it was unknown how highly concentrated the aptamers needed to be to detect the Glycophorin A protein. It is very important to note that these are initial experiments and

not intended to be the final demonstration of binding to the whole cells. The main problem with the initial experiments is the lack of consistency between the experiments. The only aptamer consistently the highest binder was aptamer 1. This is interesting as aptamer 1 was also the best binder to the protein. Throughout the three experiments, the other aptamers were constantly switching places, especially aptamer 8, which was consistently one of the strongest binders for the concentrations of 25 and 100nM but was barely stronger than the library in the 500nM experiment. It was also interesting to note that aptamers 13 and 14, which show next to no binding to the protein, had strong binding to the RBCs in all three experiments. The likely reason for this is the many different targets on the RBC surface. Aptamers 13 and 14 are likely binding to a target other than glycophorin A. Furthermore, the main issue with these experiments is reproducibility. It is therefore important that these experiments be repeated in triplicate, similar to the protein experiments, in order to verify that these results are consistent and not a fluke. Going forward however, the concentration of 100nM, the same as that used in the protein experiment, is likely sufficient based on these initial experiments. It is also important to note that these whole blood experiments will likely be more difficult in general, due to the high number of targets on the RBC surface, which represents potential binding partners to aptamers.

5.4.3 RBC Displacement Experiment

It is clear from this experiment that the antibody has no effect on the binding of the aptamers or vice versa. The only small decrease in Cy5 fluorescence was likely caused by a combination of another 15 minute wait and 3 more washes with PBS. Other than this small decrease in Cy5 fluorescence, the aptamers look similar in binding before and

after antibody introduction. Furthermore, the antibodies show binding to the RBC after being introduced. Thus, the aptamers and antibodies may not bind to the same part of the Glycophorin A protein. It may also be true that some of the aptamers are not binding to Glycophorin A at all (see my assessment about Aptamers 13 and 14) but a better way to determine this would be to conduct mass spectrometry experiments on the RBCs post aptamer binding.

5.5 Conclusion

Through the experiments in this chapter, it was concluded that several of the chosen aptamers, notably aptamers 1, 5, and 6 were possible binders to the protein Glycophorin A. Furthermore, it appears that several of these aptamers arebinders to whole red blood cells, however more research needs to be conducted. Finally it was found that the FITC labelled anti-human Glycophorin A antibody clone HI264 did not share binding sites with any of the 14 aptamers.

Chapter 6: Conclusions and Future Work

First, it has been shown that the switch to emulsion PCR removed unwanted byproduct from the aptamer pools. This was shown in both Figure 3.3 and in the fact that nearly all sequencing results were the exact correct sequence length, with very little extended by-product. Second, it has been shown that at least partial automation of the aptamer process is possible, as this selection was successful and utilized a robotic portion. Third, the data suggests that aptamers 1, 5, and 6 are very strong binders to the Glycophorin A protein as seen in Figures 5.1- 5.3. It has been shown that when binding to whole RBCs, the FITC labelled anti-human Glycophorin A antibody clone HI264 does not share binding sites with any of the selected aptamers (Figure 5.7 and Figure 5.8). There are problems with the results shown, simply in that more work needs to be done to make more solid conclusions. First, each of the aptamers should be tested against a non-glycophorin A protein to confirm that the aptamers bind only to glycophorin A and are not simply sticky in nature. While negative selection was conducted during the selection, this test is a better way to confirm GYPA specificity. Next, triplicate experiments with 100nM of each aptamer against whole red blood cells should be conducted. This is due to the promising, but somewhat conflicting RBC results collected in this work. Next, the antibody displacement test should be conducted with the plain protein on beads, to eliminate the some variables and ensure that the aptamers and antibody do not share binding sites on the protein. Finally, Mass Spectrometry could be conducted on the RBCs after binding to the aptamers, to determine that the aptamers

are in fact binding to the Glycophorin A protein, and not one of the many other targets on the RBC surface.

In terms of automation, this work was an initial experimental foray into robotic aptamers. Now that it has been shown that the more simplistic steps in the aptamer selection process can be automated effectively, steps should be taken to develop this methodology further. This would include the automation of the ePCR method and gel clean-up steps to achieve full automation. If this was possible, aptamer selection could be 100% robotic and could be achieved in the space of one day, as the robot could work 24 hours.

References

1) Ellington AD, Szostak JW. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*. **346**:818-822.

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

3) Eaton B, Gold L, Zichi DA. (1995) Let's get specific: The relationship between specificity and affinity. *Chemistry and Biology*. **2**: 633-638

4) Cao HY, Yuan AH, Shi XS, Chen W, Miao Y. (2014) Evolution of a cell-specific DNA aptamer by live cell SELEX. *Oncol Rep*. **32**:2054-2060

5) Ding F, Guo S, Xie M, Luo W, Yuan C, Huang W, et al. (2015) Cell-SELEX: in vitro selection of synthetic small specific ligands. *Methods Moleculor Biology*. **1296**: 213-224

6) Rajendran M, Ellington AD. (2008) Selection of fluorescent aptamer beacons that light up in the presence of zinc. Anal Bioanal Chem. 390: 1067-1075

7) Svobodova M, Bunka DH, Nadal P, Stockley PG, O'Sullivan CK. (2013) Selection of 2'F-modified RNA aptamers against prostate-specific antigen and their evaluation for diagnostic and therapeutic applications. *Anal Bioanal Chem*. **405**: 9149-9157

8) Bruno JG, Kiel JL. (2002) Use of Magnetic beads in selection and detection of biotoxin aptamers by electrochemiluminescence and enzymatic methods. *Biotechniques*. **32**: 178,180,182-183.

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

10) Meyer S, Maufort JP, Nie J, Stewart R, et al. (2013) Development of an efficient targeted cell-SELEX procedure for DNA aptamer reagents. *PLoS ONE*. **8**: E71798

11) Ohuchi S. (2012) Cell-SELEX Technology. *Biores Open Access*. **1**: 265-72

12) Kupakuwana GV. Crill JE, McPike MP, Borer PN. (2011) Acyclic identification of aptamers for human alpha-thromin using over-represented libraries and deep sequencing. *PLoS One*. **6**: e19395

13) Mahlknecht G, Maron R, Mancini M, Schechter B, et al. (2013) Aptamer to ErbB-2/HER2 enhances degradation of the target and inhibits tumorigenic growth. *Proc Natl Acad Sci USA*. **110**: 8170-8175
14) Kubik MF, Stephens AW, Schneider D, Marlar RA, Tasset D. (1994) High affinity RNA ligands to human alpha-thrombin. *Nucleic Acids Res.* **22**: 2619-2626

15) Ozer A, Pagano J, Lis J. (2014) New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization. *Mol Ther Nucleic Acids*. **3(8):** e183

16)Yegnasubramanian S. (2013) Explanatory chapter: next generation sequencing. *Methods in Enzymology*. 5**29**: 201-208

17) Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*. **355**: 564-566

18) Hoon S, Zhou B, Janda KD, Brenner S. (2011) Aptamer Selection by high-throughput sequencing and informatics analysis. *Biotechniques*.

19) Cho M, Xiao Y, Nie J, Stewart R, Csordas AT, Oh SS, Thomson JA, Sog HT. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high throughput sequencing. *Proceedings of the National Acadamy of Sciences of the United States of America*. **107**: 15373-15378

20) Zhu B, Murthy SK. (2013) Stem Cell Separation Technologies. *Curr Opin Chem* Eng. **2(1):** 3-7

21) Putnam DD, Namasivayam V, Burns MA. (2003) Cell affinity separations using magnetically stabilized fluidized beds - Erythrocyte subpopulation Fractionation utilizing a lectin-magnetite support. *Biotechnology and Bioengineering*. **81(6):**650–665

22) Gothard D et al. (2011) In search of the skeletal stem cell: isolation and separation strategies at the macro/micro scale for skeletal regeneration. *Lab Chip.* **11(7):** 1206-1220.

23) Didar TF, Tabrizian M. (2010) Adhesion based detection, sorting and enrichment of cells in microfluidic Lab-on-Chip devices. *Lab on a Chip* **10(22):** 3043-3053

24) Kubik MF, Stephens AW, Schneider D, Marlar RA, Tasset D. (1994) High-affinity RNA ligands to human alpha-thrombin. *Nucleic Acids Res.* **22(13):**2619-2626

25) Seligman J, Slavin S, Fabbian I. (2009) A method for isolating pluripotent/multipotent stem cells from blood using the pluripotent and germ-line DAZL gene as a marker. *Stem Cell Dev*. **18(9):** 1263-1271

26) Facer CA. (1983) Merozoites of P. falciparum require glycophorin for invasion into red cells. *Bull Soc Pathol Exot Filiales.* **76(5):** 463-469

27) Grada A, Weinbrecht K. (2013) Next Generation sequencing: methodology and application. *J Invest Dermatol*. **133(8)**: e11

28) Kircher M, Kelso J. (2010) High-throughput DNA sequencing – concepts and limitations. *Bioessays*. **32(6)**: 524-536

29) Cohen SN, Chang ACY, Hsu L. (1972) Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-factor DNA. *Proc Natl Acad Sci USA*. **69(8):** 2110-2114

30) Yanisch- Perron C, Vieira J, Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. **33(1):** 103-119

31) Vanhercke T, Ampe C, Tirry L, Denolf P. (2005) Reducing mutational bias in random protein libraries. *Anal Biochem*. **27(16)**: 6008-6013

32) Shendure J, Ji H. (2008) Next-generation DNA Sequencing. *Nat Biotechnol*. **26(10):** 1135-1145

33) Schutze T, Wilhelm B, Greiner N, Braun H, Peter F, Morl M, et al. (2011) Probing the SELEX process with next-generation sequencing. *PLoS One*. **6(12):** e29604

34) Shao K, Ding W, Wang F, Lu H, Ma D, Wang H. (2011) Emulsion PCR: A High Efficient way of PCR amplification of random DNA libraries in aptamer selection. *PLoS One*. **6(9):** e24910

35) Musheev MU, Krylov SN. (2006) Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue. *Anal Chem Acta*. **564**: 91-96

36) Huang CJ, Lin HI, Shiesh SC, Lee GB. (2012) An integrated microfluidic system for rapid screening of alpha-fetoprotein-specific aptamers. *Biosens Bioelectron.* **35(1):** 50-55

37) Shao K, Shi X. Zhu X. Cui L, Shao Q, Ma D. (2017) Construction and optimization of an efficient amplification method of a random ssDNA library by asymmetric emulsion PCR. *Biotechnol Appl Biochem*. **64(2):** 239-243

38)Jellinek D, Green LS, Bell C, Janjic N. (1994) Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor. *Biochemistry*.**33(34):**10450-10456

39) Wiegand TW, Williams PB, Dreskin SC, Jouvin MH, Kinet JP, Tasset D. (1996) High-Affinity oligonucleotide ligands to human IgE inhibit binding to Fc epsilon receptor I. *Immunology*. **157(1):** 221-230

40) Viglasky V, Hianik T. (2013) Potential uses of G-quadruplex-forming aptamers. *Gen Physiol Biophys***. 32(2):** 149-172

41) Rhodes D, Lipps H. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research*. **43(18):** 8627-8637

42) Hakimi M, Hyhlik-Durr A, von Au A, Betz M, Demirel S, Dihlmann S, Bockler D, Gross-Weissmann ML. (2013) The expression of glycophorin A and osteoprotegerin is locally increased in carotid atherosclerotic lesions of symptomatic compared to asymptomatic patients. *Int J Mol Med.* **32(2):** 331-338

43) Rahimizadeh K, Alshamaileh H, Fratini M, Chakravarthy M, Stephen M, Shigdar S, Veedu RN. (2017) Development of Cell-Specific Aptamers: Recent Advances and Insight into the Selection Procedures. *Molecules*. **22(12):** e2070

44) Wang C, Zhang M, Yang G, Zhang D, Ding H, Wang H. (2003) Single-stranded DNA aptamers that bind differentiated but not parental cells: subtractive SELEX. *J Biotechnol*. **102(1):** 15-22

45) Tucker WO, Kinghorn AB, Fraser LA, Cheung YW, Tanner JA. (2018) Selection and characterization of a DNA aptamer specifically targeting human HECT ubiquitin ligase WWP1. *Int J Mol Sci*. **19(3):** E763

46) Goecks J, Nekrutenko A, Taylor J, Galaxy Team. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. **11(8):** R86,2010-11-8-r86.

47) Bailey TL, Johnson J, Grant CE, Noble WS. (2015) The MEME Suite. *Nucleic Acids Res*.

48) Kikin O, D'Antonio L and Bagga P. (2015) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences *Nucleic Acids Research*. **34 (Web Server issue)**:W676-W682.

49) Owczarzy R, Tataurov A, Wu Y, Manthey JA, et al. (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res*. **36(Web Server Issue):** W163-W169.

50) VEGF Inhibition Study in Ocular Neovascularization (V.I.S.I.O.N.) Clinical Trial Group, D'Amico, D. J, Masonson, H. N, Patel, M, Adamis, A. P, Cunningham, E. T, Guyer, D. R, & Katz, B. (2006) Pegaptanib sodium for neovascular age-related macular degeneration: two-year safety results of the two prospective, multicenter, controlled clinical trials. *Ophthalmology* **113**: 992–1001.e6.

51)Tolle F, Wilke J, Wengel J, Mayer G. (2014) By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLoS One.* **9(12):** e114693.

52) Porter E, Polaski J, Morck M, Batey R. (2017) Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol*. **13(3):** 295-301.

53) Lin H, Zhang W, Jia S, Guan Z, Yang C, Zhu Z. (2014) Microfluidic approaches to rapid and efficient aptamer selection. *Biomicrofluidics*. **8(4):** 041501.

54) Ahn J.Y., Jo M., Dua P., Lee D.K., Kim S. (2011) A sol-gel-based microfluidics system enhances the efficiency of RNA aptamer selection. *Oligonucleotides*. **21**:93–100.

55) Bae H., Ren S., Kang J., Kim M., Jiang Y., Jin M.M., Min I.M., Kim S. (2013) Sol-gel selex circumventing chemical conjugation of low molecular weight metabolites discovers aptamers selective to xanthine. *Nucleic Acid Ther*. **23:** 443–449

56) Afgan E, Baker D, Van den Beek M, Blankenberg D, Bouvier D, Čech M,Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, and Goecks J. (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research.* **44(W1)**: W3-W10 doi:10.1093/nar/gkw343

57) Alam K, Tawiah KD, Lichte MF, Porciani D, Burke DH. (2017) A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly In Vivo. ACS Synth Biol. **6(9):** 1710-1721.

58) Sundquist W, Klug A. (1989) Telomeric DNA dimerizes by formation of quinine tetrads between hairpin loops. *Nature*. 342 (6251): p825-829.

59) Mineev KS, Bocharov EV, Goncharuk MV, Arseniev AS, Volynsky PE, Efremov RG. (2011) Spatial structure of the dimeric transmembrane domain of glycophorin A in bicelles solution. *Acta Naturae*. PMID:22649687.