University of Alberta

Novel Selection of Aptamers Against Multiple Proteins

by

Jing Zhang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Medical Sciences – Laboratory Medicine and Pathology

Department of Laboratory Medicine and Pathology

Fall 2009
Edmonton, Alberta

Your file Votre référence
ISBN: 978-0-494-71249-8
Our file Notre référence
ISBN: 978-0-494-71249-8

Examining Committee

Xing-Fang Li, Laboratory Medicine and Pathology

X. Chris Le, Laboratory Medicine and Pathology

Gregory J. Tyrrell, Laboratory Medicine and Pathology

Leluo Guan, Agricultural, Food and Nutritional Science

Abstract

Aptamers have strong binding affinity and specificity to target proteins, and can be used in a variety of applications. However, current methods for aptamer selection are tedious and difficult.

The primary objective of this thesis is to develop a new aptamer selection technique for multiple proteins. A novel single cycle selection (SCS) method was developed based on gel electrophoresis separation, followed by transferring proteins to a PVDF membrane, and incubation with an 80-nt-long ssDNA library. New aptamers against thrombin, lysozyme, and Fpg protein were generated and their sequences were identified. The secondary structures of the new aptamers against thrombin and lysozyme are similar to those of the known aptamers. An aptamer against bacterial Fpg protein was selected with K_d of 8 nM, confirming its high affinity. This thesis demonstrates that SCS is novel, rapid, and capable of producing high affinity aptamers for multiple proteins in one cycle.

Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Dr. Xing-Fang Li, for giving me the opportunity to work in her group and providing endless encouragement and suggestions. I remain grateful for her expertise, patience, and kindness.

I would like to thank Dr. X. Chris Le for his guidance and constructive suggestions throughout my whole project. My thanks also to my thesis committee members, Dr. Gregory J. Tyrrell and Dr. Leluo Guan, for their knowledge, advice, and help.

I very much appreciated the generous support from Ms. Cheryl Titus and Ms. Dianne Sergy. Many thanks to Ms. Katerina Carastathis for proofreading my thesis.

There are so many group members who supported me for the past three years. I sincerely thank Dr. Yanming Liu for sharing her expertise in microbiology and molecular biology with me. I am grateful to Dr. Hongquan Zhang for teaching and helping me with capillary electrophoresis and to Camille Hamula for helping me with molecular cloning work. Many thanks to Megan Wagner for helping me to edit my thesis and Feng Li for assisting me with the figures. And I greatly appreciate the friendship and advice of the great friends I made during my studies: Dr. Feng Qin, Dr. Narenmandula, Chuan Wang, Jessica Boyd, Birget Moe, Claire McGuigan, and many more.

Lastly and most importantly, I would like to thank my mother Chunrong Zhang and my father Guangzu Li, for giving me life, for providing unconditional love and support during my studies, and for teaching me the most important things in life. I would also like to thank my great sister, Nan Li, for her continuous encouragement and support.

Table of Content

List of Tables

List of Figures

List of Abbreviations

BAC	Bacterial Artificial Chromosome
BSA	bovine serum albumin
CE	capillary electrophoresis
CZE	capillary zone electrophoresis
FAM	carboxy-fluorescein
dNTPs	deoxynucleotide triphosphates
DMSO	dimethyl sulfoxide
K_d	dissociation constant
dsDNA	double-stranded DNA
EOF	electroosmotic flow
μ_{EOF}	electroosmotic mobility
μ_{EP}	electrophoretic mobility
E. coli	*Escherichia coli*
EB	ethidium bromide
EDTA	ethylenediaminetetraacetic acid
Fpg	Formamidopyrimidine DNA glycosylase
IS	internal standard
PI	isoelectric point
LIF	laser-induced fluorescence
Tm	melting temperature
MW	molecular wieght
1D	one-dimensional

TEMED	1, 2-bis(dimethylamino)-ethane
ori	origin of replication
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PVDF	polyvinylidene difluoride
SCS	single cycle selection
ssDNA	single-stranded DNA
SDS	sodium dodecyl sulfate
SELEX	systematic evolution of ligands by exponential enrichment
ΔG	the change in Gibbs free energy
TG	Tris glycine
2D	two-dimensional
X-gal	X-galactose
YAC	Yeast Artificial Chromosome

Chapter 1: General Introduction

1. Aptamers

Aptamers are short single-stranded DNA (ssDNA) or RNA sequences that have high specificity and affinity against targets, and were initially reported in 1990 (Ellington and Szostak; Tuerk and Gold). The binding process is based on the formation of stable three-dimensional structures including stem-loop and internal loops. Binding affinities (dissociation constants: K_d values) for typical aptamer binding are in the micromolar to low picomolar range (Ellington and Szostak, 1990; Jenison et al., 1994). A number of aptamers have been generated to a wide range of targets, including organic and inorganic molecules, nucleic acids, proteins, and peptides (Ellington, 2004).

High binding affinity and high specificity to target proteins have allowed aptamers to be used in place of antibodies (Ulrich, Martins, and Perquero, 2004). Compared to antibodies, aptamers have several advantages. The main advantage is that generating aptamers is a simple *in vitro* chemical process and does not require the use of animals. In theory, it is possible to select aptamers against any target protein, or to a specific region of the target, with specific binding properties under different binding conditions. Aptamers are easily and relatively inexpensively synthesized. They can be easily modified with a wide variety of chemical groups to achieve various functions (Nimjee, Rusconi, and Sullenger,

2005) or enhance chemical properties such as stability or resolvability.

These advantages have enabled aptamers to become useful tools in analytical, bioanalytical, therapeutic, and diagnostic applications. They can be used as molecular recognition probes in combination with analytical devices such as chromatography, mass spectrometry, molecular beacons, and some label-free biosensors (Tombelli, Minunni and Mascini, 2005; Pavski and Le, 2003; Tan, Wang and Drake, 2004; Hamula et al., 2006). Due to their high binding affinity and specificity, aptamers can be used in the purification, separation, and detection of target molecules of interest. It has also been reported that some aptamers can be used as therapeutic or diagnostic tools since they can modulate the activity of their target proteins. For example, an RNA aptamer that can specifically bind to reverse transcriptase (RT) was generated in 1992 (Tuerk, MacDougal, and Gold). This aptamer can inhibit the activity of RT and reduce the replication of the human immunodeficiency virus (HIV-1) by 90% - 99% as a result (Joshi and Prasad, 2002).

2. Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

The *in vitro* chemical method for generating aptamers specific to targets is the systematic evolution of ligands by exponential enrichment (SELEX), developed by Tuerk and Gold (1990). This method has permitted the identification of unique

RNA/ssDNA molecules from very large populations of random sequence oligomers (ssDNA or RNA library) with high affinity and specificity to the target.

2.1. General procedure of SELEX

Most SELEX methods involve three basic steps: (1) incubation of ligand sequences with a target; (2) partitioning of ligand-target complexes from unbound sequences via affinity methods; and (3) amplification of bound sequences.

These three basic steps are shown in Figure 1-1; the details of each step are described in the following sections.

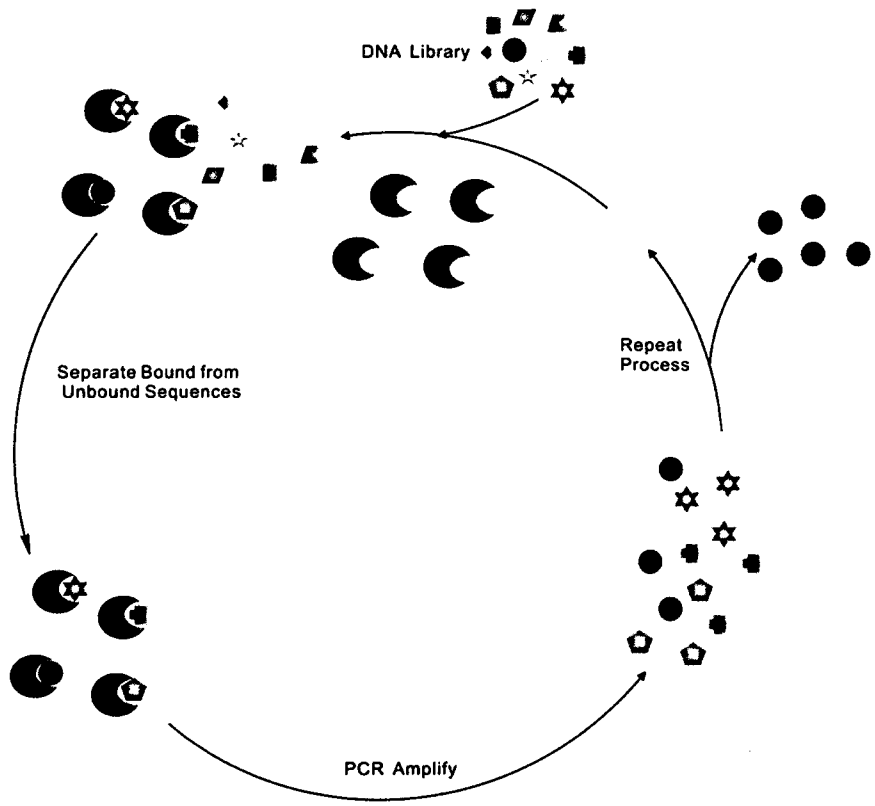


Figure 1-1: Schematic of SELEX round.

In the incubation step, nucleic acid libraries are incubated with target molecules under optimum conditions (appropriate buffer at a desired temperature). After binding, the RNA/ssDNA aptamer-target complexes are separated from nonspecific molecules; this is the partitioning step (to be described later). Bound sequences are regenerated by enzymatic amplification processes. The amplified molecules are then used in the next round of the selection processes. Selecting sequences which have the highest specificity and affinity against the target often requires 8–12 cycles of selection. The selected oligonucleotides are analyzed for their sequences and structures after cloning and sequencing. After the sequence of an aptamer is determined, that aptamer can be easily generated through nucleic acid synthesis. The aptamer's binding affinity and specificity to its target can be validated using different methods. The conventional SELEX method usually takes several weeks to months to complete, in order to achieve high efficiency and high specificity. It is a time-consuming process.

2.2. Incubation step

2.2.1. Library

The SELEX process begins with synthesizing a ssDNA or RNA library. Designing the library is very important. Each oligonucleotide comprises a central region of random sequence flanked by a 5' and a 3' region of defined primer sequences for polymerase chain reaction (PCR). Each sequence in the library is a

linear oligomer of unique sequence.

The complexity of the library is dependent on the length of the random sequence region. Any length of the randomized region may be used to provide 4^n theoretical random sequences. However, most libraries contain no more than 10^{16} random sequences because of the limitations of synthesis. To prepare a library containing one molecule each of 10^{16} random sequences, at least 16.7 nmole nucleic acids is needed. In practice, the randomized regions commonly contain 30–60 nt (Osborne and Ellington, 1997). The initial library usually contains approximately 10^{15} individual sequences. This large number of oligonucleotides permits the generation of an aptamer specific to the target of interest with high probability (Osborne and Ellington, 1997).

RNA can adopt a wide variety of conformations resulting in RNA libraries yielding aptamers with high binding affinities. However, RNA is often unstable. The stability of DNA and use of ssDNA libraries yields more stable aptamers that are often used for applications in environmental monitoring and therapy.

2.2.2. Target molecules

A variety of target molecules are used to generate aptamers using SELEX. The criterion for suitable targets is stability. In some cases, target molecules are modified with functional groups to improve stability in the partitioning step. Generally, large molecules are better targets because they offer the necessary

surface area for interaction with aptamers. Hence, proteins are the most common targets for generating aptamers.

In the incubation step, target molecules can be either incubated with a library in free solution (buffer) or they can be bound to a solid support, such as a membrane or column. Binding target molecules to solid support results in several drawbacks: (1) limiting the surface area of the target molecules for interacting with the DNA library; (2) the linker molecules used for binding the target to a solid support (column) may interact with the DNA library, resulting in high background and low efficiency of selection; (3) non-specific binding of the library to the solid surface causes high background and low efficiency. The addition of monovalent cations (such as Na^{+}) and divalent cations (such as Mg^{2+}) in buffer may reduce non-specific binding. This is another advantage of using solution incubation. Targets are usually dissolved in free solution for incubation.

In the initial incubation step, the molecular ratio of the total random oligonucleotides to target molecules is 100:1. This ensures sufficient oligonucleotide sequences for binding to the target molecules.

2.2.3. Counter-selection

In SELEX, counter-selection can be performed to eliminate non-specific aptamer sequences to the target of interest. In counter-selection, the target is replaced with an undesirable molecule for one round of selection, and the ssDNA

or RNA sequences binding the undesirable molecule are removed from the library. In the next round of selection, the desired target is used to select specific binding sequences. This increases the specificity of the aptamers for the target of interest.

2.3. Partitioning step

In the partitioning step, the bound aptamer-target complexes are separated from the unbound aptamer sequences. This is the most critical step in SELEX; the efficiency (cycle number) of SELEX is dependent on the separation efficiency of partitioning in each round. Therefore, it is vital to choose an efficient method for separation in this step.

The traditional methods of separating aptamer-target complexes from free target and unbound DNA are columns or nitrocellulose membranes (Ellington and Szostak, 1990; Tuerk and Gold, 1990; White et al., 2001). Columns and nitrocellulose membranes are used as filters to separate aptamer-target complexes and unbound sequences based on their different affinities. These techniques are simple and rapid. However, they have low separation efficiencies, thus multiple SELEX rounds have to be repeated to obtain aptamers with high affinities and specificities. Furthermore, these techniques require large amounts of both target and oligonucleotide library in the initial incubation step. The aptamers generated by these techniques have low binding affinities with K_d values in the range of nM to μM. High affinity aptamers often have K_d values in the low nM or pM range.

Photocrosslinking can be used for partitioning aptamer-target complexes from unbound sequences (Jensen et al., 1995; Golden et al., 2001). Aptamer is substituted with a fluorophore (such as 5-iodouracil or 5-bromodeoxyuridine), which facilitates covalent cross-linking of the aptamer to the target. The modified library is then irradiated by 308-nm excimer laser light during incubation with the target. Aptamers obtained via cross-linking have high binding affinities and high binding specificities with K_d values in the pM range. However, this method may be unable to generate aptamers against small target molecules since small targets may lack functional groups for cross-linking.

The separation efficiency can be improved when the target is attached to magnetic beads as a separation matrix in SELEX rounds (Stoltenberg, Reinemann, and Strehlitz, 2005). The aptamer-target complexes attached to magnetic beads can be separated from unbound sequences in solution via magnetic force. This improves the separation efficiency. Compared to traditional methods, this technique requires less quantity of the target and can yield aptamers with high binding affinity (low nM). Furthermore, aptamers selected with this method can be used to determine surface occupancy of the beads or used as a linker for conjugating other molecules to the magnetic beads.

Another partitioning method is protein gel electrophoresis combined with aptamer blotting (Noma et al., 2006). Target proteins can be separated by SDS-PAGE, transferred onto a membrane, and then treated with the aptamer on

the membrane. Aptamer-target complexes are subsequently separated by cutting the membrane. Aptamers with nM binding affinities can be selected in a few SELEX rounds.

Capillary electrophoresis (CE) provides high separation efficiency for separation of large and small molecules. The use of CE to separate aptamer-protein complexes from other unbound molecules has greatly improved the efficiency of SELEX and the affinity of selected aptamers for their protein targets (Mendonsa and Bowser, 2006). Due to the high separation efficiency (fast speed caused by high voltage), low non-specific binding, and minimal sample dilution with high detection sensitivity, CE-SELEX can generate aptamers with high affinities to their targets in only 2–4 rounds of selection (Mosing, Mendonsa, and Bower, 2005; Mendonsa and Bower, 2005). The K_d values of aptamers obtained via CE are in the low nM range. Compared to other separation methods, CE-SELEX is less labor-intensive and less time consuming to complete the whole process. Furthermore, this method can successfully generate aptamers for small targets.

2.4. Amplification (polymerase chain reaction)

Polymerase chain reaction (PCR) is a common method used in molecular biology for amplifying any piece of DNA by *in vitro* enzymatic replication (Saiki et al., 1985; Saiki et al., 1988; McPherson and MØller, 2000). PCR can amplify a

DNA sequence and generate millions of copies, and it can be adapted for use in a variety of applications. Therefore, this technique is used in this research to amplify selected sequences from SELEX rounds.

Figure 1-2 shows the process of PCR that comprises three steps: (1) Denaturation: template DNA is denatured to separate complementary stands; (2) Annealing: oligonucleotide primers hybridize to the template; (3) Extension: DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding four deoxynucleotides.

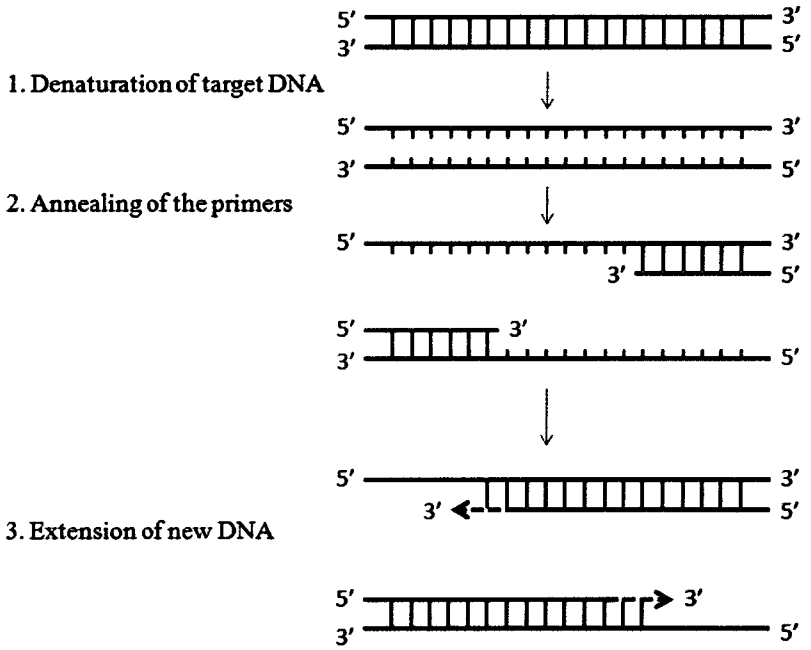


Figure 1-2: Schematic of polymerase chain reaction.

2.4.1. Reagents for PCR

To prepare a reaction mixture for PCR amplification, the reagents and components required include the DNA template containing a region to be amplified, two oligonucleotide primers (forward primer and reverse primer) that are complementary to the DNA template at the 5' or 3' ends of the amplified DNA region, four deoxynucleotide triphosphates (dNTPs) used for synthesizing new complementary DNA strands to the template strands, thermostable DNA polymerase, magnesium ions (Mg^{2+}) which are important for polymerase activity and the specificity and efficiency of the PCR, and buffer solution providing a suitable chemical environment for optimum activity of the DNA polymerase.

2.4.2. PCR procedure

In the denaturation step, double-stranded DNA (dsDNA) is separated by heating the reaction mixture to 94–98 °C for 20 to 60 seconds. Hydrogen bonds between complementary bases of the dsDNA are disrupted and ssDNA sequences are formed as a result. In the annealing step, the reaction mixture is cooled down to 50–65 °C for 20 to 60 seconds; under cool conditions, oligonucleotide primers can find and hybridize with the complementary bases in the DNA template, forming stable hydrogen bonds. In the extension step, thermostable DNA polymerase begins to extend the primers as soon as they anneal to the DNA template under a temperature that is close to the optimal polymerization

temperature. This temperature depends on the type of DNA polymerase used in PCR. *Taq* polymerase is commonly used. For this polymerase, an extension temperature of 72 °C is used for optimum activity. DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding four deoxynucleotide bases (A, T, G, C) that are complementary to the bases in the template from the 5' end to the 3' end. These three steps are usually repeated 20–45 times. The number of cycles can be adapted for desired objectives or applications. Under optimal conditions, the amount of target DNA is doubled in each cycle, resulting in the exponential amplification of the target DNA fragments. PCR is carried out in a thermal cycler that automatically regulates the temperature and time of the reaction cycles according to the preprogrammed set (optimized PCR conditions).

PCR has been widely used in molecular biology, microbiology, genetic research, clinical and environmental studies, and many other applications, due to its ability to amplify any short DNA sequence.

3. Molecular Cloning

In my research, molecular cloning is used to determine the sequences and structures of selected aptamers. Molecular cloning is a procedure of isolating a defined DNA sequence and obtaining multiple copies of the DNA sequence *in vivo* (Sambrook, Fritsch, and Maniatis, 1989).

DNA sequences are cloned into cells and a single DNA sequence from one cell can be sequenced. Molecular cloning has been widely used for DNA sequencing, gene expression, and large scale protein production.

Figure 1-3 shows the general procedures of cloning a target DNA fragment. After cleavage of a vector DNA molecule at the desired cleavage site by different types of restriction endonuclease, the amplified DNA fragment is inserted into the vector by DNA ligase and a recombinant DNA molecule is formed. In the transformation step, the recombinant molecule is taken up by a host cell. The recombinant DNA molecule containing the target DNA fragment can be multiplied with replication, amplification, and cell division. A large amount of the same target DNA fragment can be obtained after cell lysis and DNA isolation and purification.

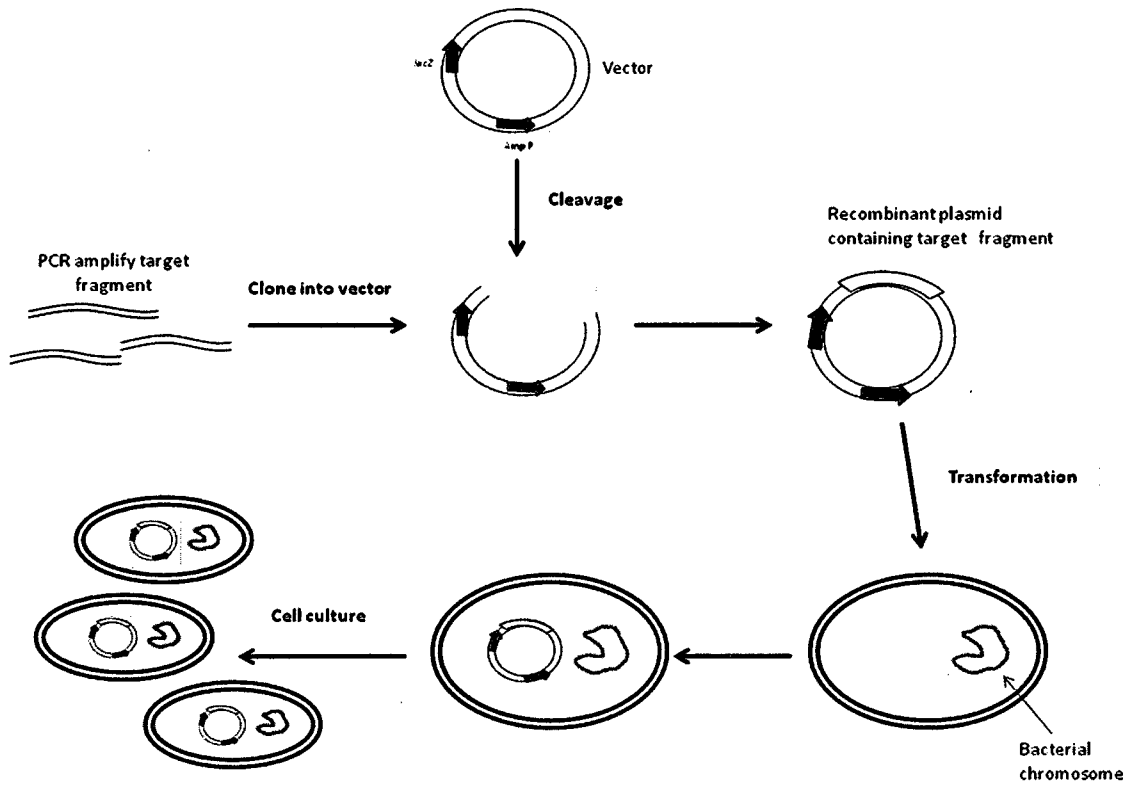


Figure 1-3: Schematic of molecular cloning of a DNA sequence.

3.1. Reagents and materials

To perform molecular cloning, the following reagents and materials are required:

3.1.1. Cloning vectors

To clone single sequences, appropriate vectors must be chosen first for inserting target DNA sequence and forming recombinant molecules. Vectors are small and autonomously-replicating DNA molecules. The type of vector depends on the length of the target DNA sequences. Plasmid from bacteria can be used for

cloning DNA up to 10 kb. Bacteriophage lambda can be used for inserting DNA sequences up to 16 kb. Cosmid is a combination of a plasmid vector with a Cos site which allows the plasmid to be packaged in a phage head, and which can carry DNA up to 45 kb. Yeast Artificial Chromosome (YAC) can clone DNA up to 2 Mb, which is replicated in yeast. However, the yeast vector is not stable in the transformation step, and the efficiency is very low. Bacterial Artificial Chromosome (BAC) is often used for the DNA sequence with a length ranging from 150 kb to 350 kb, and is more stable than YAC. The BAC vector has been used in the Human Genome Project.

Plasmids are small, closed circular, self-replicating, extra-chromosomal DNA molecules (1–200 kb). They are suitable for cloning aptamers since aptamers are usually less than 100 bp. Plasmids usually have antibiotic resistance genes in their sequences (for selection), a region called origin of replication (ori) for self-replication, and restriction sites so that target DNA sequences can be ligated in the plasmids.

3.1.2. Enzyme

A number of restriction endonucleases (EcoR1, BamH1, BglII, Dra1, HindIII, Alu1, HaeIII, and Sal1) can cleave the phosphodiester backbone of DNA sequences at specific cleavage sites, creating sticky or blunt ends. A specific restriction endonuclease is selected and used depending on the sequence of target

DNA. DNA ligase is subsequently used to reform phosphodiester bonds in the DNA backbone, and the target DNA sequence is inserted into the vector. T4 DNA ligase, a typical ligase used in molecular biology, ligates blunt ends or compatible sticky ends of target DNA sequences into the vectors.

3.1.3. Host cells

Host cells are either bacteria [such as *Escherichia coli* (*E. coli*)] or eukaryotic cells (such as mammalian cells or yeast). The uptake process via bacteria is called transformation, while the uptake process via eukaryotic cells is called transfection. Bacteria are usually used as host cells in molecular cloning for DNA sequencing, because this uptake process is faster than transfection. Mammalian cells are often used as host cells for gene expression. In transformation, *E. coli* host cells are treated with $CaCl_2$ or electroporation. This treatment results in the formation of pores in the bacterial cell wall allowing for the uptake of vectors. These treated cells are called chemically-competent or electro-competent *E. coli* cells.

3.2. Procedure of cloning aptamers

Because aptamers are short ssDNA sequences, plasmids are chosen as the vectors. Plasmid molecules and target DNA sequences are both cleaved by the same specific restriction endonuclease, in order to obtain the desired compatible sticky or blunt ends. This incubation is performed at 37 °C for 30 to 120 minutes.

Using T4 DNA ligase, the compatible ends are ligated and the target DNA sequences are inserted into the plasmid vectors. The optimum temperature for T4 ligase activity is 16 °C, and ligation takes approximately 6 to 12 hours at this temperature. In the transformation step, chemically-competent *E. coli* cells are used to clone a short DNA sequence. After heat-shock at 42 °C for 30–45 seconds, the recombinant molecules are taken up by the cells. A large number of copies of the target DNA sequences are obtained after self-replication of plasmids, amplification, and cell division.

As mentioned previously, plasmids contain antibiotic resistance genes. Bacterial cells with the plasmids also contain the antibiotic resistance genes. Thus, only cells containing the plasmids can grow and amplify on the antibiotic-containing medium. Plasmids also contain the *LacZ* gene that produces β-galactosidase. This enzyme can digest the sugar X-galactose (X-gal) and produce a blue byproduct. On medium containing X-gal, only those bacterial cells with plasmids carrying a functional *LacZ* in the polylinker (no insertion) can metabolize X-gal and produce blue colonies. Cells with recombinant plasmids have a non-functional *LacZ* gene, due to the disruption of the polylinker by the insertion of target DNA sequences. Therefore, they produce white colonies. Based on this double selection, bacterial cells containing recombinant plasmids can be selected and amplified.

After colony selection, cells without the recombinant plasmid are eliminated.

Single colonies containing the recombinant molecules are then selected and cultured individually to obtain clones of a single sequence. These cultured colonies are used to isolate and purify the desired DNA sequence. The membranes of bacterial cells are lysed by alkaline lysis (sodium hydroxide or sodium dodecyl sulfate). Plasmid DNA is subsequently precipitated and purified by isopropanol precipitation. This is often used to obtain pure DNA from large volumes of samples at room temperature. Finally, the inserted target DNA sequences can be determined based on known information of cleavage sites, upstream, and downstream sequences in plasmids.

Because molecular cloning can isolate and amplify DNA sequences, it is an ideal tool for determination of aptamer sequences. With modern DNA sequencing techniques, it is possible to develop a high throughput process for analysis of sequences of multiple aptamers.

4. Capillary Electrophoresis

Capillary electrophoresis (CE) provides high separation efficiency and requires minute sample volumes. This technique is ideal for characterizing the binding affinity of new aptamers, which has been demonstrated in a number of studies (Huang et al., 2004; Mendonsa et al. 2004; Berezovski et al., 2006). In this research, CE is used to separated aptamer-protein complexes and unbound aptamers, and detects the signals. The use of fluorescently-labeled aptamers

enables subsequent detection by laser-induced fluorescence (LIF). Based on the fluorescent output, the binding constant or dissociation constant is determined.

4.1. Capillary zone electrophoresis

The most commonly used form of CE is capillary zone electrophoresis (CZE), a simple technique that separates analyte based on a charge to size ratio (Baker, 1995; Lander, 1996; Righetti, 1996; Camilleri, 1997).

4.1.1. CZE system

Figure 1-4 shows a typical CZE system that consists of two buffer vials, a fused silica capillary, a high-voltage power supply, and two electrodes connected to the power supply. The capillary often has a very small inner diameter (5–100 μm). The small inner diameter of the capillary provides a high surface-to-volume ratio, allowing the efficient dissipation of Joule heat generated from high electric potentials used. This enables the use of high voltages, typically between 15 and 30 kV (kilovolts) for CE, leading to fast and efficient separation.

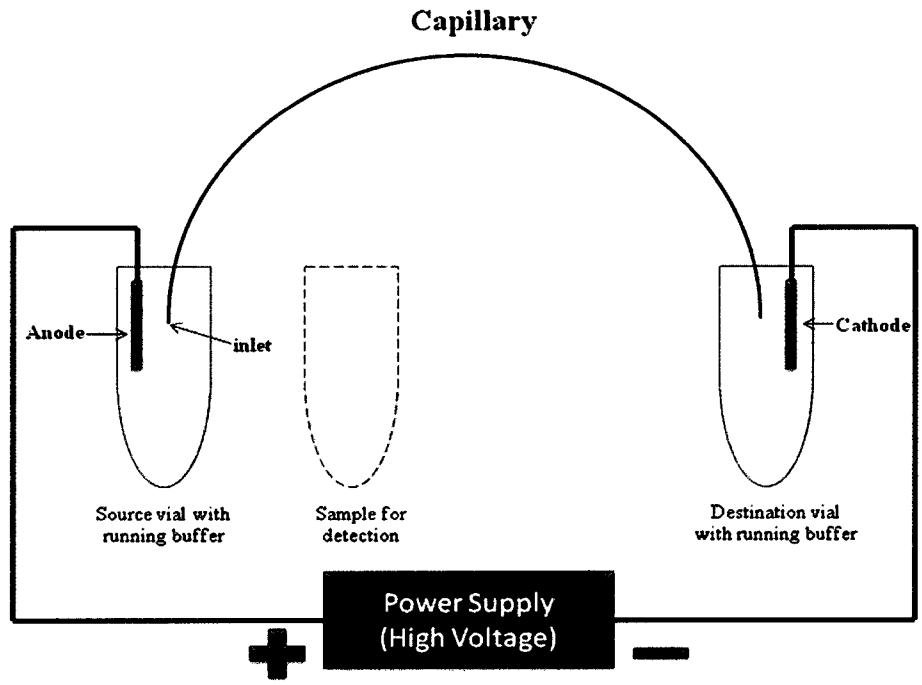


Figure 1-4: Schematic of capillary zone electrophoresis.

To perform a CZE separation, two vials are filled with a homogenous buffer of a desired pH. The selection of pH is based on the properties of analytes. An electric field is applied between the two vials. To inject a sample, the inlet vial is replaced with a sample vial, and then sample is injected electrokinetically or hydrodynamically. Electrokinetic injection of a sample is performed by applying a potential for 2–5 seconds. Hydrodynamic injection is performed simply by raising the sample vial higher than the outlet end. In this research, electrokinetic injection is used. The injected analytes migrate through the capillary with an electrophoretic mobility determined by their charge and size, and signals are detected by detectors. Different detectors can be used to achieve different

objectives, with different detection sensitivity.

Data is collected and analyzed via an integrator or a computer. The output data is displayed as an electropherogram, in which separated analytes are displayed as peaks with different migration times and signal intensity.

4.1.2. Electrophoretic mobility

Analytes are separated by their different velocity under the electric field applied in CE. The velocity of an analyte (ion) is: $v = \mu_{EP} \times E$, where v is the velocity of analyte, μ_{EP} is the eletricphoretic mobility, and E is the applied electric field. The eletrophoretic mobility is an intrinsic character of ions in a given medium, it is determined by the balancing between electric force and frictional drag through the medium; and it can be given by: $\mu_{EP} = \mathbf{q}/6\pi\eta\mathbf{r}$, where **q** is the ion charge of analyte, η is the solution viscosity, and **r** is the ion radius. Therefore, the electrophoretic mobility is dependent on the charge to mass ratio of analytes.

4.1.3. Electroosmotic flow (EOF)

In CZE, buffer is pulled through the capillary under an applied electric field without pressure. The force pumping the buffer through the capillary is known as electroosmotic flow (EOF). The generation of EOF is explained as follows. The surface of a fused silica capillary contains silanol groups. When the fused silica capillary is filled with buffer (pH > 3) under an electric field, the silanol groups

are negatively charged due to the dissociation of protons. Positively charged ions in the buffer pair with negative ions on the surface of capillary, leading to the formation of an electrical double layer, as shown in Figure 1-5. Under the application of an electric field, the cations in the diffuse portion of this double layer migrate from the anode to the cathode. This migration drags the bulk buffer from the anode to the cathode, resulting in EOF.

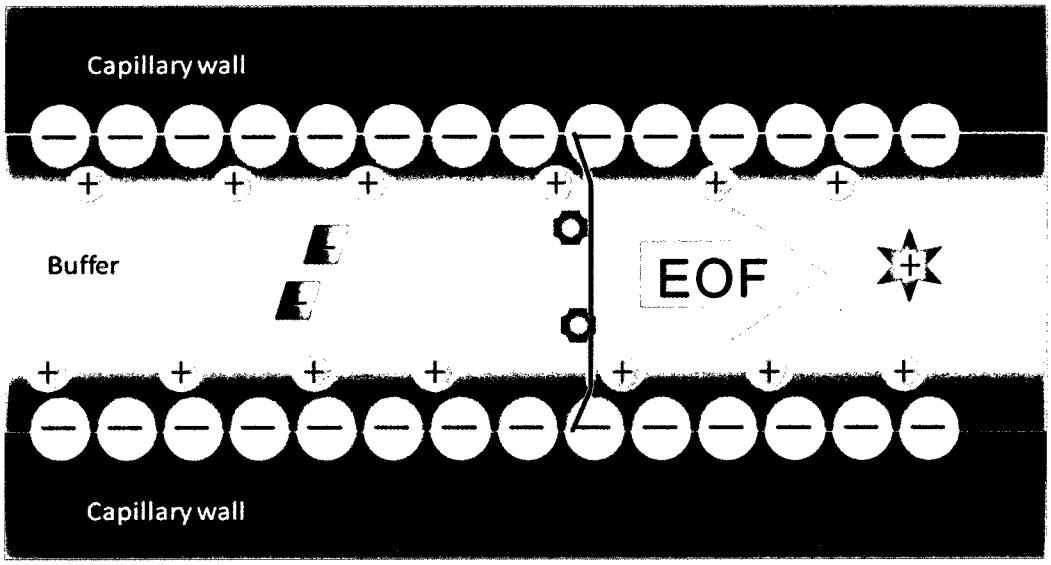


Figure 1-5: Illustration of electroosmotic flow (EOF).

As shown in Figure 1-5, the EOF profile is flat. This flat EOF flow profile enables analyte to migrate through the capillary in very narrow bands. This results in high efficiency separation. Varying several parameters can modify the EOF. Such parameters include applied voltage, buffer composition, pH, and ionic strength, temperature, and modification of the capillary wall.

In CZE, different analytes migrate through the capillary based on

electrophoretic mobility (μ_{EP}) and EOF (μ_{EOF}): $\mu = \mu_{EP} + \mu_{EOF}$. Analytes are separated based on their differences in electrophoretic mobility. Because the electrophoretic mobility of the analytes are often much smaller than the EOF, analytes moves in the same direction as the EOF regardless of their charges.

Positively charged analytes (cations) have electrophoretic mobility in the same direction as the EOF, so they move faster than the EOF. Negatively charged analytes (anions) are attracted to the anode, but they can still move toward the cathode under the EOF, so they move slower than the EOF. Neutral molecules elute with the same speed as the EOF. Thus, cations, anions, and neutrals can be separated and detected in one electrophoretic run on the CZE system, and the migration order of analytes is cations, neutrals, and finally anions.

4.2. Capillary electrophoresis with laser-induced fluorescence (CE-LIF) detection

The CE system can be coupled with a variety of detection techniques, such as ultraviolet absorbance, fluorescence, mass spectrometry, conductivity, amperometry, and refractive index. For samples with fluorescent labels, laser-induced fluorescence detection produces exceptional detection limits (Lee and Yeung, 1992; Nie, Dadoo, and Zare, 1993; Tao and Kennedy, 1996; Le et al., 1998). CE-LIF can be used for environmental and biological studies. CE-LIF has been well established for the study of affinity binding of aptamers and proteins,

and the interactions between antibodies and antigens (Le, Wan, and Lam, 2002; Zhang, Li and Le, 2008).

The approach to achieving high detection sensitivity of the CE-LIF system is to improve the collection of fluorescence from the samples and to reduce background signal. The CE-LIF system has its own characteristics for high detection sensitivity. First, lasers are used as the light source. Lasers are monochromatic and are easily filtered to produce much lower background than other light sources. Argon ion and helium-cadmium lasers are commonly used in CE. Second, the power of laser light can be optimized to improve the detection limit. Third, the laser can be focused near the diffraction limit of light. Using post-column detection with a sheath flow cuvette that was developed by Dovichi and his group (Cheng and Dovichi, 1988; Wu and Dovichi, 1989; Chen, Swerdlow et al., 1991; Zhang et al., 1991; Wu and Dovichi, 1992), the CE-LIF provides exceptionally low detection limits.

In CE with on-column LIF detection, the laser can generate large amounts of background signal due to reflection and refraction of light at the capillary walls. High sensitivity is obtained using post-column fluorescence detection due to the use of a sheath flow cuvette. In the CE-LIF system with a post-column fluorescence detector, the capillary outlet is inserted into the cuvette. Sheath flow, composed of the same separation buffer as the system, is introduced into the flow chamber at a faster speed. Thus, the two flowing streams will not mix together.

The laser is then focused at the outlet of the capillary, not on the capillary itself. The sheath flow cuvette is made of flat quartz, resulting in much less light scattering than that produced by a capillary. Therefore, the background from light scattering is greatly minimized, and much lower detection limits are obtained. The detection limits of CE-LIF combined with a post-column fluorescence detector can be as low as hundreds of molecules (Swerdlow et al., 1991; Chen and Dovichi, 1994).

4.3. In-house-built CE-LIF system

An in-house-built CE-LIF system (Figure 1-6) is used to determine binding affinity in my research.

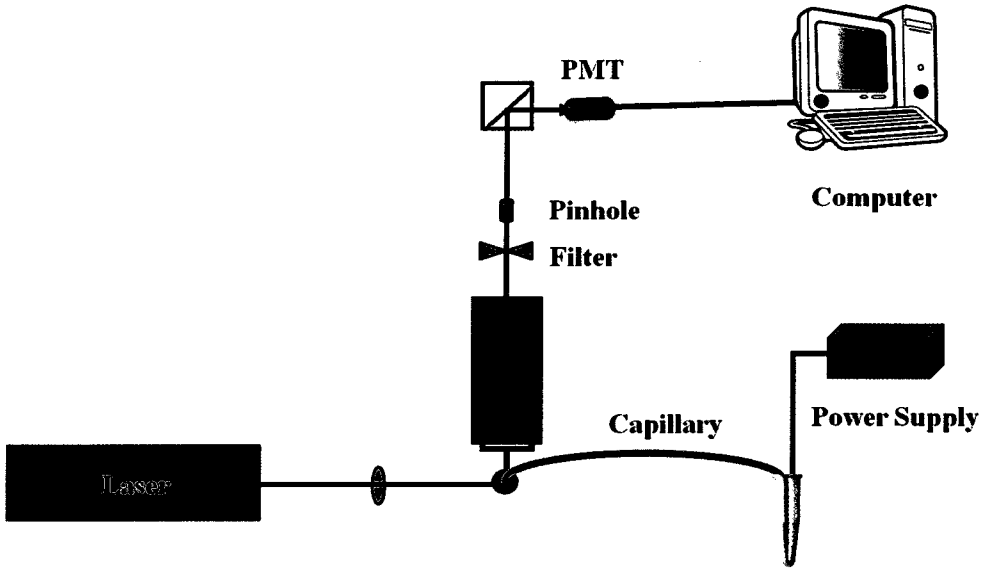


Figure 1-6: Schematic of the laboratory-built capillary electrophoresis with laser-induced fluorescence detection system.

The in-house-built CE-LIF system consists of a CE power supply (model CZE 1000R; Spellman, Plainview, NY, USA), sample and running buffer vials, waste vial, electrodes, a fused-silica capillary with 20 μm inner diameter and 140 μm outer diameter (Polymicro Technologies, Phoenix, AZ, USA), an argon ion laser (Uniphase, San Jose, CA, USA), a laser-induced fluorescence detector, and a computer for data acquisition and analysis.

To inject a sample into the CE, the capillary inlet is first placed in the sample vial, and a high voltage is applied. The sample is injected electrokinetically at 18 kV for 5 seconds. Then the capillary inlet is placed back into the running buffer vial. The analytes are separated based on their electrophoretic mobility and EOF. The analytes labeled with fluorescent probes are excited by 488 nm light from the argon ion laser. The fluorescent signals of the separated analytes are detected at a sheath flow cuvette (NSG Precision Cells, Farmingdale, NY, USA) which is used as a post-column fluorescence detection cell. After filtering through a 515 nm band-pass filter and passing through a 2 mm pinhole, the fluorescent signals are collected by a high-numerical aperture microscope objective (60×, 0.7 NA; Universe Kogaku, Oster Bay, NY, USA). By passing through the photomultiplier tube (PMT, Hamamatsu Photonics, Japan), the fluorescent signals are transferred to the computer for analysis. Data are shown as electropherograms by the software LabVIEW (National Instruments, Austin, TX, USA), and can be

analyzed using Igor Pro software (WaveMetrics, Lake Oswego, OR, USA) (Zhang, Li and Le, 2008).

5. Challenges

Studies have shown that aptamers are useful in analytical, bioanalytical, therapeutic, and diagnostic applications because of their high affinities and specificities to target molecules. To achieve high generation efficiency and select aptamers with high affinity and specificity, SELEX has been developed by combining different techniques in the partitioning step.

However, the number of aptamers generated is limited and the development of new aptamers is hampered by the tedious selection processes of available methods. It is crucial to develop a new aptamer selection method that provides simple, rapid, highly efficient, and high throughput generation of aptamers for multiple proteins. This leads to the objective of my research, which is to develop a new selection method for developing new aptamers against multiple proteins.

Chapter 2: Development of High Throughput Generation of Aptamers for Multiple Proteins

1. Introduction

Aptamers are short oligonucleotides (ssDNA or RNA) that can form three-dimensional structures that specifically bind to a wide range of targets [proteins, organic molecules, and inorganic molecules (Jayasena, 1999; Patel and Suri, 2000; Clark and Remcho, 2002; Luzi et al., 2003; You et al., 2003)] with high affinity and high specificity (Ellington and Szostak, 1990; Tuerk and Gold, 1990). The binding affinity of aptamers to proteins is similar or even higher than that of antibodies with typical dissociation constants (K_d values) of micromolar to low picomolar range (Jenison et al., 1994).

Compared to antibodies, aptamers are easier to produce and relatively inexpensive. The generation process occurs *in vitro* without the use of animals. In theory, aptamers can be generated against any target protein, and the binding target site of the protein can be determined. Once sequenced, aptamers can be synthesized at lower cost than antibodies. They can be easily modified with different chemical groups to achieve various functions (Nimjee, Rusconi, and Sullenger, 2005), or enhance chemical properties (such as stability or

resolvability). Due to these attractive features, aptamers have been used in a variety of *in vitro* and *in vivo* applications. For example, aptamers are used for protein identification, protein purification, and inhibition of receptors or enzyme activities (Mann et al., 2005). Aptamers could be useful to detect proteins from bacteria in environmental or clinical samples instead of the commonly used antibodies.

As mentioned in Section 2 of Chapter 1, the conventional approach for generating aptamers is through systematic evolution of ligands by exponential enrichment (SELEX) (Ellington and Szostak, 1990; Tuerk and Gold, 1990). In this *in vitro* procedure, target-specific aptamers are selected and synthesized from a random aptamer library. After repeating the cycles, the most specific aptamers to the target molecules are selected. It usually takes 8–12 rounds to generate aptamers with good specificity. Currently, a SELEX method containing 4–8 rounds is considered efficient.

The number of aptamers with high specificity is limited; in particular, aptamers with high specificity for detection of bacteria are rare. This is due to low throughput of the current SELEX methods that often generate one specific aptamer for one target protein at one time. To develop aptamers for detection of bacteria, it would be tedious to separate and purify multiple target proteins from culture. A capillary electrophoresis based SELEX method can be used to generate aptamers against multiple proteins. However, CE separation of proteins in

bacterial lysates may not be complete and identification of the target proteins is difficult. To avoid these problems, the objective of my research is to develop a new selection method for developing new aptamers against multiple bacterial proteins. Chapter 2 describes the design and development of this new method that makes use of proteins on membranes, after the proteins are separated with the well-established one-dimensional (1D) and two-dimensional (2D) gel separation and transferred to the membrane with Western Blot.

2. Objective

The initial goal of my MSc thesis research is to develop a new gel separation and membrane-based aptamer generation method, for the rapid and simultaneous selection of aptamers against multiple proteins. The long-term objective is to select aptamers against large number of proteins making use of 2D gel separation. The selected aptamers will be applied to detect bacterial proteins in environmental samples.

3. Project Design

To overcome the problems in the conventional SELEX and to achieve high efficiency and high throughput aptamer selection, a novel approach is based on 1D gel separation and Western Blot. The rationale of my new technique design is described as follows.

3.1. Purification and separation of target proteins

In theory, aptamers can be generated against any protein. In practice, target proteins of bacteria must be separated and purified from the cell lysate. Gel electrophoresis is a well-established and widely used method for separating a large number of proteins. Proteins from bacterial cells are first extracted and separated by 2D gel electrophoresis, which separates complex protein samples based on isoelectric point (PI) and molecular weight. Over the last decade, proteomics research has led to the establishment of a data library of 2D gel separation results, including 2D separation of bacterial proteins from various bacterial species. Obtaining sufficient amounts of pure proteins for conventional SELEX requires tedious and difficult separation, purification, and identification processes. To overcome this problem, I propose to use 2D separation to obtain the desired bacterial proteins without excess purification steps. This will be achieved by comparing my experimental 2D results with library data.

At the preliminary stage to demonstrate the proof of concept, 1D gel electrophoresis is used to develop the procedures of the new method. There are two common types of gel electrophoresis: native gel and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), which can provide better separation than native gel electrophoresis. Target proteins are separated in their denatured states using SDS-PAGE, because SDS and 2-mercaptoethanol are used

in the sample loading buffer. SDS is an anionic detergent: it can break hydrogen bonds between proteins, affect the hydrophobic interaction in a protein, and also affect the beta-sheet in secondary structure. The 2-mercaptoethanol is used to break the disulphide bridges in the tertiary structure of proteins. SDS molecules bind to denatured proteins or peptides generally in a mass ratio of 1.4:1, and confer a negative charge to the polypeptides. The denatured polypeptides become "rods" of negative charge cloud with equal charge densities per unit length. SDS-PAGE separates different proteins according to their molecular weight.

3.2. Protein transfer from gel to membrane

After proteins are separated in SDS-PAGE gels, they are denatured and their conformational structures and bioactivities are changed. My objective is to generate aptamers against proteins with native conformation, which is more useful for practical applications. Therefore, the proteins separated on SDS-PAGE are required to return to their native conformation. To achieve this, I use Western Blot, which is a well-established method to recover the proteins to their native conformations. By transferring the proteins to the membranes used in Western Blot [usually nitrocellulose membrane or polyvinylidene difluoride (PVDF) membranes], the SDS and 2-mercaptoethanol used in the SDS-PAGE are removed, resulting in the recovery of proteins of their native conformations. In this project, PVDF membrane is used because of its improved protein retention properties

under various conditions. The enhanced mechanical strength of the PVDF membrane ensures that target protein-aptamer complexes can sustain the various wash conditions required in the SELEX process. To reduce non-specifically bound aptamers in the blank region of the membrane, the membrane is blocked with bovine serum albumin (BSA) before SELEX.

3.3. Single cycle selection (SCS) of aptamers

Unlike the conventional SELEX process in which a single protein target has to be purified in advance, the novelty of my design is that multiple target proteins on the PVDF membrane are directly incubated with a random ssDNA library. The DNA-protein complexes on the membrane are washed with a number of washes in the partitioning stage. Each washing solution is collected, as are a number of elution solutions. Different fractions of elution solutions will contain ssDNA with various affinities to the target proteins. The ssDNA with highly specific affinity to their target proteins are selected from some of these fractions. In this design, simple and rapid washing and eluting steps are used to replace multiple SELEX cycles. Different aptamers against multiple proteins are obtained using single cycle membrane SELEX, which is a unique approach.

3.3.1 Optimization of the partitioning stage

The partitioning stage is the most important step in this new SCS method as described above. Unbound ssDNA/RNA are separated from the bound

ssDNA/RNA-target protein complex by a number of washes (washing step), and then the bound ssDNA/RNA are eluted from the binding complexes with a number of elutions (elution step). The efficiency of this method is dependent on the efficiency of the partitioning step. To improve efficiency, interference from non-specific or unbound ssDNA will be reduced as much as possible at this stage. To generate the most specific aptamer against the target proteins after single cycle membrane selection, optimization of this stage is essential.

3.3.2 Optimization of the PCR amplification step

To improve the efficiency of the SCS method, the aim is to obtain high quality of ssDNA from PCR amplification. The PCR conditions are optimized to eliminate or minimize non-specific amplification and to enhance the amplification of the specific ssDNA for the SCS. The parameters optimized included the cycle number, annealing temperature, and components of PCR mixture.

3.4. Sequencing of selected aptamers

After selecting aptamers with the new SCS technique, the sequences and structures of these aptamers are characterized. The sequences of aptamers are obtained using the common DNA cloning and sequencing technique, and their secondary sequences are analyzed using the software available on the website of Integrated DNA Technologies. By analyzing and comparing secondary structures, dynamic parameters (Tm and ΔG) of the selected aptamers are determined. Based

on the dynamic parameters and secondary structures, some ssDNA that have high Tm and low ΔG are selected as the aptamers. The binding affinity (K_d) is evaluated using capillary electrophoresis with laser-induced fluorescence detection.

4. Experimental

4.1. Materials

Some reagents for gel electrophoresis and Western Blot were purchased from BioRad Laboratories (Mississauga, ON). These included 30% acrylamide mix solution, 40% acrylamide mix solution, ammonium persulfate, SDS-PAGE molecular weight standards (low range), 10×Tris-glycine buffer, and PVDF membrane. Other reagents including SDS, Tris-base, PCR kit, DNA cloning and sequencing kit (TOPO TA kit), BSA, and dNTP kit were obtained from Invitrogen (Burlington, ON). MinElute PCR purification kit was purchased from Qiagen (Mississauga, ON). Tween 20 was ordered from Fisher Scientific (Nepean, ON); 1, 2-bis(dimethylamino)-ethane (TEMED) and ethylenediaminetetraacetic acid (EDTA) were ordered from EMD Chemical Inc (Gibbstown, NJ, USA).

4.2 Random DNA library and primers

The DNA library and the primers for PCR amplification of the library were synthesized by Integrated DNA Technologies (Coralville, IA, USA). The

sequences of the DNA library and primers used in this study are listed below.

The DNA library consists of 40 random bases in the middle region and 20 bases at the 5' and 3' ends:

5'-AGC AGC ACA GAG GTC AGA TG (N:25252525)(N) (N)(N) CC TAT GCG TGC TAC CGT GAA-3'

Two primers for PCR amplification of the DNA library consisted of:

Primer 1: 5'-AGC AGC ACA GAG GTC AGA TG-3' (no label)

Primer 2: 5'-TTC ACG GTA GCA CGC ATA GG-3' (no label)

Primers for DNA sequencing (provided in TOPO TA kit, Invitrogen):

M13 Forward primer: 5'-GTA AAA CGA CGG CC AG-3'

M13 Reverse primer: 5'-CAG GAA ACA GCT ATG AC-3'

T3 primer: 5'-ATT AAC CCT CAC TAA AGG GA-3'

T7 primer: 5'-TAA TAC GAC TCA CTA TAG GG-3'

Sequence of plasmid DNA from the TOPO TA cloning kit (Invitrogen):

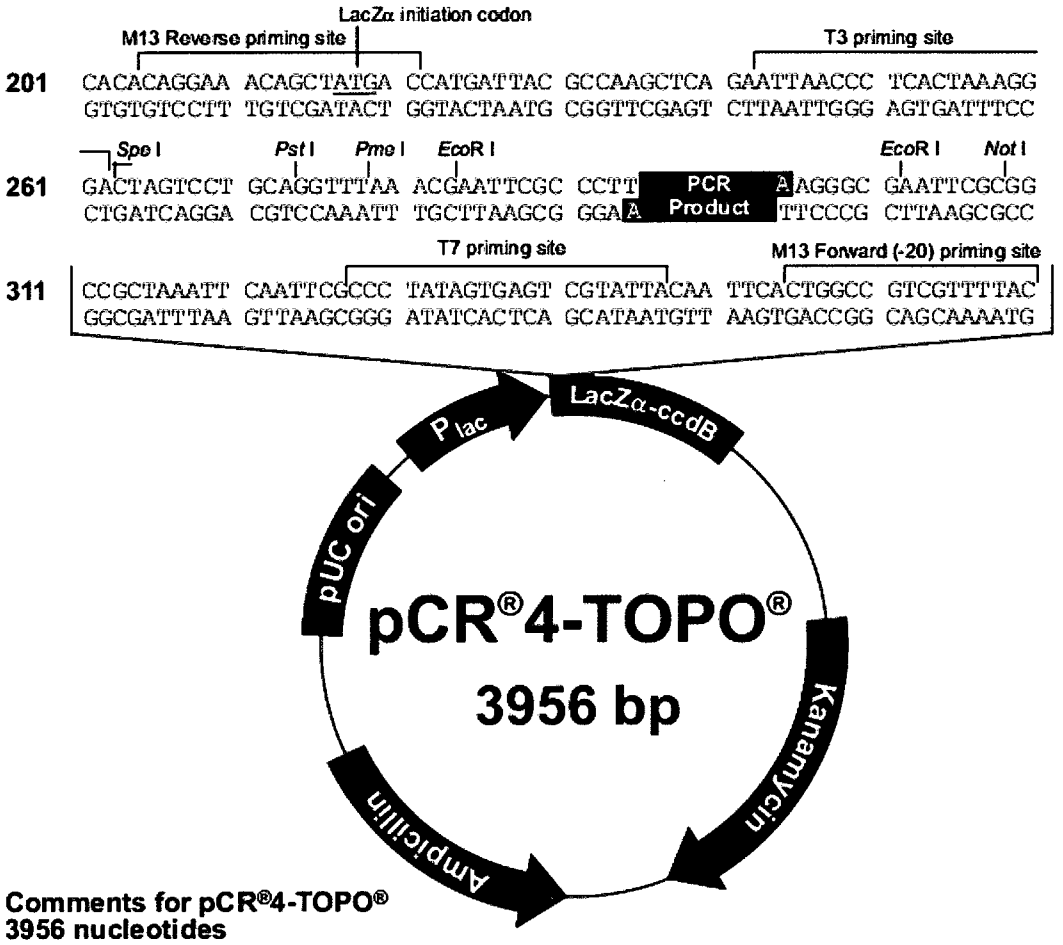


Figure 2-1: Sequences of pCR 4 -TOPO plasmid.

4.3 Standard proteins

To demonstrate the proof of principle, five standard proteins (purchased from Invitrogen) were used as the target proteins. One was the *E. coli* protein: Fpg protein (molecular weight 30.2 kDa), which stands for Formamidopyrimidine DNA glycosylase. Thrombin (molecular weight 36.7 kDa) and lysozyme (molecular weight 14.7 kDa) were used as reference proteins for SCS method

development, because their specific aptamers have already been generated to provide nM level binding affinity. BSA and ovalbumin were also used as target proteins, because a specific aptamer against them has not been generated.

4.4. SDS-PAGE and protein transfer

In this study, separation of proteins using SDS-PAGE was performed on 5% stacking gels and 12% resolving gels under the well-established conditions. The gels were freshly prepared in-house. The 5% stacking gel consisted of 4 mL 5% stacking gel mixture (2.7 mL deionized water, 0.67 mL 30% acrylamide mixture, 0.5 mL 1 M Tris-HCl with a pH 7.8, 40 μL 10% SDS solution, 40 μL 10% ammonium persulfate solution and 4 μL TEMED), and the 12% resolving gel was made of 10 mL 12% resolving gel mixture (3.3 mL deionized water, 4.0 mL 30% acrylamide mixture, 2.5 mL 1.5M Tris-HCl at pH 8.8, 0.1 mL 10% SDS solution, 0.1 mL 10% ammonium persulfate solution, and 4 μL TEMED). The 10% ammonium persulfate solution was freshly prepared for each set of experiments. This was done to ensure that all the casted SDS-PAGE gels were of good quality for separation.

The total amount of proteins used in SDS-PAGE was 0.36 nmole (0.03 nmole BSA, 0.04 nmole ovalbumin, 0.1 nmole Fpg protein, 0.05 nmole thrombin and 0.14 nmole lysozyme). The protein samples were heated at 95°C for 5 minutes and then loaded into wells with loading buffer (200 mg SDS, 2 mL glycerol, 0.5%

2-mercaptoethanol, 0.5 mL Tris-HCl at pH 6.8, and 0.1% bromophenol blue, then add deionized water to a final volume of 10 mL), the volume ratio of protein and loading buffer was 1:1. A potential of 8 V/cm was applied when protein samples were still in the stacking gel; after the dye front moved to the resolving gel, the potential was increased to 15 V/cm and the electrophoresis was finished when the bromophenol blue dye reached the bottom of the resolving gel. The 1000 mL running buffer was made of 100 mL 10×Tris-glycine buffer, 1% SDS, and 900 ml deionized water.

After separating on SDS-PAGE, the proteins were transferred to the PVDF membrane overnight under the standard conditions of 125 mAmp. During this procedure, the temperature was kept at 4 °C. The blotting buffer (1000 mL) consisted of 200 mL methanol, 100 mL 10×Tris-glycine buffer, and 700 mL deionized water. The bands containing the target proteins on the PVDF membrane were then excised with different shapes for recognizing the specific proteins and the protein-containing membrane pieces were used for the SELEX process. This was done to improve DNA binding to specific target proteins and to minimize interference from blocking protein BSA (3% BSA solution).

Theoretically, 36 nmole of DNA library was needed for the selection, to ensure the ratio of ssDNA molecules and protein molecules was 100:1. However, not all the proteins could be transferred from gel to membrane. The efficiency of the transfer is around 50–60%; 70–80% efficiency is seldom achieved. Hence the

amount of DNA library should be more than 20 nmole but less than 28.8 nmole based on the transfer efficiency. In the selection, 30 nmole of aptamer library was used in the selection.

4.5. Single cycle selection (SCS)

The process of the single cycle membrane selection method I designed involves four steps: incubation/binding, washing, elution, and PCR amplification of the selected DNA, as shown in Figure 2-2. The partitioning process includes the washing and eluting steps. The details of the processes are described as follows.

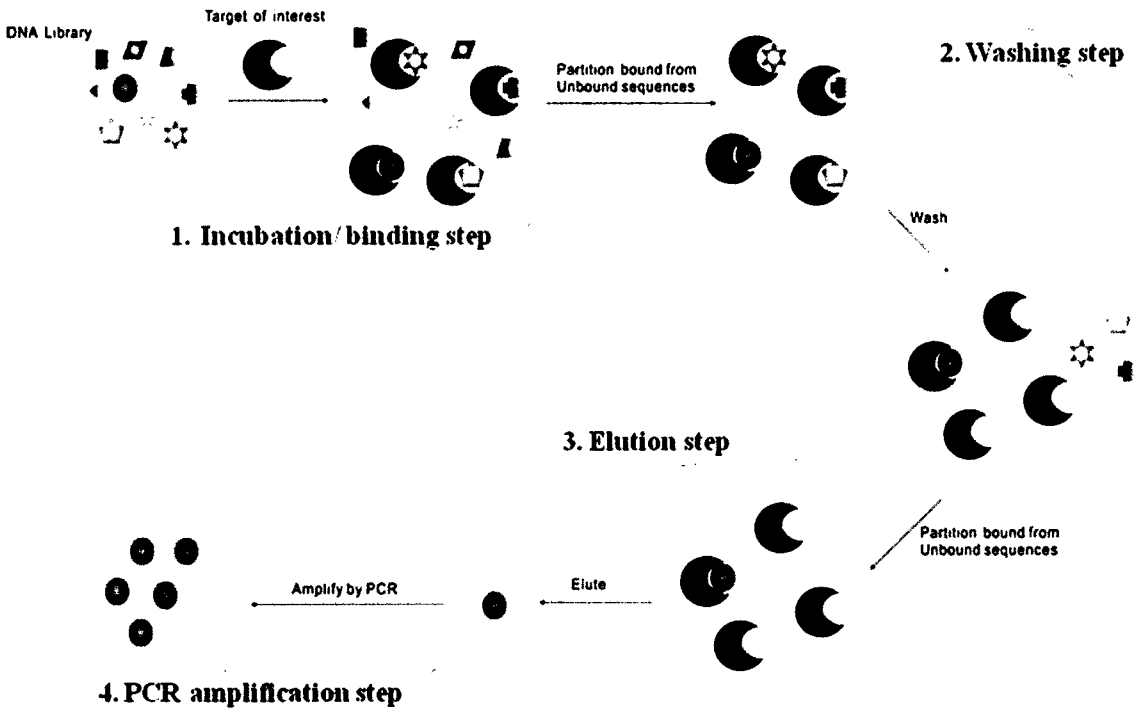


Figure 2-2: Schematic of single cycle Selection

In this method, 30 nmole of an 80-nt-long ssDNA library was dissolved with binding buffer [100 mM NaCl; 20 mM pH7.6 Tris-HCl; 2mM $MgCl_2$; 5 mM KCl; 1 mM $CaCl_2$; 0.02% Tween 20 (Stoltenburg, Reinemann, and Strehlitz, 2006)], and the five small pieces of PVDF membranes containing the five target proteins (described in Section 4.4) were immersed into this solution and incubated for 45 minutes. Based on molar concentrations, the ratio of ssDNA molecules to the target protein was 100:1, to ensure excess ssDNA molecules binding to the target protein molecules.

Theoretically, a length of 40 random bases in ssDNA should provide 4^{40} (approximate to 10^{24}) random sequences. However, most aptamer libraries contain no more than 10^{16} random sequences, because of limitations of synthesis. The amount of 1 mole ssDNA contains 6.02×10^{23} molecules, thus the amount of 30 nmole ssDNA I used in the experiments contains $30\times10^{-9}\times6.02\times10^{23}=1.806\times10^{16}$ (approximate to 2×10^{16}) molecules, which ensures that at least each random sequence has one molecule during the binding step.

After incubation with the DNA library, the membrane pieces were then washed 20 times with clean binding buffer. Washing solutions were collected separately and precipitated with ethanol. The amounts of DNA in washing solutions were determined by UV absorbance detection. After washing with a number of washes (will be stated in the Results and Discussion section), the 5

pieces of membranes containing the 5 target-aptamer complexes were separated into 5 tubes. The bound ssDNA were eluted from the complexes on the specific membrane with elution buffer [40 mM at pH 8.0 Tris-HCl, 10 mM EDTA, 3.5M urea, 0.02% Tween 20 (Stoltenburg, Reinemann, and Strehlitz, 2006)] 10 times and collected in 10 fractions. DNA in each elution solution was purified with ethanol precipitation and amplified by PCR. Parameters that may affect PCR amplification were examined and optimized. The optimized PCR mixture was composed of 10 μL 10×PCR buffer, 5.6 μL 50 mM $MgCl_2$, 5 μL 20 mM forward primer, 5 μL 20 mM reverse primer, 4 μL 10 mM dNTPs, 63 μL autoclaved deionized water, 5 μL DMSO, 0.5 μL 5 unit/μL Taq polymerase and 2 μL template DNA. After 19 PCR cycles, amplified PCR products were analyzed with 12% DNA PAGE gel at 120 volts for 1 hour. The 12% DNA PAGE gel was made of 2.5 mL 40% acrylamide mixture, 5.5 mL deionized water, 2 mL 5×TBE buffer, 62 μL 10% ammonium persulfate, and 10 μL TEMED. Finally the DNA PAGE gels were stained with ethidium bromide (EB) to confirm that only full-length (80 bp) PCR products were obtained.

4.6. DNA cloning and sequencing

Before cloning, the amplified 80-bp PCR products were further purified by a MinElute PCR purification kit (Qiagen) to eliminate interference from left primers, self-dimers, and unspecific PCR products. The PCR purification was

performed using the recommended protocol. Using the buffer and column in the kit, fragments ranging from 70 bp to 4 kb were retained in the column, while primers, polymerase, and salts were removed. Then the desired PCR products were eluted from the column. Before DNA cloning, purified PCR products were precipitated from the buffer and dissolved in autoclaved deionized water.

Cloning was performed using a TOPO TA kit (purchased from Invitrogen). *Taq* polymerase-amplified PCR products were directly inserted into plasmid vectors. The linear plasmid vector in the kit contained single 3' thymidine (T) overhangs, and *Taq* polymerase-amplified PCR products had a single deoxyadenosine (A) on each 3' end. Based on the principle of complementary base pairing, PCR products were ligated with the vector efficiently.

DNA cloning and sequencing were performed following the instructions provided by the supplier. To perform the ligation, 1 μL purified PCR product, 1 μL salt solution (1.2 M NaCl, 0.06 M $MgCl_2$), 3 μL autoclaved deionized water, and 1 μL TOPO vector (10 ng/μL plasmid DNA in 50% glycerol, 50 nM Tris-HCl with pH 7.4, 1 mM EDTA, 2mM DTT, 0.1% Triton X-100, 100 μg/mL BSA, 30 μM phenol red) were mixed gently and incubated for 5 minutes at room temperature. An aliquot of 2 μL of the recombinant plasmid vectors with inserted target DNA fragments were added to one vial of DH5α^{TM}-T1^{R} Chemically Competent cells (from the kit produced by Invitrogen) and then incubated for 30 minutes on ice. These cells were heat-shocked for 30 seconds in a 42 °C water

bath, and then the tubes containing cells were immediately transferred to ice. An aliquot of 250 μL of S.O.C. medium (2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added to each tube, then the tubes were capped tightly and shaken at 200 rpm horizontally at 37°C for one hour. Selective plates (LB medium containing 50 μg/mL ampicillin) were pre-warmed in 37 °C incubator for 30 minutes, and then 20–50 μL of the transformed bacterial cells were cultured on each selective plate. The plates were then incubated at 37 °C for 12–16 hours. After overnight culturing, single colonies were selected from the plates and then cultured in LB broth medium with 50 μg/mL ampicillin for another overnight culturing at 37°C for 12–14 hours. When the broth medium became turbid, the cultured cells were ready for plasmid DNA isolation and purification.

The cells in broth medium were collected after 5 minute centrifuging at 13,000 rpm. By using the PureLink™ Quick Plasmid Miniprep Kit (Invitrogen), plasmid DNA containing the inserted fragment was precipitated and purified. Plasmid DNA was eluted from the membrane in the column (provided in the kit) by TE buffer (10 mM Tris-HCl at pH 8.0, 0.1 mM EDTA). However, the DNA in TE buffer was not suitable for sequencing, so DNA samples were precipitated using ethanol precipitation, and dried DNA samples were dissolved in autoclaved deionized water. Sequences were analyzed using T3 primer, T7 primer, or M13 forward/reverse primer.

5. Results and Discussion

5.1. Characterization of retention of proteins on PVDF membrane

The retention of proteins on the PVDF membrane affects the efficiency of the single cycle membrane selection method. It is important to achieve efficient retention of the aptamer-protein complexes on the PVDF membrane, while efficiently removing non-specific DNA molecules during the washing steps at the partitioning stage. Therefore, I carefully evaluated the retention of proteins on the PVDF membrane. The amounts of the proteins on the PVDF membrane with and without washes were determined using Coomassie Brilliant Blue staining as described below.

5.1.1. Retention property of PVDF membrane using BSA as target protein

BSA was used as a target protein and transferred from the gel onto the PVDF membrane using the same conditions that would be used in the SELEX experiments. The proteins on the PVDF membrane without washing are shown in Figure 2-3. In parallel experiments, the proteins on the PVDF membrane were washed 5 times with binding buffer followed by 3 times with elution buffer, and the proteins retained on the membrane are shown in Figure 2-4. Comparing the intensity of the stained protein (BSA) on the two PVDF membranes with and without washing steps, it is clear that the intensities of BSA bands are similar with

and without washings. This indicates that washing steps do not significantly affect the retention of proteins on the membrane.

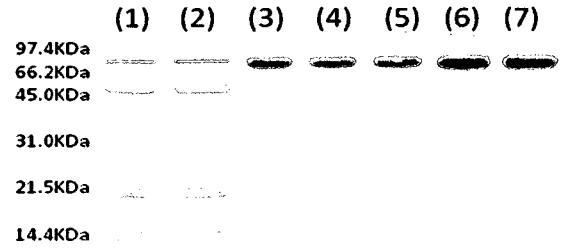


Figure 2-3: PVDF membrane stained with Coomassie Brilliant Blue directly after transfer (Western Blot). Lanes 1 and 2 were SDS-PAGE molecular weight standards (low range) and lanes 3 to 7 were all BSA standards (MW~67 kDa) with two different loadings (5 µg for lanes 3 to 5 and 15 µg for lanes 6 and 7).

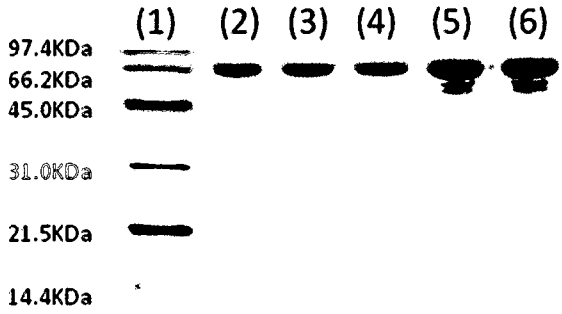


Figure 2-4: PVDF membrane stained with Coomassie Brilliant Blue after 5 washes with binding buffer and 3 washes with elution buffer. Lane 1 was SDS-PAGE molecular weight standard (low range) and lanes 2 to 6 were all BSA standards (MW~67 kDa) with two different loadings (5 µg for lanes 2 to 4 and 15 µg for lanes 5 and 6).

5.1.2. Retention property of PVDF membrane using 5 standard proteins

The retention of BSA on PVDF membrane is not sufficient. To develop a general membrane-based SELEX method, it is important that proteins of different sizes and properties are retained efficiently on the membrane. Thus, the retention properties of different proteins were further examined using protein ladders and the five standard proteins that were used in the method development. In addition, the strength of washings was greatly increased to 30 times with binding buffer and 20 times with elution buffer. Figure 2-5 shows the proteins retained on the membrane: there is no significant loss of the proteins on the membrane after the total of 50 times washing and eluting. This demonstrates the robustness and flexibility of proteins on PVDF membranes for SELEX. Therefore, transferring proteins from gel separation onto PVDF membranes is a feasible approach to obtain target proteins for high throughput SELEX. Furthermore, this indicates the possibility of single cycle membrane SELEX using different numbers of washes and elutions instead of many cycles of SELEX to improve the specificity of the selected ssDNA to the target proteins.

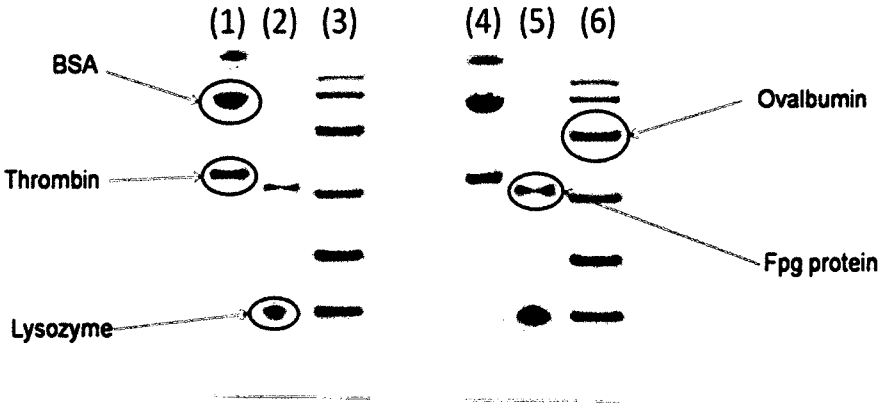


Figure 2-5: PVDF membrane with or without washing and eluting steps. Lane 3 and lane 6 were the low range SDS-PAGE molecular weight standards (ovalbumin included); lane 1 and lane 4 were the proteins BSA and thrombin; lane 2 and lane 5 were the Fpg protein and lysozyme. The PVDF membrane on the right was stained with Coomassie Brilliant Blue directly after the transfer. The membrane on the left was stained with Coomassie Brilliant Blue after 30 washes with binding buffer and 20 washes with elution buffer.

5.2. Optimization of PCR conditions

5.2.1. Optimization of PCR cycle number

The number of PCR cycles affects the specificity of the amplification products. In general, non-specific amplifications increase when the PCR cycles are increased over 20. For example, Musheev and Krylov (2006) reported that dsDNA products were generated when PCR was performed for 19 cycles. When PCR cycles were increased to above 20, nonspecific amplification of other

unwanted products, such as ss-ds DNA byproducts, became severe. This led to the reduction of desired dsDNA products.

To obtain specific amplification of the desired PCR products, I examined the number of PCR cycles ranging from 19 to 24 for the same 80-bp length DNA samples while keeping all other parameters constant. The amplified products were separated by 12% DNA PAGE. And results are shown in Figure 2-6.

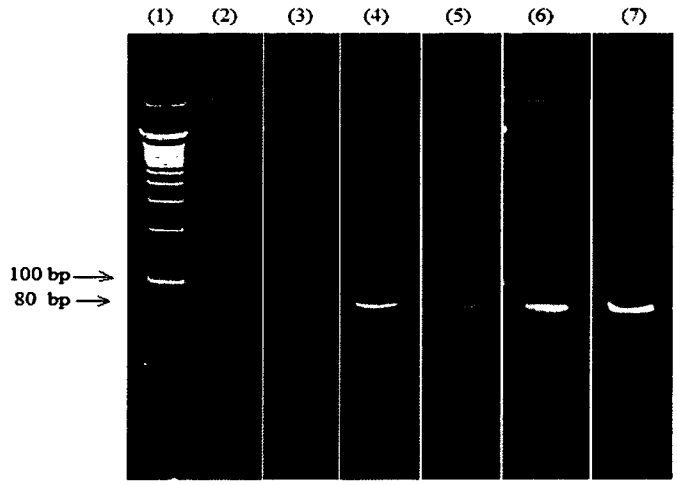


Figure 2-6: 12% polyacrylamide gel of PCR products: lane 1, 100 bp DNA ladder; lanes 2–7, PCR products applied with PCR cycle numbers 19–24.

Figure 2-6 clearly shows a clean band of dsDNA products (on lanes 2–3) from cycle numbers 19 and 20. As expected, non-specific amplifications clearly increased (lanes 4–7), when the PCR cycles were increased from 21 to 24. Thus, PCR amplification with 19 and 20 cycles was used in future experiments.

5.2.2. Function of adding dimethyl sulfoxide (DMSO) in PCR mixture

In the PCR process, DNA with high GC contents may form stem-loop secondary structures and induce the generation of shorter sequences of products. Kang, Lee, and Gorenstein have reported that the addition of DMSO in PCR can reduce the production of shorter DNA products (2005). PCR was performed with and without DMSO in the mixtures under the same conditions and the results are shown in Figure 2-7.

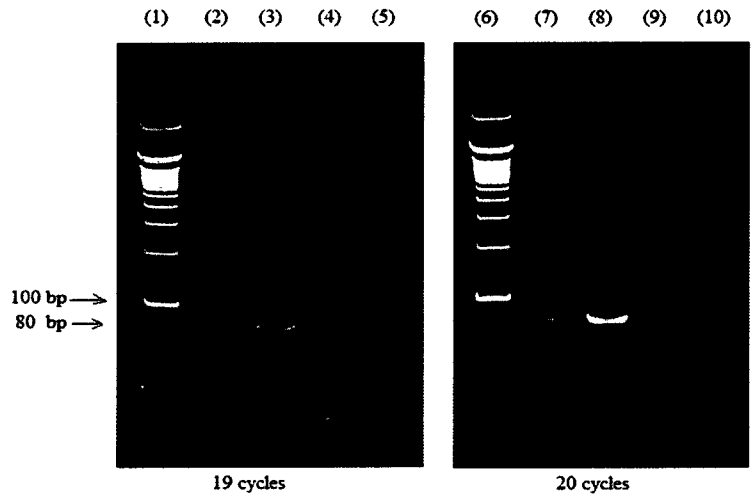


Figure 2-7: 12% polyacrylamide gel of PCR products with and without DMSO (5% PCR mixture volume). Lanes 1 and 6: 100 bp DNA ladder; lanes 2–5 (applied 19 cycles) and lanes 7–10 (applied 20 cycles), PCR products under different conditions: without DMSO (lanes 2 and 7); with DMSO (lanes 3 and 8); without template and DMSO (lanes 4 and 9); with DMSO, without template (lanes 5 and 10).

Figure 2-7 clearly shows that addition of DMSO to the PCR mixtures generated much cleaner bands of the desired products. As a result, DMSO was included in all PCR amplifications performed in this study, to improve the

specificity of full-length products. The PCR results from 20 cycles (Figure 2-7, lanes 7–8) showed non-specific amplifications after the addition of DMSO. Therefore, PCR amplification with 19 cycles was chosen for further experiments. When the DNA concentration was very low, the number of PCR cycles could be adjusted to achieve the best amplification results.

5.2.3. Annealing temperature

The annealing temperature in PCR is another important parameter affecting PCR results. Two annealing temperatures (55 °C and 56 °C) were tested and compared in this study. The temperature of 56 °C was found to increase specific amplifications, and was therefore selected as the annealing temperature for this study.

5.3. Optimization of partitioning step

The number of washes in the partitioning stage was optimized to achieve a balance between the washing and the eluting steps. The DNA-protein complexes would be retained on membrane while free or unbound DNA would be removed at the washing step. At the eluting step, ssDNA that was specifically bound to the target proteins could be eluted from the binding complexes.

To determine the optimum number of washes and elutions, all washing and elution solutions were collected individually. After ethanol precipitation, ssDNA

from each fraction of washing and elution solutions were dissolved in autoclaved deionized water individually. The 20 washing solutions were sequentially numbered as washing solution 1–20. For each target protein, there were 10 elution fractions that were sequentially numbered elution solution 1–10. The five target proteins resulted in five groups of 10 elution fractions.

5.3.1. UV absorbance detection

UV absorbance detection of OD_{260} is commonly used to estimate DNA concentration. To determine the best dynamic range of this detection (Smartspec 3000; BioRad), DNA samples with known concentration were detected. Based on the OD_{260} values (1 unit of OD_{260} equals 33 μg/mL ssDNA), ssDNA concentrations detected by the instrument were calculated for comparison.

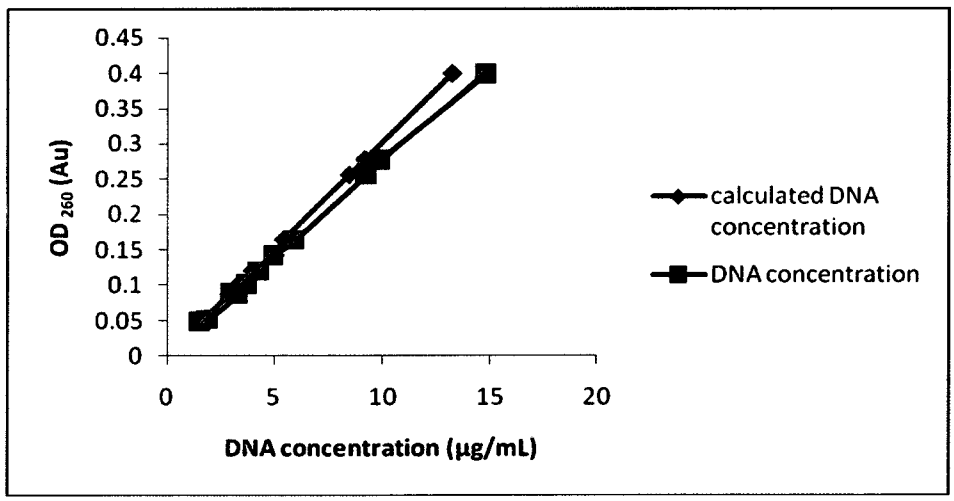


Figure 2-8: Calibration of UV absorbance detection by comparing the real concentration of DNA samples and the calculated concentration.

The results shown in Figure 2-8 demonstrate that the calculated

concentrations match completely the actual concentrations when OD_{260} is in the range of 0.05 to 0.2 (Au), corresponding to concentrations of DNA from 1.65 to 6.6 μg/mL. Based on this, the lowest concentration of ssDNA for detection is 1.65×10^{-3} g/L, which equals 62.5 nM of 80-nt-long ssDNA. This calculation is based on 1 M of 80-nt ssDNA being equivalent to 26400 g/L. The OD_{260} measurement cannot detect such low DNA concentration in the 20 washing solutions and 10 elution solutions from each target protein. When those DNA samples were amplified, the gel bands could be clearly observed in the washing solutions and elution solutions. Thus, PCR amplification was used to determine the number of washes in the partitioning step, instead of UV absorbance detection.

5.3.2. PCR results of 20 washing solutions and 10 elution solutions

Negative controls were included in each set of PCR experiments. No amplification band appeared in the negative control confirming that there was no contamination in PCR amplification.

5.3.2.1. PCR results of washing solutions 11–20

The PCR products from the washing solutions 11–20 were separated on gel and shown in Figures 2-9 and 2-10.

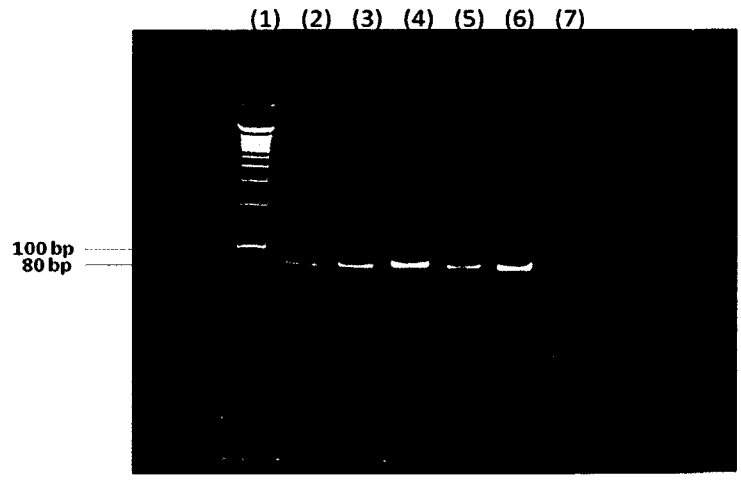


Figure 2-9: PCR results of washing solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: washing solutions 11–15 after washing step; lane 7: negative control.

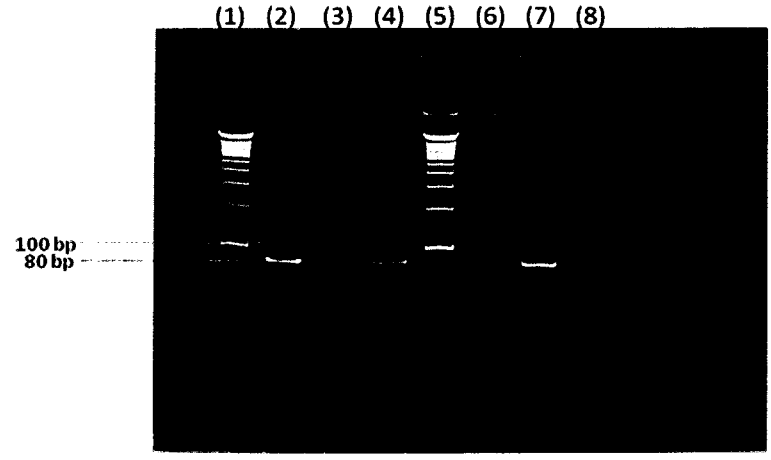


Figure 2-10: PCR results of washing solutions 16–20. Lane 1 and lane 5: 100 bp DNA ladder; lanes 2–4: washing solutions 16–18; lane 6 and 7: washing solutions 19 and 20; lane 8: negative control.

Figures 2-9 and 2-10 show the amplification products of washing solutions 11–20. In Figure 2-9 (lanes 4 and 6), non-specific amplification products were also observed along with the expected products (80 bp) in washing solutions 13

and 15. This suggests that DNA collected in these washing solutions had very high concentrations to generate non-specific amplification under the optimized PCR conditions. These results demonstrated that 20 washes in the washing step did not sufficiently remove non-specific binding DNA.

5.3.2.2. PCR results of elution solutions 5–10

After 20 washes, membrane pieces with 5 different target proteins were separated into 5 tubes. Each membrane piece was eluted 10 times with the elution buffer (1 mL). The individual elution solutions were collected separately and numbered sequentially as elution solution 1–10. The elution solutions 5–10 in each group were precipitated separately with ethanol and amplified by PCR. Figures 2-11, 2-12, 2-13, 2-14, and 2-15 show the PCR products from the elution solutions 5–10 for the five different proteins (BSA, thrombin, lysozyme, Fpg protein, and ovalbumin).

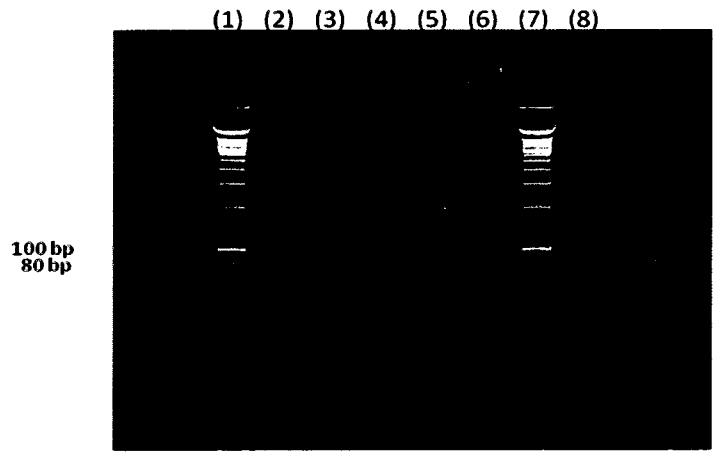


Figure 2-11: PCR results of elution solutions 5–10 (target protein: BSA). Lane 1 and 7:

100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 8: negative control.

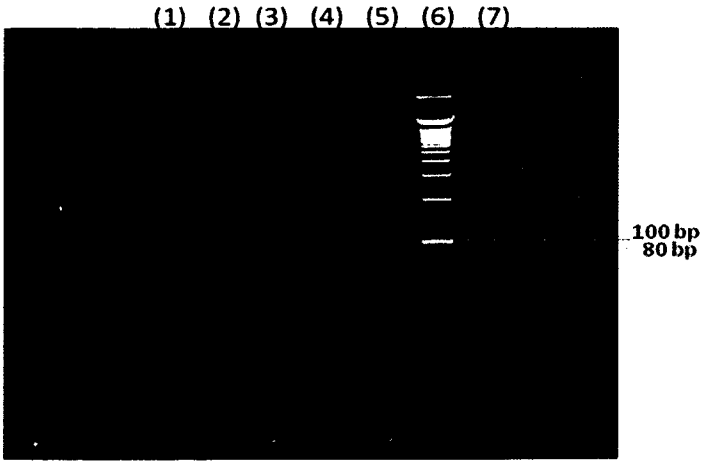


Figure 2-12: PCR results of elution solutions 5–10 (target protein: thrombin). Lanes 1–5: elution solutions 6–10, lane 6: 100 bp DNA ladder, lane 7: negative control.

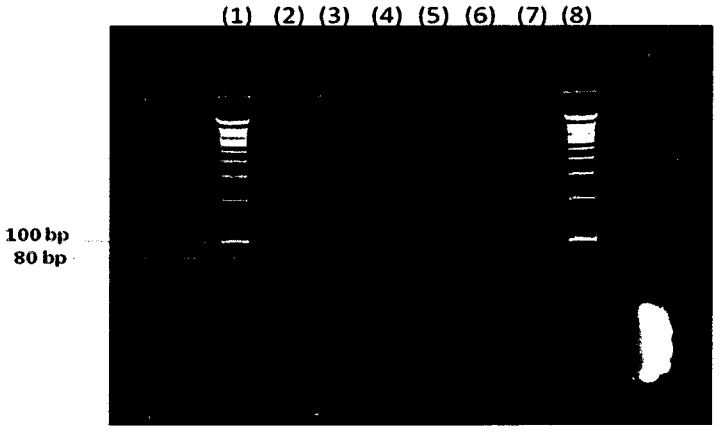


Figure 2-13: PCR results of elution solutions 5–10 (target protein: lysozyme). Lane 1 and 8: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

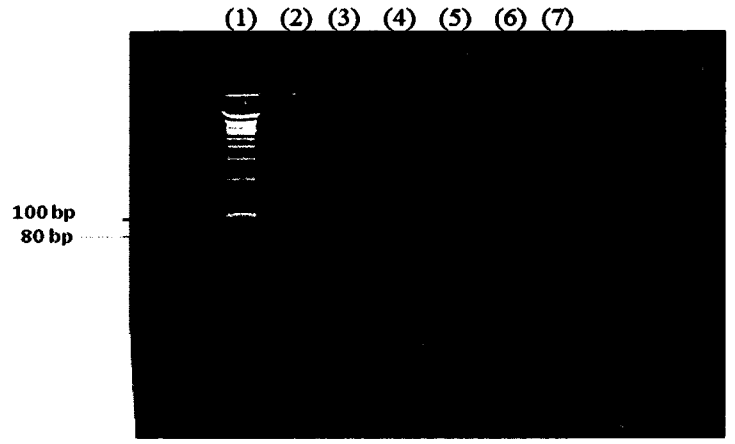


Figure 2-14: PCR results of elution solutions 5–10 (target protein: Fpg protein). Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

There is one band in lane 7 (Fig 2-14), the negative control, but the product was not the 80-bp products. Based on the position of the band in the gel, this band was most likely amplified from the primer, not the template DNA. Thus the band detected in the negative control was not due to contamination in this PCR experiment.

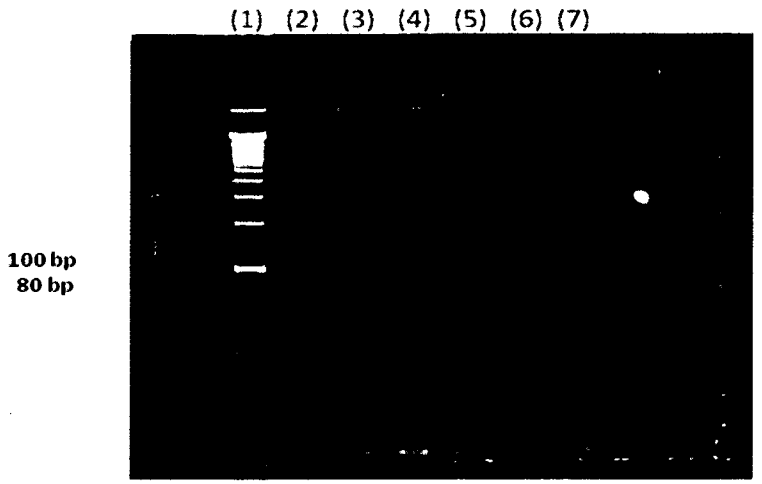


Figure 2-15: PCR results of elution solutions 5–10 (target protein: ovalbumin). Lane 1:

100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

Specific amplified PCR products were obtained from the elution solutions 5–10 from the five target proteins. Although the bands in Figures 2-11, 2-14 and 2-15 are weak, they can still be observed. DNA molecules that specifically bind to the target proteins could still be obtained after 10 elutions, when the membrane had already been washed 20 times in the washing step. These results show that 20 washes and 10 elutions were not sufficient.

To further optimize the number of washes and elutions at the partitioning stage, 30 washes and 15 elutions were performed after single cycle SELEX. The PCR results of these washing and elution solutions were analyzed and compared, described as follows.

5.3.3. PCR results of 30 washing solutions and 15 elution solutions

After single cycle membrane selection, ssDNA and proteins embedded on the membrane were washed 30 times with binding buffer and 15 times with elution buffer. All the solutions were collected individually, and ssDNA were precipitated and dissolved in the same volume of autoclaved deionized water for PCR amplification.

5.3.3.1. PCR results of washing solutions 16–30

Washing solutions 1–30 were collected and the fractions 16–30 were PCR amplified. Figures 2-16 and 2-17 show the PCR results of washing solutions

16–30. It can be observed that when the number of washes is increased, more DNA was collected in washing solutions 24–30 (Figure 2-17) compared to the washing solutions 16–23 (Figure 2-16), suggesting that DNA was removed from the membrane with the target proteins with increasing numbers of washes.

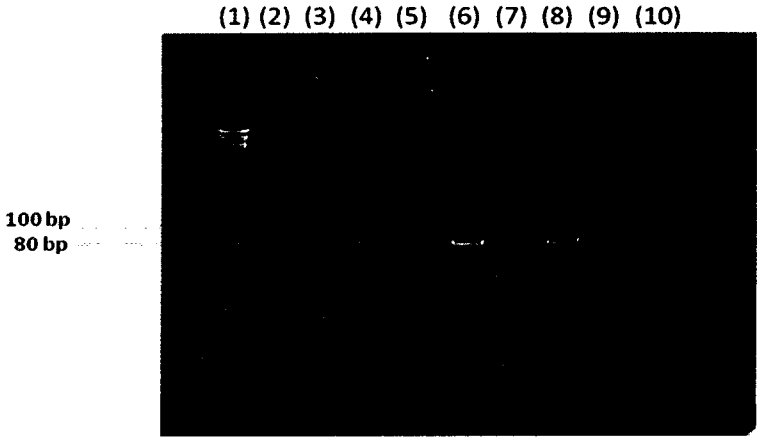


Figure 2-16: PCR results of washing solutions 16–23. Lane 1: 100 bp DNA ladder; lanes 2–9: washing solutions 16–23; lane 10: negative control.

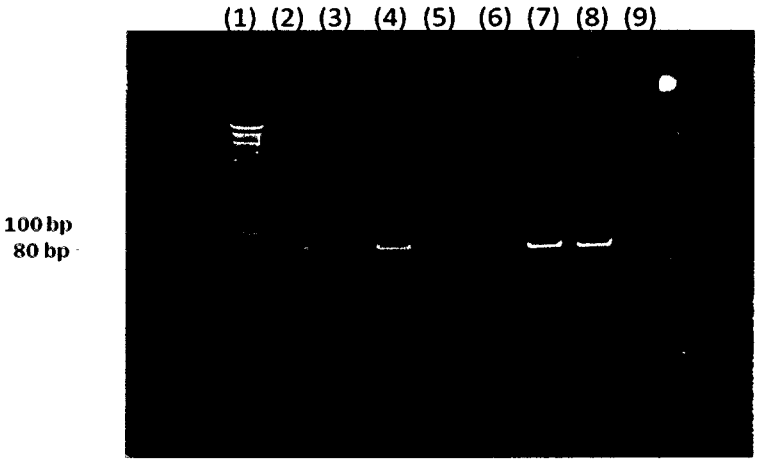


Figure 2-17: PCR results of washing solutions 24–30. Lane 1: 100 bp DNA ladder; lanes 2–8: washing solutions 24–30; lane 9: negative control.

In Figure 2-17, the band of washing solution 25 (lane 3) is very weak, but the band of washing solution 26 (lane 4) is darker than washing 25. Then the bands of washing solutions 27 and 28 (lanes 5 and 6) are becoming weaker. However, the bands of washing solutions 29 and 30 (lanes 7 and 8) are getting darker again. These results (Figure 2-16 and 2-17) suggest that when increasing washes in washing step, some DNA sequences that specifically bind to their target proteins may also be released from the binding complexes. This could cause the loss of some specific DNA sequences for the target proteins. Taking all this together, 25 washes were chosen for the washing step, to remove free or unbound DNA from binding complexes.

5.3.3.2. PCR results of elution solutions 6–15

After 30 washes, 15 elution solutions for each target protein were sequentially collected and numbered as elution solution 1–15. The elution solutions 6–15 from each group were amplified under the same PCR conditions. The results are shown in Figures 2-18 to Figure 2-27.

5.3.3.2.1. DNA samples may specifically bind to BSA

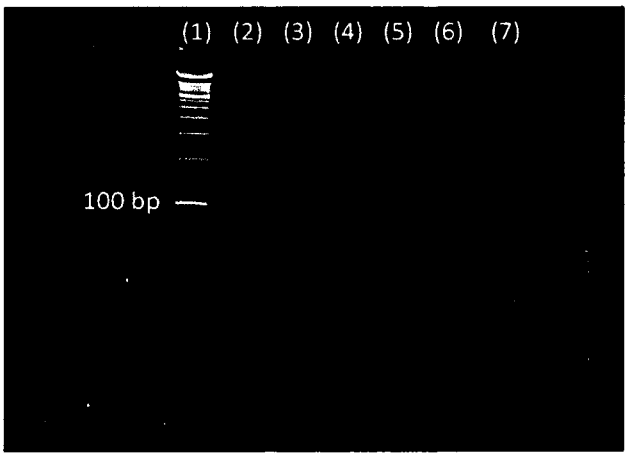


Figure 2-18: PCR results of elution solutions 6–10. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

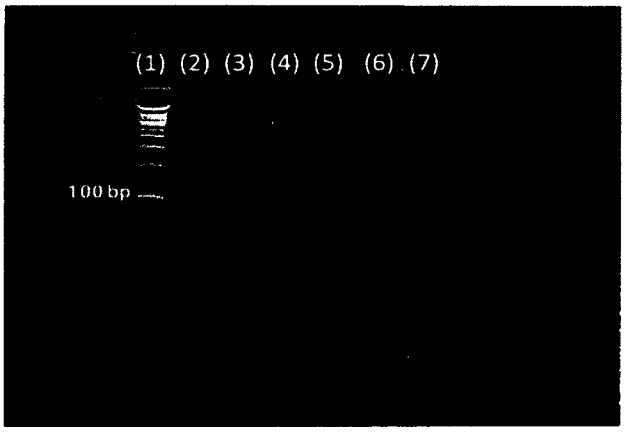


Figure 2-19: PCR results of elution solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 11–15; lane 7: negative control.

5.3.3.2.2. DNA samples may specifically bind to ovalbumin

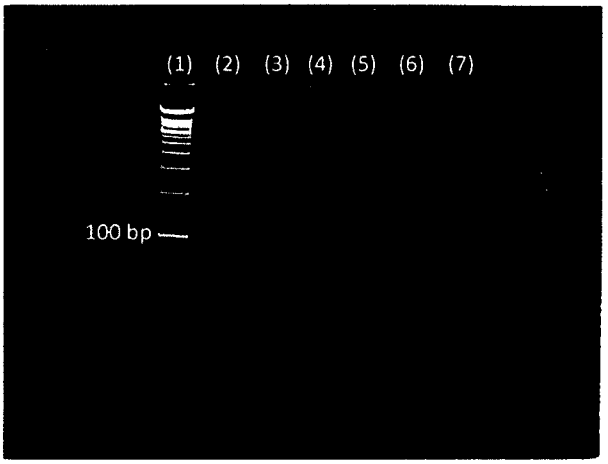


Figure 2-20: PCR results of elution solutions 6–10. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

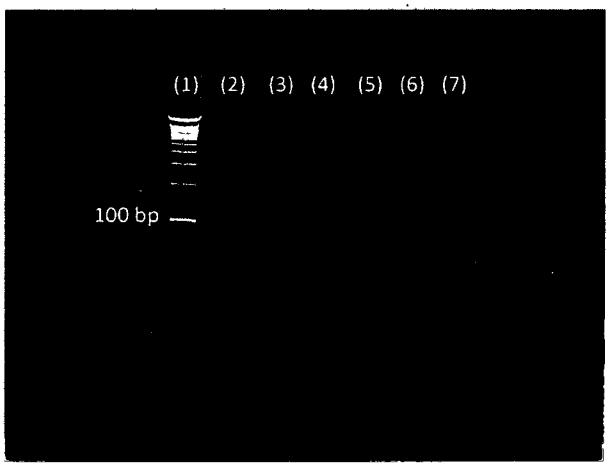


Figure 2-21: PCR results of elution solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 11–15; lane 7: negative control.

Figures 2-18 to 2-21 show that BSA and ovalbumin have very low binding affinities to ssDNA sequences, and even after PCR amplification, no clear band can be observed from 12% DNA PAGE. Thus, using BSA as the blocking buffer would not affect the incubation of ssDNA molecules and target proteins.

5.3.3.2.3. DNA samples may specifically bind to thrombin

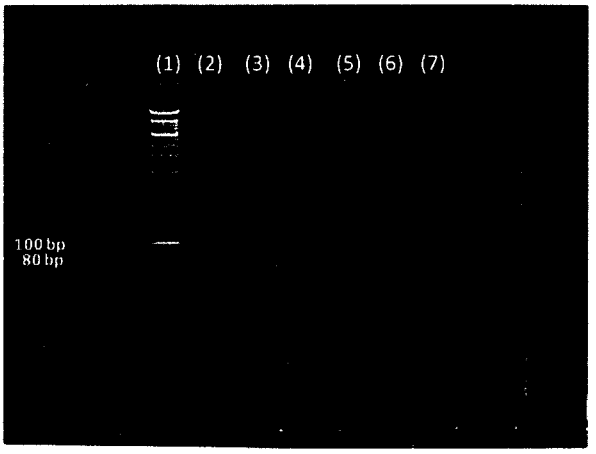


Figure 2-22: PCR results of elution solutions 6–10. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

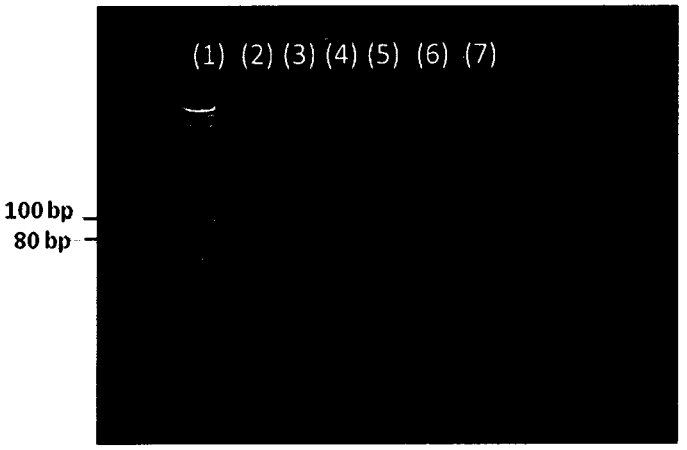


Figure 2-23: PCR results of elution solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 11–15; lane 7: negative control.

5.3.3.2.4. DNA samples may specifically bind to lysozyme

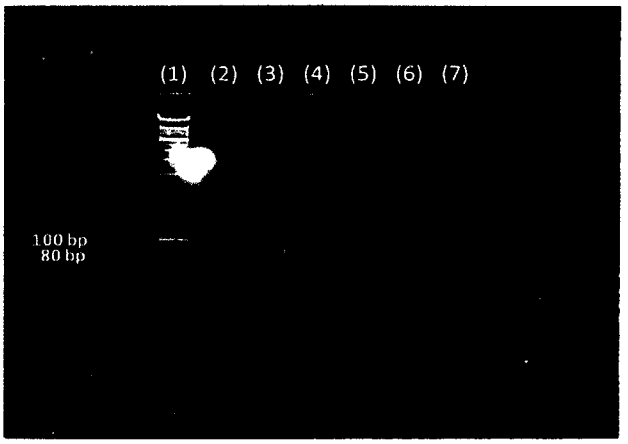


Figure 2-24: PCR results of elution solutions 6–10. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

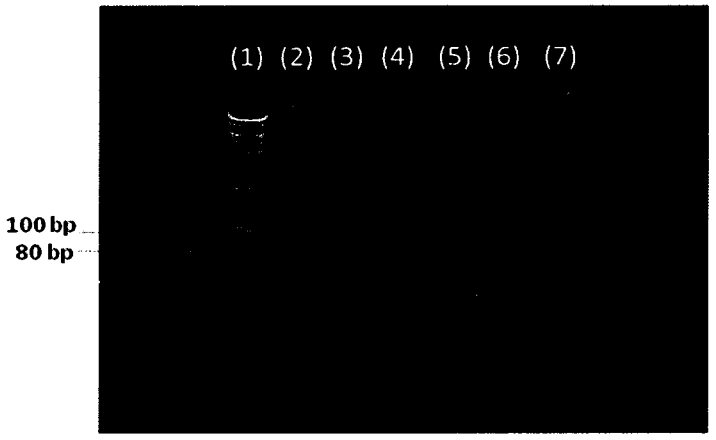


Figure 2-25: PCR results of elution solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 11–15; lane 7: negative control.

5.3.3.2.5. DNA samples may specifically bind to Fpg protein

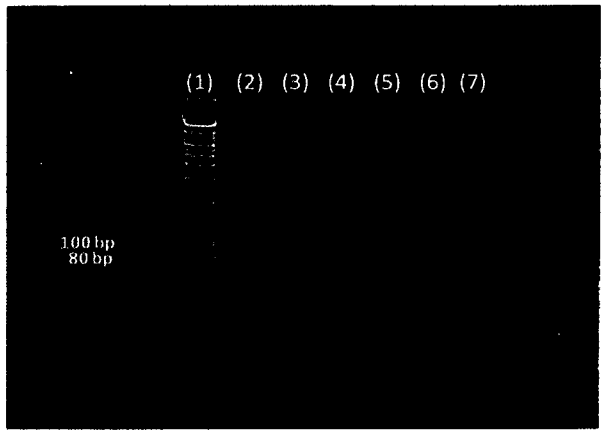


Figure 2-26: PCR results of elution solutions 6–10. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 6–10; lane 7: negative control.

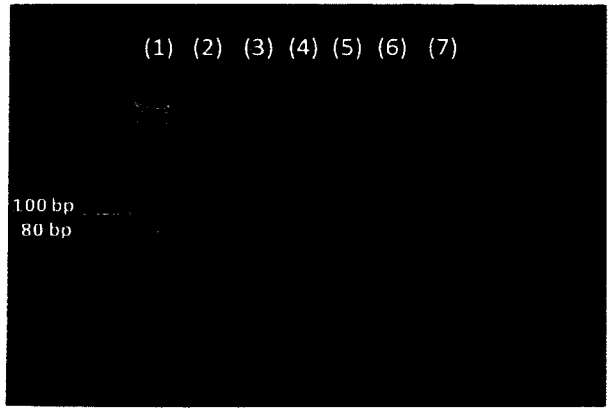


Figure 2-27: PCR results of elution solutions 11–15. Lane 1: 100 bp DNA ladder; lanes 2–6: elution solutions 11–15; lane 7: negative control.

In a comparison of all PCR results of last 5 elution solutions (11–15) from the 5 groups (Figures 2-19, 2-21, 2-23, 2-25, and 2-27), no clear product band can be observed. This confirms that after 30 washes in the washing step, 10 elution fractions are sufficient, and there is no need to wash the membrane 15 times with elution buffer to elute DNA from binding complexes.

However, if 25 washes in washing step are used (based on the conclusion shown in Section 5.3.3.1), it is better to wash the membrane 15 times with elution buffer in the eluting step, to ensure that as much ssDNA as possible can be eluted from the binding complexes and collected.

For the PCR results of elution solutions 6–10, it is difficult to observe amplification products in the groups of DNA which may specifically bind to BSA (Figure 2-21) or ovalbumin (Figure 2-23). These two figures demonstrate that BSA and ovalbumin may have very low binding affinity to ssDNA. Therefore, the remaining three groups of ssDNA (which may specifically bind to thrombin, Fpg protein, or lysozyme) were chosen to do DNA cloning and sequencing.

5.4. DNA sequencing

Several single-stranded DNA (ssDNA) sequences from three groups of elution solutions were amplified by PCR, and then amplified double-stranded DNA products were inserted into plasmid DNA. Bacteria with inserted recombinant plasmids were cultured on selective LB medium. Several single colonies from each selective medium were picked, and in total 80 colonies were chosen for DNA sequencing.

Around 60 sequences of ssDNA which may specifically bind to their target proteins (thrombin, Fpg protein, or lysozyme) were obtained after DNA sequencing.

Based on the primer sequences, sequences of the selected aptamers are determined from raw data and grouped into two groups. And then common bases of sequences in the same group are compared, with the assumption that the sequences which may have strong binding affinity to their target proteins should have many common bases.

5.4.1. Sequences of ssDNA that may specifically bind to thrombin

All the sequences are grouped by their primer sequences.

(1)

5'-AGC AGC ACA GAG GTC AGA TG **TTA TTC TGC GGA TCG GTT GTA GGC GGC CGG ATG GGT GT GG** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CAG AGG AGC CCC GCG TGT ATG CAT ATA TTG TAT CGG GG GT** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **GCT TAA CGA CTG TGC GTC TAG ATA GTC CAT GGC ATT AT AT** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **GAT CTA GTT GTG TGC CTG CCC TTG TCT TGA TGT GAG CT GC** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **ACT AAT CTG TGC GAA AAG TAA CTG CTT TTT GAT CAT CC CA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **ACC TAT CAG TGT CCT TGG TTA TAA TAG CTC CGG CAG TC TA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TCG AAC ATG AAT TGT CAA GAT TTG TGG TTT ATG ATA TG CT** CC TAT GCG TGC TAC CGT GAA-3'

Common bases compared:

```
..................TTATTCTGCGGATCGGTTGTAGGCGGCCGGATGGGTGTGG
.......CAGAGGAGCCCCGCGTGTATGCATATATTGTATCGGGGGT
GCTTAACGACTGTGCGTCTAGATAGTCCATGGCATTATAT
.........GATCTAGTTGTGTGCCTGCCCTTGTCTTGATGTGAGCTGC
.......ACTAATCTGTGCGAAAAGTAACTGCTTTTTGATCATCCCA
....................ACCTATCAGTGTCCTTGGTTATAATAGCTCCGGCAGTCTA
..............TCGAACATGAATTGTCAAGATTTGTGGTTTATGATATGCT
```

(2)

5'-TTC ACG GTA GCA CGC ATA GG **CCA CAC CCA TCC GGC CGC CTA CAA CCG ATC CGC AGA AT AA** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **ACC CCC GAT ACA ATA TAT GCA TAC ACG CGG GGC TCC TC TG** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **ATA TAA TGC CAT GGA CTA TCT AGA CGC ACA GTC GTT AA GC** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **GCA GCT CAC ATC AAG ACA AGG GCA GGC ACA CAA CTA GA TC** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGG GAT GAT CAA AAA GCA GTT ACT TTT CGC ACA GAT TA GT** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TAG ACT GCC GGA GCT ATT ATA**

ACC AAG GAC ACT GAT AG GT CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **AGC ATA TCA TAA ACC ACA AAT**

CTT GAC AAT TCA TGT TC GA CA TCT GAC CTC TGT GCT GCT-3'

Common bases compared:

```
·····CCACACCCATCCGGCCGCCTACAACCGATCCGCAGAATAA
·····ACCCCCGATACAATATATGCATACACGCGGGGCTCCTCTG
·····ATATAATGCCATGGACTATCTAGACGCACAGTCGTTAAGC
·····GCAGCTCACATCAAGACAAGGGCAGGCACACAACTAGATC
·····TGGGATGATCAAAAAGCAGTTACTTTTCGCACAGATTAGT
TAGACTGCCGGAGCTATTATAACCAAGGACACTGATAGGT
·····AGCATATCATAAACCACAAATCTTGACAATTCATGTTCGA
```

5.4.2. Sequences of ssDNA that may specifically bind to lysozyme

All the sequences are grouped by their primer sequences.

(1)

5'-AGC AGC ACA GAG GTC AGA TG **ACT TAC AAT ATT TTA AGG GTT**

TGC TAA TCG GGG ATG GG CT CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **AAA CAG TAA TCC CGT GGT CAC**

AGA ACG CAG CGC TTA GT AC CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CTT CGT TCG ATT TGA CGC CTT**

CCT CTC TCC TAG CGA GG CG CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TGA ATC ATC CAG CAT CTC ATA**

CTA AGC TAC GTA TAC CT TA CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **AAC CTC GAT GCC AGT GAT TTA ATT CGA GCC TAT TGT TT GC** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CAG GAG TTC ATA CCT TGA CTG TCG TTA GTA GCG GGA AA CA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TTC CGC AAA TTC TGT GTA AAA CAT CGC GCC AGT ATC TT TT** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **ACG TGG TTA ACT TAA CTC CCT TCG CGG TGG CTG GGT CT GA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TTC TTC ACG CTA CGG GAC ACA GTT ACG CGC GGT TCT GC CA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **GAG CAA GCG ACC AGT ATA TTG ATT ATC GTG CTA GAT GT CA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TAG CAT TTA AGA CTG CGT TGG GCG GGG GCT GGC CGG GT CA** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **ATT TCA CTC AAG AAC ACT TAT GGA GCG CAA TGA ACT CT GG** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CTG AAT GAC TGG CGC GGG CTC CGT CTG GAC CTG CTC TT GG** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CCG ATA GTC ACC ACT ATC CGC GGC CCT TGC GGT GGA TG TC** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TAC GTA ACT TCA CGC TCC TTA TCT TGA CGC ACT TGG CC CG** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **GTG GCT CAA CTC TAT GCA CCG ATG TCG GTC TAT CTG AT CG** CC TAT GCG TGC TAC CGT GAA-3'

5'-AGC AGC ACA GAG GTC AGA TG **GTT GAC TAA AGC ACA CAC TAT CCA CCC TCC CGC GGG TC CA** CC TAT GCG TGC TAC CGT GAA-3'

Common bases compared:

```
..........ACTTACAATATTTTAAGGGTTTGCTAATCGGGGATGGGCT
......AAACAGTAATCCCGTGGTCACAGAACGCAGCGCTTAGTAC
......................CTTCGTTCGATTTGACGCCTTCCTCTCTCCTAGCGAGGCG
.........TGAATCATCCAGCATCTCATACTAAGCTACGTATACCTTA
............AACCTCGATGCCAGTGATTTAATTCGAGCCTATTGTTTGC
....................CAGGAGTTCATACCTTGACTGTCGTTAGTAGCGGGAAACA
............TTCCGCAAATTCTGTGTAAAACATCGCGCCAGTATCTTTT
..............ACGTGGTTAACTTAACTCCCTTCGCGGTGGCTGGGTCTGA
..........TTCTTCACGCTACGGGACACAGTTACGCGCGGTTCTGCCA
.........GAGCAAGCGACCAGTATATTGATTATCGTGCTAGATGTCA
·TAGCATTTAAGACTGCGTTGGGCGGGGGCTGGCCGGGTCA
..........ATTTCACTCAAGAACACTTATGGAGCGCAATGAACTCTGG
......................CTGAATGACTGGCGCGGGCTCCGTCTGGACCTGCTCTTGG
.......CCGATAGTCACCACTATCCGCGGCCCTTGCGGTGGATGTC
TACGTAACTTCACGCTCCTTATCTTGACGCACTTGGCCCG
..........GTGGCTCAACTCTATGCACCGATGTCGGTCTATCTGATCG
.....GTTGACTAAAGCACACACTATCCACCCTCCCGCGGGTCCA
```

(2)

5'-TTC ACG GTA GCA CGC ATA GG **AGC CCA TCC CCG ATT AGC AAA CCC TTA AAA TAT TGT AA GT** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **GTA CTA AGC GCT GCG TTC TGT**

GAC CAC GGG ATT ACT GT TT CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **CGC CTC GCT AGG AGA GAG GAA**

GGC GTC AAA TCG AAC GA AG CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TAA GGT ATA CGT AGC TTA GTA**

TGA GAT GCT GGA TGA TT CA CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **GCA AAC AAT AGG CTC GAA TTA**

AAT CAC TGG CAT CGA GG TT CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGT TTC CCG CTA CTA ACG ACA**

GTC AAG GTA TGA ACT CC TG CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **AAA AGA TAC TGG CGC GAT GTT**

TTA CAC AGA ATT TGC GG AA CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TCA GAC CCA GCC ACC GCG AAG**

GGA GTT AAG TTA ACC AC GT CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGG CAG AAC CGC GCG TAA CTG**

TGT CCC GTA GCG TGA AG AA CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGA CAT CTA GCA CGA TAA TCA**

ATA TAC TGG TCG CTT GC TC CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGA CCC GGC CAG CCC CCG CCC**

AAC GCA GTC TTA AAT GC TA CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **CCA GAG TTC ATT GCG CTC CAT**

AAG TGT TCT TGA GTG AA AT CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **CCA AGA GCA GGT CCA GAC GGA GCC CGC GCC AGT CAT TC AG** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **GAC ATC CAC CGC AAG GGC CGC GGA TAG TGG TGA CTA TC GG** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **CGG GCC AAG TGC GTC AAG ATA AGG AGC GTG AAG TTA CG TA** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **CGA TCA GAT AGA CCG ACA TCG GTG CAT AGA GTT GAG CC AC** CA TCT GAC CTC TGT GCT GCT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TGG ACC CGC GGG AGG GTG GAT AGT GTG TGC TTT AGT CA AC** CA TCT GAC CTC TGT GCT GCT-3'

Similar bases in the sequences were compared:

```
.....AGCCCATCCCCGATTAGCAAACCCTTAAAATATTGTAAGT
.....GTACTAAGCGCTGCGTTCTGTGACCACGGGATTACTGTTT
.....CGCCTCGCTAGGAGAGAGGAAGGCGTCAAATCGAACGAAG
TAAGGTATACGTAGCTTAGTATGAGATGCTGGATGATTCA
.....GCAAACAATAGGCTCGAATTAAATCACTGGCATCGAGGTT
.....TGTTTCCCGCTACTAACGACAGTCAAGGTATGAACTCCTG
.....AAAAGATACTGGCGCGATGTTTTACACAGAATTTGCGGAA
.....TCAGACCCAGCCACCGCGAAGGGAGTTAAGTTAACCACGT
.....TGGCAGAACCGCGCGTAACTGTGTCCCGTAGCGTGAAGAA
.....TGACATCTAGCACGATAATCAATATACTGGTCGCTTGCTC
.....TGACCCGGCCAGCCCCCGCCCAACGCAGTCTTAAATGCTA
.....CCAGAGTTCATTGCGCTCCATAAGTGTTCTTGAGTGAAAT
CCAAGAGCAGGTCCAGACGGAGCCCGCGCCAGTCATTCAG
.....GACATCCACCGCAAGGGCCGCGGATAGTGGTGACTATCGG
.....CGGGCCAAGTGCGTCAAGATAAGGAGCGTGAAGTTACGTA
CGATCAGATAGACCGACATCGGTGCATAGAGTTGAGCCAC
.....TGGACCCGCGGGAGGGTGGATAGTGTGTGCTTTAGTCAAC
```

5.4.3. Sequences of ssDNA that may specifically bind to Fpg protein

All the sequences are grouped by primer sequence.

(1)

5'-AGC AGC ACA GAG GTC AGA TG **CGG CGG AGT AGC GCG TTG ATT ATG CCA AGT CGG TCC GT GC** CCT ATG CGT GCT ACC GTG AA-3'

5'-AGC AGC ACA GAG GTC AGA TG **CTG GAT TTC ATG ATC GCC ATG TCT CGA TTG TGG GCC GA AA** CCT ATG CGT GCT ACC GTG AA-3'

5'-AGC AGC ACA GAG GTC AGA TG **ACT CGG GGA ACC GGG GAG AGT GAA GCC TCG CGC AAT CA TA** CCT ATG CGT GCT ACC GTG AA-3'

5'-AGC AGC ACA GAG GTC AGA TG **TAC AGC GGC GCA ACG ATA ACG GCT CAC TAA ACC CGA CG AT** CCT ATG CGT GCT ACC GTG AA-3'

5'-AGC AGC ACA GAG GTC AGA TG **AGA GAT TAG CGT AGG GAA GCC GAT GCC GAG CCT CGC CC AA** CCT ATG CGT GCT ACC GTG AA-3'

5'-AGC AGC ACA GAG GTC AGA TG **AGT TGG CCG AAG CCA ACC TTC AAA GCG CAG ACG CCA TC AT** CCT ATG CGT GCT ACC GTG AA-3'

Comparing common bases in these sequences:

.....CGGCGGAGTAGCGCGTTGATTATGCCAAGTCGGTCCGTGC

.....CTGGATTTCATGATCGCCATGTCTCGATTGTGGGCCGAAA

.....ACTCGGGGAACCGGGGAGAGTGAAGCCTCGCGCAATCATA

.....TACAGCGGCGCAACGATAACGGCTCACTAAACCCGACGAT

AGAGATTAGCGTAGGGAAGCCGATGCCGAGCCTCGCCCAA

.....AGTTGGCCGAAGCCAACCTTCAAAGCGCAGACGCCATCAT

(2)

5'-TTC ACG GTA GCA CGC ATA GG **GCA CGG ACC GAC TTG GCA TAA TCA ACG CGC TAC TCC GC CG** CAT CTG ACC TCT GTG CTG CT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TTT CGG CCC ACA ATC GAG ACA TGG CGA TCA TGA AAT CC AG** CAT CTG ACC TCT GTG CTG CT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TAT GAT TGC GCG AGG CTT CAC TCT CCC CGG TTC CCC GA GT** CAT CTG ACC TCT GTG CTG CT-3'

5'-TTC ACG GTA GCA CGC ATA GG **ATC GTC GGG TTT AGT GAG CCG TTA TCG TTG CGC CGC TG TA** CAT CTG ACC TCT GTG CTG CT-3'

5'-TTC ACG GTA GCA CGC ATA GG **TTG GGC GAG GCT CGG CAT CGG CTT CCC TAC GCT AAT CT CT** CAT CTG ACC TCT GTG CTG CT-3'

5'-TTC ACG GTA GCA CGC ATA GG **ATG ATG GCG TCT GCG CTT TGA AGG TTG GCT TCG GCC AA CT** CAT CTG ACC TCT GTG CTG CT-3'

Comparing common bases in the sequences:

.....GCACGGACCGACTTGGCATAATCAACGCGCTACTCCGCCG

.....TTTCGGCCCACAATCGAGACATGGCGATCATGAAATCCAG

.....TATGATTGCGCGAGGCTTCACTCTCCCCGGTTCCCCGAGT

.....ATCGTCGGGTTTAGTGAGCCGTTATCGTTGCGCCGCTGTA

.....TTGGGCGAGGCTCGGCATCGGCTTCCCTACGCTAATCTCT

ATGATGGCGTCTGCGCTTTGAAGGTTGGCTTCGGCCAACT

After comparing all the common bases of sequences, it is difficult to assume which sequences have stronger binding affinities. In order to identify specific sequences (aptamers) that have specific affinity for the target protein, I examined the secondary structures of these sequences; and these results are described in Chapter 3.

6. Conclusions

Five standard proteins were separated by one-dimensional gel electrophoresis, and after being transferred to PVDF membrane, they were incubated with an 80-nt-long ssDNA library. After a single cycle membrane selection with optimized conditions, different groups of sequences which may bind to their specific target proteins were generated simultaneously. Further characterization of the secondary structures and binding affinity of these selected sequences will be described in Chapter 3.

Chapter 3: Characterization of the Selected ssDNA After Single Cycle Selection

1. Introduction

The new single cycle selection (SCS) method is described in Chapter 2. To evaluate the usefulness of this new technique, I will characterize the structures and binding affinity and specificity of the aptamers selected using the new method. Aptamers can specifically bind to their target proteins, similarly to antibody-antigen binding (Nimjee, Rusconi, and Sullenger, 2005). The affinity and specificity of the aptamers against the target protein are dependent on the equilibrium of binding and dissociation processes, represented by their binding constant K and dissociation constant K_d, respectively, as shown in the equations below. To evaluate the specificity and utility of a new aptamer, it is essential to determine its binding affinity or dissociation constant (K_d value). I will focus on the determination of dissociation constant K_d in this chapter.

$$\mathbf{A\,(aptamer) + P\,(protein)} \underset{\mathbf{K_d}}{\overset{\mathbf{K}}{\rightleftharpoons}} \mathbf{A-P\,(binding\ complex)}$$

Equation 1:

$$K_d = \frac{[A][P]}{[A-P]}$$

[A], [P], and [A–P] represent the concentrations of aptamer, protein, and aptamer-protein complex at equilibrium, respectively; as a result, the unit of K_d is molar (M).

The dissociation constant represents the stability of the aptamer-protein complex, in other words, the binding affinity of the aptamer against the target protein. When 50% of the aptamer binds to the target protein, K_d equals the concentration of the free protein without the aptamer. Thus the lower the K_d is, the stronger the aptamer binds to the protein – high binding affinity.

Capillary electrophoresis (CE) is a common method for detecting binding the affinity or dissociation constant. CE with laser-induced fluorescence detection (CE-LIF) can provide high sensitivity detection with very small sample volume (~ nL). Aptamer-protein complexes and free aptamers have different electrophoretic mobility under the same conditions, thus the complexes and free aptamers can be separated with CE. The aptamer is labeled with a fluorescent probe and the fluorescent signals from the complexes and free aptamers can be separated and determined by CE-LIF. In this study, the selected aptamers (ssDNA) described in Chapter 2 are labeled with 6-carboxy-fluorescein (6-FAM) on the 5' end, because this fluorescent label can be excited by the argon ion laser. The fluorescent signals (peak areas) from the complexes and free aptamers are determined, corresponding to their concentrations. Thus, the K_d values of the aptamer-protein complexes can be determined based on the fluorescence signals. The method used for the

determination of K_d is described as follows.

Based on **Equation 1**, K_d = [Protein]×[Aptamer]/[Complex]

where [Aptamer], [Protein], and [Complex] are the concentrations of aptamer, protein, and aptamer-protein complex at equilibrium.

A parameter R is defined as $R = [Aptamer]_{eq} / ([Aptamer]_0 - [Aptamer]_{eq})$

where $[Aptamer]_{eq}$ is the concentration of free aptamers at equilibrium, and $[Aptamer]_0$ is the initial concentration of the aptamer prior to the binding.

Then $[Aptamer]_0 = k \times \mathbf{a}_0$, $[Aptamer]_{eq} = k \times \mathbf{a}_{eq}$, where k is a constant for the conversion, $\mathbf{a}_0$ is the fluorescent signal of the aptamers before binding, and $\mathbf{a}_{eq}$ is the fluorescent signal of the free aptamers at equilibrium.

Thus the parameter R can be defined as $R = \mathbf{a}_{eq} / (\mathbf{a}_0 - \mathbf{a}_{eq})$

Based on these equations, K_d can be determined by **Equation 2** (Huang et al., 2004):

$K_d = \{[Protein]_0 \times (1+R) - [Aptamer]_0\} / (1+1/R)$

where [Protein] represents the concentration of protein, [Aptamer] is the concentration of aptamer (ssDNA), and $[Protein]_0$ and $[Aptamer]_0$ represent the initial concentrations of the protein and the aptamer before binding.

2. Objective

The objectives of this chapter are to determine the binding affinity of the ssDNA sequences that have been selected using the new single cycle selection

(SCS) method (Chapter 2), and to validate and the demonstrate advantages and utility of the new SCS method for generation of new aptamers for multiple proteins.

3. Experimental

3.1. Materials

Reagents purchased from Invitrogen (Burlington, ON) include: bovine serum albumin (BSA), lysozyme, and Fpg protein.

The 10×Tris Glycine (TG) buffer was supplied by BioRad Laboratories (Mississauga, ON).

Fluorescein and 1 M Tris-HCl solution were produced by Sigma-Aldrich (Oakville, ON). Selected ssDNA sequences with a 6'-FAM label on 5' end were synthesized by Integrated DNA Technologies (Coralville, IA, USA).

3.2. Secondary structure analysis

Secondary structures of the selected ssDNA sequences were analyzed using the IDT SciTools OligoAnalyzer (version 3.1, Integrated DNA Technologies). In order to estimate which DNA sequences may have stronger binding affinities with stable structures during the binding process, two dynamic parameters, ΔG and Tm, were evaluated. ΔG (the change in Gibbs free energy) represents the lowest energy needed to form secondary structures for a sequence; the lower ΔG

represents a more stable structure. Tm is the melting temperature of DNA sequences. When the temperature of the experiments exceeds the Tm of a DNA sequence, the secondary structure of this DNA sequence will become unstable. A desirable aptamer should have low ΔG and high Tm. By analyzing ΔG and Tm, the selected ssDNA sequences could be compared and their properties estimated.

3.3. Determination of binding affinity using capillary electrophoresis with laser-induced fluorescence (CE-LIF)

Based on the secondary structure analysis and evaluation of two dynamic parameters, three ssDNA sequences labeled with 6-FAM were synthesized. K_d values of these fluorescently labeled DNA samples were determined using the in-house built CE-LIF system that has been described in Section 4.3 of Chapter 1. An uncoated fused-silica capillary with 20 μm inner diameter, 140 μm outer diameter, and 30 cm length (Polymicro Technologies, Phoenix, AZ, USA) was used for separation of the aptamer-protein complexes and unbound aptamer.

3.3.1. Sample preparation

The incubation buffer consisted of 10 mM Tris-HCl with pH 7.4 and 1 mM $MgCl_2$. The fluorescently labeled ssDNA samples were prepared in the incubation buffer with specific concentrations that will be described below. Fluorescein was prepared at 25 nM with the incubation buffer immediately prior to the CE analysis to avoid photo bleaching, and it was used as the internal standard (IS). BSA (0.5

mg) was dissolved in 1 mL incubation buffer, and this solution was used as the blocking buffer to reduce the adsorption of the target protein onto the capillary wall. The CE running buffer for the binding experiments was 1×TG buffer that was directly diluted from 10×TG buffer.

Mixture samples for determination of aptamer-protein binding consisted of ssDNA, target protein, BSA (blocking buffer), and IS in the incubation buffer. To obtain a desired final concentration of the ssDNA and target protein in a mixture sample, the initial concentrations of ssDNA and target protein were 10 times higher than the desired concentration of the mixture sample. The incubation mixture consisted of 2 μL of an initial concentration of ssDNA, 2 μL of an initial concentration of target protein, 2 μL of 25 nM IS, 2 μL of 0.5 mg/ml BSA, and 12 μL of incubation buffer. The final volume of a mixture sample for binding detection was 20 μL. The mixture was incubated for 20 minutes in the dark at room temperature.

3.3.2. Capillary electrophoresis detection

As mentioned above, the in-house built CE-LIF system was used for the binding affinity detection. Electrophoresis separation was performed with a potential of 18 kV at room temperature. Samples were injected electrokinetically into the capillary using 18 kV for 5 seconds.

Before CE analysis, the capillary was prepared by running 20 mM NaOH

solution for 20 minutes, at which time the current reached a steady baseline, followed by a series of washes. The capillary was first washed with deionized water for 10 minutes until the baseline became stable with near zero signal readout of the current wave; then it was washed by running 1×TG buffer, when the baseline was stable with low readout (approximately 0.1–0.12 unit) of the current wave. Then the capillary was ready for sample analysis. Because CE can only detect fluorescence from fluorescently labeled DNA samples, ssDNA samples with the same concentration were incubated with different concentrations of the target protein to determine the K_d value. Two or three mixtures at each concentration level of the protein were prepared, each sample was injected three to four times, and fluorescent signals were recorded to determine the binding affinity.

The fluorescent label (6'-FAM label) on the 5' end of ssDNA was excited by 488 nm light and signals were collected after filtering through a 515-nm band-pass filter. After analysis of 10 samples, the capillary was washed with 20 mM NaOH solution for 10 minutes, followed by deionized water for 10 minutes, and then by the CE running buffer until the stable baseline was established.

4. Results and Discussion

4.1. Secondary structure analysis

Secondary structures of the selected ssDNA are dependent on a number of

parameters. To determine the ΔG and Tm of these 80-nt-long ssDNA sequences, the important parameters include the concentrations of oligonucleotides, Na^+, Mg^{2+}, and dNTPs; experimental temperature; suboptimality; maximum folding bases; and folding start and stop positions of sequences. These parameters were considered in the secondary structure analysis. The specific conditions of these parameters (same as the conditions of SCS) are summarized in Table 3-1.

4.1.1. Conditions

Table 3-1: Conditions of secondary structure analysis for 80-nt-long ssDNA sequences.

Oligonucleotide type:	ssDNA	**Oligo concentration:**	30 μM
Na+ concentration:	100 mM	**Mg2+ concentration:**	2 mM
dNTPs concentration:	0 mM	**Temperature:**	25 °C
Start position:	1	**Stop position:**	80
Maximum folding:	20	**Suboptimality:**	50%

4.1.2. Secondary structure analysis of the sequences selected against thrombin

A total of fourteen sequences that were selected against thrombin are presented in Section 5.4.1 of Chapter 2. They are organized into two groups (group 1 and 2) based on their primer sequences.

4.1.2.1. Secondary structures of sequences

Every sequence had several possible secondary structures with different dynamic parameters, and different sequences had different secondary structures.

For example, 4 different secondary structures with 4 different sequences are shown in Figure 3-1. Figure 3-2 shows the 4 different possible secondary structures of the same DNA sequence.

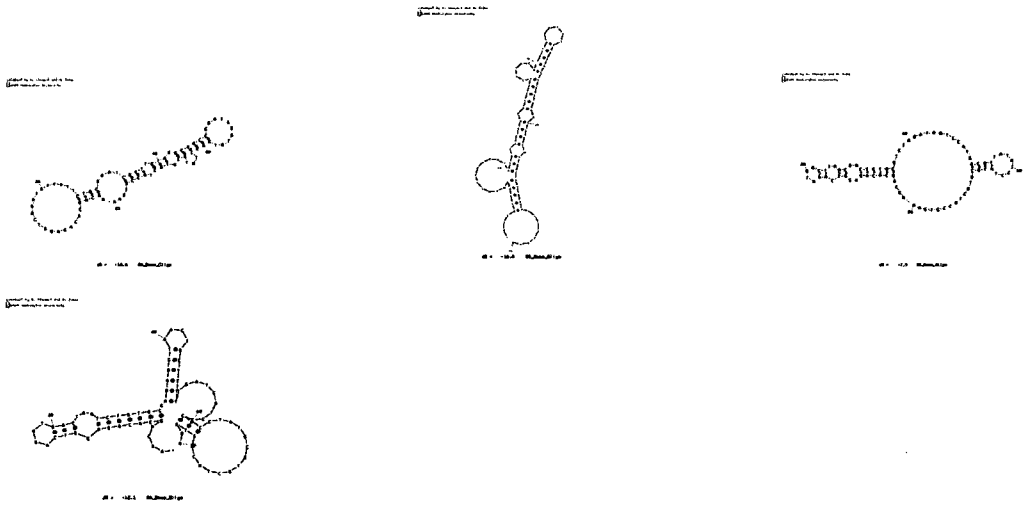


Figure 3-1: Secondary structure of 4 different sequences that may specifically bind to thrombin.

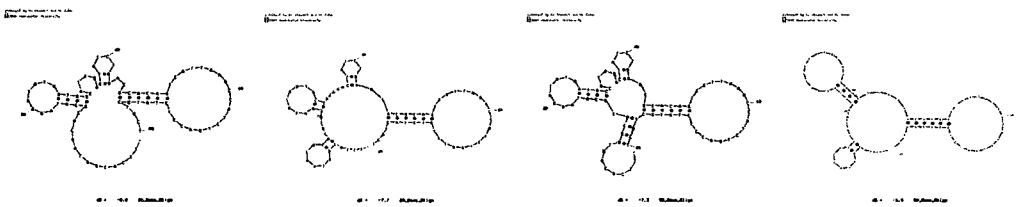


Figure 3-2: Four possible secondary structures of sequence 3 in group 2.

4.1.2.2. Dynamic parameters of sequences

As mentioned above, each sequence had several possible secondary structures with different dynamic parameters. By comparing the two dynamic parameters, the most stable secondary structure of each sequence was determined. The two dynamic parameters (ΔG and Tm) of these secondary structures are summarized

in Tables 3-2 and 3-3.

Table 3-2: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 1 that were selected against thrombin.

Group 1	ΔG (kcal/mole)	Tm (°C)
Sequence 1	-10.6	52.1
Sequence 2	-15.97	63.5
Sequence 3	-7.91	46.7
Sequence 4	-7.78	43.5
Sequence 5	-12.12	51.1
Sequence 6	-7.44	53.5
Sequence 7	-3.19	38.8

Table 3-3: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 2 that were selected against thrombin.

Group 2	ΔG (kcal/mole)	Tm (°C)
Sequence 1	-9.31	53.7
Sequence 2	-11.66	55.8
Sequence 3	-8.75	49.9
Sequence 4	-8.26	44.2
Sequence 5	-12.06	51.3
Sequence 6	-9.51	52.1
Sequence 7	-3.54	38.1

4.1.2.3. Secondary structures of selected ssDNA

For the ssDNA sequences that may specifically bind to thrombin, sequences 2 and 5 from group 1 and sequences 2 and 5 from group 2 were selected based on the low ΔG and high Tm. The secondary structures of these sequences are shown in Figures 3-3 to 3-6.

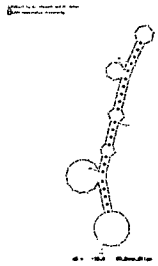


Figure 3-3: Secondary structure of sequence 2 from group 1.

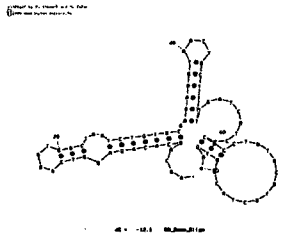


Figure 3-4: Secondary structure of sequence 5 from group 1.

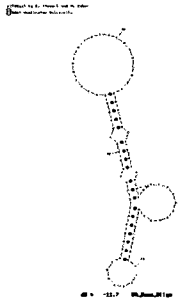


Figure 3-5: Secondary structure of sequence 2 from group 2.

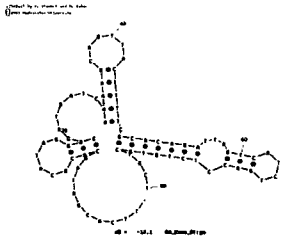


Figure 3-6: Secondary structure of sequence 5 from group 2.

An aptamer that is specific to thrombin has been reported by Zhang, Li and Le (2008). This aptamer consists of a 38-nt-long sequence: 5'-CAG TCC GTG

GTA GGG CA GGT TGG GGT GAC TTC GTG GAA-3', and it has strong binding affinity to thrombin with a K_d value of 0.5 nM.

I attempted to analyze the secondary structure of this sequence; however, the sequence was too short to form a stable secondary structure in the binding buffer. In order to compare it with the structures I have obtained for thrombin, the two groups of primer sequences of the ssDNA library that I used were added to the 5' and 3' ends of the 38mer. Then the two 78-nt-long sequences were analyzed to estimate their secondary structures.

Sequence thrombin-1:

5'-AGC AGC ACA GAG GTC AGA TG CAG TCC GTG GTA GGG CA GGT TGG GGT GAC TTC GTG GAA CC TAT GCG TGC TAC CGT GAA-3'

Sequence thrombin-2:

5'-TTC ACG GTA GCA CGC ATA GG CAG TCC GTG GTA GGG CA GGT TGG GGT GAC TTC GTG GAA CA TCT GAC CTC TGT GCT GCT-3'

The most stable secondary structures of these two sequences are shown in Figures 3-7 and 3-8.

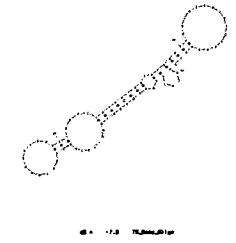


Figure 3-7: Secondary structure of 78-nt-long sequence thrombin-1.

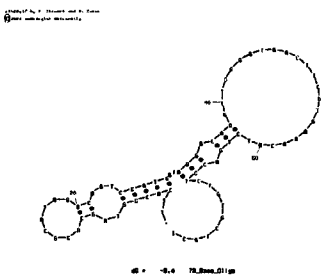


Figure 3-8: Secondary structure of 78-nt-long sequence thrombin-2.

Similarities in the structures in Figures 3-3 and 3-7 were found. Also, similar structures in Figures 3-5 and 3-8 are also identified under the same binding conditions. Thus sequence 2 from group 1 and sequence 2 from group 2 that I selected against thrombin have similar secondary structures to those of the previously reported aptamers. This suggests that my new method is capable of selecting sequences that may have strong binding affinities to target thrombin.

Bock's research (1992) found that the key sequence for binding thrombin was **GGTTGGTGTGGTTGG**. By comparing the sequences that described in Section 5.4.1 of Chapter 2, sequence 1 (5'-AGC AGC ACA GAG GTC AGA TG TTA TTC TGC GGA TC**G GTT G**TA GGC GGC CGG ATG **GGT GT GG** CC TAT GCG TGC TAC CGT GAA-3') and sequence 7 (5'-AGC AGC ACA GAG GTC AGA TG TCG AAC ATG AAT TGT CAA GA**T TTG TGG TTT** ATG ATA TG CT CC TAT GCG TGC TAC CGT GAA-3') from group 1 also contain such sequences. This further supports that the new method described in Chapter 2 is capable of generating new sequences that may have strong binding affinities to thrombin. I will further determine the binding affinities of these 4 sequences to the

target protein (thrombin). Similar approaches were used to determine secondary structures for the ssDNA sequences selected against the other target proteins.

4.1.3. Secondary structure analysis of sequences selected against lysozyme

A total of thirty-four sequences that were selected against lysozyme are described in Section 5.4.2 of Chapter 2. They are grouped into two groups (group 1 and 2) by the different primer sequences. Possible secondary structures and the two dynamic parameters ΔG and Tm of each sequence were analyzed.

4.1.3.1. Dynamic parameters of sequences

Tables 3-4 and 3-5 present the two dynamic parameters of the most stable secondary structure of each sequence.

Table 3-4: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 1 that were selected against lysozyme.

Group 1	ΔG (kcal/mole)	Tm (°C)
Sequence 1	-4.3	37.4
Sequence 2	-6.84	42.1
Sequence 3	-8.53	56.2
Sequence 4	-6.7	42.8
Sequence 5	-6.5	49.4
Sequence 6	-5.73	44.2
Sequence 7	-7.22	51.1
Sequence 8	-10.67	51.4
Sequence 9	-7.2	42.6
Sequence 10	-5.95	39.8
Sequence 11	-6.48	44.6
Sequence 12	-6.89	39.2
Sequence 13	-10.92	48.9
Sequence 14	-9.07	44.6

Sequence 15	-7.39	52.2
Sequence 16	-9.7	54.5
Sequence 17	-7.35	46.7

Table 3-5: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 2 that were selected against lysozyme.

Group 2	ΔG (kcal/mole)	Tm (°C)
Sequence 1	-4.77	39.4
Sequence 2	-8.13	43.4
Sequence 3	-11.18	54.6
Sequence 4	-9.7	46.3
Sequence 5	-8.71	48.1
Sequence 6	-6.72	41.9
Sequence 7	-6.98	53.7
Sequence 8	-11.75	46.5
Sequence 9	-7.97	44.5
Sequence 10	-7.52	43.3
Sequence 11	-5.31	37.1
Sequence 12	-7.91	41.5
Sequence 13	-8.7	50.5
Sequence 14	-12.14	44.5
Sequence 15	-8.99	46.1
Sequence 16	-12.43	60.7
Sequence 17	-10.24	51.0

4.1.3.2. Secondary structures of selected ssDNA

For ssDNA that were selected against lysozyme, sequences 8 and 13 from group 1 and sequences 3, 8, 14, and 16 from group 2 were chosen based on their low ΔG and high Tm. The secondary structures of these sequences are shown in Figures 3-9 to 3-14.

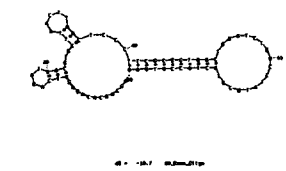


Figure 3-9: Secondary structure of sequence 8 from group 1.

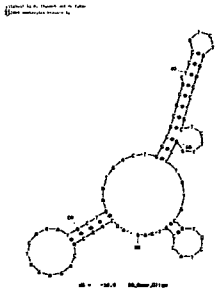


Figure 3-10: Secondary structure of sequence 13 from group 1.

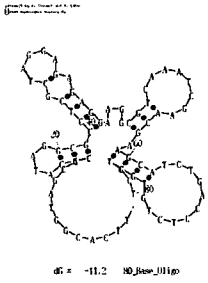


Figure 3-11: Secondary structure of sequence 3 from group 2.

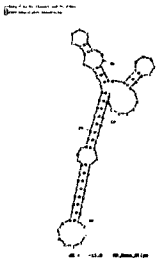


Figure 3-12: Secondary structure of sequence 8 from group 2.

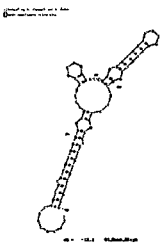


Figure 3-13: Secondary structure of sequence 14 from group 2.

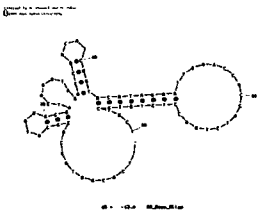


Figure 3-14: Secondary structure of sequence 16 from group 2.

An aptamer specific to lysozyme has been generated (Cox and Ellington, 2001). It contains 42 nt (5'-ATC TAC GAA TTC ATC AGG GCT AAA GAG TGC AGA GTT ACT TAG-3') and has a K_d of 31 nM.

Similarly to the situation described above on thrombin aptamer, because this sequence is also too short to form a stable secondary structure for secondary structure analysis, two groups of primer sequences were added to the aptamer sequence to have two 82-nt-long sequences, which were analyzed for their secondary structures.

Sequence lysozyme-1:

5'-AGC AGC ACA GAG GTC AGA TG ATC TAC GAA TTC ATC AGG GCT AAA GAG TGC AGA GTT ACT TAG CC TAT GCG TGC TAC CGT GAA-3'

Sequence lysozyme-2:

5'-TTC ACG GTA GCA CGC ATA GG ATC TAC GAA TTC ATC AGG GCT AAA GAG TGC AGA GTT ACT TAG CA TCT GAC CTC TGT GCT GCT-3'

The most stable secondary structures of these two sequences are shown in Figures 3-15 and 3-16.

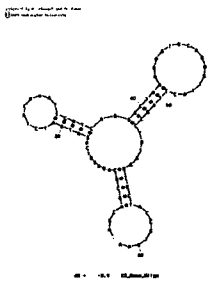


Figure 3-15: Secondary structure of sequence lysozyme-1.

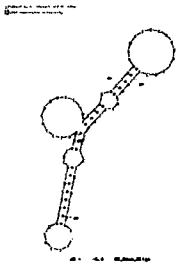


Figure 3-16: Secondary structure of sequence lysozyme-2.

The stable secondary structures that I selected for lysozyme are shown in Figures 3-10, 3-12, and 3-13, and they have similar secondary structures with the one in Figure 3-16. Once again, this supports that the new selection method I developed in Chapter 2 is capable of generating sequences that may be specific to lysozyme. For example, sequence 13 from group 1 and sequences 8 and 14 from group 2 have a similar secondary structure to that of the known aptamer specific to lysozyme. Further confirmation of the binding affinities of these three

sequences will be determined using CE-LIF.

4.1.4. Secondary structure analysis of the sequences selected against Fpg protein

Twelve sequences were selected against Fpg protein and were described in Section 5.4.3 of Chapter 2. These sequences are organized into two groups (group 1 and 2) according to their primer sequences and named as sequence 1 to 6 in each group. Each sequence was analyzed for its possible secondary structures and dynamic parameters.

4.1.4.1. Dynamic parameters of sequences

The two dynamic parameters of the most stable secondary structure of each sequence are listed in Tables 3-6 and 3-7.

Table 3-6: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 1 that were selected against Fpg protein.

Group 1	ΔG (kcal/mole)	Tm (°C)
Sequence 1	-7.09	45.5
Sequence 2	-9.36	50.2
Sequence 3	-9.0	48.9
Sequence 4	-9.26	44.7
Sequence 5	-9.75	50.8
Sequence 6	-13.88	48.0

Table 3-7: Results of the ΔG and Tm of the most stable secondary structure of each ssDNA sequence in group 2 that were selected against Fpg protein.

Group 2	ΔG (kcal/mole)	Tm (°C)

Sequence 1	-8.55	51.6
Sequence 2	-8.45	46.4
Sequence 3	-10.47	58.4
Sequence 4	-5.66	39.7
Sequence 5	-7.26	46.9
Sequence 6	-12.18	50.0

4.1.4.2. Secondary structures of selected ssDNA

Comparing the two dynamic parameters in Tables 3-6 and 3-7 and the secondary structures, the ssDNA that may bind specifically to Fpg protein are selected based on low ΔG and high Tm. As a result, three ssDNA sequences were selected: sequence 6 of group 1, and sequences 6 and 3 of group 2. These three sequences were used to further determine the binding affinity. The secondary structures of these three sequences are shown in Figures 3-17 to 3-19.

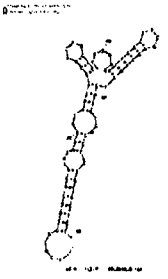


Figure 3-17: Secondary structure of sequence 6 from group 1.

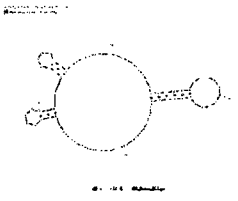


Figure 3-18: Secondary structure of sequence 3 from group 2.

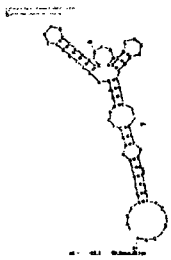


Figure 3-19: Secondary structure of sequence 6 from group 2.

No aptamer that is specific to Fpg protein has been previously reported. Here I will demonstrate the binding affinity of the three sequences that were selected against Fpg protein using CE-LIF. These ssDNA sequences were synthesized with a fluorescent label by Integrated DNA Technologies; sequence 6 of group 1, and sequences 6 and 3 of group 2 were named Jing 1, 2, and 3 (aptamer 1, 2, and 3).

4.2. Determination of binding affinity of the selected aptamer using capillary electrophoresis with laser-induced fluorescence detection

To determine the binding affintiy with CE, three fluorescently labeled sequences (Jingl, 2, and 3) which were selected against Fpg protein were synthesized by Integrated DNA Technologies. Because no aptamer is available for Fpg protein, here I will focus on determining the binding affinity of the ssDNA sequences selected for this protein.

4.2.1. CE results of three selected aptamers

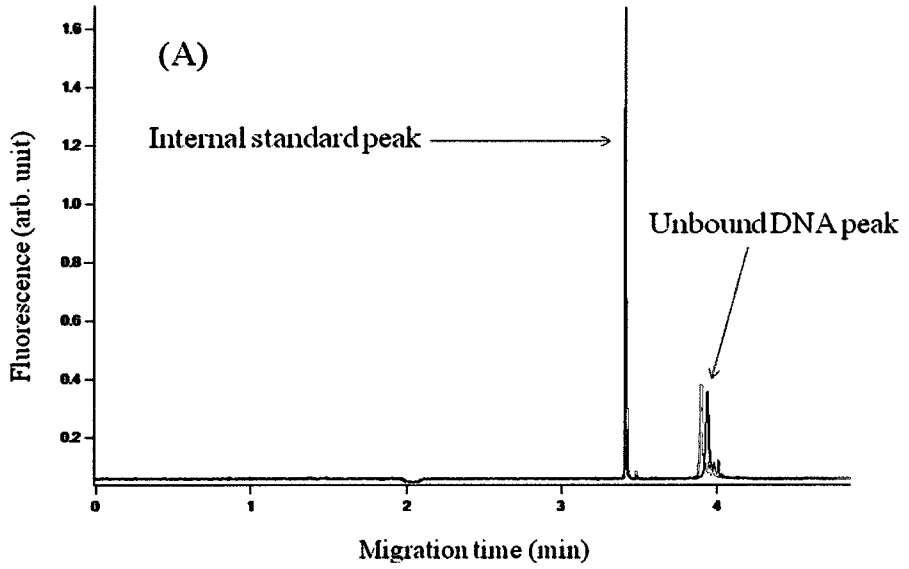
To determine the migration time of internal standard (IS), unbound ssDNA,

and aptamer-protein complex, the three fluorescently labeled ssDNA were incubated with and without the target Fpg protein in incubation buffer. The incubated samples were then analyzed with CE.

To prepare blank samples (without target protein), 5 nM aptamer, 2.5 nM IS, and 50 μg/mL BSA were mixed in the incubation buffer (10 mM Tris-HCl at pH 7.4). To confirm whether the three aptamers can bind to the target Fpg protein, 5 nM aptamer, 2.5 nM IS, 50 μg/mL BSA and a high concentration Fpg protein (500 nM) were incubated in the incubation buffer (10 mM Tris-HCl at pH 7.4) and then the mixtures were analyzed with CE-LIF. Each samples was analyzed twice.

4.2.1.1. CE results of aptamer 1 (Jing 1)

Aptamer 1 (IDT: Jing 1) is the selected sequence 6 from group 1: 5'-AGC AGC ACA GAG GTC AGA TG AGT TGG CCG AAG CCA ACC TTC AAA GCG CAG ACG CCA TC AT CCT ATG CGT GCT ACC GTG AA-3'.



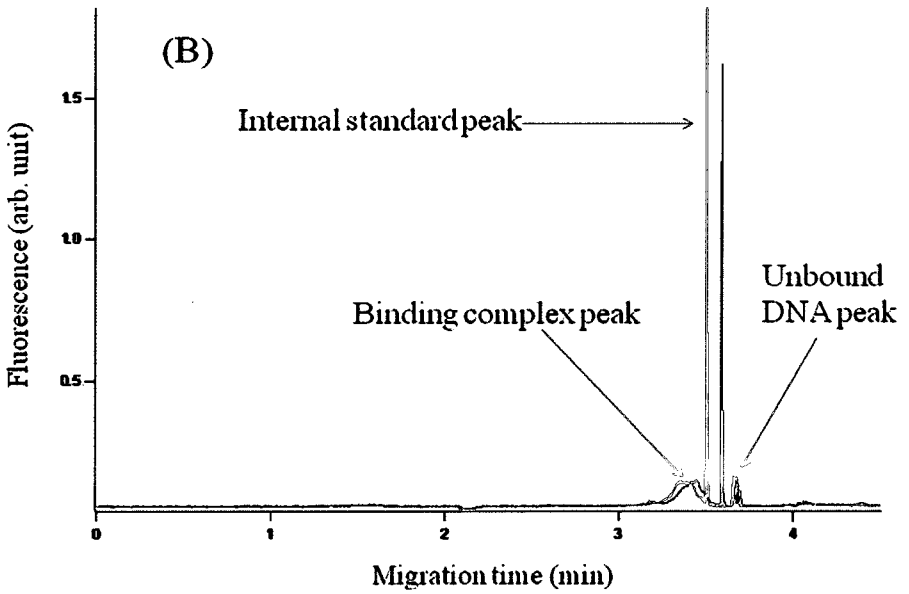
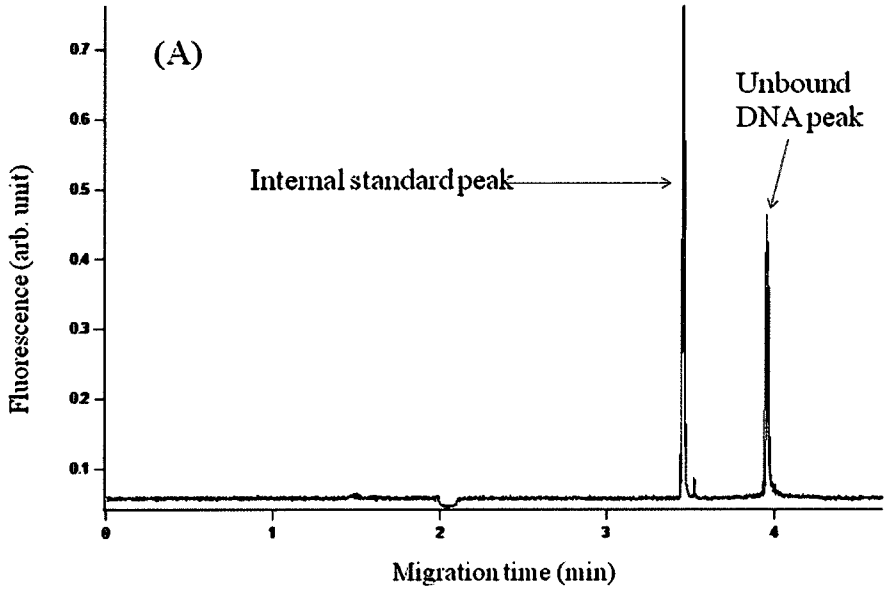


Figure 3-20: CE result of aptamer 1 with and without target Fpg protein. (A): 5 nM aptamer 1 + 2.5 nM IS + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). (B): 500 nM Fpg protein + 5 nM aptamer 1 + 2.5 nM internal standard + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). Blue and red traces are from two separate injections.

4.2.1.2. CE results of aptamer 2 (Jing 2)

Aptamer 2 (IDT: Jing 2) is the selected sequence 6 from group 2: 5'-TTC ACG GTA GCA CGC ATA GG ATG ATG GCG TCT GCG CTT TGA AGG TTG GCT TCG GCC AA CT CAT CTG ACC TCT GTG CTG CT-3'.



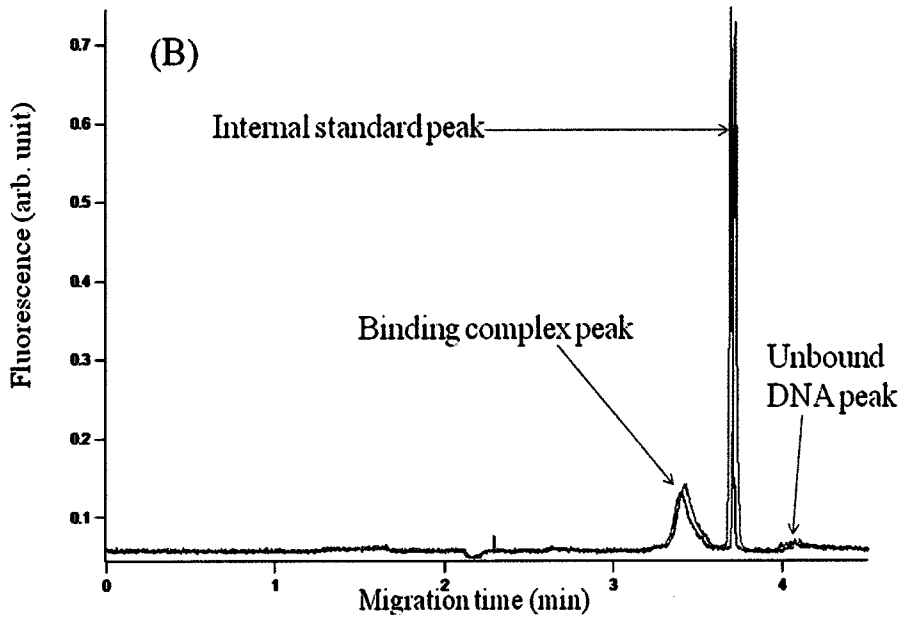


Figure 3-21: CE result of aptamer 2 with and without target Fpg protein. (A): 5 nM aptamer 2 + 2.5 nM IS + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). (B): 500 nM Fpg protein + 5 nM aptamer 2 + 2.5 nM internal standard + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). Blue and red traces are from two separate injections.

4.2.1.3. CE results of aptamer 3 (Jing 3)

Aptamer 3 (IDT: Jing 3) is the selected sequence 3 from group 2: 5'-TTC ACG GTA GCA CGC ATA GG TAT GAT TGC GCG AGG CTT CAC TCT CCC CGG TTC CCC GA GT CAT CTG ACC TCT GTG CTG CT-3'.

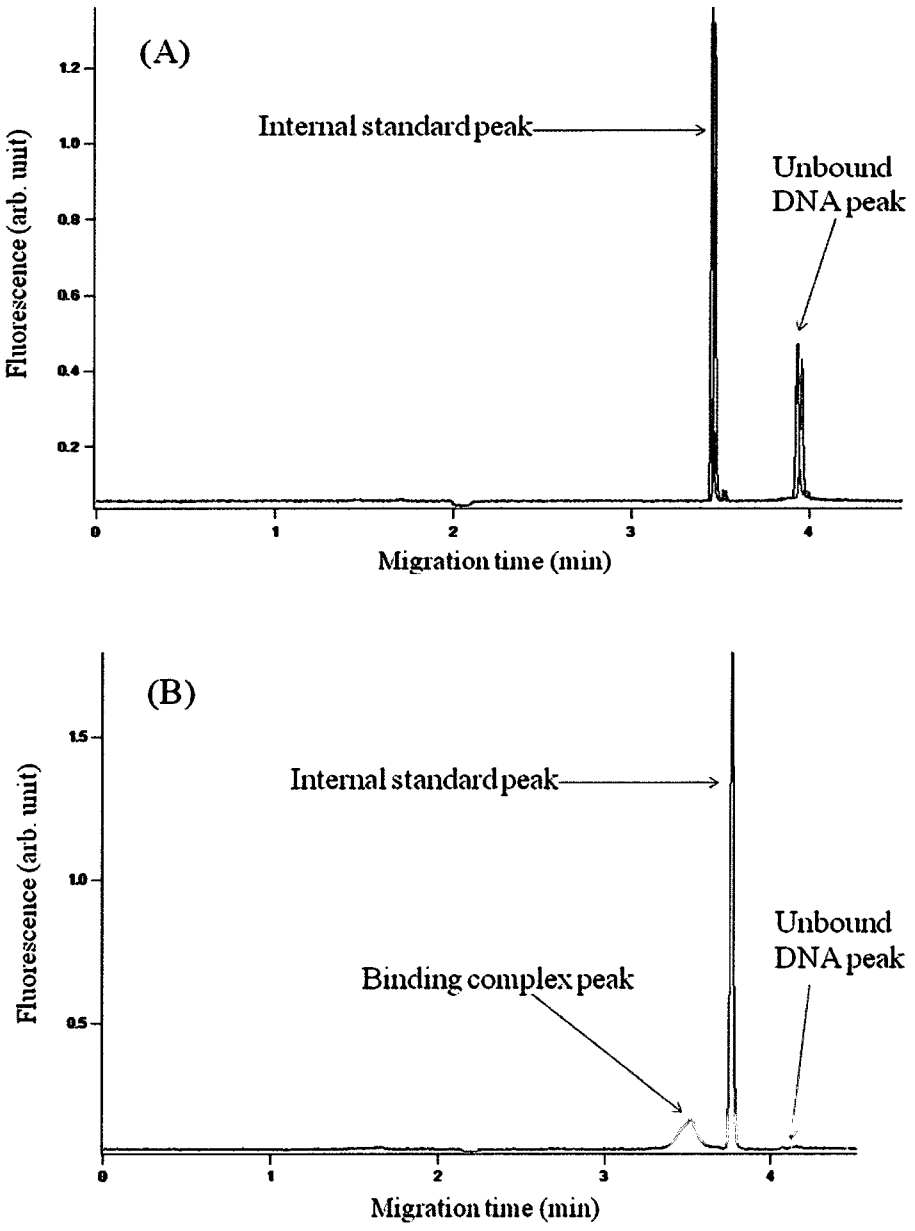


Figure 3-22: CE result of aptamer 3 with and without target Fpg protein. (A): 5 nM aptamer 3 + 2.5 nM IS + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). (B): 500 nM Fpg protein + 5 nM aptamer 3 + 2.5 nM internal standard + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). Blue and red traces are from two separate injections.

Figures 3-20(A), 3-21(A), and 3-22(A) clearly show the migration time of the internal standard and unbound DNA. The internal standard (fluorescein) migrates at 3.5 minutes, and the unbound DNA migrates more slowly, at around 4 minutes.

Figures 3-20(B), 3-21(B), and 3-22(B) show that all three aptamers can bind to the target Fpg protein. The migration time of the aptamer-protein complex is also determined. The aptamer-protein complex migrates faster than the internal standard and its migration time is around 3.3 minutes. These results demonstrate the complete separation of the aptamer-protein complex from the free aptamer and internal standard.

4.2.2. CE results of mixtures of high concentrations of aptamer and Fpg protein

To find out which aptamer has the strongest binding affinity to target protein, a high concentration (500 nM) of each aptamer was incubated with a high concentration (100 nM) of its target Fpg protein and the mixtures were analyzed on CE.

4.2.2.1. CE results of blank samples

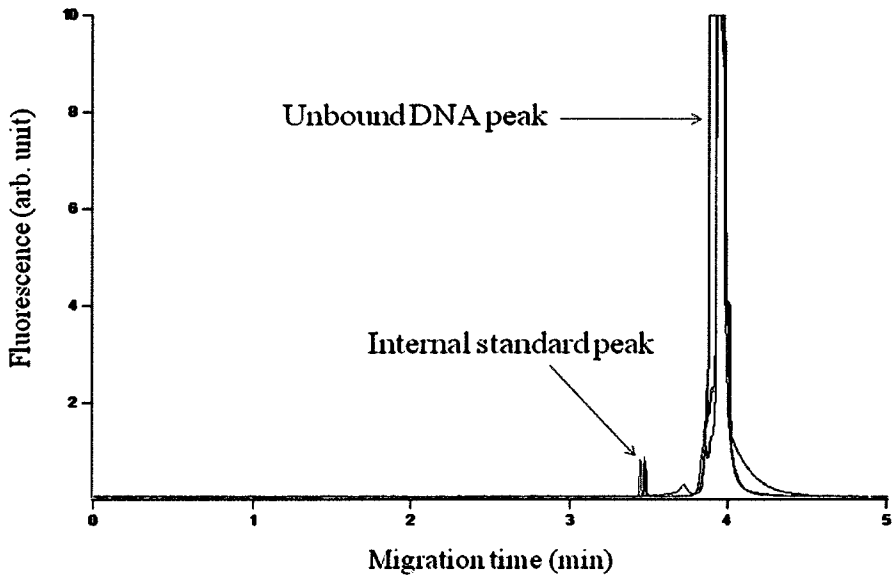


Figure 3-23: CE results of three ssDNA blank samples. 500 nM aptamer (Jing 1/Jing 2/Jing 3) + 2.5 nM IS + 50 μg/mL BSA in incubation buffer (10 mM Tris-HCl at pH 7.4). Peaks with three different colors (red, blue, to black) were the results of three ssDNAs from aptamers 1, 2, and 3 (Jing 1, 2 and 3). The highest peaks around 4 minutes were the unbound DNA peaks.

4.2.2.2. CE results of ssDNA samples incubated with Fpg protein