DEVELOPMENT OF A CYTOCHROME C BIOSENSOR

TOWARDS THE DEVELOPMENT

OF A CYTOCHROME C BIOSENSOR

By

JENNIFER ANN LEE, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Science

MASTER OF SCIENCE (2012) McMaster University

(Biochemistry & Biomedical Sciences) Hamilton, Ontario

TITLE: Towards the Development of a Cytochrome c Biosensor

AUTHOR: Jennifer Ann Lee, B.Sc. (Carleton University)

SUPERVISOR: Professor Yingfu Li (Simon Fraser University)

NUMBER OF PAGES: ix, 90

ABSTRACT

Cytochrome c (Cyt c) is a heme-containing protein that is a component of the electron transport chain as well as the mitochondrial apoptotic pathway. It is released from the mitochondrial intermembrane space to the cytosol during apoptosis and is also thought to be a biomarker for cancer and liver disease. Therefore, an efficient Cyt c biosensor would be a very useful tool for studying apoptosis. Here we show the process of development of Cyt c-dependent aptazymes, derived by *in vitro* selection. These aptazymes consist of 3 components: 1) a substrate with a cleavage site that consists of a single ribonucleotide flanked by a fluorophore and quencher; 2) a DNAzyme (catalytic DNA) motif capable of cleaving the substrate; 3) an aptamer, a short piece of single-stranded DNA that can specifically bind Cyt c. When Cyt c is absent, the aptamer occludes the catalytic core of the DNAzyme and the fluorophore of the intact substrate is quenched. However, when Cyt c is present, the aptamer binds Cyt c, allowing the DNAzyme to cleave the embedded ribonucleotide, separating the fluorophore and quencher, resulting in a fluorescent signal. Simulations of *in vitro* selection of Cyt c-dependent aptazymes were also performed. The simulations revealed several methods that can improve the success rate of future *in vitro* selections of aptazymes.

Further analysis of the previously derived DNAzyme DEC22-18 was also performed. A detailed understanding of this DNAzyme will allow it to be developed into a biosensor.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALL ABBREVIATIONS AND SYMBOLS

Apaf-1: Apoptotic Protease-Activating Factor-1

BSA: Bovine Serum Albumin

Cyt c: Cytochrome c

DABCYL: 4-(4- dimethylaminophenylazo) benzoic acid

DNA: Deoxyribonucleic Acid

DNAzyme: Deoxyribozyme

dNTP: Deoxyribonucleotide triphosphate

DTT: Dithiothreitol

EDTA: ethylenediaminetetraacetic acid

ELISA: Enzyme-Linked Immunosorbent Assay

EndoG: Endonuclease G

FNA: Functional Nucleic Acid

ΔG: Gibbs free energy

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

MRE: Molecular Recognition Element

PCR: Polymerase Chain Reaction

RNA: Ribonucleic Acid

RNase: Ribonuclease

SELEX: Systematic Evolution of Ligands by EXponential enrichment

TBAF: Tetra-*n*-butylammonium fluoride

CHAPER ONE

INTRODUCTION

1.1   APOPTOSIS AND CYTOCHROME C

Fundamentally, cancer is a disease in which cells continuously divide, having lost their regulatory mechanisms that control cell proliferation (Jones & Thompson, 2009). A regulatory mechanism, known as programmed cell death (apoptosis), is blocked in cancer cells, thus preventing induction of the apoptotic pathways. This blockage often occurs upstream of the "point of no return" step of apoptosis, the permeabilization of the outer mitochrondrial membrane. Once a cell has passed this step, it no longer has the ability to prevent cell death from occurring. Several cancer treatments can fail due to this blockage, so a better understanding of the apoptotic pathways would allow the development of drugs that may directly induce the "point of no return" step, bypassing the blockage and leading to more effective cancer treatments (Fulda, 2010).

Three classes of proteins in the Bcl-2 family regulate apoptosis. The first class includes Bax and Bak, which form pores in the outer mitochondrial membrane, permeabilizing it. Anti-apoptotic members, including Bcl-2 and Bcl-XL, form the second class and inhibit apoptosis. The third class of proteins, called BH3-only proteins, including Bad, Bid and Bik, are activated by different forms of cellular stress (B. Leber, Lin & Andrews, 2010). In many cancers, it has been found that the dysregulation or blockage of the apoptotic pathways is caused by overexpression of anti-apoptotic members of the Bcl-2 family (Certo *et al.*, 2006).
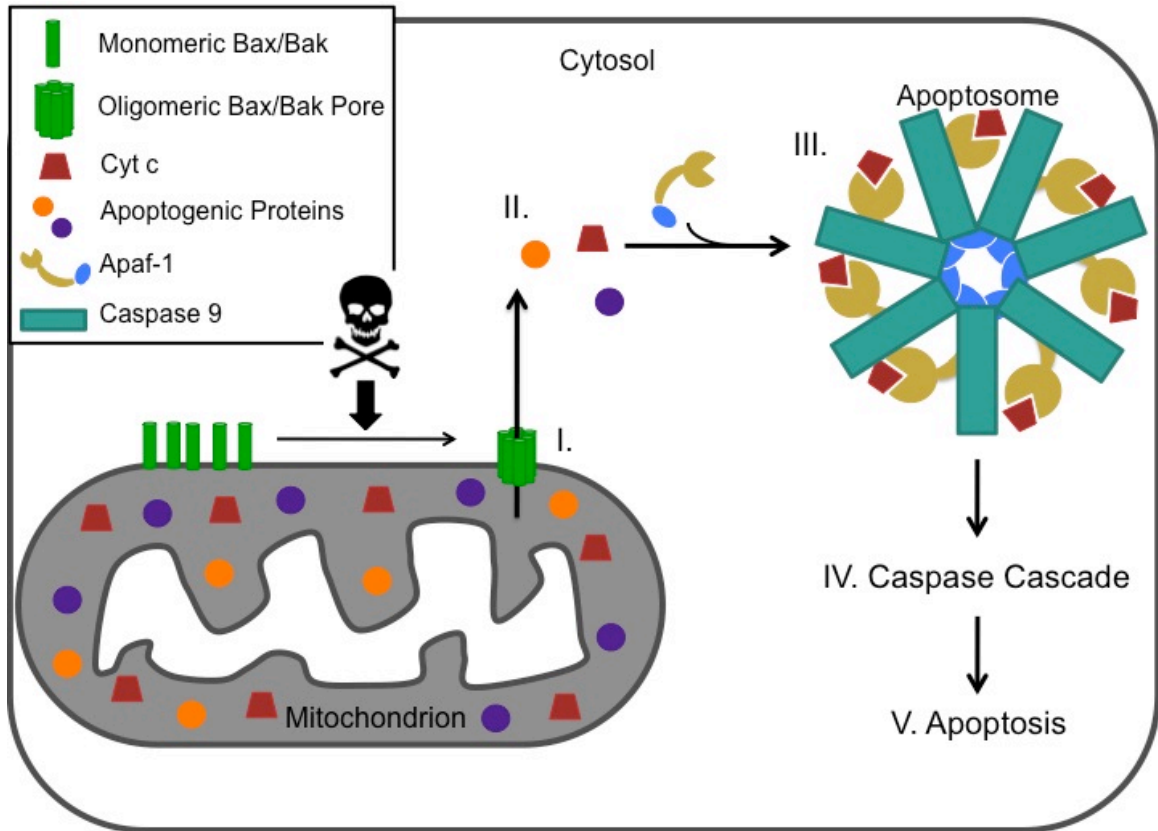
Figure 1. The mitochondrial apoptotic pathway. When an apoptotic stimulus is detected, tBid (or another activated BH3-only protein, not shown) interacts with Bax and/or Bak (green bars). **I)** This causes them to fully insert into the outer mitochondrial membrane and oligomerize to form a pore, permeabilizing the membrane. **II)** This allows Cyt c (red trapezoid) and other apoptogenic proteins (orange and purple circles) to move from the intermembrane space to the cytosol. **III)** Once in the cytosol Cyt c binds Apaf-1 (yellow and blue); seven of these Cyt c-Apaf-1 complexes then oligomerize into a heptameric complex called the apoptosome. Caspase 9 (teal rectangles) is recruited by the apoptosome, allowing their autoactivation. **IV)** Activated Caspase 9 then activates other caspases causing a caspase cascade. **V)** Leading to proteolytic cell death (Alberts *et al*., 2008, p.1125).

The mitochondrial pathway of apoptosis is triggered by several forms of cellular stress, including viral infection, DNA damage, the presence of unfolded proteins and nutrient deprivation (Chipuk & Green, 2008; Ow, Green, Hao & Mak, 2008; Youle & Strasser, 2008). When an apoptotic stimulus is detected, truncated-Bid (tBid, or possibly

another activated BH3-only protein) interacts with Bax and/or Bak, which are peripherally bound to the outer mitochondrial membrane. This interaction causes Bax and/or Bak to fully insert into the membrane (Figure 1). The integral membrane Bax or Bak can then recruit other Bax or Bak proteins, respectively, and oligomerize to form a pore. This subsequently permeabilizes the outer mitochondrial membrane, resulting in the "point of no return" for the cell (I in Figure 1). These pores allow cytochrome c (Cyt c) and other apoptogenic proteins in the intermembrane space of the mitochondria to be released into the cytosol (Leber, Lin & Andrews, 2007).

After Cyt c reaches the cytosol, it binds to apoptotic protease-activating factor-1 (Apaf-1). In a healthy cell, Apaf-1 is in a conformation that prevents its oligomerization (Ledgerwood & Morison, 2009). However, when Cyt c binds to Apaf-1 in the presence of dATP/ATP, it induces a conformational change that allows Apaf-1 to oligomerize (Jiang & Wang, 2004; Ledgerwood & Morison, 2009). Seven Apaf-1 proteins, each with a bound Cyt c molecule, form a wheel-like heptameric complex called the "apoptosome". Cyt c, while essential to the formation of the apoptosome, may not remain associated with the apoptosome after its formation (Ow *et al*., 2008). Once formed, the apoptosome recruits and allows autoactivation of caspase 9, which in turn activates other caspases. These caspases, which are a class of proteases, cleave various substrates in the cell and are ultimately responsible for proteolytic cell death (Kroemer, 1999).

An important aspect of studying apoptosis is detecting if apoptosis has occurred. One of the more popular methods to confirm that apoptosis has occurred is the detection of Cyt c. It is a 14.3 kDa, highly conserved protein that contains a single heme group

(Kroemer, 1999). In a normal cell, Cyt c is anchored to the outer surface of the inner

mitochondrial membrane, which is made up of cardiolipin, a mitochondria-specific

phospholipid (Kagan *et al*., 2005; Ow *et al*., 2008). Cyt c is a component of the electron

transport chain, shuttling electrons between Cytochrome c_1 of Complex III and

Cytochrome c Oxidase of Complex IV. Cytochrome c_1 reduces the iron of Cyt c's heme

group to Fe^{2+}, while Cytochrome c Oxidase oxidizes it back to Fe^{3+} (Voet, D. & Voet, J.

G., 2004, p.815). However, during the early stages of apoptosis, Cyt c interacts with a

cardiolipin molecule to form cardiolipin oxygenase, which oxidizes other cardiolipin

molecules of the membrane. This frees the Cyt c molecules from the membrane and

allows them to move to the intermembrane space (Kagan *et al*., 2005).

    Cyt c is a popular biomarker for several reasons. It is only released into the

cytosol past the "point of no return", so it is a reliable marker of apoptosis. Furthermore,

Cyt c is thought to be a serum biomarker in liver disease and cancer (Barczyk *et al*., 2005;

Ben-Ari *et al.,* 2003). Current cell-free Cyt c assays are conducted by performing an

Enzyme-Linked ImmunoSorbent Assay (ELISA) or by Western blotting. In both these

assays, antibodies are used to quantify the amount of Cyt c, in a process that is time

consuming and involves multiple steps (Alberts *et al*., 2008, pp.518, 588-589; Ben-Ari *et

al*., 2003). Furthermore, there is currently no reliable method of labeling Cyt c within

living cells. There have been attempts to label Cyt c with Green Fluorescent Protein

(GFP); however the Cyt c-GFP fusions do not behave exactly like native Cyt c

(Goldstein, Waterhouse, Juin, Evan & Green 2000; Heiskanen, Bhat, Wang, Ma &

Nieminen, 1999; Roucou, Prescott, Devenish & Nagley, 2000; Unkila, McColl, Thomensius, Heiskanen & Distelhorst, 2001).

A more efficient method would be to use a mix-and-read assay using a Cyt c biosensor that could facilitate high-throughput analysis of Cyt c in cell-free samples and provide an alternative to the time consuming ELISA and Western blotting assays. The biosensor could also be developed into an intracellular probe for Cyt c, allowing cytosolic Cyt c to be detected.

1.2   FUNCTIONAL NUCLEIC ACIDS

The proposed Cyt c biosensor to be developed would be composed almost entirely of deoxyribonucleic acid (DNA). It is quite stable compared to ribonucleic acid (RNA) since the sugar residues of DNA lack a 2'-OH group, making it much more resistant to hydrolytic degradation (Li & Breaker, 2001). DNA is best known for its genetic role in cells. However, in recent years it has been found that single-stranded DNA molecules can form intricate structures to carry out various functions such as molecular recognition and catalysis. Single-stranded RNA molecules can also perform these functions. Together these nucleic acid molecules are generally referred to as functional nucleic acids (FNAs) (Li & Breaker, 2001).

One class of FNA is known as "aptamers". Aptamers are single-stranded DNA, RNA or modified nucleic acids that can fold into defined secondary and tertiary structures to bind a specific target. Aptamers have been isolated for many targets including small molecules, proteins, viruses, bacteria, and mammalian cells (Chiuman & Li, 2007). The binding event of an aptamer to its target does not usually generate a detectable signal, so aptamers on their own are not considered sensors (Chiuman & Li, 2007).

Another class of FNA is nucleic acid enzymes, which include ribozymes and deoxyribozymes (DNAzymes). Ribozymes have been found in nature, including the hammerhead, hairpin, hepatitis delta virus (HDV), group II intron, ribonuclease (RNase) P, and varkud satellite (VS) (Doudna & Cech, 2002). Synthetic ribozymes have also been developed, catalyzing reactions such as forming carbon-carbon, peptide and amide bonds, performing acyl transfer reactions and producing acyl-Coenzyme A (Bagheri & Kashani-

Sabet, 2004). At present, no DNAzymes have been found in nature. Various DNAzymes have been developed to catalyze many different chemical reactions, including DNA phosphorylation, DNA ligation, and RNA cleavage (Breaker, 2004; Li & Breaker, 2001).

All synthetic FNAs are developed through an iterative process known as "*in vitro* selection" or Systematic Evolution of Ligands by EXponential enrichment (SELEX) (Navani, Mok & Li, 2009). Today the term SELEX is usually applied to the selection of aptamers, while the term *in vitro* selection is applied to the selection of DNAzymes and ribozymes, but the process of selecting for any FNA is essentially the same.

The process of *in vitro* selection can be generalized as follows: a starting library of $\sim 10^{14}$ unique single-stranded nucleic acid sequences is synthesized, including constant regions to allow PCR amplification. The sequences with the desired activity are separated from the unwanted ones. These selected sequences are then amplified by PCR and used for the next round of selection (Chiuman & Li, 2009). After every round of selection, the enriched pool contains an increased number of sequences exhibiting the desired activity. Once the enriched library pool reaches the desired level of activity, standard cloning and sequencing protocols are used to obtain the identities of sequences within the enriched library pool. Further testing of these sequences is then conducted (Chiuman & Li, 2009).

In technical terms, a biosensor is composed of a molecular recognition element (MRE) for binding a biological target of interest and a signal transducer that produces a signal once the MRE binds the target (Navani & Li, 2006). Aptamers are as sensitive and selective as antibodies (Breaker, 2004), so aptamers make excellent MREs and control

elements to modulate the catalysis of ribozymes and DNAzymes, the signal transducers.

Incorporation of an aptamer and nucleic acid enzyme in such a way that the aptamer has

control of the nucleic acid enzyme's catalytic activity is known as an allosteric nucleic

acid enzyme or aptazyme (Chiuman & Li, 2009). Aptazymes are able to bind their target

and generate a signal by catalysis.

Figure 2. Examples of various aptazymes. Thick black lines represent the nucleic acid enzyme, thick blue lines represent the aptamer and thick green lines represent the substrate. Dashed lines are Watson-Crick base pairs while dots are base mispairs. The black triangle indicates the cleavage site. **A)** An ATP-dependent aptazyme developed by rational design. P2 of the hammerhead ribozyme was replaced by the ATP aptamer, making the ribozyme's activity dependent on ATP (Tang & Breaker, 1997). **B)** A uranyl-dependent aptazyme developed by *in vitro* selection. The activity of this aptazyme, selected from a fully randomized library, is dependent on uranyl (Liu *et al*., 2007). **C)** An FMN-dependent aptazyme developed by a hybrid approach. P2 of the hammerhead ribozyme was replaced with the FMN aptamer and a randomized communication bridge, in red. *In vitro* selection was used to select for an aptazyme whose activity is dependent on FMN (Soukup & Breaker, 1999).

There are three main methods of engineering aptazymes: rational design, *in vitro* selection or a hybrid approach. The ATP allosteric hammerhead ribozyme was developed by rational design (Figure 2A) (Tang & Breaker, 1997). Many other aptazymes have been developed by this method and the list includes a mercury ion-dependent RNA cleaving deoxyribozyme (Liu & Lu, 2007), and ATP-, theophylline-, and FMN-dependent ligating ribozymes (Robertson & Ellington, 2000). In these instances, a stem-loop region important for the stability of the nucleic acid enzyme's active conformation was replaced by the relevant aptamer.

Many other aptazymes were developed by *in vitro* selection of the aptamer region. One example is a uranyl (UO_2^{2+})-dependent RNA cleaving deoxyribozyme (Figure 2B). A starting library consisting of a 50-deoxyribonucleotide random region surrounded by two constant regions, one of which contained a single ribonucleotide (ribo-A), was constructed and a total of ten rounds of selection were conducted to isolate the aptazyme (Liu *et al.*, 2007).

The hybrid approach of developing aptazymes is a combination of rational design and *in vitro* selection. The FMN-dependent hammerhead ribozyme mentioned above was developed by a hybrid method (Figure 2C). P2 of the ribozyme was replaced with two 4-ribonucleotide random regions, the communication bridge. The two ends of the FMN aptamer were then fused to the two ends of the ribonucleotide random regions to construct the library. Six rounds of *in vitro* selection were conducted to isolate the bridge (Soukup & Breaker, 1999). An ATP-dependent ligating DNAzyme with a

communication bridge design was also developed using this hybrid approach (Levy &

Ellington, 2002).

In general, developmental approaches that utilize *in vitro* selection result in more

sensitive aptazymes than rational design. This is likely due to the fact that the *in vitro*

selection and hybrid approaches allow for the optimization of the function of the

aptazymes during selection, while optimization must be done mostly by trial and error

when utilizing a rationally designed approach (Levy & Ellington, 2002).

1.3   EXISTING PROTEIN-DEPENDENT APTAZYMES

There are relatively few protein-dependent aptazymes compared to the large number and variety of small molecule- and ion-dependent aptazymes. This may be due to the fact that proteins can denature, are larger and are more susceptible to changes of pH and salt concentrations compared to small molecules and ions.

Some of the first protein-dependent aptazymes were rationally designed group I introns dependent on HIV-1 Rev and bacteriophage λN. The RNA region bound by the protein was inserted into the ribozyme so that when the relevant protein bound the protein-binding region it stabilized the ribozyme, leading to an increased rate of splicing (Atsumi, Ikawa, Shiraishi & Inoue, 2001).

In 2002 Hartig and Famulok developed an HIV-1 Reverse Transcriptase (HIV-1 RT)-dependent hammerhead ribozyme by rational design. An HIV-1 RT aptamer replaced P2 of the ribozyme, and was designed so that cleavage was abolished in the presence of HIV-1 RT (Hartig & Famulok, 2002). The Famulok group later used this aptazyme to screen a library of synthetic chemicals where a novel small molecule inhibitor of HIV-1 RT was identified. If a chemical displaced the aptazyme bound to HIV-1 RT, the aptazyme reverted to its active conformation and cleavage activity would return, indicating the inhibitor's presence (Yamazaki et al., 2007).

The Famulok group did a screen for HIV-1 Rev inhibitors by the same method. They first designed two HIV-1 Rev-dependent hammerhead ribozymes, one whose cleavage activity was abolished when bound by HIV-1 Rev and another whose cleavage activity was restored (Figures 3A and B, respectively). They then used these aptazymes

to screen a library of known antibiotics for compounds that could displace HIV-1 Rev

from its aptamer. By this method they found a novel HIV-1 Rev inhibitor (Hartig *et al.*,

2002). HIV-1 Tat-dependent hammerhead ribozymes were also developed by replacing
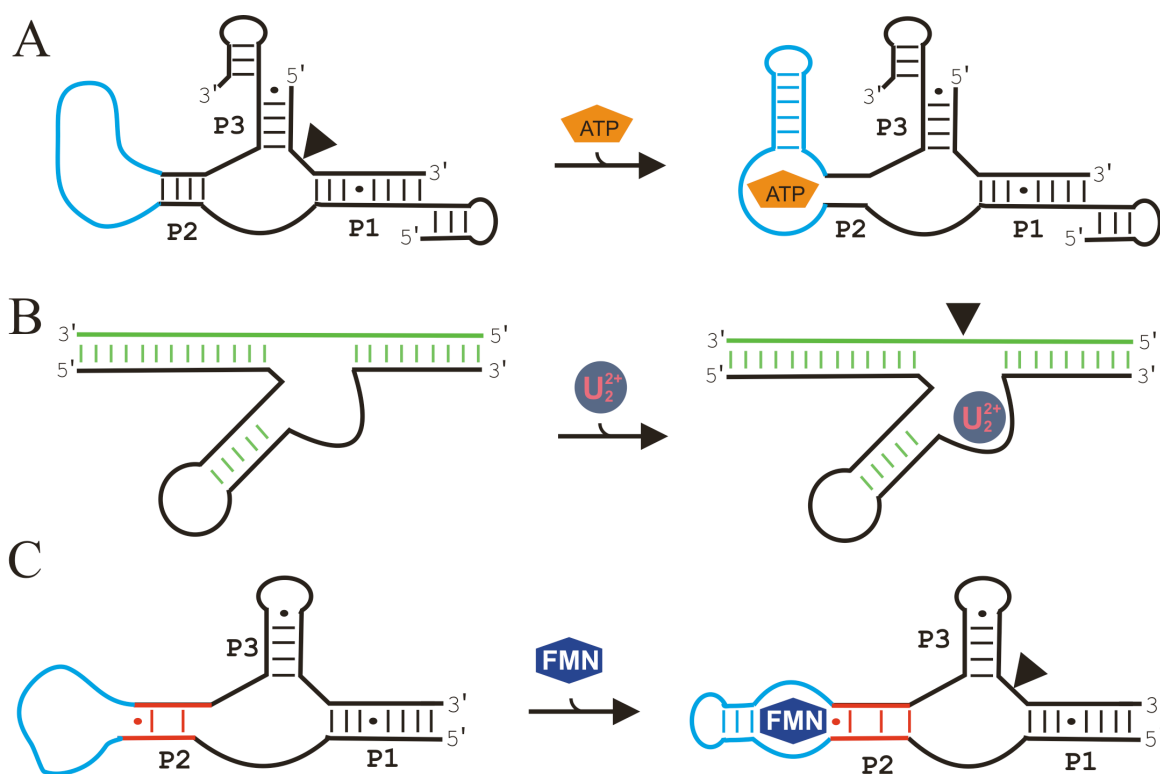
P2 (Wang & Sen, 2002).
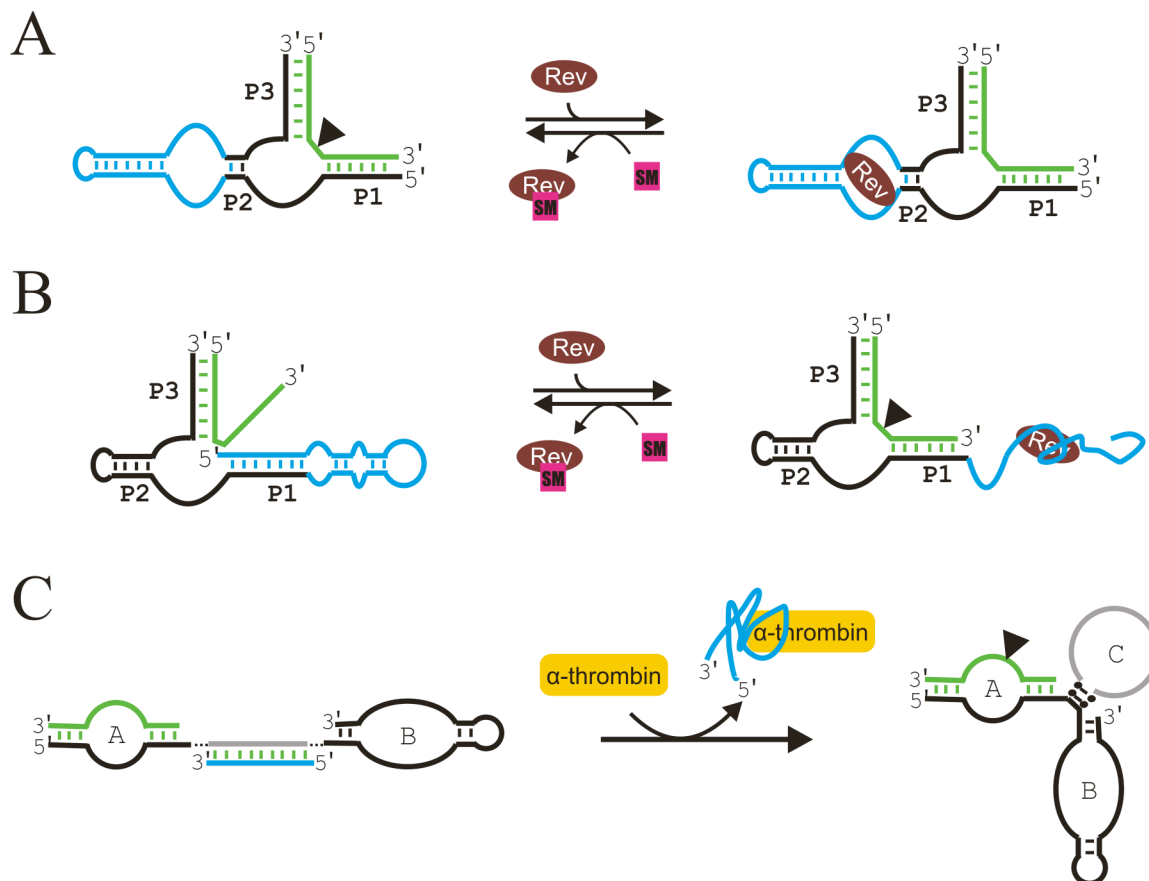
Figure 3. Examples of protein-dependent aptazymes. Thick black lines represent the nucleic acid enzyme, thick blue lines represent the aptamer and thick green lines represent the substrate. Dashed lines are Watson-Crick base pairs. The black triangle indicates the cleavage site. **A)** The HIV-1 Rev (Rev)-dependent aptazyme whose catalytic activity is suppressed in the presence of Rev. Rev is bound by a small molecule (SM) which displaces Rev from the aptamer (Hartig *et al.*, 2002). **B)** The Rev-dependent aptazyme whose catalytic activity is increased in the presence of Rev. Rev is displaced from the aptamer in the same manner as in A (Hartig *et al.*, 2002). **C)** An aptazyme dependent on α-thrombin. The thick grey line is the complement of the α-thrombin aptamer's sequence. When no α-thrombin is present the aptamer hybridizes with its complement, suppressing activity of the hairpin ribozyme (Najafi-Shoushtari & Famulok, 2007).

The Ellington group developed another HIV-1 Rev-dependent aptazyme. They first developed a ligating ribozyme dependent on the HIV-1 Rev arginine-rich motif (ARM) by *in vitro* selection (Robertson, Knudsen & Ellington, 2004). One stem-loop of the ribozyme was replaced with a 50-ribonucleotide random region and seven rounds of selection were conducted, yielding a single aptazyme sequence. A partially randomized sequence library (each position of the original random region had a 70% chance of being the original ribonucleotide and a 10% chance of being one of the other three ribonucleotides) was synthesized. Further *in vitro* selection with the HIV-1 Rev ARM and the full HIV-1 Rev protein yielded aptazymes dependent on these proteins after six rounds of selection each (Robertson *et al*., 2004).

Other aptazymes targeting proteins unrelated to HIV-1 have also been developed. The Ellington group used the same starting library as above to develop ligating ribozymes dependent on tyrosyl transfer RNA synthase (Cyt18) and hen egg white lysozyme. Nine rounds of *in vitro* selection were conducted for Cyt18-dependent aptazymes while eleven rounds were conducted for lysozyme-dependent aptazymes (Robertson & Ellington, 2001).

An aptazyme for the unphosphorylated form of ERK2, a protein kinase, was rationally designed by replacing part of P2 of the hammerhead ribozyme with the RNA-binding domain specific to unphosphorylated ERK2. Catalytic activity of the aptazyme increased in the presence of unphosphorylated ERK2 only. An aptazyme for the phosphorylated form of ERK2 designed in the same manner gave the same result, except catalytic activity only increased in the presence of phosphorylated ERK2 (Vaish *et al*.,

2002). Hepatitis C virus (HCV) replicase- and helicase-dependent aptazymes have also been developed by rational design (Cho, Kim, Lee, Kim & Kim, 2005).

Very little work has been done on aptazymes targeting proteins that do not have RNA- or DNA-binding regions. One has already been mentioned: the hen egg white lysozyme-dependent aptazyme by the Ellington group (Robertson & Ellington, 2001). Another is an α-thrombin-dependent aptazyme composed of a hairpin ribozyme and an α-thrombin aptamer (Figure 3C). The α-thrombin aptamer forms a duplex with a modified domain C, abolishing cleavage activity. But when α-thrombin is present the aptamer moves away from domain C and binds the thrombin, allowing the ribozyme to regain cleavage activity (Najafi-Shoushtari & Famulok, 2007). This aptazyme was used to detect protein-protein interactions: the aptamer bound to α-thrombin was displaced by hirudin, an α-thrombin inhibitor, which led to a decrease in cleavage activity of the aptazyme (Hartig *et al*., 2002). An α-thrombin-dependent aptazyme using a hammerhead ribozyme was also developed using this same design, except that P2 was modified to form a duplex with the α-thrombin aptamer. Again, when α-thrombin was present, the hammerhead ribozyme regained cleavage activity (Hartig *et al*., 2002; Hartig & Famulok, 2008).

1.4   THE MGZ DNAZYME

The MgZ DNAzyme (see Figure 4) was developed by Chiuman and Li (Chiuman & Li, 2007). This DNAzyme uses the divalent magnesium ion as a cofactor and cleaves a single ribo-A (rA) unit embedded in an otherwise all-DNA molecule. MgZ was chosen for this work for several reasons: DNA is more stable than RNA; ionic magnesium is readily found in cells and cell lysates so no exotic divalent metal ions would have to be added for the DNAzyme to function; RNA cleavage usually results in two shorter fragments and the process is relatively easy to monitor for the purpose of biosensing;
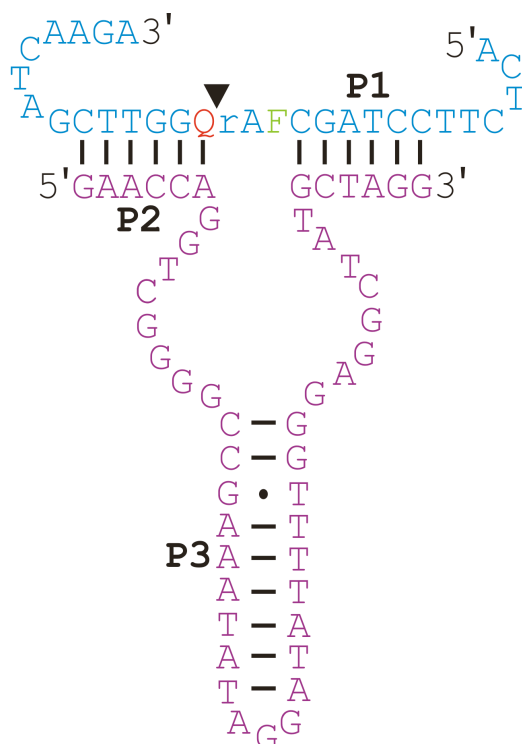
Figure 4. The MgZ DNAzyme. MgZ (purple) cleaves the lone ribo-A (rA) of the substrate (blue) to produce two fragments (the black triangle indicates where the cleavage occurs). The rA is flanked by a fluorescein-labeled T (F, in green) and a DABCYL quencher-labeled T (Q, in red) (Chiuman & Li, 2007). Short black lines indicate base pairing, while a black dot indicates mispairing.

it is relatively easy to obtain efficient DNAzymes that catalyze RNA cleavage, which has

been demonstrated in many previous *in vitro* selection experiments (Chiuman & Li,

2007).

As shown in Figure 4, MgZ's single rA unit is flanked by a fluorescein-labeled T

(F, green) and a 4-(4- dimethylaminophenylazo) benzoic acid (DABCYL, a fluorescence

quencher)-labeled T (Q, red). In this configuration, the fluorescein molecule is not able

to emit strong fluorescence in the uncleaved substrate, due to the nearby quencher.

However, when MgZ cleaves the rA, the F and Q moieties are separated, resulting in an

increase in fluorescence (Chiuman & Li, 2007). This arrangement minimizes the

background fluorescence since the F and Q moieties are such a short distance apart and

also minimizes any false positives since they can only separate when the rA linkage is

cleaved (Mei, Liu, Brennan, & Li, 2003).
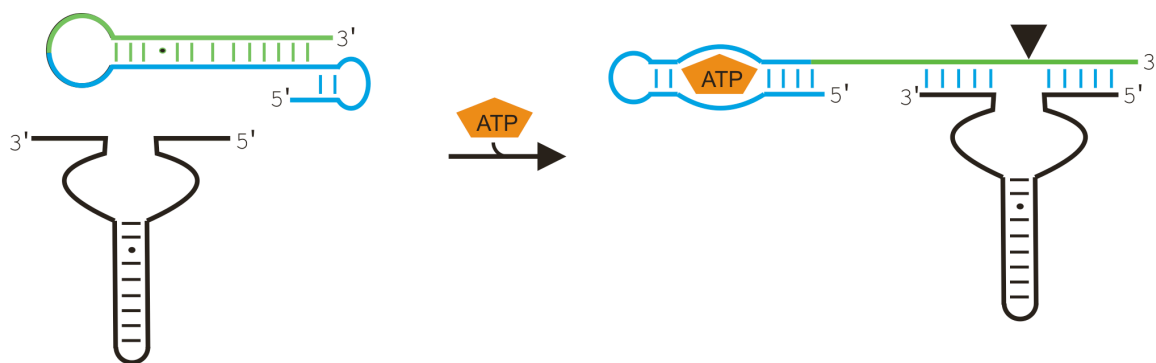
Figure 5. An allosteric MgZ DNAzyme. Thick black lines represent MgZ, thick blue lines represent the ATP aptamer and thick green lines represent the substrate. Dashed lines are Watson-Crick base pairs while dots are base mispairs. The black triangle indicates the cleavage site. The ATP-dependent aptazyme developed by Chiuman and Li utilizes MgZ as the signal transducer. In the presence of ATP, the ATP aptamer exposes the substrate, allowing it to be cleaved by MgZ (Chiuman & Li, 2007).

Two aptazymes have already been rationally designed utilizing MgZ as the signal transducer. A modified DNA ATP aptamer was annealed to the 5' end of the substrate to convert MgZ into an ATP-dependent aptazyme with a detection limit of 5-10 μM (Chiuman & Li, 2007). As seen in Figure 5, in the absence of ATP, the aptamer hybridizes with the substrate, preventing catalysis by MgZ. In the presence of ATP the aptamer binds it, allowing MgZ to catalyze the cleavage of the rA linkage. As well, MgZ was converted into an ADP-dependent aptazyme using the same design scheme with an RNA ADP aptamer with a detection limit of 2 μM (Chiuman & Li, 2007).

1.5   PROJECT OBJECTIVE

The goal of this project was to develop Cyt c-dependent aptazymes through *in vitro* selection. The aptazymes were to be composed of an aptamer (the MRE) that bound Cyt c and a fluorescence-signaling DNAzyme, MgZ, which transduced the Cyt c-aptamer interaction into a fluorescent signal that could be conveniently detected. These aptazymes could then be used to develop alternative methods of Cyt c detection in cell-free lysates as well as in live cells. These improved methods of detection could facilitate an increased understanding of apoptosis as well as the diseases associated with apoptotic dysregulation.

A DNA library was designed in which P3 of MgZ (see Figure 4) was replaced with a random-sequence domain of 39-40 nucleotides. It was previously determined that most of the nucleotides in P3 are not required for the function of MgZ (Chiuman & Li, 2007); therefore it was considered a suitable location for the aptamer domain. An *in vitro* selection approach was used to select for aptazymes that only cleaved the rA unit when both Cyt c and Mg^{2+} were present. The resulting sequences from *in vitro* selection were analyzed. Finally, simulations of the *in vitro* selection process were performed to better understand the process.

CHAPTER TWO

METHODS

2.1   OLIGONUCLEOTIDES, PROTEIN AND CHEMICALS

All unmodified DNA sequences as well as primer 3 with the hexaethyleneglycol spacer were synthesized by Integrated DNA Technologies (Coralville, IA). All modified sequences, which include the DNA/RNA chimeric substrate labeled with fluorescein and DABCYL and the randomized library pool were synthesized by the Keck Oligonucleotide Synthesis Facility at Yale University (New Haven, CT). The fluorescein and DABCYL groups were attached by using a fluorescein-dT phosphoramidite and a DABCYL-dT phosphoramidite, respectively, from Glen Research (Sterling, VA). The single riboadenosine located internally in the substrate had a tert-butyldimethylsilyl (TBDMS) protecting group at the 2'-OH. It was deprotected by incubation in tetra-*n*-butylammonium fluoride (TBAF) at 60 ºC for 12 hours.

Bovine heart cytochrome c was purchased from BioVision (Mountain View, CA). Tris, urea and boric acid were purchased from Bioshop Canada (Burlington, ON). Glycerol was purchased from Caledon Laboratories (Georgetown, ON). All other chemicals were purchased from Sigma-Aldrich (Oakville, ON). Polynucleotide kinase, T4 DNA ligase, proteinase K, bovine serum albumin (BSA) and dNTPs were purchased from Fermentas (Burlington, ON), *Thermus thermophilus* DNA polymerase was purchased from Biotools (Madrid, Spain) and porcine pancreas elastase was purchased from Calbiochem by EMD (Gibbstown, NJ). [γ-^{32}P]ATP and [α-^{32}P]dGTP were acquired from GE Healthcare (Uppsala, Sweden). Nanosep 3K Centrifugal Filters were purchased

from Pall Corporation (Ann Arbor, MI). Water was purified with a Milli-Q Synthesis

A10 water purification system and then autoclaved. All oligonucleotides were purified

by 10% denaturing PAGE, dissolved in double-deionized water and then quantified using

a NanoVue spectrophotometer from GE Healthcare (Uppsala, Sweden).

2.2   FIRST *IN VITRO* SELECTION

The starting library for the *in vitro* selection of Cyt c-dependent aptazymes is shown in Figure 6. The library sequence contains the following individual elements: the 5' primer-binding site (black), MgZ (purple) with the embedded random sequence (yellow), the 3' primer-binding site (black), and the substrate (blue). Every position of the random region had an equal chance of being T, G, A or C. Following synthesis and purification, this library was used for the *in vitro* selection process to derive Cyt c-dependent aptazymes (see Figure 7).
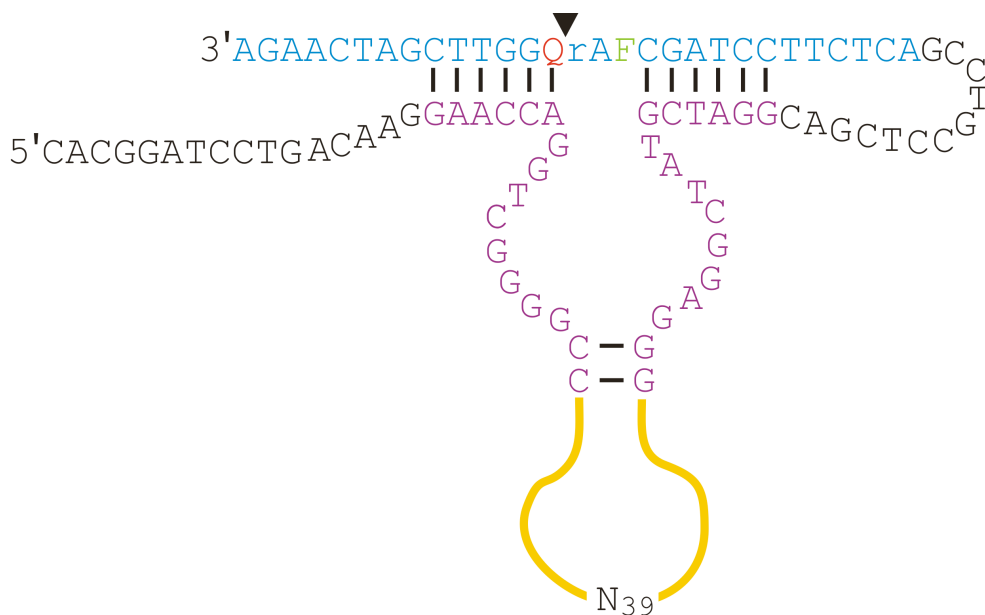
Figure 6. The starting library for the first *in vitro* selection of Cyt c-dependent aptazymes. In blue is the substrate, in black are the primer binding regions, in purple is MgZ and in yellow is the 39-deoxyribonucleotide random region. Every position of the random region had an equal chance of being T, G, A or C. The black lines indicate base pairing. The black triangle indicates the cleavage site.

I. Starting Library

VI. Ligation

II. Negative Selection

discarded

V. PCR 2

IV. PCR 1

III. Positive Selection

discarded
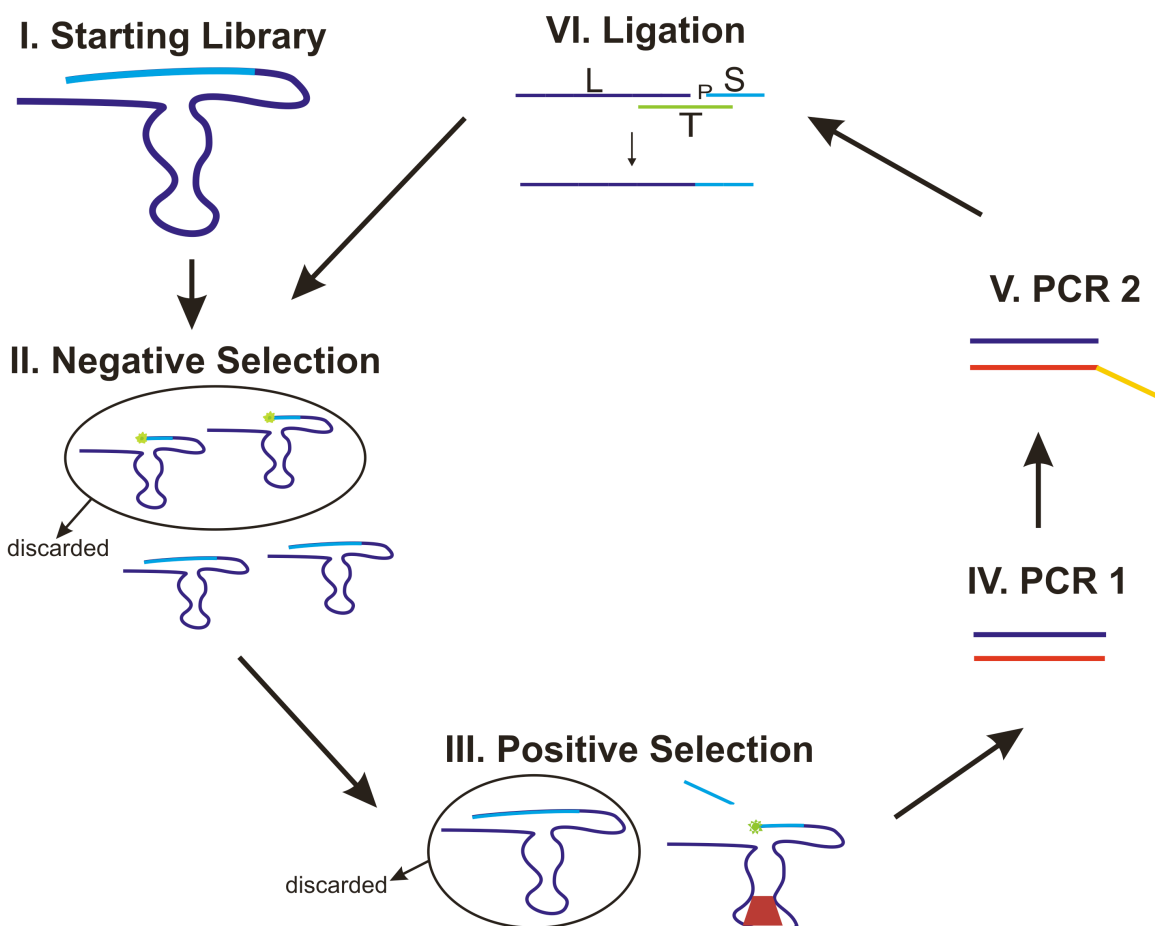
Figure 7. The *in vitro* selection process to select for Cyt c-dependent aptazymes.
I) The starting library. **II)** The library is first incubated in selection buffer only and uncleaved sequences are isolated by 10% denaturing PAGE. **III)** The remaining uncleaved sequences are then incubated in selection buffer containing Cyt c (red trapezoid). The cleaved sequences are isolated by 10% denaturing PAGE. Two PCR reactions are carried out. **IV)** PCR1 amplifies the cleaved sequences from the positive selection. **V)** PCR2, seeded by a small amount of PCR1 product, uses a blocked primer to make the antisense strand longer than the sense strand, allowing the sense and antisense strands to be separated by 10% denaturing PAGE. **VI)** The purified sense strand (L) is then ligated to 5'-phosphorylated substrate (S) in the presence of ligation template (T) and the now enriched library pool is purified by 10% PAGE. The enriched library pool then goes on to the next round of *in vitro* selection, starting at II.

The starting library consisted of ~10^{14} unique DNA molecules (500 pmol). It was

heated at 90 ºC for 45 seconds and allowed to cool to room temperature (~23 ºC) over ten

minutes. The library was incubated for 5 hours in selection buffer only (150 mM NaCl,

10 mM MgCl$_2$, 0.01% Tween-20 and 50 mM 4-(2-hydroxyethyl)-1-

piperazineethanesulfonic acid (HEPES) at pH 7.5 at 23 ºC), quenched to a final

concentration of 30 mM ethylenediaminetetraacetic acid (EDTA) at pH 8 at 23 ºC and

then purified by 10% denaturing PAGE to remove any cleaved sequences. This is called

the "negative selection" step, which is intended to remove any catalytic DNA sequences

that could cleave themselves without Cyt c. The uncleaved sequences from this step were

again heated at 90 ºC and cooled and then used to conduct the "positive selection" step,

where they were incubated for 10-30 minutes in selection buffer containing 0.5-10 μM

Cyt c and quenched to a final concentration of 30 mM EDTA at pH 8 at 23 ºC. The

cleaved sequences were purified by 10% denaturing PAGE. Next, two PCR reactions

were carried out. The first reaction (PCR1) used the standard primers and served as the

main amplification step. A primer containing a hexaethyleneglycol spacer and 20-A

poly-A tail was used in the second PCR reaction (PCR2). The hexaethyleneglycol spacer

prevents the poly-A tail from being amplified, making the non-DNAzyme-coding

(antisense) strand 20 nucleotides longer than the coding (sense) strand. The two strands

were separated by 10% denaturing PAGE. The purified sense strand was ligated to the 5'

phosphorylated substrate and the ligated constructs were purified again by 10%

denaturing PAGE. The enriched pool of purified, ligated constructs then went on to the

next round of selective amplification.

The constructs were labeled with [α-^{32}P]dGTP during the PCR2 step and [γ-

^{32}P]ATP during the ligation step for easy tracking. All PAGE gels during selection had a

phosphorimage taken of them using either a Storm 820 from the Molecular Dynamics

division of Amersham Pharmacia Biotech (Piscataway, NJ) or a Typhoon 9200 Variable

Imager from GE Healthcare (Uppsala, Sweden). All PAGE gels containing fluorescent

DNA also had a fluorescent image taken of them using a Typhoon 9200 Variable Imager.

The cleavage percentages of all negative and positive selections were determined from

the PAGE gels' phosphorimages using ImageQuant software.

First *In Vitro* Selection	
Primers	
Name	Sequence
P1	5' - CACGGATCCTGACAAG - 3'
P2	5' - CGGACGGAGCTG - 3'
P3	5' - AAAAAAAAAAAAAAAAAAAA-L-CGGACGGAGCTG - 3'
Ligation Templates	
Name	Sequence
Template	5' - CTAGGAAGAGTCGGACGGAGAGCTG - 3'
Template + T	5' - CTAGGAAGAGTTCGGACGGAGAGCTG - 3'
Second *In Vitro* Selection	
Primers	
Name	Sequence
P1	5' - GAACAGTCGTACGCAC - 3'

Table 1. Primers and ligation templates for the two *in vitro* selections for Cyt c-dependent aptazymes. Sequences that are not listed under the second selection are identical to the first selection.

The primer and ligation template sequences used during *in vitro* selection are

provided in Table 1. Primers 1 and 2 were used for PCR1 and primers 1 and 3 were used

for PCR2. Primer 3 had the same sequence as primer 2 except for the hexaethyleneglycol

spacer and 20-nucleotide poly-A tail. Two ligation templates were used during ligation of

the substrate and the PCR-generated sense strand of the enriched library pool, since it is

known that during the elongation step of PCR, an extra A is often added to the 3' end of

double-stranded DNA. Therefore, a second template with a T insertion was included to

increase the yield of the ligation reaction.

Modifications were made to this protocol during the course of 20 rounds of selection. Starting at Round 5, the negative selection step was repeated twice. Starting at Round 6, after the two negative selection steps, the remaining uncleaved sequences were equally divided into two fractions. The first fraction was incubated with the selection buffer containing Cyt c, while the other fraction was only incubated with the selection buffer. In rounds 14 and 15 an alternative condition was used for the negative selection: samples were heated at 90 ºC for 1 minute, before being slowly cooled to 25 ºC over 4½ minutes and incubated at 25 ºC for 10 minutes, in a thermocycler. Twelve cycles were conducted in Round 14 and eight cycles in Round 15. In rounds 17-20, the negative selection was changed to 7 iterations of heating at 90 ºC for 45 seconds and incubation at room temperature for an hour.

In Round 14, the positive selection step was again modified. The remaining uncleaved sequences were divided into three fractions, one with selection buffer, and two with selection buffer and Cyt c. Once quenched to a final concentration of 30 mM EDTA at pH 8 at 23 ºC, one fraction containing Cyt c was treated with 2.5 units of proteinase K, an enzyme that degrades proteins, while the two others were left untreated. Proteinase K was added to remove the Cyt c, allowing the sequences to migrate properly through the 10% denaturing PAGE gel. In rounds 16, 17 and 19, dithiothreitol (DTT), a strong reducing agent (Voet, D. & Voet, J. G., 2004, p.167), was added in the positive selection step with Cyt c to reduce the iron of Cyt c's heme group from Fe^{3+} to Fe^{2+}. A subtle conformation change occurs when the oxidation state of Cyt c's heme changes (Banci *et al.*, 1997). Cyt c exists in both forms in cells so it would be beneficial if the eventual Cyt

c aptazymes could interact with both. The addition of DTT in some rounds of selection

and not others was to allow the enriched library pool to be exposed to both forms. To

increase the diversity of the enriched library pool, hypermutagenic PCR was used in the

PCR2 step of rounds 9 and 11. Unequal concentrations of dNTPS were used to generate

a ~1000-fold higher mutation rate compared to regular PCR (B & M Labs; Vartanian,

Henry & Wain-Hobson, 1996).

At the end of Round 20, sequences from the enriched library both with and

without the addition of proteinase K were cloned using the InsTAclone PCR cloning kit

from Fermentas and sequenced by Functional Biosciences (Madison, WI). A few

sequences from the original library were also cloned and sequenced (see Appendix I for

all analyzed sequences). The cleavage activities of two selected sequences were tested in

the presence of selection buffer only and in the presence of selection buffer and the

following: 1 μM Cyt c; 1 μM Cyt c + 2.5 mM DTT; 2.5 mM DTT only; 1 μM BSA; 1

μM BSA + 2.5 mM DTT; 15% glycerol, the concentrated solvent of the Cyt c and BSA

stocks; and 15% glycerol + 2.5 mM DTT. All reactions were quenched to a final

concentration of 30 mM EDTA at pH 8 at 23 ºC. No proteinase K was added to the

reactions before being separated by 10% denaturing PAGE. The cleavage percentages

were determined by the same method as above.

2.3   SECOND *IN VITRO* SELECTION

The starting library for the second *in vitro* selection of Cyt c-dependent aptazymes is shown in Figure 8. The library sequence contains the following individual elements: the 5' primer-binding site (black), the partially randomized MgZ (purple) with the embedded fully randomized sequence (yellow), the 3' primer-binding site (black), and the substrate (blue). Every position of the MgZ sequence had a 70% chance of being the original nucleotide and a 10% chance of being any of the other 3 nucleotides. The second *in vitro* selection process to derive Cyt c-dependent aptazymes was similar to the first with a few notable exceptions.
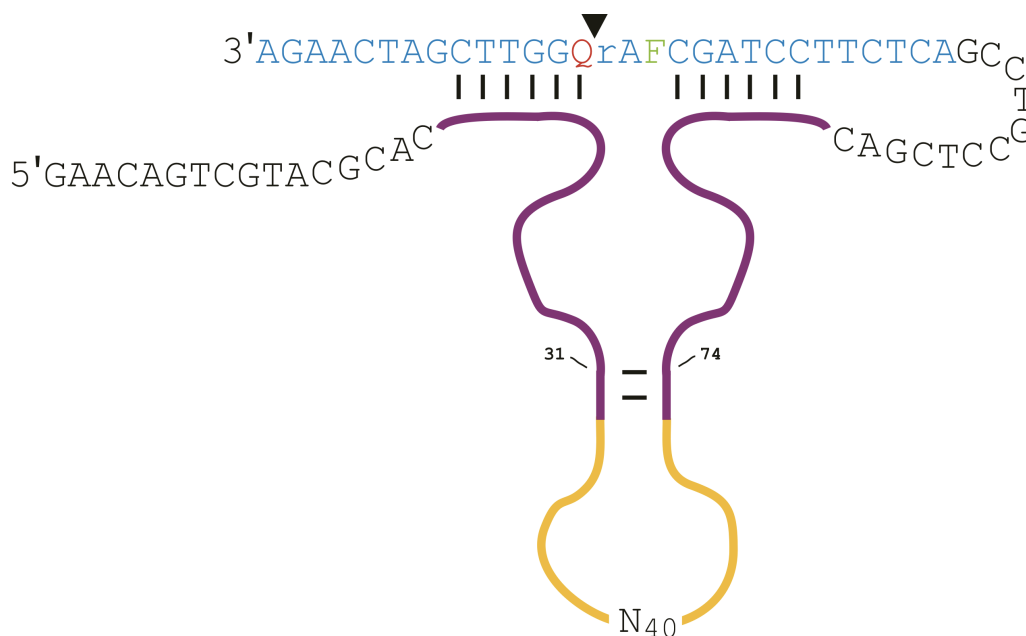
Figure 8. The starting library for the second *in vitro* selection of Cyt c-dependent aptazymes. In blue is the substrate, in black are the primer binding regions, in purple is the partially randomized MgZ and in yellow is the 40-deoxyribonucleotide random region. Every position of the MgZ sequence had a 70% chance of being the original nucleotide and a 10% chance of being any of the other 3. Every position of the random region had an equal chance of being T, G, A or C. The black lines indicate base pairing. The black triangle indicates the cleavage site.

For the negative selection, the library was first incubated for 12 hours in selection buffer only (100 mM NaCl, 25 mM KCl, 10 mM MgCl$_2$, 50 mM HEPES at pH 7.5 at 23 ºC) in a thermocycler, which was programmed to heat the reaction at 90 ºC for 45 seconds every 2 hours. The temperature was then returned to 25 ºC by slowly stepping down the temperature, remaining at 75 ºC for 2 minutes and then 50 ºC for 2 minutes. The reaction was quenched as before to a final concentration of 30 mM EDTA at pH 8 at 23 ºC and the uncleaved sequences were isolated by 10% denaturing PAGE.

For the positive selection, the uncleaved sequences from the negative selection were divided into three fractions: one with selection buffer only (to determine the amount of background cleavage), one with selection buffer and Cyt c and one with selection buffer, Cyt c and DTT. After an incubation period of 10-30 minutes, all reactions were quenched to a final concentration of 30 mM EDTA at pH 8 at 23 ºC and the two reactions containing Cyt c were treated with proteinase K. Cleaved sequences from all positive selection reactions were separated by 10% denaturing PAGE and the cleaved sequences from the reactions containing Cyt c and Cyt c + DTT were isolated for amplification. Cleavage percentages of the negative and positive selections were determined as before.

As in the first *in vitro* selection, two PCR reactions were carried out. The only modification to the amplification step was to change the sequence of Primer 1 (see Table 1), to minimize any chance of contamination from the first *in vitro* selection. The ligation step to generate the enriched pool for the next round of selection was unchanged.

Again, modifications to this protocol were made throughout the process. After six rounds of *in vitro* selection following the process described above, *in vitro* selection was

restarted at the amplification step of Round 3. The negative selection step was modified to consist of three consecutive incubations in selection buffer only. The first incubation was for 30 minutes, the second for 2 hours and the third for 5 hours. After each incubation the uncleaved sequences in the reaction were isolated by 10% denaturing PAGE. In rounds 8-1 and 10-1, the 2-hour incubation was replaced by the 12-hour incubation in the thermocycler described above. The positive selection remained mostly unchanged; however in rounds 7-1, 9-1 and 11-1 the positive selection reactions containing Cyt c were treated with 3 units of elastase instead of 2.5 units of proteinase K. The alternation between the two enzymes was to minimize any chance of sequences being selected for proteinase K-dependence instead of Cyt c-dependence.

After nine rounds of selection using this modified protocol, *in vitro* selection was started again at the amplification step of Round 7-1 and the negative and positive selection steps were again modified. The negative selection step was reverted to a single 5-hour incubation in selection buffer only and the uncleaved sequences were isolated by low-pH 10% denaturing PAGE. These denaturing PAGE gels used a running buffer of pH 6.2 and contained 2-(N-morpholino)ethanesulfonic acid (MES) instead of the usual running buffer at pH 8.0, which contained boric acid. The positive selection step remained the same as before, except that after incubation the cleaved sequences were isolated by low-pH 10% denaturing PAGE. After two rounds of selection using the above, modified protocol, sequences from Round 6, Round 12-1 and Round 9-2 were cloned using the InsTAclone PCR cloning kit and sequenced by Functional Biosciences (see Appendix II for all analyzed sequences).

2.4   ANALYSIS OF SECOND *IN VITRO* SELECTION

mFold, an online folding simulation tool (Zucker, 2003), was used to determine the theoretical Gibbs free energy (ΔG) and secondary structure of the random region with the two G-C base pairs connecting it to MgZ of several sequences (bases 31-74 of each sequence). The theoretical ΔG and secondary structure were obtained for all identified sequences from Round 6, the two most abundant sequences from Round 12-1 and the original MgZ sequence. Three sequences from Round 6 and the two most abundant sequences from R12-1 were chosen for further examination.

250 pmol of substrate was 5' phosphorylated and ligated to the full sequence of interest to form the exact same sequence as that during *in vitro* selection. 50 pmol of substrate was 5' phosphorylated with [γ-^{32}P]ATP and ligated to the sequence for tracking purposes. Each kinetic assay was performed in triplicate. Each reaction contained a final concentration of 100 nM of sequence and 2x selection buffer was added to a final volume of 6 µL to start the reaction. Reactions were incubated for 0 min, 6 seconds, 30 seconds, 1 minute, 5 minutes, 10 minutes, 30 minutes, 1 hour, 2.5 hours, 5 hours and 16 hours. All reactions were quenched with 1 µL of 3 M EDTA, pH 8 at 23 ºC. 6 µL of 2x loading buffer was then added to the reaction and the cleaved and uncleaved sequences of each reaction were separated by 10% denaturing PAGE.

A phosphorimage was taken of the 10% denaturing PAGE gels and the cleavage percentage of each reaction was determined using ImageQuant software. The fraction cleaved and standard deviation of each time point was determined and plotted on a graph

of time *versus* fraction cleaved. Using GraphPad Prism 4 a one-phase exponential with

equation

$$Y = Y_{MAX}\left(1 - e^{-k_{obs}t}\right)$$

where Y_{MAX} is the maximum fraction cleaved, k_{obs} is the observed kinetic rate and t is

time in minutes, was fitted to each sequence's data set.

The Cyt c solution used in the kinetic assays was created as follows: Cyt c was

dissolved in double-distilled water to a concentration of 1 mg/mL, the optimal

concentration cited by the 3K column manufacturer. 300 µL of this solution was treated

with 12 µL of 1M DTT, the same Cyt c-to-DTT ratio used during *in vitro* selection, for

ten minutes to reduce the iron in the heme of Cyt c. The 3K column was washed with

500 µL of double-distilled water and centrifuged for 5 minutes at 10,000 *g*. 312 µL of the

reduced Cyt c solution was then loaded onto the column and centrifuged until almost all

the liquid had migrated through the column. Two 500 µL double-distilled water washes

were sent through the column to wash away any remaining DTT.

The reduced Cyt c on the column was dissolved in 600 µL double-distilled water

and removed. A final concentration of 5 µM each of reduced Cyt c and oxidized Cyt c

was combined for a total concentration of 10 µM Cyt c in 30% glycerol. The kinetic

assays of a subset of sequences in the presence of Cyt c was performed using the same

method as mentioned previously, except that the final volume of each reaction was 10 µL.

1 µL of the Cyt c solution was added to the reaction just before the 2x selection buffer

was added, for a final concentration of 1 µM.

2.5   *IN VITRO* SELECTION SIMULATIONS

The simulations of the *in vitro* selection of Cyt c-dependent aptazymes under various conditions were performed using Microsoft Excel. Several assumptions and conditions were imposed to simplify the simulations: the number of sequences in the starting library pool of all selections was 1×10^{14}; the number of sequences in the enriched library pool of all subsequent rounds of selection was 1×10^{13}. In each round of selection the Y_{MAX} of the library pool was 0.6, the incubation time of the negative selection was 600 minutes and positive selection's incubation time was 10 minutes. In each round of selection the library pool contained five groups of sequences: fast catalyzing DNAzymes (group A) with a k_{obs} of 0.1/min; medium catalyzing DNAzymes (group B) with a k_{obs} of 0.01/min; slow catalyzing DNAzymes (group C) with a k_{obs} of 0.001/min; Cyt c-dependent aptazymes with a k_{obs} of 0.001/min in the absence of Cyt c and a k_{obs} in the presence of Cyt c of 0.5/min, 0.1/min and 0.01/min, depending on the simulation; and inactive sequences with a k_{obs} of 3×10^{-7}/min, the rate of hydrolysis of the rA linkage (Li & Breaker, 1999).

For the first negative selection step of each round, or the positive selection step in the simulations without a negative selection step, the number of cleaved sequences in each group was calculated by

$$F \times NS \times Y$$

where F indicates the fraction of the library pool that the group occupies, NS indicates the total number of sequences in the library pool and Y is the fraction cleaved at time t, calculated using the one-phase exponential equation introduced above. For all subsequent

negative and positive selection steps in each selection round, the number of cleaved

sequences in each group was calculated by

$$\left(\left(F \times NS\right) - NS_C\right) \times Y$$

where NS_C is the number of sequences cleaved during the previous negative selection

steps of that particular round. The subtraction of the cleaved sequences in previous steps

of that round allows only the remaining uncleaved sequences to be considered. See

Appendix III for an example of one of the simulations.

CHAPTER THREE

RESULTS & DISCUSSION

3.1   FIRST *IN VITRO* SELECTION

The cleavage percentages of all negative and positive selections are listed in Table 2. A graphical representation of the cleavage percentages of all the positive selections is in Figure 9. The percent cleavage observed in the negative selection was found to be increasing in successive rounds, thus the negative selection step was repeated twice from Round 5. Starting at Round 6, the portion not incubated in the presence of Cyt c was equivalent to the level of background cleavage in the positive selection. The background cleavage is caused by sequences that can cleave in the absence of Cyt c and were not removed during the negative selection. The cleavage activity in the presence of Cyt c can be obscured if the background cleavage is too high, so attempts were made to reduce non-specific cleavage.

Other modifications to the negative selection step were attempted to keep the background cleavage as low as possible. For these schemes, it was thought that repeated heating and slow cooling would allow the self-cleaving sequences to have multiple chances to fold properly and cleave, lowering the background cleavage in the positive selection. However, both these schemes failed to reduce the background cleavage below the levels achieved by the two five-hour incubations, as can been seen in Table 2. Starting in Round 5, the fraction incubated with Cyt c had a band develop just underneath the bottom of the well in the 10% denaturing PAGE gel. This band was posited to be composed of large complexes of DNA and Cyt c, which were unable to migrate through

the PAGE gel. A portion of those sequences may have been bound to Cyt c and would

therefore be of interest, so allowing those sequences to migrate through the gel was

imperative. To achieve this, proteinase K was added to digest all the Cyt c before the

	Negative Selection		Positive Selection				
	% Cleavage			% Cleavage		Cyt c conc.	Incub.
Round	1st	2nd	NO Cyt c	Cyt c	Cyt c + Pro K	(µM)	Time (min)
1	1.4	--	--	1.0	--	1	30
2	1.2	--	--	0.5	--	1	30
3	24	--	--	1.4	--	1	30
4	44	--	--	4.2	--	1	30
5	48	20	--	3.5	--	1	30
6	43	19	3.1	2.9	--	1	30
7	35	19	5.6	5.2	--	0.5	30
8	28	16	2.1	3.5	--	0.5	10
9[a]	27	18	7.6	7.5	--	0.5	30
10	23	11	4.7	8	--	0.5	15
11[a]	24	19	6.9	8.6	--	0.5	12
12	24	13	4.8	6.2	--	0.5	10
13	25	14	5.4	7.6	--	0.5	10
14	30[b]	--	17	20	17	1	10
15	42[b]	--	16	30	20	10	30
16	23	10	13[d]	15[d]	14[d]	1	10
17	27[c]	--	19	23, 20[d]	--	1	10
18	17[c]	--	18	31	20	1	10
19	30[c]	--	14[d]	19[d]	16[d]	1	10
20	21[c]	--	23	34	25	1	10

Table 2. Cleavage percentages and conditions of the first *in vitro* selection for Cyt c-dependent aptazymes. The negative selection originally consisted of one or two 5-hour incubations at room temperature. 'Cyt c + Pro K' indicates that proteinase K was added to the positive selection reaction after incubation with Cyt c. 'a' indicates that hypermutagenic PCR was done in the PCR2 step of that round. 'b' indicates that the negative selection consisted of multiple iterations of heating and cooling, performed in a thermocycler. 'c' indicates that the negative selection consisted of seven iterations of heating at 90 ºC and then incubation at room temperature for 1 hour. 'd' indicates that 0.5 µM DTT was added to that positive selection reaction. All cleavage percentages were calculated from the respective phosphorimages of the 10% denaturing PAGE gels using ImageQuant software.
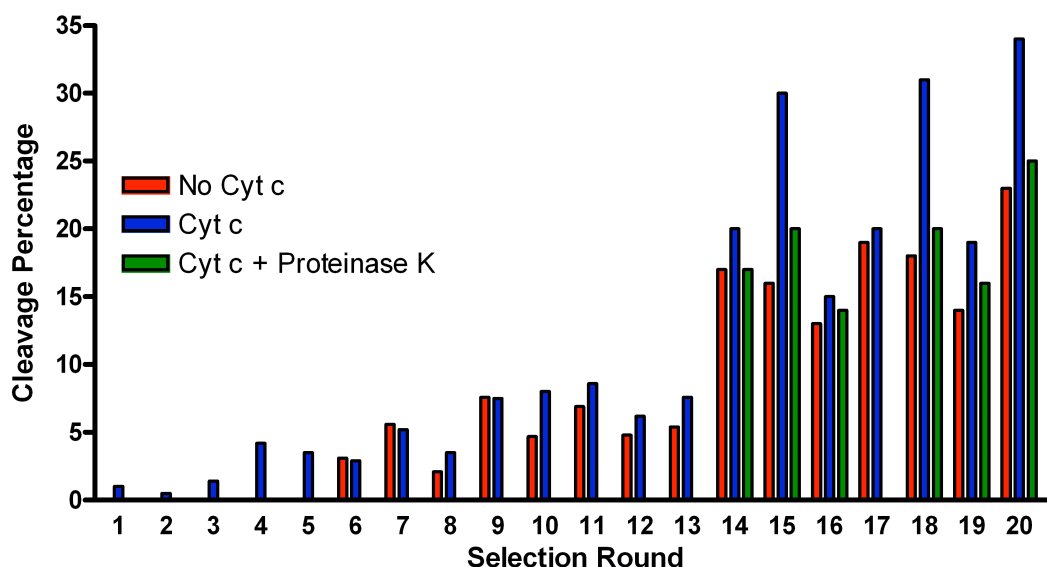
Figure 9. Graphical representation of positive selection cleavage percentages of the first *in vitro* selection of Cyt c-dependent aptazymes. Please see Table 2 for selection conditions for each selection round. All cleavage percentages were determined from the respective phosphorimages of the 10% denaturing PAGE gels using ImageQuant software.

cleaved sequences were isolated by 10% denaturing PAGE, starting in Round 14. The

removal of Cyt c allowed the sequences to migrate properly through the 10% denaturing

PAGE gel. Unsuccessful attempts were made to extract these sequences separately by

conventional elution and by electroelution.

As can be seen in Figure 9, the cleavage activities with and without the presence

of Cyt c were the same until Round 10, when the cleavage activities became higher in the

presence of Cyt c than in its absence. The cleavage activities in the presence of Cyt c but

treated with proteinase K were similar to those in the absence of Cyt c. This may indicate

that both cleaved and uncleaved sequences were interacting with Cyt c in the top band.

The diversity of the enriched library pool was increased by hypermutagenic PCR in

rounds 9 and 11. It was posited that increasing the diversity of the enriched library pool would allow beneficial mutations to accumulate more quickly.

Sequencing of the starting library confirmed that all 39 nucleotides in the random region were sufficiently randomized (see Appendix I). It was also found that the enriched library pool from Round 20, both with and without the addition of proteinase K, was dominated by a single sequence, named C1. 55% of all the sequences were identical to C1. The remaining sequences had five or less point mutations or were missing a base. 11% of the sequences had the two most common point mutations, and this sequence was named C1-2PM. MgZ was fully conserved in the vast majority of the sequences. In the other sequences, MgZ was almost fully conserved except for a few point mutations.

The cleavage activities of C1 and C1-2PM ligated to the substrate were tested in the presence and absence of Cyt c and DTT. Preliminary results demonstrated that C1 had a slightly higher cleavage activity in the presence of Cyt c, with or without DTT, while C1-2PM showed no difference. Also, the band just below the bottom of the well seen in the positive selection 10% denaturing PAGE gels was not present in these PAGE gels. This may be due to the decrease in the absolute amount of DNA or Cyt c or both, as the test reaction volume decreased compared to the positive selection while the relative amounts of DNA and Cyt c stayed the same.

Further cleavage activity tests of C1 ligated to substrate were conducted in the presence of Cyt c, BSA and glycerol with and without DTT. The averaged triplicates with the standard deviations shown as error bars are seen in Figure 10. From
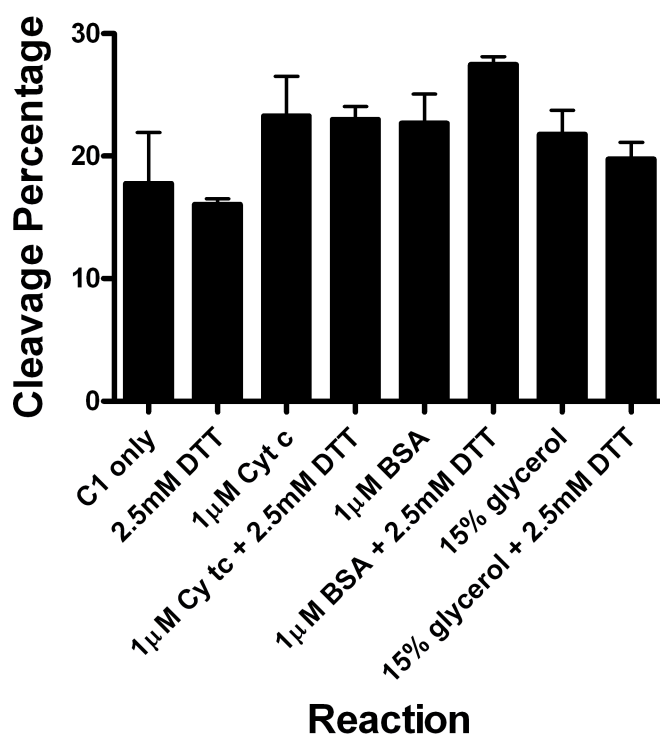
Figure 10. Cleavage activity of C1 in the presence of various compounds. C1 was ligated to the substrate, which was ^{32}P-labeled. All reactions used the sequence at a final concentration of ~250 nM. The compounds were added to C1, then 2x selection buffer was added to start the reaction. All reactions were incubated at room temperature for 3 hours and quenched with a final concentration of 30 mM EDTA at pH 8 at 23 ºC. The cleaved fraction of each reaction was separated by 10% denaturing PAGE and the cleavage percentage of each reaction was determined using ImageQuant software. All reactions were completed in triplicate; the error bars indicate the standard deviation of the triplicates. C1 was not significantly activated by the presence of 1 μM Cyt c or any other compound tested.

Figure 10, it can be seen that the cleavage activity of C1 was slightly higher in the

presence of Cyt c, but it was also slightly higher in the presence of BSA. Furthermore,

the background cleavage activity, C1 on its own and with DTT only, was ~15%. It would

be difficult to use a detection method with such high background activity as well as a

detection method with only a 0.3-fold increase above background in the presence of Cyt

c. Small amounts and/or small concentration changes of Cyt c would be difficult to

detect. C1 also did not bind Cyt c in any significant way, as no shift was detected when

C1 was incubated with Cyt c and isolated by 8% native PAGE.

3.2   SECOND *IN VITRO* SELECTION

The cleavage percentages of the negative and positive selections are listed in

Table 3. A graphical representation of the cleavage percentages of the positive selections

is in Figure 11. By Round 6, the background cleavage, caused by sequences that were not

removed during the negative selection even though they could cleave without the

presence of Cyt c, had increased dramatically. As was mentioned previously, the

cleavage activity due to the presence of Cyt c can be obscured if the background cleavage

is too high. Therefore changes were made to the selection protocols in an attempt to keep

the background cleavage as low as possible.

In vitro selection was restarted at the end of Round 3 since the background

cleavage had not increased significantly by Round 3. The negative selection was

modified to be three consecutive incubations of increasing length, each followed by

isolating the uncleaved sequences by 10% denaturing PAGE. Isolation by 10%

denaturing PAGE was deemed the most reliable way of forcing the sequences to refold

into different conformations between incubations. This was determined since heating and

cooling alone did not adequately facilitate the removal of Cyt c-independent sequences.

In Round 11-1, there was a 1% increase in cleavage activity above the background

in the presence of both oxidized and reduced forms of Cyt c. However, this increase in

cleavage activity was not seen in Round 12-1 and so was not a true signal of Cyt c-

dependent sequences. Once a true signal of target-dependence emerges, the cleavage

activity of the enriched library pool in the presence of the target should increase during

successive rounds of selection as those target-dependent sequences begin to dominate the

enriched library pool.

Round #	Negative Selection			Positive Selection				
	% cleavage			No Cyt c (%)	Cyt c (% [conc])	Cyt c + DTT (% [conc])	Incub. Time	+ Proteinase K/Elastase
1	2.4			0.26	0.27 [10μM]	0.27 [10μM/0.5mM]	30 mins	Pro K
2	2.8			0.53	1.1 [10μM]	1 [10μM/0.5mM]	30 mins	Pro K
3	1.6			0.22	0.2 [1μM]	0.36 [1μM/0.5mM]	15 mins	Pro K
4	1.9			0.51	0.58 [1μM]	0.72 [1μM/0.5mM]	15 mins	none
5	12			0	0 [1μM]	3.6 [1μM/0.5mM]	10 mins	Pro K
6	32			20	18 [1μM]	18 [1μM/0.5mM]	10 mins	Pro K
Round #	Negative Selection (% cleavage)			Positive Selection				
	1st (30mins)	2nd (2hrs)	3rd (5hrs)	No Cyt c (%)	Cyt c (% [conc])	Cyt c + DTT (% [conc])	Incub. Time	+ Proteinase K/Elastase
4-1	0.5	0.44	0.35	--	-- [1μM]	-- [1μM/0.5mM]	15 mins	Pro K
5-1	0.82	0.86	0.52	--	-- [1μM]	-- [1μM/0.5mM]	10 mins	none
6-1	1.8	1.5	1.0	--	-- [1μM]	-- [1μM/0.5mM]	10 mins	Pro K
7-1	1.2	1.7	1.7	0.4	0.52 [1μM]	0.36 [1μM/0.5mM]	10 mins	Elastase
8-1	3.0	13*	1.6	0.4	0.35 [1μM]	0.25 [1μM/0.5mM]	10 mins	Pro K
9-1	4.8	4.6	3.4	1.5	1.2 [1μM]	1.1 [1μM/0.5mM]	10 mins	Elastase
10-1	7.7	23*	2.4	1.4	1.6 [1μM]	1.6 [1μM/0.5mM]	10 mins	Pro K
11-1	11	10	8.6	3.4	4.4 [1μM]	4.3 [1μM/0.5mM]	10 mins	Elastase
12-1	12	15	16.0	7.0	6.0 [1μM]	6.5 [1μM/0.5mM]	10 mins	Pro K
Round #	Negative Selection (% cleavage)			Positive Selection				
	3 hours	5 hours		No Cyt c (%)	Cyt c (% [conc])	Cyt c + DTT (% [conc])	Incub. Time	+ Proteinase K/Elastase
8-2	10	--		1.2	1.5 [1μM]	1.1 [1μM/0.5mM]	10 mins	none
9-2	22	8.7		2.2	2.5 [1μM]	2.2 [1μM/0.5mM]	10 mins	Pro K

Table 3. Cleavage percentages and conditions of the second *in vitro* selection for Cyt c-dependent aptazymes. The negative selection of rounds 1 to 6 was a 12-hour incubation in a thermocycler. The negative selection of rounds 4-1 to 12-1 consisted of 3 incubations. In rounds 8-1 and 10-1 the second incubation was a 12-hour incubation in a thermocycler, denoted by '*'. The negative selection of rounds 8-2 and 9-2 was a 5-hour incubation. All positive and negative selection reactions in rounds 8-2 and 9-2 were separated by low-pH denaturing 10% PAGE. For the Cyt c and Cyt c + DTT reactions of the positive selections, the value not in brackets is the cleavage percentage and the value in brackets is the concentration of Cyt c and the concentration of Cyt c/DTT, respectively. 'Incub. Time' indicates the incubation time of the positive selection. '+Proteinase K/Elastase' indicates whether one of these proteins was added to the Cyt c and Cyt c + DTT reactions after incubation. All cleavage percentages were calculated from the respective phosphorimages of the 10% denaturing PAGE gels using ImageQuant software. Round 4-1 is seeded by Round 3 and Round 8-2 is seeded by Round 7-1.
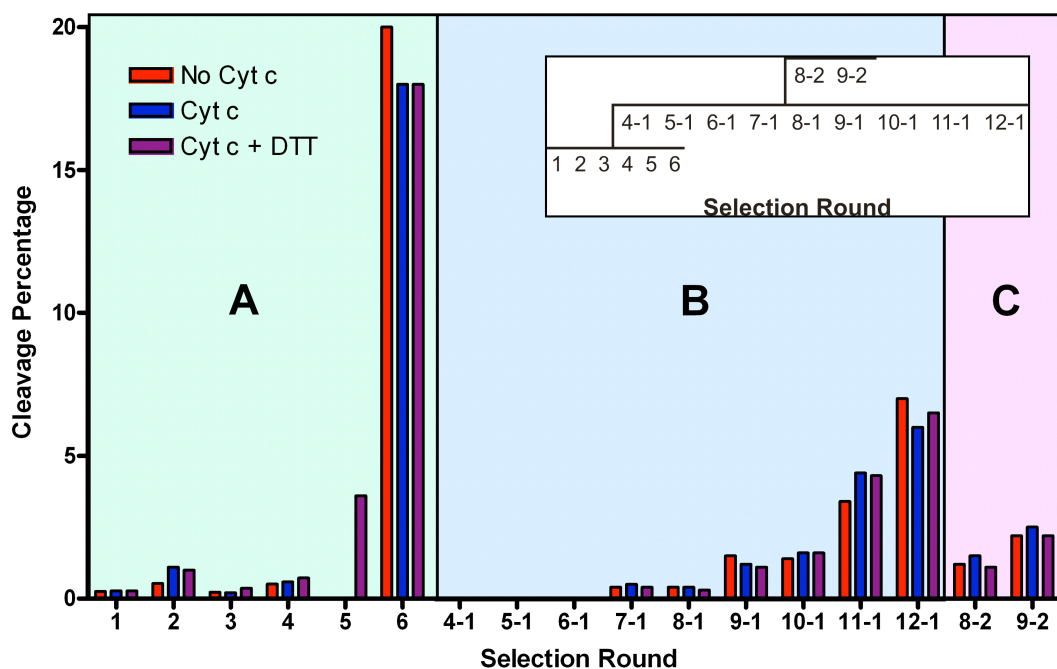
Figure 11. Graphical representation of positive selection cleavage percentages of the second *in vitro* selection of Cyt c-dependent aptazymes. Please see Table 3 for selection conditions for each selection round. All cleavage percentages were determined from the respective phosphorimages of the 10% denaturing PAGE gels using ImageQuant software. Block A) negative selection was one 12-hour incubation, heated at 90 ºC for 45 seconds every 2 hours, in a thermocycler. Block B) negative selection consisted of three incubations of 30 minutes, 2 hours and 5 hours. After each incubation sequences were separated by 10% denaturing PAGE. Block C) negative selection was one 5-hour incubation; sequences of both negative and positive selections were separated by low-pH 10% denaturing PAGE.

In Round 12-1, since no increase in cleavage activity in the presence of Cyt c was observed above the background cleavage and the background cleavage was again increasing steadily, *in vitro* selection was restarted at the end of Round 7-1. The background cleavage in Round 7-1 was ~0.5%. The negative selection was reverted to one 5-hour incubation in selection buffer only, but the uncleaved sequences were then isolated by low-pH denaturing 10% PAGE.

When RNA-cleaving DNAzymes dependent on Mg^{2+} were being selected by *in vitro* selection, background cleavage activity by sequences independent of Mg^{2+} was a major obstacle. The cleavage activities of these sequences were found to be suppressed by low pH conditions (Chiuman & Li, 2007), thus isolation by low-pH denaturing 10% PAGE was used during the positive and negative selection steps. This crucial change allowed MgZ to be selected. It was hypothesized that this same change might suppress the background cleavage of the current *in vitro* selection, so for rounds 8-2 and 9-2 the uncleaved sequences after the negative selection and the cleaved sequences after the positive selection were isolated by low-pH denaturing 10% PAGE.

After Round 9-2 it was thought that a lack of sequence diversity could be having an adverse effect on the *in vitro* selection, so sequences from rounds 6, 12-1, and 9-2 were cloned and sequenced (see Appendix II). Round 6 had no dominant sequences, Round 12-1 had 2 dominant sequences (R1201 and R1202 composing 50% and 15% of the library pool, respectively) and R9-2 had one dominant sequence very similar to R1201, comprising 21% of the library pool. MgZ was highly conserved in all the selection rounds that were sequenced.

3.3   KINETIC STUDIES OF SELECTED SEQUENCES

The random region with the two central G-C base pairs of MgZ (bases 31-74, see

Figure 8) of each sequence was isolated to determine ΔG and secondary structure, seen in

Table 4, using mFold (Zucker, 2003). This was the region selected for interaction with

Cyt c and therefore of the most interest. The three sequences chosen from Round 6

represented the widest possible range of ΔG values. The sequences with the lowest and

highest ΔG values from Round 6, as well as a sequence with a midrange ΔG value (R606,

R626 and R617, respectively) and the two most abundant sequences from R12-1 (R1201

and R1202) were chosen for further study. As seen in Table 4, the total number of base

pairs in the random region and the number of base pairs in the central stem connecting the

random region to MgZ was also determined for each sequence. It is understood that the

more base pairs there are in a given sequence, the more stable that sequence will be.

Since the central stem of the random region brings the two halves of MgZ together to

perform catalysis, the more base pairs there are in the central stem, it follows that there

will be more background cleavage.

To validate this hypothesis, a modified sequence of each chosen sequence was

designed so that it had between nine and eleven base pairs in its central stem (all

sequences with the 'mod' prefix, see Table 4, also see Appendix II). The ΔG values and

secondary structures of these modified sequences were also determined. To compare the

actual activities of the chosen sequences to each other and to their modified counterparts,

a kinetic assay was performed on each full sequence containing primer binding regions,

MgZ and random region ligated to the substrate. A one-phase exponential curve was
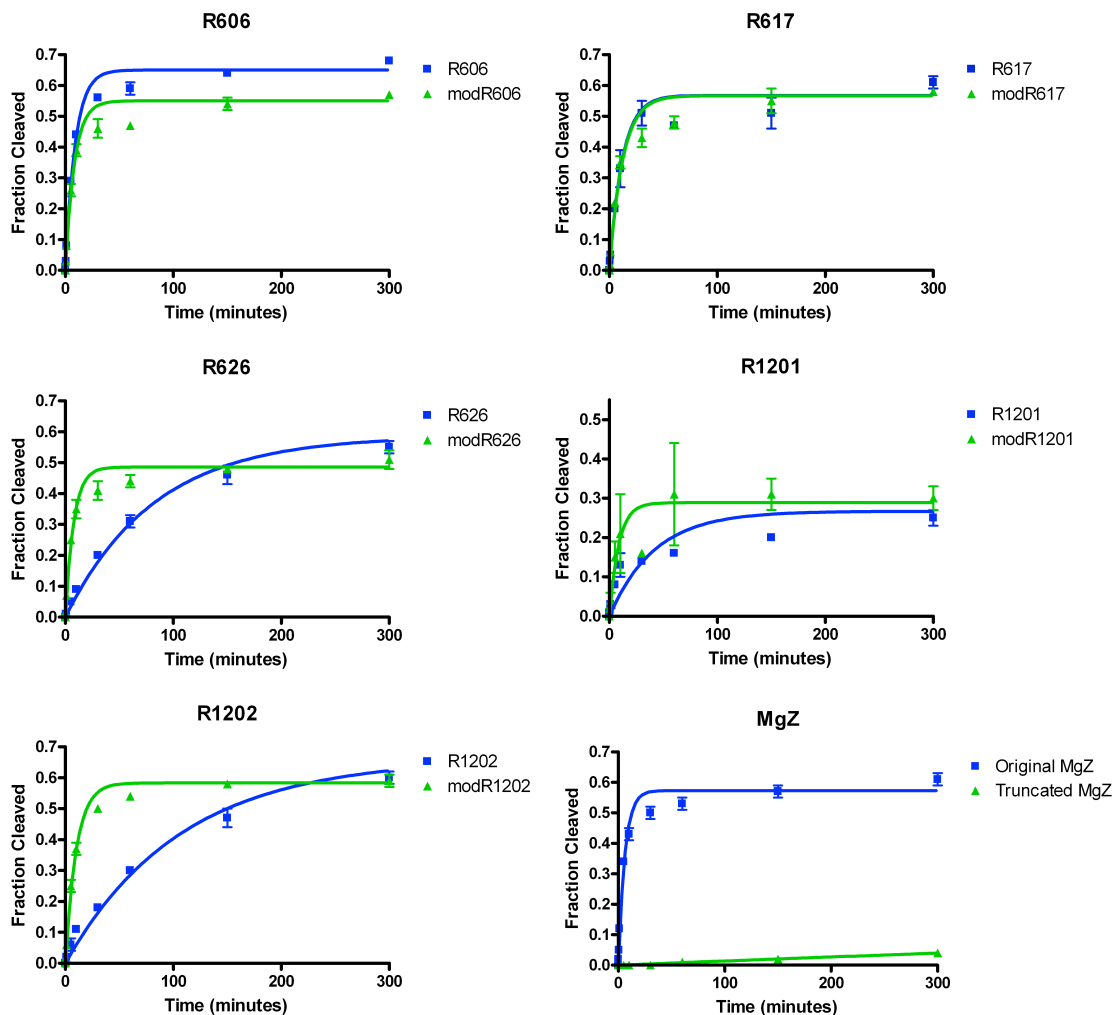
Figure 12. Kinetic curves of original and modified sequences from *in vitro* selection.
Each sequence was ligated to the substrate, which was ^{32}P-labeled. All reactions used the sequence at a final concentration of 100 nM. 2x selection buffer was added to start the reaction. All reactions were incubated at room temperature for a prescribed amount of time and quenched with 1 µL of 3 M EDTA at pH 8 at 23 ºC. All reactions were completed in triplicate. The cleaved fraction of each reaction was separated by 10% denaturing PAGE and the fraction cleaved of each reaction was determined using ImageQuant software. Each triplicate set of data was normalized to 0 fraction cleaved at 0 minutes and the cleavage fraction value is the average of the three values, with the standard deviations of the average values shown as error bars. Each set of averaged values was fitted with a one-phase exponential with the equation $Y = Y_{MAX}(1 - e^{-k_{obs}t})$ where Y_{MAX} is the maximum fraction cleaved, k_{obs} is the observed kinetic rate and t is time in minutes. The modified sequences of R626, R1201 and R1202 displayed a significant increase in k_{obs} compared to their original sequences.
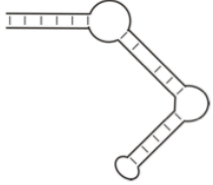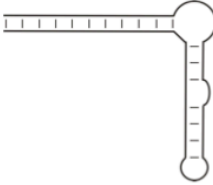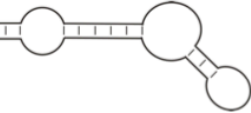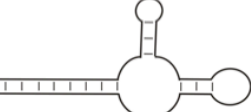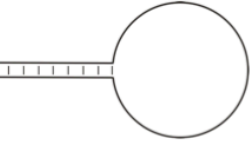
45

Name	Predicted Secondary Structure	ΔG	Base Pairs		k_{obs}	Y_{MAX}	+ 1 µM Cyt c	
			Total #	in Stem			k_{obs}	Y_{MAX}
R606		-14.39	16	6	0.11 ± 0.02	0.65 ± 0.02	0.11 ± 0.01	0.60 ± 0.01
modR606		-19.14	18	11	0.12 ± 0.03	0.55 ± 0.03		
R617		-7.82	12	3	0.08 ± 0.02	0.57 ± 0.03		
modR617		-11.17	15	9	0.08 ± 0.02	0.57 ± 0.02		
R626		-1.16	3	0	0.012 ± 0.001	0.59 ± 0.02	0.015 ± 0.001	0.62 ± 0.01
modR626		-7.32	9	9	0.13 ± 0.02	0.49 ± 0.02		

Table 4. Secondary structure, stability, k_{obs} and Y_{MAX} of various sequences from the second *in vitro* selection. The secondary structures pictured are bases 31-74 of each sequence (see Figure 8). All the secondary structures, ΔG and base pair values were predicted using mFold (Zucker, 2003). All k_{obs} (in min^{-1}) and Y_{MAX} values were determined by fitting a one-phase exponential curve to averaged triplicate kinetic data of each sequence.
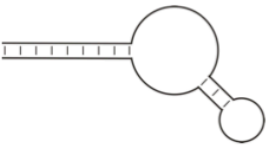
Name	Predicted Secondary Structure	ΔG	Base Pairs		k_{obs}	Y_{MAX}	+ 1 μM Cyt c	
			Total #	in Stem			k_{obs}	Y_{MAX}
R1201		-3.53	10	7	0.02 ± 0.01	0.27 ± 0.03	0.016 ± 0.006	0.29 ± 0.03
modR1201		-8.61	11	9	0.12 ± 0.04	0.29 ± 0.02		
R1202		-8.50	8	0	0.0094 ± 0.0009	0.66 ± 0.02	0.014 ± 0.001	0.65 ± 0.02
modR1202		-10.58	12	9	0.10 ± 0.01	0.58 ± 0.02		
Original MgZ		-4.03	9	9	0.16 ± 0.03	0.57 ± 0.02		
Truncated MgZ		N/A	2	2	0.0004 ± 0.0001	0.31 ± 0.07		

Table 4 continued. Secondary structure, stability, k_{obs} and Y_{MAX} of various sequences from the second *in vitro* selection. In Original and Truncated MgZ, secondary structure pictured is P3 of MgZ (see Figure 4). All other sequences, secondary structures pictured are bases 31-74 (see Figure 8). All the secondary structures, ΔG and base pair values were predicted using mFold (Zucker, 2003). All k_{obs} (in min^{-1}) and Y_{MAX} values were determined by fitting a one-phase exponential curve to averaged triplicate kinetic data of each sequence.

fitted to each set of data and the observed kinetic rate (k_{obs}) and maximum fraction cleaved (Y_{MAX}) were determined (see Figure 12, Table 4).

The k_{obs} in general followed the same trend as the ΔG values: the more stable the random region, the higher the k_{obs}, which supports the theory that the more stable this region is, the more background cleavage there will be. Two notable exceptions were the k_{obs} of R606 compared to modR606 and R617 compared to modR617: despite the increase in ΔG in the modified sequences, no increase in k_{obs} wa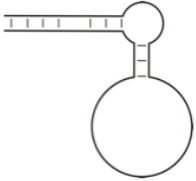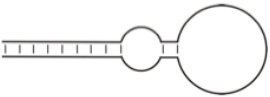s observed. This may indicate that the random regions of R606 and R617 were already quite stable and could not be improved by increasing the number of base pairs in the central stem.

Another notable exception to this trend was the original MgZ sequence. Despite the original sequence's central stem having a relatively low ΔG value it had the highest k_{obs} of all sequences analyzed. This may be due to the fact that this sequence was shorter and less complicated than the other sequences, making it comparatively more stable. To determine how important the central stem was to the catalytic activity of MgZ, a truncated MgZ sequence was created where there were only the two G-C base pairs in the central stem. The k_{obs} of this sequence was several-fold lower than any other sequence analyzed, indicating that the central stem loop was important for cleavage activity, but that the sequence itself was not. This was hypothesized when MgZ was first selected and analyzed (Chiuman & Li, 2007), and is further confirmed here.

The next question to arise was whether any of the selected sequences' k_{obs} or Y_{MAX} values would increase in the presence of Cyt c. R1201 had a lower Y_{MAX} than all the other sequences tested, so the possibility existed that in the presence of Cyt c this

value and/or its k_{obs} value would increase. R626 and R1202 had random regions with

very few base pairs; perhaps Cyt c would somehow stabilize those regions, increasing the

k_{obs} values. R606 was chosen to act as a sort of negative control; since its random region

was already very stable, no difference in the k_{obs} or Y_{MAX} values was expected.
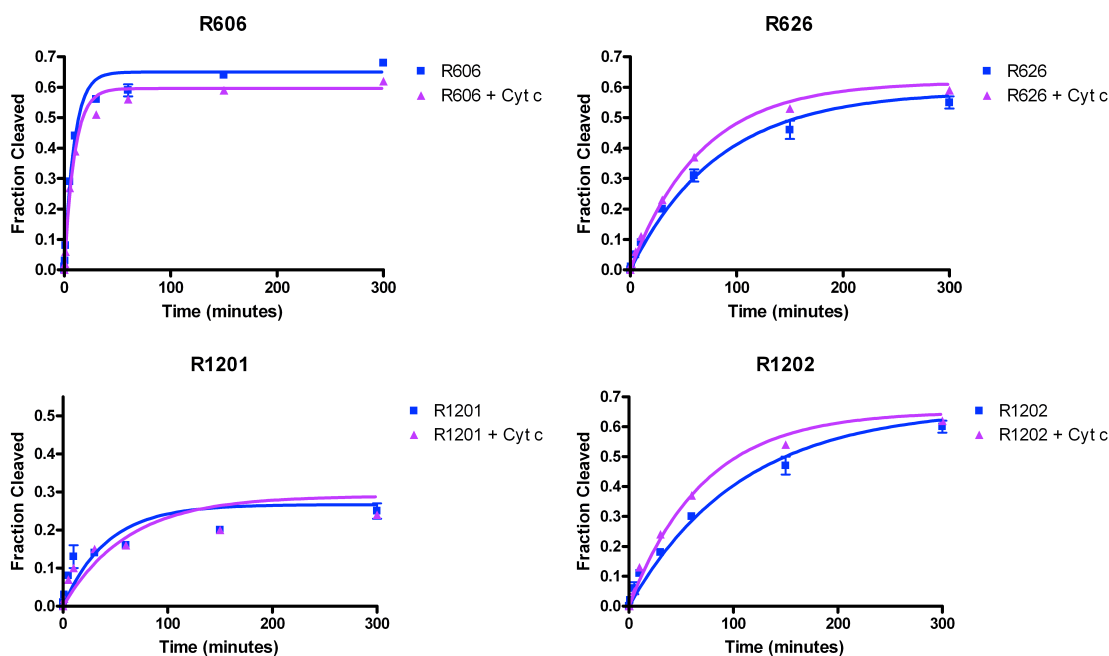
Figure 13. Kinetic curves of original sequences from *in vitro* selection in the presence of Cyt c. The Cyt c used was a 1:1 mixture of Cyt c whose heme was reduced (Fe^{2+}) and oxidized (Fe^{3+}). Each sequence was ligated to the substrate, which was ^{32}P-labeled. All reactions used the sequence at a final concentration of 100 nM. Cyt c to a final concentration of 1 µM was added and then 2x selection buffer was added to start the reaction. All reactions were incubated at room temperature for a prescribed amount of time and quenched with 1 µL of 3 M EDTA at pH 8 at 23 ºC. All reactions were completed in triplicate. The cleaved fraction of each reaction was separated by 10% denaturing PAGE and the fraction cleaved of each reaction was determined using ImageQuant software. Each triplicate set of data was normalized to 0 fraction cleaved at 0 minutes and the cleavage fraction value is the average of the three values, with the standard deviations of the average values shown as error bars. Each set of averaged values was fitted with a one-phase exponential with the equation $Y = Y_{MAX}(1 - e^{-k_{obs}t})$ where Y_{MAX} is the maximum fraction cleaved, k_{obs} is the observed kinetic rate and t is time in minutes.

Since the second *in vitro* selection was performed with both reduced and oxidized Cyt c, it was possible that any of these sequences would be able to interact with either or both forms of Cyt c. Therefore, a solution of an equal mixture of both forms of Cyt c was used (see section 2.4 for more information). A kinetic study of each of the four chosen sequences in the presence of Cyt c was conducted. A one-phase exponential was again fitted to each data set and the k_{obs} and Y_{MAX} values determined (see Figure 13, Table 4).

Compared to the several-fold increase of the modified sequences to their original counterparts, no several-fold increases in the k_{obs} and no large increases of Y_{MAX} values of sequences with and without Cyt c were observed. As was hypothesized, there was no difference between the k_{obs} values of R606 with and without Cyt c and the Y_{MAX} values were nearly identical. There was also no significant difference between the k_{obs} and Y_{MAX} values of R626 and R1201 with and without Cyt c. R1202 displayed the only significant difference in k_{obs} values with and without Cyt c; the k_{obs} value in the presence of Cyt c was 0.5-fold higher than in the absence of Cyt c. No difference was seen in the Y_{MAX} values. But R1202 would be difficult to utilize as a detection method of Cyt c for the same two reasons as C1 from the first *in vitro* selection: its high background activity as well as its modest increase above background in the presence of Cyt c.

3.4   SIMULATIONS OF *IN VITRO* SELECTION

It is clear that the two *in vitro* selections for Cyt c-dependent aptazymes failed to achieve the desired result. They can, however, provide valuable information as to why the selections were unsuccessful. This knowledge would then guide future *in vitro* selections, increasing the likelihood of their success. One method of gaining a better understanding of previous *in vitro* selections is by attempting to model, or simulate them. These simulations have been performed based on a set of assumptions derived from the results of the kinetic studies above, as well as the literature.

The number of sequences in the library pool of each selection round, specifically 1 x 10^{14} sequences for the first round and 1 x 10^{13} sequences for all subsequent rounds, is based on what was actually used as well as what was used in previous simulations of *in vitro* selection (Schlosser & Li, 2005). The Y_{MAX} value of 0.6 is the approximate average of the Y_{MAX} values from the kinetic data in Table 5. The k_{obs} values used in the simulations are also based on the kinetic data:

- Group A: the fast-cleaving sequences have a k_{obs} of 0.1/min, since two of the fastest-cleaving sequences, R606 and R617, have k_{obs} values of 0.11/min and 0.8/min, respectively;

- Group B: the medium-cleaving sequences have a k_{obs} of 0.01/min, based on the k_{obs} of R626 at 0.012/min;

- Group C: the slow-cleaving sequences have a k_{obs} of 0.001/min, which is based on the slowest-cleaving sequence tested, Truncated MgZ with a k_{obs} of 0.0004/min;

- Group D: the k_{obs} of Cyt c-dependent sequences in the absence of Cyt c was also set at 0.001/min, since a low rate of background cleavage is required for the sequences to be used as sensitive biosensors;

- Group E: the k_{obs} of the inactive sequences is the spontaneous rate of cleavage of the rA linkage of the substrate, 3×10^{-7}/min (Li & Breaker, 1999).

Realistic percentages for each group occupied within the starting library pool were also chosen. It was assumed that the majority of sequences in the starting library pool would be inactive, so group E occupied 99.99% of it. Of the sequences that were active, most would not be Cyt c-dependent. It is thought that of the entire starting library pool, very few sequences would display the characteristics that were being sought (Cho, Lee & Ellington, 2009). So a ratio of groups A:B:C:D was set at 50,000,000:50,000,000:50,000,000:1, which corresponds to 6 sequences in group D in the starting library pool.

An aptazyme for cyclic GMP (cGMP) based on the hammerhead ribozyme displays a 510-fold increase of k_{obs} in the presence of *versus* the absence of cGMP (Koizumi, Soukup, Kerr & Breaker, 1999). So a 500-fold increase of k_{obs} was chosen as the highest-fold increase for group D sequences in the presence of Cyt c. From a survey of RNA- and DNA-cleaving aptazymes in the presence of their target, the average-fold increase of k_{obs} is 87. Therefore a 100-fold increase of group D's k_{obs} in the presence of Cyt c was chosen for another set of simulations. From this same survey it was noted that very few aptazymes were reported displaying a <10-fold increase of k_{obs}, so this value

was chosen as the lowest-fold increase in the presence of Cyt c for another set of

simulations.

Using these assumptions, many simulations using a variable number of negative

selection steps per round (from zero to five), as well as the various-fold increases in k_{obs}

of group D in the presence of Cyt c, were conducted (see Appendix III for an example of

one of the simulations). The results of one of the simulations are shown in Figure 14. In

this simulation, one negative selection step per round was utilized as well as a 500-fold

increase of group D's k_{obs} in the presence of Cyt c. From Figure 14 it can been seen that
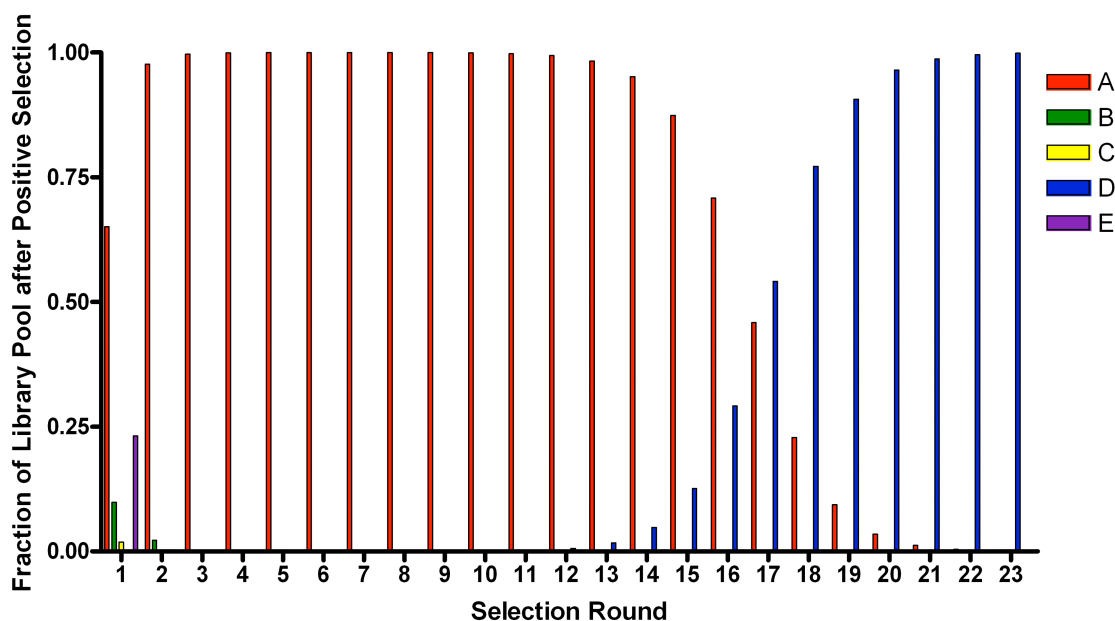
Figure 14. A simulation of *in vitro* selection of Cyt c-dependent aptazymes. This
simulation has one negative selection step per selection round and assumes a 500-fold
increase in k_{obs} for group D in the presence of Cyt c. Group A: fast catalyzing DNAzymes
with a k_{obs} of 0.1/min; Group B: medium catalyzing DNAzymes with a k_{obs} of 0.01/min;
Group C: slow catalyzing DNAzymes with a k_{obs} of 0.001/min; Group D: Cyt c –
dependent aptazymes with a k_{obs} of 0.001/min in the absence of Cyt c and a k_{obs} in the
presence of Cyt c of 0.5/min, 0.1/min and 0.01/min, depending on the simulation; and
Group E: inactive sequences with a k_{obs} of 3 x 10^{-7}/min, the rate of spontaneous cleavage
of the substrate's rA linkage (Li & Breaker, 1999).

groups B, C, D and E were minimized in the library pool after 2 rounds of selection, while group A quickly dominated the library pool. Group A continued to dominate the library pool for several rounds. Starting at Round 12, group D started to occupy a larger portion of the library pool until finally it occupied a majority of the library pool at Round 17. The selection round at which group D overtook group A and occupied a majority of the library pool for each simulation is shown in Table 5.

# Negative Selections	Group D Fold Increase in Presence of Cyt c		
	500	100	10
0	Never	Never	Never
1	17	30	>60/Never
2	11	15	>60/Never
3	8	10	>60/Never
4	7	8	35
5	6	6	16

Table 5. Results of the simulations of the *in vitro* selection for Cyt c-dependent aptazymes under various conditions. Each value in the table indicates the selection round in which group D overtook group A and occupied a majority of the library pool under the specified conditions. 'Never' indicates that this situation would never occur under those conditions. '>60/Never' indicates that this situation may or may not occur under those conditions, but if it did it would take over 60 rounds of selection and would therefore be impractical.

These simulations reveal an important concept for the *in vitro* selection of Cyt c-dependent aptazymes: more negative selection steps per selection round would allow the Cyt c-dependent sequences to occupy a majority of the library pool in fewer rounds. But at the same time, the more negative selection steps that are performed, the higher the likelihood that the few Cyt c-dependent sequences in the library pool would be lost. It is clear from Table 5 that there is an optimal number of negative selection steps that should be performed depending on the increase in activity of the Cyt c-dependent sequences in the presence of Cyt c.

Judging from the two existing *in vitro* selections for Cyt c-dependent aptazymes, it is likely that four or five negative selection steps per round would be needed for Cyt c-dependent sequences to overtake the library pool. In the second *in vitro* selection three negative selection steps were performed for eight consecutive rounds with no indications that any Cyt c-dependent sequences were starting to overtake the library pool. But if these simulations are applied broadly, an *in vitro* selection for aptazymes with a simpler target would require only two or three negative selection steps per selection round.

These simulations indirectly point to another possible reason why the two *in vitro* selections for Cyt c-dependent aptazymes were unsuccessful: Cyt c may be an unsuitable target. As mentioned in the introduction, the majority of protein-dependent aptazymes have DNA- or RNA-binding protein targets. These proteins already bind to DNA or RNA so developing aptazymes based on these targets is relatively simple. Cyt c, on the other hand, is not a DNA- or RNA-binding protein. A 118-nucleotide DNA aptamer has been selected for it, but the aptamer has a binding affinity of 0.8 μM (Chinnapen & Sen, 2002), which is relatively poor compared to most aptamers (Cho *et al*., 2009). This may indicate that it would be very difficult to find a DNA or RNA sequence that binds tightly and specifically to Cyt c, making the selection for Cyt c-dependent aptazymes all the more challenging.

CHAPTER FOUR

CONCLUSIONS & FUTURE RESEARCH

Despite unsuccessful attempts to develop Cyt c-dependent aptazymes, valuable insights into MgZ and the *in vitro* selection for aptazymes have come about. It was confirmed that MgZ requires a stable central stem for cleavage activity, but the exact sequence is unimportant. This indicates that rational design could be used to convert MgZ into an aptazyme by replacing the central stem with an aptamer. Also, the *in vitro* selection simulations revealed a method of improving future *in vitro* selections for aptazymes utilizing MgZ: increasing the number of negative selection steps will decrease the total number of rounds of selection needed before the target-dependent sequences occupy the majority of the library pool.

From these conclusions, two possibilities exist for the development of Cyt c-dependent aptazymes. One possibility is to develop a more sensitive Cyt c aptamer and then develop a Cyt c-dependent aptazyme by rational design. However, since there are easier ways to transform an aptamer into a biosensor, such as structure switching (Lau, Coombes & Li, 2010), this possible method of development is unlikely to be used. A more likely possibility is developing a Cyt c-dependent aptazyme by using a more knowledge-based *in vitro* selection. From the simulations, it is clear that increasing the number of negative selection steps per round to four or five would be advantageous, as it would decrease the number of *in vitro* selection rounds needed. This, along with other strategies utilized in the first two *in vitro* selections, such as digesting the Cyt c before

isolating the cleaved sequences by 10% PAGE in the positive selection, would increase the likelihood of successful development of Cyt c-dependent aptazymes.

While the development of Cyt c-dependent aptazymes should absolutely be further pursued, it is possible that Cyt c may be a poor target for aptazyme development. If Cyt c must be abandoned as an aptazyme target, other proteins released from mitochondria during apoptosis should be considered, including SMAC/DIABLO, HtrA2/Omi and, in particular, Endonuclease G (EndoG) (Irvine *et al.*, 2005; Ledgerwood & Morison, 2009). EndoG is a non-specific nuclease that cleaves both DNA and RNA and in healthy cells resides in the mitochondrial intermembrane space. It is released into the cytosol during apoptosis and is responsible for DNA and RNA degradation (Schäfer *et al.*, 2004). EndoG is larger than Cyt c at 28 kDa, but like Cyt c it is highly conserved in a wide range of organisms (Irvine *et al.*, 2005; Schäfer *et al.*, 2004).

Performing *in vitro* selection to find a substrate that EndoG can cleave efficiently, where the substrate produces an increase in fluorescence when cleaved, is possible. This method is currently being utilized in the Li lab to find signaling substrates for several RNases. A biosensor of this nature for EndoG could then be developed into a novel apoptotic detection method.

CHAPTER FIVE

FURTHER STUDY OF THE 22-18 DNAZYME

5.1   INTRODUCTION

The selection of fast-acting functional nucleic acid enzymes are important

precursors to DNA- and RNA-based biosensors. DNAzymes are especially suited for the

development of robust biosensors due to DNA's superior resistance to chemical and

enzymatic degradation relative to RNA (Mei *et al.*, 2003). As was demonstrated in

Chapter 3, it is often necessary to have a thorough understanding of a DNAzyme's

secondary structure before more complex biosensors, including aptazymes, can be

developed.

The DEC22-18 DNAzyme was the first DNAzyme selected with a chimeric

substrate composed of all DNA with a single ribo-A flanked by a fluorescein-labeled T

and a DABCYL quencher-labeled T. The *in vitro* selection of an RNA-cleaving

DNAzyme was performed in the presence of Mg^{2+} and several other divalent metal ions,

since the presence of these metal ions has been shown to lead to the isolation of a diverse

range of DNAzymes (Wang, Billen & Li, 2002). DEC22-18 was the result of this

selection; it is composed of 109 nucleotides (including the substrate domain) and was

found to exhibit an observed catalytic activity (k_{obs}) of 1.0/min. Sequence truncation

performed on the 3' end of the sequence determined that removing the last 29 bases

resulted in a DNAzyme exhibiting a k_{obs} of 2.1/min, denoted E1 (Mei *et al.*, 2003).

A reselection of E1 was performed to determine the functionally essential and

non-essential nucleotides of the DNAzyme. The sequence of E1, excluding the primer

binding arms and substrate, was partially randomized so that each base had a 76% chance

of being the original nucleotide and an 8% chance of being each of the other three

nucleotides. After four rounds of *in vitro* selection with the partially randomized library,

the enriched library was cloned and sequenced and all sequences from the reselection

were compared to the original sequence of E1 (Kandadai & Li, 2005). The determination

of the functional importance of each base is as follows: if a nucleotide is fully conserved,

then it must be functionally essential since no sequences with any other bases in that

position survived. If a nucleotide is highly conserved, then it must be functionally

important, but not absolutely essential. If a nucleotide is not conserved, *i.e.* other

nucleotides have taken its place, it is not functionally important, since it does not seem to

make a difference which nucleotide takes its place.
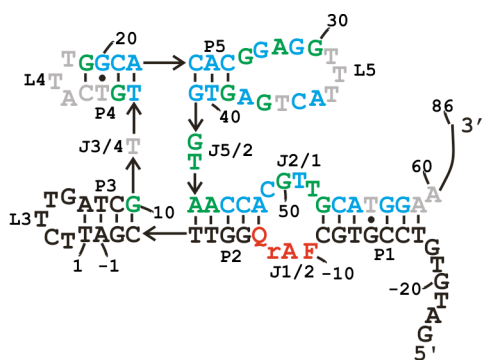
Figure 15. The primary and secondary structure of the 22-18 DNAzyme. The fluorescein-labeled T (F), ribo-A (rA) and DABCYL quencher-labeled T (Q) of the substrate are in red. Essential nucleotides are in green, important nucleotides are in blue and non-essential nucleotides are in grey. Areas of base pairing are labeled P1 - P5, loops of nucleotides are labeled L3 - L5 and junctions linking different areas of base pairing are labeled to indicate which two areas they link together.

Figure 15 shows E1 with its bases colour-coded to indicate functionally essential

nucleotides in green, important ones in blue, and non-essential ones in grey. This figure

also demonstrates E1's putative secondary structure, which contains five putative

stretches of duplex DNA, denoted P1 to P5 as well as several loops (L) and junctions (J).

Many constructs containing mutations and deletions to all areas of this DNAzyme were

conducted to determine which components of the secondary structure were essential, as

well as to corroborate the information from the reselection. All constructs' kinetic

activities were determined and rated on a scale explained in section 5.2.

Most of these constructs have already been produced and their kinetic activities

rated by S. A. Kandadai (results not published). His work has led to the following

conclusions: all of L3 and all but three nucleotides of P3 can be removed without a

decrease in activity. The number of base pairs in P1 can be decreased to just two when

they are removed from the open ends of the DNAzyme towards J1/2 without a significant

loss in activity. P4 is necessary for optimal activity, while L4 is unnecessary. L5 cannot

be shortened or mutated and P2 cannot be mutated without a significant loss of activity.

The absence of the DABCYL-modified T, rA or fluorescein-modified T completely

abolishes catalytic activity. These conclusions support the involvement of many or all of

the functionally essential nucleotides that reside in these areas.

The goal of this project was to complete the analysis of E1. The catalytic

activities of further constructs with modified P1 regions were tested. Furthermore, to

confirm the necessity of the functionally essential nucleotides that were not tested by the

previous kinetic assays, the catalytic activity of constructs with one fully conserved

nucleotide mutated to another nucleotide were determined. The completion of these tests

provides a complete understanding of which areas and nucleotides are functionally

essential and which areas can be modified. The areas that can be modified are potential

locations to insert aptamers, allowing E1 to be developed into a biosensor.

5.2   METHODS

All unmodified DNA sequences were synthesized by Integrated DNA

Technologies (Coralville, IA). Modified substrate sequences were synthesized by the

Keck Oligonucleotide Synthesis Facility at Yale University (New Haven, CT). The

fluorescein and DABCYL groups were attached by using a fluorescein-dT

phosphoramidite and a DABCYL-dT phosphoramidite, respectively, from Glen Research

(Sterling, VA). The single riboadenosine located internally in the substrates had a

TBDMS protecting group at the 2'-OH. It was deprotected by incubation in TBAF at 60

ºC for 12 hours. All sequences were purified by 10% denaturing PAGE, redissolved in

double-deionized water and then quantified using a NanoVue spectrophotometer from GE

Healthcare (Uppsala, Sweden). $[\gamma\text{-}^{32}P]ATP$ was acquired from GE Healthcare (Uppsala,

Sweden). Water was purified with a Milli-Q Synthesis A10 water purification system and

then autoclaved. Tris, urea and boric acid were purchased from Bioshop Canada

(Burlington, ON). All other chemicals were purchased from Sigma-Aldrich (Oakville,

ON).

Polynucleotide kinase and T4 DNA ligase were purchased from Fermentas

(Burlington, ON). The nucleic acid enzyme sequences were split into two portions for 5'

phosphorylation: 250 pmol of each was phosphorylated following manufacturer's

instructions; 50 pmol of each sequence was phosphorylated using $[\gamma\text{-}^{32}P]ATP$. These

sequences were then ligated to the appropriate substrate sequence following

manufacturer's instructions. The constructs were purified by 10% denaturing PAGE and

the non-radioactive portion of each construct was dissolved in 100 µL double-deionized

water and quantified using a NanoVue spectrophotometer. The radioactive portion was

dissolved in 50 µL of double-deionized water.

Each DNAzyme construct was diluted to a concentration of 1.0 pmol/µL and 1 µL

of the radioactively labeled sequence was added for tracking purposes. The kinetic

analysis of each construct was completed as follows: 2.5 pmol of a construct was heated

at 90 ºC for 30 seconds and then allowed to fully cool to room temperature over five

minutes. An equal volume of 2x reaction buffer (800 mM NaCl, 200 mM KCl, 15 mM

$MgCl_2$, 2 mM $CoCl_2$, 10 mM $MnCl_2$, 0.5 mM $NiCl_2$, 2.5 mM $CdCl_2$ and 100 mM HEPES,

pH 6.8 at 23 ºC) was added to start the reaction. Reactions were incubated for 0, 0.1, 1,

10, 100, and 1000 minutes and then quenched with 21 µL of solution containing 1x

PAGE loading solution and 5 mM EDTA, pH 8 at 23 ºC. All reactions were completed in

duplicate.

The cleaved and uncleaved products of each reaction were separated by 10%

denaturing PAGE and the phosphorimage of each gel was taken. The cleavage

percentages of each reaction were determined from the phosphorimages using

ImageQuant software. The cleavage activity of each DNAzyme construct was rated as

follows: a construct exhibiting ≥10% cleavage in 1 min was given "+++++"; a construct

exhibiting ≥10% cleavage in 10 min, 100 min and 1000 min were scored "++++", "+++"

and "++", respectively; those with less than 10% cleavage in 1000 min were given a "+";

molecules that produced no cleavage after incubation for 1000 min were denoted by a "-".

5.3   RESULTS & DISCUSSION

The P1 region was modified in several ways to determine which regions of P1 were required for optimal activity, and which regions were dispensable (see Figure 16). In constructs E44 and E46, most of P1 was replaced with fully base-pairing sequences. In construct E45 most of P1 was replaced with a non-base-pairing sequence. All three of these constructs demonstrated moderate cleavage activity, indicating that the specific sequence of P1 was required for optimal activity. Constructs E48 and E49 had one and two modified base pairs of P1 closest to J1/2, respectively. These constructs retained optimal cleavage activity, indicating that while most of the sequence of P1 could not be modified without reducing cleavage activity, these two base pairs could be.
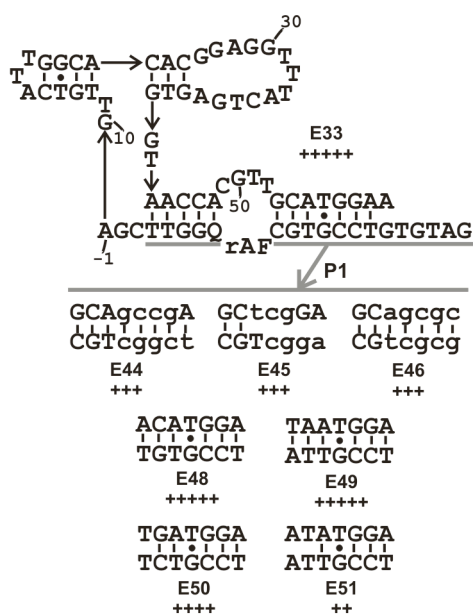
Figure 16. Cleavage activities of various modifications of P1. For each construct, all reactions used a final concentration of 0.5 µM. 2x selection buffer was added to start the reaction. All reactions were incubated at room temperature for a prescribed amount of time and quenched with solution containing 1x PAGE loading solution and 5 mM EDTA, pH 8 at 23 ºC. All reactions were completed in duplicate. The cleaved fraction of each reaction was separated by 10% denaturing PAGE and the fraction cleaved of each reaction was determined using ImageQuant software. The cleavage activity of each construct was rated using the index listed.

Construct E50 had one less base pair closest to J1/2 than E1, while E51 had two less base pairs than E1. E50 had a slight loss of cleavage activity, while E51 had a significant loss of cleavage activity. This indicates that while the exact sequence of those two base pairs did not affect activity, they could not be removed without a loss of activity. Overall, the sequence of P1 played a role in the optimal activity of E1.

```
                                                30
          T GGCA ──▶ CAC GG AG G T
        T   | • | |        | | |        G  T
          A CTGT T    GTG A GTC A T
                   G                          E33
                 ↗10 G                       ++++
                     T
                     ↓ G     GT T
                  AACCA C |    GCATGGAA
                  | | | | | 50  | | | • | | |
               AGCTTGG Q      CGTGCCTGTGTAG
                 ↙                rAF
                -1
```

Name	Mutation	Activity	Name	Mutation	Activity
E55	G26 to T	++	E56	A28 to T	+
E57	G30 to T	+++	E58	G37 to T	+
E59*	G39 to C	-	E60	G42 to T	-
E61	T43 to G	+	E62	G50 to T	+
E63	T52 to G	++	E64	G53 to T	++

Figure 17. Cleavage activities of mutated functionally essential nucleotides. Each construct contains only the single point mutation listed. For each construct, all reactions used a final concentration of 0.5 µM. 2x selection buffer was added to start the reaction. All reactions were incubated at room temperature for a prescribed amount of time and quenched with solution containing 1x PAGE loading solution and 5 mM EDTA, pH 8 at 23 ℃. All reactions were completed in duplicate. The cleaved fraction of each reaction was separated by 10% denaturing PAGE and the fraction cleaved of each reaction was determined using ImageQuant software. The cleavage activity of each construct was rated using the index listed.

Figure 17 contains the rated cleavage activities of several constructs, each of which contained a single point mutation of a fully conserved nucleotide. It can be seen that all the fully conserved nucleotides were required for optimal cleavage activity, but to varying degrees. G_{39} and G_{42} were absolutely essential nucleotides; their mutations caused complete loss of activity. The mutation of T_{43}, A_{28}, G_{37} and G_{50} caused a drastic

but not total loss of activity, while the mutation of G_{26}, T_{52} and G_{53} caused a significant loss of activity. The mutation of G_{30} only caused a moderate loss of activity. These results confirm that all the functionally essential nucleotides identified in the reselection were indeed necessary for optimal activity. However, their mutation results in varying degrees of loss of cleavage activity, indicating that some of these nucleotides were more functionally important than others.

5.4   CONCLUSIONS

Complete analysis of 22-18 yielded several conclusions. First, its various loops, stems and junctions comprise a novel secondary structure not seen before in the literature (Mei *et al*., 2003). Second, the functionally essential nucleotides previously identified by S. A. Kandadai were confirmed to be essential to optimal cleavage activity. Third, a thorough study of various elements of 22-18's secondary structure revealed two potential locations to place an aptamer and convert it into a biosensor.
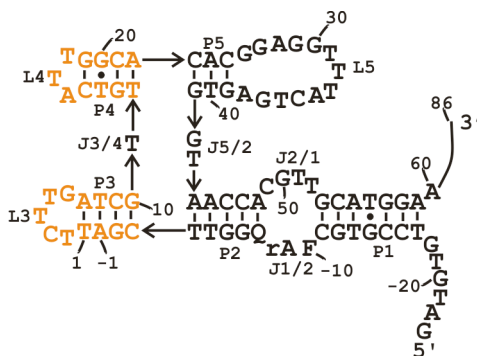
Figure 18. Potential locations on 22-18 for replacement by an aptamer. The nucleotides highlighted in orange, P3-L3 and P4-L4, have been identified as potential locations for aptamer replacement to convert 22-18 into an aptazyme. Just one location could be replaced with an aptamer, or both at once could be replaced by two different aptamers.

P4-L4 and P3-L3, highlighted in orange in Figure 18, could be replaced with an aptamer to yield an allosteric DNAzyme, or aptazyme. These two sites are potential aptamer replacement locations because the removal of P3-L3, as well as the modification of P4, was previously determined to cause the complete loss of cleavage activity of 22-18. This allows the aptamer in the place of those sites to have allosteric control over 22-18's activity, including low background activity in the absence of the target. 22-18 could also potentially be converted into a two-target aptazyme, whereby an aptamer for one target replaces one location, while an aptamer for another target replaces the other location. The

cleavage activity of 22-18 would only be restored in the presence of both targets. P4-L4

and/or P3-L3 could be replaced by an aptamer through several methods, including

rational design, *in vitro* selection and a hybrid approach, covered in detail in Chapter 1.

CHAPTER SIX

REFERENCES

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2008). *Molecular biology of the cell* (5th ed.). New York, NY: Garland Science.

Atsumi, S., Ikawa, Y., Shiraishi, H. & Inoue, T. (2001). Design and development of a catalytic ribonucleoprotein. *EMBO Journal, 20*, 5453-60.

B & M Labs. *Biotools DNA polymerase.* Retrieved 02/22, 2010, from <http://www.biotools.eu/biotools_DNA_poly.html>.

Bagheri, S. & Kashani-Sabet, M. (2004). Ribozymes in the age of molecular therapeutics. *Current Molecular Medicine, 4*, 489-506.

Banci, L., Bertini , I., Gray, H. B., Luchinat, C., Reddig, T., Rosato, A., *et al*. (1997). Solution structure of oxidized horse heart cytochrome c, *Biochemistry, 36*(32), 9867-9877.

Barczyk, K., Kreuter, M., Pryjma, J., Booy, E. P., Maddika, S., Ghavami, S., *et al*. (2005). Serum cytochrome c indicates in vivo apoptosis and can serve as a prognostic marker during cancer therapy. *International Journal of Cancer, 116*(2), 167-173.

Ben-Ari, Z., Schmilovotz-Weiss, H., Belinki, A., Pappo, O., Sulkes, J., Neuman, M. G., *et al*. (2003). Circulating soluble cytochrome c in liver disease as a marker of apoptosis. *Journal of Internal Medicine, 254*(2), 168-175.

Breaker, R. R. (2004). Natural and engineered nucleic acids as tools to explore biology. *Nature, 432*, 838-845.

Certo, M., Moore, V. D. G., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S. A., *et al*. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell, 9*(5), 351-365.

Chinnapen, D. J. F. & Sen, D. (2002). Hemin-stimulated docking of cytochrome *c* to a hemin-DNA aptamer complex. *Biochemistry, 41*, 5202-5212.

Chipuk, J. E. & Green, D. R. (2008). How do BCL-2 proteins induce mitochondrial outer membrane permeabilization? *Trends in Cell Biology, 18*(4), 157-164.

Chiuman, W. & Li, Y. (2007). Simple fluorescent sensors engineered with catalytic DNA 'MgZ' based on a non-classic allosteric design. *PLoS One, 2*, e1224.

Chiuman, W. & Li, Y. (2009). Chapter 5: Fluorescent ribozyme and deoxyribozyme sensors. In Y. Li & Y. Lu (Eds.), *Functional nucleic acids for analytical applications* (pp. 131-153). New York, NY: Springer Science + Business Media.

Cho, E. J., Lee J.-W. & Ellington, A. D. (2009). Applications of aptamers as sensors. *Annual Review of Analytical Chemistry, 2*, 241-264.

Cho, S., Kim, J., Lee, B., Kim, J. & Kim, B. (2005). Bis-aptazyme sensors for hepatitis C virus replicase and helicase without blank signal. *Nucleic Acids Research, 33*, e177.

Doudna, J. A. & Cech, T. R. (2002). The chemical repertoire of natural ribozymes. *Nature, 418*, 222-8.

Fulda, S. (2010). Exploiting mitochondrial apoptosis for the treatment of cancer. *Mitochondrion, 10*(6), 598-603.

Goldstein, J. C., Waterhouse, N. J., Juin, P., Evan, G. I. & Green, D. R. (2000). The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant. *Nature Cell Biology, 2*, 156-162.

Hartig, J. S. & Famulok, M. (2002). Reporter ribozymes for real-time analysis of domain-specific interactions in biomolecules: HIV-1 reverse transcriptase and the primer-template complex. *Angewandte Chemie International Edition, 41*, 4263-6.

Hartig, J. S. & Famulok, M. (2008). Screening of molecular interactions using reporter hammerhead ribozymes. *Methods in Molecular Biology, 429*, 251-63.

Hartig, J. S., Najafi-Shoushtari, S. H., Grüne, I., Yan, A., Ellington, A. D. & Famulok, M. (2002). Protein-dependent ribozymes report molecular interactions in real time. *Nature Biotechnology, 20*, 717-22.

Heiskanen, K. M., Bhat, M. B., Wang, H.-W., Ma, J. & Nieminen, A.-L. (1999). Mitochondrial depolarization accompanies cytochrome *c* Release during apoptosis in PC6 cells. *Journal of Biological Chemistry, 274*, 5654-5658.

Irvine, R. A., Adachi, N., Shibata, D. K., Cassell, G. D., Yu, K., Karanjawala, Z. E., *et al*. (2005). Generation and characterization of endonuclease G null mice. *Molecular and Cellular Biology, 25*(1), 294-302.

Jiang, X. & Wang, X. (2004). Cytochrome C-mediated apoptosis. *Annual Review of Biochemistry, 73*, 87-106.

Jones, R. G. & Thompson, C. B. (2009). Tumor suppressors and cell metabolism: A recipe for cancer growth. *Genes & Development, 23*, 537-548.

Kagan, V. E., Tyurin, V. A., Jiang, J., Tyurina, Y. Y., Ritov, V. B., Amoscato, A. A., *et al*. (2005). Cytochrome c acts as a cardiolipin oxygenase required for release of proapoptotic factors. *Nature Chemical Biology, 1*, 223-232.

Kandadai, S. A. & Li, Y. (2005). Characterization of a catalytically efficient acidic RNA-cleaving deoxyribozyme. *Nucleic Acids Research, 33*(22), 7164-7175.

Koizumi, M., Soukup, G. A., Kerr, J. N. Q. & Breaker, R. R. (1999). Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nature Structural Biology, 6*(11), 1062-1071.

Kroemer, G. (1999). Cytochrome c. *Current Biology, 9*(13), R468-R468.

Lau, P. S., Coombes, B. K. & Li, Y. (2010). A general approach to the construction of structure-switching reporters from RNA aptamers. *Angewandte Chemie International Edition, 49*(43), 7938-7942.

Leber, B., Lin, J. & Andrews, D. W. (2010). Still embedded together binding to membranes regulates bcl-2 protein interactions. *Oncogene, 29*, 5221-5230.

Leber, B., Lin, J. & Andrews, D. W. (2007). Embedded together: The life and death consequences of interaction of the bcl-2 family with membranes. *Apoptosis, 12*(5), 897-911.

Ledgerwood, E. C. & Morison, I. M. (2009). Targeting the apoptosome for cancer therapy. *Clinical Cancer Research, 15*, 420-424.

Levy, M. & Ellington, A. D. (2002). ATP-dependent allosteric DNA enzymes. *Chemistry & Biology, 9*, 417-26.

Li, Y. & Breaker, R. R. (1999). Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *Journal of the American Chemical Society, 121*, 5364-5372.

Li, Y. & Breaker, R. R. (2001). *In vitro* selection of kinase and ligase deoxyribozymes. *Methods, 23*(2), 179-90.

Liu, J., Brown, A. K., Meng, X., Cropek, D. M., Istok, J. D., Watson, D. B., *et al*. (2007). A catalytic beacon sensor for uranium with parts-per-trillion sensitivity and millionfold selectivity. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 2056-2061.

Liu, J. & Lu, Y. (2007). Rational design of "turn-on" allosteric DNAzyme catalytic beacons for aqueous mercury ions with ultrahigh sensitivity and selectivity. *Angewandte Chemie International Edition, 46*(40), 7587-90.

Mei, S. H., Liu, Z., Brennan, J. D. & Li, Y. (2003). An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling. *Journal of the American Chemical Society, 125*, 412-420.

Najafi-Shoushtari, S. H. & Famulok, M. (2007). DNA aptamer-mediated regulation of the hairpin ribozyme by human α-thrombin. *Blood Cells, Molecules, and Diseases, 38*, 19-24.

Navani, N. K. & Li, Y. (2006). Nucleic acid aptamers and enzymes as sensors. *Current Opinion in Chemical Biology, 10*(3), 272-281.

Navani, N. K., Mok, W. K. & Li, Y. (2009). Chapter 22: *In vitro* selection of protein-binding DNA aptamers as ligands for biosensing applications. In A. Rasooly & K. E.

Herold (Eds.), *Methods in molecular biology: Biosensors and biodetection* (pp. 399-415). Totowa, NJ: Humana Press.

Ow, Y.-L. P., Green, D. R., Hao, Z. & Mak, T. W. (2008). Cytochrome c: Functions beyond respiration. *Nature Reviews Molecular Cell Biology, 9*, 532-542.

Robertson, M. P. & Ellington, A. D. (2000). Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Research, 28*, 1751-1759.

Robertson, M. P. & Ellington, A. D. (2001). *In vitro* selection of nucleoprotein enzymes. *Nature Biotechnology, 19*, 650-5.

Robertson, M. P., Knudsen, S. M. & Ellington, A. D. (2004). *In vitro* selection of ribozymes dependent on peptides for activity. *RNA, 10*, 114-27.

Roucou, X., Prescott, M., Devenish, R. J. & Nagley, P. (2000). A cytochrome c–GFP fusion is not released from mitochondria into the cytoplasm upon expression of bax in yeast cells. *FEBS Letters, 471*(2-3), 235-239.

Schäfer, P., Scholz, S. R., Gimadutdinow, O., Cymerman, I. A., Bujnicki, J. M., Ruiz-Carrillo, A., *et al*. (2004). Structural and functional characterization of mitochondrial EndoG, a sugar non-specific nuclease which plays an important role during apoptosis. *Journal of Molecular Biology, 338*, 217-228.

Schlosser, K. & Li, Y. (2005). Diverse evolutionary trajectories characterize a community of RNA-cleaving deoxyribozymes: A case study into the population dynamics of *in vitro* selection. *Journal of Molecular Evolution, 61*, 192-206.

Soukup, G. A. & Breaker, R. R. (1999). Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 3584-3589.

Tang, J. & Breaker, R. R. (1997). Rational design of allosteric ribozymes. *Chemistry & Biology, 4*, 453-9.

Unkila, M., McColl, K. S., Thomensius, M. J., Heiskanen, K. & Distelhorst, C. W. (2001). Unreliability of the cytochrome *c*-enhanced green fluorescent fusion protein as a marker of cytochrome *c* Release in cells that overexpress bcl-2. *Journal of Biological Chemistry, 276*, 39132-39137.

Vaish, N. K., Dong, F., Andrews, L., Schweppe, R. E., Ahn, N. G., Blatt, L., *et al*. (2002). Monitoring post-translational modification of proteins with allosteric ribozymes. *Nature Biotechnology, 20*, 810-5.

Vartanian, J.-P., Henry, M. & Wain-Hobson, S. (1996). Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Research, 24*, 2627-5.

Voet, D. & Voet, J. G. (2004). *Biochemistry* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Wang, D. Y. & Sen, D. (2002). Rationally designed allosteric variants of hammerhead ribozymes responsive to the HIV-1 tat protein. *Combinatorial Chemistry & High Throughput Screening, 5*, 301-12.

Wang, W., Billen, L. P. & Li, Y. (2002). Sequence diversity, metal specificity, and catalytic proficiency of metal-dependent phosphorylating DNA enzymes. *Chemistry & Biology, 9*(4), 507-517.

Yamazaki, S., Tan, L., Mayer, G., Hartig , J. S., Song, J., Reuter, S., *et al*. (2007). Aptamer displacement identifies alternative small-molecule target sites that escape viral resistance. *Chemistry & Biology, 14*, 804-12.

Youle, R. J. & Strasser, A. The BCL-2 protein family: Opposing activities that mediate cell death. *Nature Reviews Molecular Cell Biology, 9*, 47-59.

Zucker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415.

CHAPTER SEVEN

APPENDICES

7.1   APPENDIX I: FIRST *IN VITRO* SELECTION SEQUENCES

All sequences analyzed from the first *in vitro* selection. Five sequences of the original library were sequenced. Thirty-nine sequences from the enriched library after Round 20 without the addition of Proteinase K and forty-five sequences with the addition of Proteinase K were analyzed. The first occurrence of sequence C1 is R20C1. The first occurrence of sequence C1-2PM is R20C6.   The sequences are colour-coded as follows: the primer binding sequences are in black, MgZ is in purple and the randomized region is in yellow.

1st *in vitro* selection: Original Library Sequences (5' - 3')

CACGGATCCTGACAAGGAACCAGGTCGGGCGCGGAGAGACGAAAGAGTAAGGCATTCAGACATACATTGCGGGAGGCTATGCTAGGCAGCTCCGTCCG

CACGGATCCTGACAAGGAACCAGGTCGGGCCGCTGACGGGCTCAGGGCGGGTTGGATTGTGGGGTAATAGGGAGGCTATGCTAGGCAGCTCCGTCCG

CACGGATCCTGACAAGGAACCAGGTCGGGGCCGAAAATGATACCTCTACAGGGTGAAACAGCGGGTCTGGAGGAGGCTATGCTAGGCAGCTCCGTCCG

CACGGATCCTGACAAGGAACCAGGTCGGGCCATAGGGATTGGGGTCTGTCTTTGGAAGGCGGGGGGAAATCGGGAGGCTATGCTAGGCAGCTCCGTCCG

CACGGATCCTGACAAGGAACCAGGTCGGGCCGAGCGATGGTGGGCGTTGGAGATGGAGGAGGTGCGAGGAGGGAGGCTATGCTAGGCAGCTCCGTCCG

1st *in vitro* selection: Round 20, Cyt c Sequence (5' - 3')

Name	Sequence (5' - 3')
R20C1	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C2	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C3	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C4	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C5	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C6	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C7	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C8	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C9	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C10	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C11	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C12	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACAATCGAACGAGGCACCCAAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C13	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG

1st *in vitro* selection: Round 20, Cyt c

Name	Sequence (5' - 3')
R20C14	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCCATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C15	CACGGATCCTGCACNTGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C16	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C17	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGAGCGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C18	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C19	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C20	CACGGATCCTGCCNTNGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C21	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C21	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C22	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C23	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTANGCAGCTCCGTCCG
R20C24	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C25	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C26	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C27	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C28	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGCTGAGTAACGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C29	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C30	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C31	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C32	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C33	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAATGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C34	CACGGATCCTGACAAGTAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAAAGGGCTCCGTCCG
R20C35	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAGCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C36	CACGGATCCTGACANGGAACCAGGTCGGGGCCATAAGGGCCGCTGAGTAACGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C37	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C38	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20C39	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGCCGCCGAGNAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG

1st *in vitro* selection: Round 20, Cyt c + Proteinase K

Sequence (5' - 3')

Name	Sequence (5' - 3')
R20CP1	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP2	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAAGCGGCTCCGTCCG
R20CP3	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP4	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCGCCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP5	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP6	CACGGATCCTGACAAGAAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTANGCAGCTCCGTCCG
R20CP7	CACGGATCCTGACANGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP8	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCNGCTCCGTCCG
R20CP9	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP10	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP11	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGCTGAGTAACGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP12	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP13	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP14	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP15	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGCTGAGTAATGATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP16	CACGGATCCTGACAAGAAACCAGTCGGGGCCATAAGGGCGCCGAGTAGCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP17	CACGGATCCTGACANGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP18	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP19	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGCAACAATCGAACGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP20	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP21	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP22	CACGGATCCTGACAAGGAACCAGTCGGGGCCATAAGGGCGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG

1st *in vitro* selection: Round 20, Cyt c + Proteinase K

Name	Sequence (5' - 3')
R20CP23	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP24	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP25	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP26	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP27	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP28	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAATGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP29	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP30	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP31	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP32	CACGGATCCTGACANGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP33	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP34	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGCTGAGTAACGATCGAATGAGGCACCCGAATGGGAGGCTATGCTANGCAGCTCCGTCCG
R20CP35	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP36	CACGGATCCTGACAAGAAACCAGGTCGGGGCCATAAGGGGCCGAGTGACGATCGAACTAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP37	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP38	CACGGATCCTGACANGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP39	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP40	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAGCGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP41	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP42	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAATGATCGAACGAGGCACCCGAGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP43	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAATGATCGAACGAGGCACCCGAGTGGGAGGCTACNNCTAANCGGGCTCCGTCCG
R20CP44	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R20CP45	CACGGATCCTGACAAGGAACCAGGTCGGGGCCATAAGGGGCCGAGTAACGATCGATCGAATGAGGCACCCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG

7.2   APPENDIX II: SECOND *IN VITRO* SELECTION SEQUENCES

All sequences analyzed from the second *in vitro* selection. Thirty-one sequences were analyzed from the enriched library after Round 6. Twenty-five sequences from the enriched library after Round 12-1 were analyzed. Twenty sequences were analyzed from the enriched library after Round 9-2. The sequences are colour-coded as follows: the primer binding sequences are in black, MgZ is in purple and the randomized region is in yellow. For the modified sequences (with the 'mod' prefix), the modified bases are in blue.

2nd *in vitro* selection: Round 6

Name	Sequence (5' – 3')
R601	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGAGCGCGGGAAAGAAGTAACAAGACTTTGGGGGCGATGGGGAGGCTTTGCTAGGCAGCTCCGTCCG
R602	GAACAGTCGTACGCACTAACCAGGTCGGGGCCGGGAGAAGCACCGGGGTCAACGTACTGGGGCTCCTCTTCGGGAGGCTATGCTAGGCAGCTCCGTCCG
R603	GAACAGTCGTACGCACTAACCAGGTCGGGGCCGCGGGAGCGTTTCGCCGTATATGGCCATAATTGCGCACACGGGAGGCTATGCTAGGCAGCTCCGTCCG
R604	GAACAGTCGTTCGCACGAACCAGGTCGGGGCCAATTCACGAGACAGACGTAGGGTGTGGGGGCTAGAATTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R605	GAACAGTCGTACGCACGAACCAGGTCGGGGCCTGAGCATGGGCCAGGTGGAACCCACCTGGCTGCACGAAAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R606	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGTGCACGCCGCTCTCGGCTATGGAATAGCGAGGGAGCACGGGAGGCTAAGCTAGGCAGCTCCGTCCG
R607	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGCTATGAATCAGAGTCCAGGCACAACCTGATGGAAAAGAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R608	GAACAGTCGTACGCACGAACCAGGTCGGGGCCCGCGGTGTTGTATTAGCGATAGTGGCCATGCTAACACCAGGGGAGGCTATGCTAGGCAGCTCCGTCCG
R609	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGATACCAGTGCAAGAATTCAACCACACTGGTCAGATAATTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R610	GAACAGTCGTACGCACGAACCAGGTCGGGGCCATCTCACGGGTAACACGCATGAGACCGAGTGCCACATGATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R611	GAACAGTCGTACGCACGAACCAGGTCGGGGCCCGGAAAATCCAGATACGTACGGATGGCGTATAAACAGCAGGGGAGGCTATGCTAGGCAGCTCCGTCCG
R612	GAACAGTCGTACGCACAAACCAGGTCGGGGCCAACACGAAATAGAAGTACTGGGAACTGCCGATAGTTGTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R613	GAACAGTCGTACGCACGAACCAGGTCGGGGCCTACCCATAATGTCCACAAGATAGGAGGACATGTACGGGTAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R614	GAACAGTCGTACGCACGAACCAGGTCGGGGCCACATCATGAAGGCATGCTTTACACGTAAATGCGGAGTAGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R615	GAACAGTCGTACGCACGAACCAGGTCGGGGCCTGCGTGTGGGAGCGCGGTCGGAAAACGTGCACACGTAGGGAGGCTAAGCTAGGCAGCTCCGTCCG
R616	GAACAGTCGTACGCACGAACCAGGTCGGGGCCTTAGTTTGAGGGGTGGGTGGCAAAGTTCCCAAAACGTGAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R617	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGGGTGGCACAGGGACTGAGTCTCGCAGGCAGTGCCAATTCGGGAGGCTATGCTAGGCAGCTCCGTCCG
R618	GAACAGNCNTTTGCACGAAACAGGTCGGGGCCACCTCGTGGAAAGGCTACAGTCCGGTCGCCTCAGAGAGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R619	GAACAGTCGNACGNACGAACCAGGTCGGGGCCTGGCGCACTCGGACGCCTGGACATTGGCGAGTAGAAGTCAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R620	GAACAGTCGTACGCACATACCAGGTCGGGGCCGGTTGGCGAGAGGGTGAAAAAGTCCACATGAAGCGAGACCGGGAGGCTATGCTAGGCAGCTCCGTCCG
R621	GAACAGTCGTACGCACGAAACAGGTCGGGGCCAACGCAAGTATGGAGTCATCAGTGGACAGGCTTGCGCCTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R622	GAACAGTCGTACGCACAAACCAGGTCGGGGCCAAGGCTGAGATATACGGCCGTCAGCGTACTATACGGCCCATGGGAGGCTATGCTAGGCAGCTCCGTCCG

2nd *in vitro* selection: Round 6

Name	Sequence (5' - 3')
R623	GAACAGTCGTACGCACGAACCAGTCGGGGCCTGAGCGGGGGCCAACGGCAAGGAGGCCCTCCAAACTCAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R624	GAACAGTCGTACGCNCGAATCAGTCGGGGCCGCCGGAGGCGCGCTTGCAACACAGTCAAGAGCCCCGTGGATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R625	GAACAGTCGTACGCACGAACCAGTCGGGCCCTTGGCAAAAGGGGCTCACGTTGTACCATGAGCCGCAAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R626	GAACAGTCGTACGCACGAACCAGTCGGGGCCAGAGGAGATGGAAGGAGTAAAGCGAATGAGAAAGACAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R627	GAACAGTCGTACGCACGAACCAGTCGGGGCCAGACGTGATAAGACAGACACTGGAGGACTAAAACGGCTCTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R628	GAACAGTCGTACGCACGAACCAGTCGGGGCCACCAGAACCCCGCTTAGGCGCCCAGGGGAACTTTCTGGTAGGGAGGCTATGCTAGGCAGCTCCGTCCG
R629	GAACAGTCGTACGCACGAACCAGTCGGGGCCAAGGGGTGTTAGCTCACGCGTGTGTAATTTGAGCGCGTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R630	GAACAGTCGTACGCACGAACCAGTCGGGGCCATTCCGGAGGGGCATGAGATATGATCCGCAGCGGGGGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R631	GAACAGTCGTACGCACGAACCAGTCGGGGCCTAAGGGTTGGGGATAGCGTGCGATGAGCCAAACCTTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
modR606	GAACAGTCGTACGCACGAACCAGTCGGGGCCGTGCTCCCCGCTCTCGGCTATGGAATAGCGAGGGGAGCACGGGAGGCTAAGCTAGGCAGCTCCGTCCG
modR617	GAACAGTCGTACGCACGAACCAGTCGGGGCCGAATTGGACAGGGACTGAGTCTCGCAGGCAGTGCCAATTCGGGAGGCTATGCTAGGCAGCTCCGTCCG
modR626	GAACAGTCGTACGCACGAACCAGTCGGGGCCATTGTCTATGAAGGAGGTAAAGGCAATGGAGAAAGACAATGGGAGGCTATGCTAGGCAGCTCCGTCCG

2nd *in vitro* selection: Round 12-1

Sequence (5' - 3')

Name	Sequence (5' - 3')
R1201	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1202	GAACAGTCGTACGCCACTAACCAGGTCGGGGCCACACTTGCGTAGTGTGCGCCTGGGAAGCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG
R1203	GAACAGTCGTTCGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGCTTACTGGGAGGCTATGCTAGGCAGCTCCGNCCG
R1204	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1205	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGAGGCGGCTTACTGGGAGGCTATGCTAGGCANNGCCGTCCG
R1206	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1207	GAACAGTCGTACGCACGAACCAGGAGGGCCCATGGTAGATGCCGCTAGGTTCAATGCGCCGGGATCGAATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1208	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCNCCGTCCG
R1209	GAACAGTCGTACGCACGAACCAGGTGGGGCCCGCATAGCAAAGAACGCTATATTAGATATGCCGTCCATGCGGAGAGGCTATGCTAGGCAGCTCCGTCCG
R1210	GAACAGTCGTACGCACGAACCAGGTCGGGGCCACACTTGCGTAGTGTGCGCCTGGGGAGCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG
R1211	GAACAGTCGTACGCCACTAACCAGGTCGGGGCCACACTTGCGTAGTGTGCGCCTGGGGAGCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG
R1213	GAACAGTCGTACGCACGAACCAGGTCGGGGCCCAGCGCGCGGAAAGCGAGCATACCACTAGCGACGCGTTTGGGAGGCTATGCTAGGCAGCTCCGTCCGCG
R1214	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1215	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1216	GAACAGTCGTACGCACGAACCAGGAGGGGCCCAAGGAGAAAGCTATCGAAAGTGGTAGGCCATGCGTCCTTGGGAGGCTATGTGCGTCCTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1217	GAACAGTCGTACGCCACTAACCAGGTCGGGGCCACACTTGCGTAGTGTGCGCCTGGGGAGCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG
R1218	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1219	GAACAGTCGTACGCACGAACCAGGTGGGGCCCCATAAGCAGTAGATTCACACCGAAAATGCAGCTGGTTATGGGAGGCTACGCTAGGCAGCTCCGTCCG
R1220	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTNACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1221	GAACAGTCGTNCGCACTAACCAGGTCGGGGCCACACTTGCGTAGTGTGCGCCTGGGAACCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG
R1222	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1223	GAACAGTCGTACGCACGAACCAGGTCGGGGCCACACAGTCCGAGCGTAAGAAGGGAGGCAGCTTGCGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1224	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAGTTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R1225	GAACAGTCGTACGCACGAACCAGGAGAGGGCCCAAGGAGAAAGCTATCGGTATGGTAGGCCATGTGCGTCCTTGGGAGGCTATGCTAGGCAGCTCCGTCCG
modR1201	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGTAAGCATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
modR1202	GAACAGTCGTACGCCACTAACCAGGTCGGGGCCACCCTTTCGTAGTGTGCGCCTGGGAAGCGAAGTAAAGGTGGAGGCTAAGCTAGGCAGCTCCGTCCG

2nd *in vitro* selection: Round 9-2

Sequence (5' - 3')

Name	Sequence (5' - 3')
R901	GAACAGTCGTACGCACGAAGCATGTAGGAGCCCGGTATAGACCGTTTGCATGCACATATAGCAGAAAGGGAATTGAGGGTAGGCTAGCAGCTCCGTCCG
R902	GAACAGTCGTACGCACGAAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R903	GAACAGTCGTACGCACCAATCAGGTTGTGGCCTGTATTAAGGCGATCGATAGTATAATCTAACTTAGGGGGTGGGAGGTAGTTCAATGCAGCTCCGTCCG
R904	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAAATTCCGACGAGCAATTAAATGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R905	GAACAGTCGTACGCACGAAGCGAAAGGCCCACAAAATACCGTAGTAATACTCGAGGGGCTCGAGAGAATGTTGCAGGCTATGCTTACCAGCTCCGTCCG
R906	GAACAGTCGTNCGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R907	GAACAGTCGTACGCACGAACTAGGGCGGGGAAGTCACTAGGAAACACGAAAGTGATGGGTGTGACGGGATTCCGGTTATGCTGCCAGCNCCGTCCG
R908	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGAGAAACGACGAAGAAGCGTAAGGGAGAGATGTGATGGGGAGGCTATGCTAGGCAGCTCCGTCCG
R909	GAACAGTCGTACGCACAAACCTGGTCGGGGCCAGTAGAGATCATAATAGACACAGAACGTCATTCATGGCATCTGGGAGGCTAGGCTAGGCAGCTCCGTCCG
R910	GAACAGTCGTACGCACTAACCAGGTCGGGGCCACACTTGCGTAGTGTGNGCCTGGAAGCGAAGTAAAGGTGGGAGGCTAAGCTAGGCAGCTCCGTCCG
R911	GAACAGTCGTACGCACCAACCAGGGCGAGGTAGCCAAAGCGCAACACCCGGCCATTGCCTAGATGGTGTTCTGGGATTGTACTCGAGGCAGCTCCGTCCG
R912	GAACAGTCGTACGCACCAACCAGTCCGCAGCTTAGGGGCCAAGAAGGGGAGTGAGGATGAGACAGGAAAAGCGCGGCTCAGGTAGGCAGCTCCGTCCG
R913	GAACAGTCGTACGCACGAACCAAGCCGTGGCCTCCGGCAGTGACAATTCCTTGTTCAGGCAAAGTTGCGGGGACGCTCTATGATCCAGCTCCGTCCG
R914	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R915	GAACAGTCGTACGCACTTACAATGTTGGGTTTAACTTAACTGGGCAGATAGCGCTACGCGAATTGGAGAGAGAAACAGTGCTAGGCAGCTCCGTCCG
R916	GAACAGTCGTACGCACTAACCAGGATGCGGCACTAGGGATCGCGGGCACGCGACACCGAAAGGCGGAAAGACGGAAGCGTCTGTGATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R917	GAACAGTCGTACGCACGAACCAGGTCGGGGCCACACGACACGAAGGCGGAAAAGACGGAAGCGTCTGTGATGGGAGGCTATGCTAGGCAGCTCCGTCCG
R918	GAACAGTCGTNCGCACTATCCAGCTCGGTGTAGCGGAAAGCCCGGAAACCAATACGGTGTGGAAGAGAGTGGGTAGGCTATTCTTGGCAGCTCCGTCCG
R919	GAACAGTCGTACGCACGAACCAGGTCGGGGCCAGCAAGAATTCCGACGAGCAATTAAAAGAGGCGGCTTACTGGGAGGCTATGCTAGGCANCTCCGTCCG
R920	GAACAGTCGTACGCACGTACCAGGTCGGGGCCGCCATGGAGACGACAAGCCACGGAAGAGACGTACATGGGGGAGGCTATGCTAGGCAGCTCCGTCCG
R921	GAACAGTCGTACGCACGAACCAGGTCGGGGCCACACAGCTCCGAGCGTAAGAAGAAGGGAGGCAGCTTGCGTGGGAGGCTATGCTAGGCAGCTCCGTCCG
R922	GAACAGTCGTACGCACGAACCAGGTCGGGGCCGTTGGTTGGTCGGCTCTCTTGTTGAATCAGAAAAGCAACGGGAGGCTATGCTAGGCAGCTCCGTCCG

86

7.3   APPENDIX III: *IN VITRO* SELECTION SIMULATION EXAMPLE

All *in vitro* selection simulations were performed in the same manner as the
example to be worked through here. For this example, two negative selection steps per
round and a 500-fold increase in cleavage activity of group D are assumed. To start Y,
the fraction cleaved, at 10 minutes (the positive selection) and 600 minutes (the negative
selection) must be determined for each group of sequences. This is determined with the
one-phase exponential equation

$$Y = Y_{MAX}\left(1 - e^{-k_{obs}t}\right)$$

For all groups, $Y_{MAX} = 0.6$. For group A, $k_{obs} = 0.1/\text{min}$

At t = 10 minutes

$$Y = 0.6\left(1 - e^{(-(0.1 \times 10))}\right) = 0.397$$

At t = 600 minutes

$$Y = 0.6\left(1 - e^{(-(0.1 \times 600))}\right) = 0.6$$

The fraction cleaved of all other groups is calculated this way. Note that for the negative
selection group D has $k_{obs} = 0.001/\text{min}$, while for the positive selection group D has $k_{obs} =$
0.5/min. These are all the calculated fraction cleaved values:

at t=10 mins (+ve selection)				
Group A	Group B	Group C	Group D	Group E
0.379272335	0.057097549	0.0059701	0.595957232	1.8E-06

at t=600 mins (-ve selection)				
Group A	Group B	Group C	Group D	Group E
0.6	0.598512749	0.270713018	0.270713018	0.00010799

Next, for performing the simulations it is easiest to work with the initial ratios of
each group within the starting library as a percentage. Group A occupies

$$\frac{50{,}000{,}000}{50{,}000{,}000 + 50{,}000{,}000 + 50{,}000{,}000 + 1} \times 0.01\% = 0.0033\%$$

The initial percentages of each group in the starting library:

Initial %				
A	B	C	D	E
0.003333333	0.003333333	0.003333333	6.66667E-11	99.99

Now performing the simulation can begin. The number of sequences cleaved in each group during each negative and positive selection step is determined. For the first negative selection step of each round the number of sequences cleaved is calculated as follows:

$$F \times NS \times Y$$

where F is the fraction of the library pool that the group occupies and NS is the total number of sequences in the library pool. For Round 1, NS $= 1 \times 10^{14}$. In Round 1, group A's F $= \frac{0.0033\%}{100} = 0.000033$ (it must be a fraction, not a percentage). So the number of group A sequences cleaved during the first negative selection of Round 1 is

$$(0.000033)(1 \times 10^{14})(0.6) = 1999999987 \text{ sequences}$$

The number of cleaved sequences for all the other groups is calculated in exactly the same way, and looks like this:

# Sequences Cleaved after -ve selection				
A	B	C	D	E
1999999987	1995042482	902376721.8	18.04753444	10797948156

For all subsequent selection steps of each round, the number of cleaved sequences is calculated as follows:

$$\left((F \times NS) - NS_C\right) \times Y$$

88

where NS_C is the number of sequences cleaved during previous selection steps. The number of cleaved sequences in each group must be removed from the total number of sequences in each group since they are already cleaved and cannot cleave again. So the number of sequences cleaved in group A during the second negative selection step is

$$\left((0.000033)\left(1 \times 10^{14}\right) - 1999999987\right)(0.6) = 799999994.7 \text{ sequences}$$

The number of cleaved sequences for all the other groups after the second negative selection step looks like this:

# Sequences cleaved after 2nd -ve selection				
A	B	C	D	E
799999994.7	800984122.5	658091595.8	13.16183192	10796782082

The positive selection step is calculated using the exact same equation as was just used. However, the number of cleaved sequences of both negative selection steps is subtracted from the total number of sequences and the positive selection Y value for each group is used. The number of group A sequences cleaved during the positive selection is

$$\left((0.000033)\left(1 \times 10^{14}\right) - 1999999987 - 799999994.7\right)(0.379) = 202278577.5 \text{ sequences}$$

The number of cleaved sequences for all the other groups after the positive selection looks like this:

# Sequences Cleaved after +ve selection				
A	B	C	D	E
202278577.5	30678896.09	10584180.9	21.1310343	179942860

Now the fraction that each group occupies within the total number of cleaved sequences of the positive selection is calculated, since the cleaved sequences in the positive selection are what will be carried into the next round of selection. It is simply calculated by the number of cleaved sequences of a group divided by the total number of cleaved sequences

within the positive selection. The fraction of group A cleaved sequences within the

positive selection is

$$\frac{202278577.5}{202278577.5 + 30678896.09 + 10584180.9 + 21.13 + 179942860} = 0.478$$

The fraction of each group within the cleaved sequences of the positive selection looks

like this:

	Fraction after +ve selection			
A	B	C	D	E
0.477652808	0.072443958	0.02499308	4.9898E-08	0.42491011

These fractions are then used for calculating the number of sequences cleaved in the next

round of selection. The calculations are exactly the same as Round 1; only the total

number of sequences for all rounds except for Round 1 is 1 x 10^{13}. The first negative

selection step of Round 2 looks like this:

	# Sequences Cleaved after -ve selection				
Selection Round	A	B	C	D	E
1	1999999987	1995042482	902376721.8	18.04753444	10797948156
2	2.86592E+12	4.33586E+11	67659508400	135080.4007	458861617.9

All other simulations are performed in exactly the same manner as this example.

The only differences between simulations are the number of negative selection steps per

round and the Y value of group D for the positive selection step (the increase in k_{obs} of

Cyt c-dependent sequences in the presence of Cyt c is a variable).

Note: Doing the simulations in Microsoft Excel or other spreadsheet software allows one

to simulate many, many rounds of *in vitro* selection very easily. Once the calculations for

Round 2 are inputted, it is a simple case of cutting and pasting, or dragging to copy, to

simulate more rounds of selection.