

**IMPROVING DNAZYME CATALYSIS THROUGH SYNTHETICALLY
MODIFIED DNAZYMES AND PROBING DNA POLYMERASE FUNCTION TO
IMPROVE SELECTION METHODOLOGY**

by

CHRISTOPHER JOHN HIPOLITO

B.Sc., The University of British Columbia, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

October 2010

© Christopher John Hipolito, 2010

Abstract

DNA enzymes (DNAzymes) are part of a growing field of nucleic acid-based catalysts. Unlike ribozymes, DNAzymes have no apparent precedence in nature to date and can only be discovered through *in vitro* selection. These selections can accommodate both natural and chemically-functionalized nucleotides. Functionalities provide DNAzymes with enhanced novel function or catalytic rates normally not attainable using natural nucleotides. The overall purpose of this thesis is to evaluate and improve the utility of modified nucleotides as components in a DNAzyme selection. This dissertation discusses an updated synthesis of the functionalized 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate, dA^{ime}-TP, that was based on the synthesis used to produce the phosphoramidite analog. dA^{ime}-TP bears the imidazole which was crucial in the discovery of several divalent metal cation-independent DNAzymes. A DNAzyme selection using a derivative nucleoside triphosphate, 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine triphosphate, investigated shortening the linker of the appended imidazole group. The most efficient clone, Dz20-49, was determined to have a catalytic rate of $3.5 \pm 0.4 \cdot 10^{-3} \text{ min}^{-1}$. In the context of eight different DNAzyme selections, the replication of modified DNA is examined, and it was found that DNA templates modified with 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine are poorly amplified compared to an oligonucleotide template of the same sequence modified with 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine. Also examined was the use of 5'-modified dUTP derivatives as substrates in PCR. The resulting doubly-modified dsDNA was used for restriction enzyme digestions and cloning. Restriction sites containing doubly-modified dsDNA were found to be resistant to restriction enzyme digestion. The doubly-modified amplicons were ligated into a vector and transfected into cells. Plasmids copied from the modified dsDNA were sequenced. Fidelity appeared to be maintained through PCR and cell-mediated replication. Due to the limitations of incorporation and read-through of modified nucleotides, steps were taken towards the directed evolution of *Thermus aquaticus* DNA polymerase I, commonly referred to as Taq, for improved incorporation and read-through of modified nucleotides. Short patch compartmentalized self-replication (spCSR) was chosen for the directed evolution. Three unnatural

nucleoside triphosphates targeted for use in the polymerase evolution included 1-(2-deoxy-2-fluoro- β -D-arabinofuranosyl)thymine 5'-triphosphate, 5-aminoallyl-2'-deoxycytidine triphosphate and 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine triphosphate.

Preface

The work presented in this thesis is the result of work performed by the author and several scientific collaborations. Contributions from each scientist and the resulting publications are listed in detail below.

The work in Chapter 3 has been accepted as: Hipolito, C. J., Hollenstein, M., Lam, C. H., and Perrin, D. M. The Interface of Sequence Space and Chemical Space in Combinatorial Selection of Protein-Inspired Modified DNAs: Effects of shortening functional linker length in a DNase selection utilizing 8-imidazolyl modified deoxyadenosine. The syntheses of the modified nucleotides (dA^{imm}TP, dA^{imp}TP, dU^{ga}TP) were performed by Curtis Lam. Dr. Marcel Hollenstein provided invaluable insight into the realm of *in vitro* selection. All other work in this chapter was performed by the author.

The cloning work described in Chapter 4 resulted in multiple publications. The mercury sensor DNase selection, Dz10-13 selection, was published in Hollenstein, M., Hipolito, C. H., Lam, C. H., Dietrich, D., and Perrin, D. M. A highly selective DNase sensor for mercuric ions. *Angew. Chem. Int. Ed.* 2008. The Dz9-86 selection was published in Hollenstein, M., Hipolito, C. H., Lam, C. H., and Perrin, D. M. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M²⁺). *Nucleic Acids Res.* 2009. The Dz10-66 selection was published in Hollenstein, M., Hipolito, C. H., Lam, C. H., and Perrin, D. M. A DNase with Three Protein-Like Functional Groups: Enhancing Catalytic Efficiency of M²⁺-Independent RNA Cleavage. 2009. The studies on the amplification of the modified templates were also included in the manuscript to be published as: Hipolito, C. J., Hollenstein, M., Lam, C. H., and Perrin, D. M. The Interface of Sequence Space and Chemical Space in Combinatorial Selection of Protein-Inspired Modified DNases: Effects of shortening functional linker length in a DNase selection utilizing 8-imidazolyl modified deoxyadenosine. All synthesized modified nucleotides described in Chapter 4 were produced by Curtis Lam. The ATP sensor DNase and the stachyose sensor DNase selections were performed by Curtis Lam. The mercury sensor DNase selection and the RNaseA mimic DNase selections were

performed by Dr. Marcel Hollenstein. Cloning and sequencing for these selections were performed by the author. All other experiments presented in this chapter were performed by the author.

Chapter 5 includes experiments that have not been previously published and are integral for a future research article. The experiments were performed exclusively by the author. Although their work is not explicitly presented in this chapter, Mr. Jefferson Chan, Dr. David Kwan, and Mr. Sherman Farahani (during their undergraduate program) all helped in some aspect toward the final goal of the work in this chapter.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents	vi
List of Tables	xii
List of Figures.....	xiv
List of Abbreviations	xxvii
Acknowledgements	xxx
Dedication	xxxi
Chapter 1: Nucleic Acids, Nucleic Acid Enzymes and <i>Thermus aquaticus</i> DNA	
Polymerase I	1
1.1 Nucleic acids	2
1.1.1 DNA	3
1.1.2 RNA	4
1.1.3 Ribonucleic acid enzymes and the RNA world	7
1.1.4 Cleavage of RNA: applications and optimizations	8
1.1.4.1 RNA degradation by <i>S. mansoni</i> hammerhead ribozyme	10
1.1.4.2 RNA degradation by RNaseA.....	11
1.2 <i>In vitro</i> selection of nucleic acid enzymes	13
1.2.1 SELEX and generalized combinatorial selection	14
1.2.1.1 <i>In vitro</i> selection of ribozymes	16
1.2.1.2 <i>In vitro</i> selection of DNAzymes	17
1.2.2 Ribophosphodiester bond cleavage by DNAzymes.....	18
1.2.3 Divalent metal cation-independent cleavage	20
1.2.4 Ribophosphodiester bond cleavage by modified DNAzymes	22
1.2.5 Expanding the chemical landscape	25
1.3 Enzymic recognition of modified DNA.....	26
1.3.1 Incorporation of modified nucleotides.....	27
1.3.2 Transcription of a modified DNA template	28
1.3.3 Amplification of doubly-modified dsDNA.....	29
1.3.4 Restriction enzyme digestion of modified dsDNA	32
1.4 Evolved polymerases	33
1.4.1 <i>Thermus aquaticus</i> DNA polymerase I.....	33

1.4.2 Evolution of Taq and Stoffel Fragment	36
1.4.2.1 Complementation	36
1.4.2.2 Phage display	39
1.4.2.3 Compartmentalized self-replication	41
1.5 Thesis objectives	45
Chapter 2: Chemically Synthesized Modified Nucleobase Analog	47
2.1 Introduction	48
2.2 Objective of this work	50
2.3 Materials and methods	50
2.3.1 General methods	50
2.3.2 Nuclear magnetic resonance, mass spectrometry and matrix assisted laser desorption ionization instruments	50
2.3.3 Oligonucleotides	51
2.3.4 Synthesis of (6 <i>N</i> -benzoyl-8-(2-(4-imidazolyl) aminoethyl)adenyl))-5'- <i>O</i> -(4,4'- dimethoxytrityl)-2'-deoxy-β-D-ribofuranose and (6 <i>N</i> -benzoyl-8-(2-(4- imidazolyl)ethylamino)adenyl)) -3'- <i>O</i> -(2-cyanoethyl- <i>N,N</i> - diisopropylphosphoramidyl)-5'- <i>O</i> -(4,4'-dimethoxytrityl)-2'-deoxy-β-D-ribofuranose	51
2.3.5 Synthesis of (6 <i>N</i> -benzoyl-8-(2-(4-imidazolyl) aminoethyl)adenyl))-3'- <i>O</i> - (methoxyacetyl)-2'-deoxy-β-D-ribofuranose (2.8):	52
2.3.6 Synthesis of 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA ^{ime} TP) 2.1	53
2.3.7 Chemical synthesis of an 8-modified dA containing oligonucleotide	54
2.3.8 PCR amplification using an 8-modified dA oligonucleotide	54
2.4 Results	55
2.4.1 MALDI spectra of unmodified and modified oligonucleotide primer	55
2.4.2 PCR using a modified oligonucleotide primer	56
2.4.3 Synthesis of dA ^{ime} TP	57
2.4.4 Enzymic incorporation of dA ^{ime} monophosphate	59
2.5 Discussion	60
2.5.1 Modified oligonucleotide synthesis	60
2.5.2 Triphosphate synthesis	61
Chapter 3: A DNzyme Selection used to Evaluate Effects of Shortening Linker Length between Imidazole and 8-Modified 2'-Deoxyadenosine	63
3.1 Introduction	64
3.2 Objective of this work	68

3.3 Materials and methods	68
3.3.1 Chemicals and reagents.....	68
3.3.2 Cells, plasmids, enzymes, and protein	69
3.3.3 Oligonucleotides	69
3.3.4 Buffers and cocktails.....	70
3.3.5 Detection of radioactive DNA	70
3.3.6 <i>In vitro</i> selection	71
3.3.7 General TA cloning method.....	77
3.3.8 Cloning and sequencing of generation 20.....	78
3.3.9 Screening of the clones	78
3.3.10 Kinetics of native Dz20-49 and modified dA ^x replacements	79
3.3.11 37 °C kinetics of Dz20-49	80
3.3.12 pH dependence.....	81
3.4 Results.....	81
3.4.1 Progress of the selection	81
3.4.2 Activity of the clones	83
3.4.3 Predicted secondary structure of Dz20-49	87
3.4.4 Dz20-49 kinetics	88
3.4.4.1 Room temperature (24 °C) kinetics	88
3.4.4.2 Modified dA requirement	90
3.4.4.3 Aminoethyl and aminopropyl replacement.....	90
3.4.4.4 37 °C kinetics.....	92
3.4.4.5 pH rate profile	93
3.5 Discussion	94
3.5.1 General discussion	94
3.5.2 Selection for ribophosphodiester bond cleaving activity.....	95
3.5.3 Dz20-49's secondary structure	96
3.5.4 dA ^x replacement studies	96
3.5.5 Temperature dependence	97
3.5.6 pH dependence of Dz20-49	99
Chapter 4: Enzymic Recognition of Modified DNA	101
4.1 Introduction.....	102
4.1.1 Modified DNA as a template	102
4.1.2 Restriction enzyme digestion of modified DNA	105

4.2 Objective of this work.....	106
4.3 Materials and methods	107
4.3.1 Chemicals and reagents.....	107
4.3.2 Enzymes	107
4.3.3 Oligonucleotides	108
4.3.3 TA cloning and sequencing of the final DNAzyme generations	108
4.3.4 Modified template synthesis for PCR, purification and standardization	109
4.3.5 PCR using modified templates.....	111
4.3.6 Production of doubly-modified dsDNA.....	111
4.3.7 Restriction digests of doubly-modified dsDNA	112
4.3.8 Transfection of doubly-modified dsDNA	113
4.4 Results	113
4.4.1 Sequences of the final DNAzyme generations	113
4.4.2 Quantification of various dA-modified templates based on Dz20-49	126
4.4.3 Amplicons from modified template PCR	129
4.4.4 Production of doubly-modified dsDNA.....	132
4.4.5 Restriction digests of doubly-modified dsDNA	134
4.4.6 Sequences originating from doubly-modified dsDNA	136
4.5 Discussion	137
4.5.1 General analysis of the sequences of the final DNAzyme generations	137
4.5.2 Evaluation of modified DNA as a template in PCR	141
4.5.3 Production of doubly-modified dsDNA.....	142
4.5.4 Restriction digest of modified dsDNA	144
4.5.5 Processing of dU ^{ph} and dU ^{ga} <i>in vivo</i>	145
Chapter 5: Towards the Evolution of <i>Thermus aquaticus</i> DNA Polymerase I.....	148
5.1 Introduction.....	149
5.2 Objective of this work.....	153
5.3 Materials and methods	155
5.3.1 Chemicals and reagents.....	155
5.3.2 Oligonucleotides	155
5.3.3 Recombinant Taq.....	157
5.3.4 Stoffel Fragment	157
5.3.5 E615G site-directed mutagenesis.....	157
5.3.6 Expression.....	158

5.3.7 Control experiments.....	158
5.3.7.1 Visualization of single cell encapsulation.....	158
5.3.7.2 Chemical detection of cross-reaction.....	159
5.3.7.3 Chemical evidence of whole cell isolation	161
5.3.8 Shuffled Taq library	163
5.3.9 Construction of the Taq library via cassette mutagenesis.....	164
5.3.10 Active site library preselection using unmodified nucleoside triphosphates.....	165
5.3.11 First amplification of preselection library products.....	166
5.3.12 Preselected library activity.....	167
5.3.13 O-helix library construction.....	168
5.3.14 CSR with modified nucleoside triphosphates	168
5.3.15 First amplification of CSR products	169
5.3.16 Generation 1 activities	170
5.3.17 CSR using modified nucleoside triphosphates: Round 2.....	170
5.4 Results.....	171
5.4.1 Expression of wild-type Taq and Stoffel Fragment.....	171
5.4.2 Site-directed mutagenesis	172
5.4.3 CSR controls	174
5.4.3.1 Reverse micelle size in relation to <i>E. coli</i>	174
5.4.3.2 Determination of cross-reaction between droplets	177
5.4.3.3 Isolation of individual cells through emulsification	181
5.4.5 Shuffled Taq library.....	187
5.4.6 CSR preselection of the active site using natural nucleoside triphosphates ..	191
5.4.7 Combined library construction	196
5.4.8 CSR using unnatural nucleosides.....	198
5.4.9 CSR using unnatural nucleoside triphosphates and first amplification product	199
5.4.10 Activity of generation 1	201
5.4.11 First amplification product of round 2	202
5.5 Discussion.....	203
5.5.1 General discussion	203
5.5.2 Taq libraries	204
5.5.3 Site-directed mutagenesis	205
5.5.4 Visualization of the compartmentalized cell.....	205
5.5.5 CSR controls to measure cross reaction between droplets	206

5.5.6 Physical separation of cells using an emulsion.....	207
5.5.7 Library construction, expression and CSR	210
5.5.8 Analysis of the isolated and active site preselection CSR clones.....	211
5.5.9 The O-helix library	212
5.5.10 CSR using modified nucleoside triphosphate	212
Chapter 6: Summary, Conclusions and Future Directions.....	215
6.1 Summary	216
6.2 Conclusions.....	218
6.3 Future directions	219
References.....	220
Appendix A: NMR spectra for dA^{imm}TP	231

List of Tables

Table 2.1 HPLC Gradients.....	54
Table 3.1 Decant (supernatant) collected after the indicated duration of incubation from each round of selection. Fresh cleavage buffer was added after each collection.	76
Table 3.2 Amount of purified First amplification product added to the second amplification reaction to increase stringency according to round.	76
Table 3.3 Generation 20 sequences and k_{obs}	86
Table 4.1 Individual clone sequences from the two modifications, N40 selection	115
Table 4.2 Individual clone sequences from the Dz10-13 (mercury sensor) selection	116
Table 4.3 Individual clone sequences from the Dz9-86 (3modifications, N20) selection	117
Table 4.4 Individual clone sequences from the all RNA-cleaving Dz12-91(3 modifications, N20) selection.....	119
Table 4.5 Individual clone sequences from the stachyose sensor selection	121
Table 4.6 Individual clone sequences from the ATP sensor selection	122
Table 4.7 Individual clone sequences from the Dz10-66 (3modifications, N40) selection	123
Table 4.8 Individual clone sequences from the Dz20-49 (3 modifications, N40) selection	124

Table 4.9 Summary of the percentage of dU-modified dsDNA cleaved by the indicated restriction enzyme after two hours of incubation at 37 °C.	136
Table 5.1 Solution phase cross reaction controls.....	160
Table 5.2 Emulsion phase cross reaction controls	160
Table 5.3 Taq and Stoffel Fragment expression control reactions.	161
Table 5.4 Solution phase whole cell PCR controls.....	162
Table 5.5 Compositions of the emulsion phase whole cell isolation controls.....	163

List of Figures

Figure 1.1 Diagram of a nucleotide monomeric unit. Adenosine monophosphate for R = OH or 2'-deoxyadenosine monophosphate for R = H. The subunits phosphate, sugar, and nucleobase groups are marked with grey ovals.....	2
Figure 1.2 ChemBioDraw 12's cartoon representation of B-form double helix DNA illustrating the major groove and minor groove (left) and a close-up of the nucleobase pairings and the positions of the bases that face each groove (right). 3	3
Figure 1.3 Nucleobase numbering – shown is adenosine paired with either uridine or thymidine.	5
Figure 1.4 Secondary RNA structures: a. double-stranded; b. single-stranded; c. hairpin loop; d. bulge; e. internal loop; f. junction.	6
Figure 1.5 Representation of the secondary structure of <i>Schistosoma mansoni</i> hammerhead ribozyme. Shown above is the <i>trans</i> -cleaving ribozyme. The loop-loop interaction is indicated by the double-headed arrow and the cleavage site is indicated by the single-headed arrow.	7
Figure 1.6 A Cartoon representation the crystal structure of the <i>S. mansoni</i> hammerhead ribozyme. PDB code 2GOZ build pymol 0.99rc6 using the cartoon setting. B Proposed Mg ²⁺ -mediated cleavage mechanism of a ribophosphodiester bond by the hammerhead ribozyme. Ribozyme is shown in green and the substrate is shown in red.....	10
Figure 1.7 A Pymol's cartoon representation of the NMR Structure of RNase A. B Proposed mechanism of ribophosphodiester bond cleavage.	12
Figure 1.8 General <i>in vitro</i> selection scheme. Colored triangles represent individual clones.	14

- Figure 1.9 The method for selecting a ribophosphodiester bond-cleaving DNAzyme that was developed by Joyce and coworkers. This selection scheme became the general format for many subsequent self-cleaving DNAzyme selections..... 19
- Figure 1.10 Primary sequence and secondary structure of *cis*-cleaving Dz9₂₅-11. 8-Modified-dA is represented by a red A. 5-modified dU is represented by a blue U. Embedded ribonucleotide is indicated by an rC. 24
- Figure 1.11 Primary sequence and secondary structure of *cis*-cleaving DNAzymes 9-86 and 10-66. 8-Modified dA, 5-modified dC and 5-modified dU are represented by boldface A, C and U, respectively. Embedded ribonucleotide is indicated by a red rC. The modified U: G mismatch is indicated with a middle dot (•)..... 26
- Figure 1.12 Polymerase chain reaction (PCR). Each cycle involves a step carried out at each of three different temperatures. Green lines represent the original DNA template. Black arrows represent primers. Black lines represent the newly synthesized DNA. 1. Template DNA is denatured at high temperature. 2. Temperature is lowered and primers anneal to the single-stranded template DNA. 3. Temperature is raised to the optimal extension temperature for the thermostable polymerase. Both the original DNA and the newly synthesized DNA are used as template for the subsequent cycle. The theoretical amount of DNA doubles with each cycle. 31
- Figure 1.13 PyMol's cartoon representation of the crystal structure of *Thermus aquaticus* DNA polymerase I. PDB entry 1TAU..... 34
- Figure 1.14 Line representation of the active site polypeptide region contained within the Motif A and the bound dideoxycytidine triphosphate from the crystal structure of Taq. PDB entry 1TAU. Green, blue, red, and orange lines represent carbon, nitrogen, oxygen and phosphorus, respectively. Amino acid residues 605 to 617 are shown. Asp610 and the triphosphate from the dideoxycytidine triphosphate are coordinated to magnesium cations, which are represented by green spheres.37

Figure 1.15 PyMol's cartoon and line representation of the O-helix and a bound dideoxycytidine triphosphate from the crystal structure of Taq. PDB entry 1TAU. Green, blue, red, and orange lines represent carbon, nitrogen, oxygen and phosphorus, respectively. Shown is the α -helix composed of amino acid residues 659-671. For clarity, only the side chains of residues Y659, F663, L667, R670 and R671 are shown. Residues Y659 and F663 are implicated in the base pairing process, and residues L667, R670, and R671 provide electrostatic stabilization for the negatively-charged triphosphate.	37
Figure 1.16 Genetic complementation by Taq polymerase to <i>E. coli</i> recA718/polA12 with a thermosensitive DNA polymerase I. <i>E. coli</i> recA718/polA12 is only capable of growth when supplemented with an active Taq polymerase gene.	39
Figure 1.17 A The synthetic propynylisocarbostyryl (PICS) nucleobase analog self-pair. B Cartoon representation of a phage particle displaying both pIII-Stoffel and pIII-Acid peptide.	41
Figure 1.18 Compartmentalized self-replication. The polymerase gene library is cloned into an expression vector and subsequently transformed into an <i>E. coli</i> cell. After the polymerase library is expressed, the cells are isolated and encapsulated with PCR reagents using a water and oil emulsion. The emulsion is then thermocycled, destroying the <i>E. coli</i> cell wall and allowing amplification to occur. Post CSR, the amplicons enriched in genes encoding active mutants can be collected and purified for the next round of selection.	43
Figure 2.1 The structure of 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA ^{ime} TP) 2.1	48

Figure 2.2 Literature synthesis of (6 <i>N</i> -benzoyl-8-(2-(4-imidazolyl)ethylamino)adenyl)-3'- <i>O</i> -(2-cyanoethyl- <i>N,N</i> -diisopropylphosphoramidyl)-5'- <i>O</i> -(4,4'-dimethoxytrityl)-2'-deoxy- β -D-ribofuranose. The triple arrows branching off from 2.6 represent a synthetic route that affords the corresponding modified nucleoside triphosphate analog.....	49
Figure 2.3 MALDI spectrum of unmodified primer oligonucleotides ODN 2.2 showing a mass to charge ratio of 6720.3 (left) and 3'-dA ^{ime} -modified primer oligonucleotide ODN 2.1 showing a mass to charge ratio of 6829.4 (right). Values differ by a mass corresponding to a 2-(4-imidazolyl)aminoethyl group. .	56
Figure 2.4 PCR amplification using Taq polymerase and ODN 2.1 , a primer bearing an 8-modified dA at the 3' end. Lane 1: PCR amplification product. Lane 2: Invitrogen 1kb Molecular Weight Ladder.	57
Figure 2.5 Synthesis of (8-(2-(4-Imidazolyl)ethylamino)adenyl)- 5'-triphosphate-2'-deoxy- β -D-ribofuranose	58
Figure 2.6 A) Polyacrylamide gel showing a single and a double incorporation of the dA ^{ime} nucleoside monophosphate by terminal transferase (left). B) Representative MALDI spectrum of the incorporation reaction (right). Mass to charge ratios found are 6377.4, 6799.7, 7222.8. Peak values differ by a mass corresponding to a single neutrally charged dA ^{ime} monophosphate.....	60
Figure 3.1 The two modified nucleoside triphosphates propargylamino-modified deaza-dATP (left) and imidazolyl-dUTP (right) used by Sidorov <i>et al.</i>	65
Figure 3.2 Chemical structure of 2.1 (dA ^{ime} -TP), 3.1 (dATP), 3.2 (dA ^{imm} -TP), 3.3 (dA ^{imp} -TP), 3.4 (alkynyl-linked imidazole dATP).....	67

Figure 4.1 DNAzyme selection processes that affect fitness of a sequence in a self-cleaving DNAzyme selection include both incorporatability and read-through for amplification.	103
Figure 4.2 Modified dUTP derivatives dU ^{ga} TP (left), dU ^{ph} TP (center) and dU ^{aa} TP (right) used in PCR, restriction digests and transfection.	104
Figure 4.3 Bar graph depicting the number of sequences as a function of the number of dA ^{ime} modifications present in the N40 region, which after the base deletions and additions during the selection consists of a range of 39 - 41 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA ^{ime} and dU ^{aa}	115
Figure 4.4 Bar graph depicting the number of sequences as a function of the number of dA ^{ime} modifications present in a N40 region, which after the base deletions during the selection consists of a range of 35 - 39 bases, for a selection for a mercury sensor DNAzyme modified with dA ^{ime} and dU ^{aa} . (Hollenstein <i>et al.</i> , 2008)	116
Figure 4.5 Bar graph depicting the number of sequences as a function of the number of dA ^{ime} modifications present in a N20 region, which after the base deletions and additions during the selection consists of a range of 18 - 23 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA ^{ime} , dU ^{ga} and dC ^{aa} . (Hollenstein <i>et al.</i> , 2009a)	118
Figure 4.6 Bar graph depicting the number of sequences as a function of the number of dA ^{ime} modifications present in a N20 region, which after the base deletions and additions during the selection consists of a range of 19 - 26 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA ^{ime} , dU ^{ga} and dC ^{aa} capable of cleaving at a region composed of twelve ribonucleotides.	120

Figure 4.7 Bar graph depicting the number of sequences as a function of the number of dU ^{ph} modifications present in a N40 region, which after the base additions during the selection consists of a range of 40 - 42 bases, for a selection for a stachyose sensor DNAzyme modified with dU ^{ph} only.....	121
Figure 4.8 Bar graph depicting the number of sequences as a function of the number of dU ^{ph} modifications present in a N40 region, which after the base additions during the selection consists of a range of 40 - 43 bases, for a selection for a ATP sensor DNAzyme modified with dU ^{ph} only.....	122
Figure 4.9 Bar graph depicting the number of sequences as a function of the number of dA ^{ime} modifications present in a N40 region, which after the base deletions and additions during the selection consists of a range of 38 - 43 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA ^{ime} , dU ^{ga} and dC ^{aa} .(Hollenstein <i>et al.</i> , 2009b)	123
Figure 4.10 Bar graph depicting the number of sequences as a function of the number of dA ^{imm} modifications present in a N40 region, which after the base deletions and additions during the selection consists of a range of 39 – 45 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA ^{imm} , dU ^{ga} and dC ^{aa}	125
Figure 4.11 Calibration curve used to determine modified DNA concentration. A Exposed screen showing the radioactive intensity of calibration standards at various dilutions and 1 µl and 10 µl of synthesized modified and unmodified templates. B Calibration curve produced by autoradiography for determining modified template concentration. Graph fitted to equation $y = -0.9344x + 20.5035$, $R^2 > 0.99$	127

Figure 4.12 Amplification of Dz20-49 templates with varying modifications using Vent (exo-) polymerase. **A** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 37, 39, 41, 43, 45 cycles of PCR, respectively, using unmodified template. Lanes 8-12: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imm}-modified template. **B** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{im}-modified template. Lanes 8-12: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imp}-modified template. 130

Figure 4.13 Comparison of the amplicons produced by Vent (exo-) using modified templates. Lanes 1 + 6: NEB Low Molecular Weight Ladder. Lane 2: Amplicon produced after 43 cycles of PCR using template containing no modifications. Lanes 3-5: Amplicons produced after 43 cycles of PCR using templates containing modified dC, modified dU and one of dA^{imm}, dA^{ime}, dA^{imp}, respectively. 130

Figure 4.14 Amplification of Dz20-49 templates with varying modifications using Taq polymerase. **A** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 27, 29, 31, 33, 35 cycles of PCR, respectively, using unmodified template. Lanes 8-12: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imm}-modified template. **B** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{im}-modified template. Lanes 8-12: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imp}-modified template. 131

Figure 4.15 Comparison of the amplicons produced by Taq using modified templates. Lanes 1 + 6: NEB Low Molecular Weight Ladder. Lane 2: Amplicon produced after 33 cycles of PCR using template containing no modifications. Lanes 3-5: Amplicons produced after 33 cycles of PCR using templates containing modified dC, modified dU and one of dA^{imm}, dA^{ime}, dA^{imp}, respectively. 131

Figure 4.16 Amplicon sequence and modified PCR products. **A** Amplicon sequence showing foreign sequence introduced by PCR primers in *italics*, primer binding regions in lowercase, and restriction digest sites shown in boldface. **B** Lane 1: NEB Low Molecular Weight DNA Marker. Lanes 2-5: dU^{ga}-modified amplicon, dU^{ph}-modified amplicon, dU^{aa}-modified amplicon and unmodified amplicon, respectively. **C** Gel purified PCR products. Lane 1 and 6: NEB Low Molecular Weight DNA Marker. Lanes 2-5: dU^{ga}-modified amplicon, dU^{ph}-modified amplicon, dU^{aa}-modified amplicon and unmodified amplicon, respectively. ... 133

Figure 4.17 Restriction digests of modified dsDNA. **A** Lanes 1, 6 + 11: NEB Low Molecular Weight DNA Marker. Lanes 2 -5: Untreated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 -10: Kpn I treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. **B** Lanes 1, 6, 11 + 16: NEB Low Molecular Weight DNA Marker. Lanes 2 – 5: Hind III-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 – 10: Sal I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 12 – 15: Xba I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. **C** Lane 1, 6, 11 + 16: NEB Low Molecular Weight DNA Marker. Lanes 2 – 5: Bam HI-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 – 10: Eco RI-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 12 – 15: Sma I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. 135

Figure 4.18 Multiple cloning site sequences obtained from plasmids originating from dU^{ga}-modified or dU^{ph}-modified DNA. The dT to dA base mutation is indicated with an underlined A. 137

Figure 5.1 Structure of 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)thymine triphosphate (2'F-araTTP). 150

Figure 5.3 General scheme depicting site-directed mutagenesis based on Strategene's Quikchange Kit.....	173
Figure 5.4 Cycle-dependent production site-directed mutant plasmid. Lane 1: Invitrogen's 1 kb Plus Molecular Weight Ladder. Lanes 2-6 show cycles 4, 8, 12, 16, 20. Lane 7 shows Dpn I-treated amplicon.....	174
Figure 5.5 Formation of reverse micelles. A A surfactant-containing oil phase sits on top of a suspension of cells. Included in the two phases is a stir bar. B The two phases are stirred. C Because of the added surfactants, the two phases form an emulsion consisting of water-in-oil reverse micelles.	176
Figure 5.6 Sixty times magnified view of <i>E. coli</i> cells expressing EGFP. Cells were illuminated using white light to visualize the borders of the aqueous compartment. Cells were also irradiated with UV light to induce fluorescence and to visualize the <i>E. coli</i> cells. Aqueous compartments contain one or no <i>E. coli</i> cells.	176
Figure 5.7 Solution phase PCRs of Reaction A, Reaction B and a homogeneous mixture of the two reactions. In the mixture, the shorter product is exclusively produced.	178
Figure 5.8 Compartmentalized self-replication of targets of differing size. Two PCRs were prepared separately with each amplifying a target of different lengths. Prior to thermocycling the reactions were mixed; then emulsified. Lane 1 + 5: Invitrogen's 1kb Plus Molecular Weight Marker. Lane 2: Reaction A targeting a 160 bp fragment with no polymerase added. Lane 3: Reaction B targeting a 0.8 kb fragment with polymerase added. Lane 4: 1:1 mixture of Reaction A and Reaction B.	179

Figure 5.9 Scheme showing the possible products formed from stable or unstable reverse micelles. Reaction B produced a 0.8 kb amplicon represented by the green lines. Reaction A contains no polymerase and fails to yield any amplicon. If the mixture of the two emulsions stays intact and no components can cross the oil phase barrier, only the larger 0.8 kb amplicon will be produced. If there is exchange of components, either primers or polymerase, the shorter 160 bp amplicon will be produced..... 180

Figure 5.10 Compartmentalized self-replication of targets of differing size. Two PCRs were prepared separately with each amplifying a target of different lengths. Prior to thermocycling the reactions were emulsified and then mixed. Lane 1 + 5: Invitrogen's 1kb Plus Molecular Weight Marker. Lane 2: PCR targeting a 160 bp fragment with no polymerase added. Lane 3: PCR targeting a 0.8 kb fragment with polymerase added. Lane 4: 1:1 mixture of the two PCRs. 181

Figure 5.11 Solution phase PCR. Top. Representation of the two control plasmids pStoff and pTaq and the condition that support amplification of their polymerase gens. Bottom. Whole cell PCRs using *E. coli* cells containing either a plasmid bearing a Stoffel Fragment gene or a Taq gene. Lane 1: A PCR mixture containing Stoffel Fragment-expressing cells with a 0.8 kb target. Lane 2: A PCR mixture containing only Taq-expressing cells with a 0.8 kb amplicon target. Lane 3: A PCR mixture containing only Stoffel Fragment-expressing cells. Lane 4: A PCR mixture containing only Taq-expressing cells. Lane 5: A 1:1 mixture of the Taq-expressing cells PCR mixture and the Stoffel Fragment-expressing cells PCR mixture. Lane 6: Invitrogen's 1 kb molecular weight ladder. 184

Figure 5.12 Compartmentalized self-replication of two different genes. **A** Scheme showing how the emulsification isolates cells prior to thermocycling. Successful isolation of the individual cells followed by thermocycling as represented would result in only the Taq gene being amplified and the reverse micelles containing cells expressing Stoffel Fragment would have no amplicon. **B** *E. coli* cells expressing either Stoffel Fragment or a Taq were emulsified and thermocycled. Lanes 1 + 5: Invitrogen's 1 kb molecular weight ladder. Lane 2: Stoffel Fragment-expressing cells were emulsified and thermocycled. Lane 3: Taq expressing cells emulsified and thermocycled. Lane 4: 1:1 ratio of Taq expressing cells and Stoffel expressing cells mixed prior to emulsion and thermocycling. 186

Figure 5.13 Time-dependent fragmentation of DNA by DNase I. Lanes 1-5: 1, 2, 3, 4, and 5 minutes of digestion, respectively..... 189

Figure 5.14 Polymerase-mediated reassembly of gene fragments. Lane 1 Invitrogen 1kb Plus Ladder. Lane 2: Taq-mediated shuffling with exogenous DNA fragments. Lane 3: Taq-mediated shuffling 4: Pfu-mediated shuffling. 189

Figure 5.15 Amplification of the shuffled Taq gene. Lane 1: Invitrogen's 100 bp Ladder. Lane 2: Taq-reassembled Taq reamplified shuffled library. Lane 3: Pfx-reassembled Taq reamplified shuffled library. 190

Figure 5.16 Activity of the shuffled Taq library. Lane 1: Invitrogen's 100 bp Ladder. Lane 2: Whole cell PCR using the shuffled library..... 191

Figure 5.17 **A.** Important biomolecules present in the spCSR selection reaction. Plasmid is represented by the green circle. Polymerase is represented by the green chevron. Green and purple arrows represent primers. Green and purple lines represent spCSR selection product. Note that the selection products are not one defined size. **B.** The biomolecules from A purified using Qiagen's PCR Purification Kit. Polymerase is removed, but dNTPs, primers and plasmids are present in trace amounts. **C.** The components of B are treated with ExoSAP-It to remove trace dNTPs from the selection and primers. **D.** New primers represented by purple arrows and Vent (exo-) represented by the blue chevron are added to the remaining plasmid and amplicon in a PCR reaction mixture. **E.** The reaction is thermocycled, and selection products are amplified exclusively. **F.** The new amplicon is agarose gel purified to remove plasmid and primer. . 193

Figure 5.18 Activities of preselected clones of the active site library. Lanes 1 and 12: NEBs Low Molecular Weight Markers. Lanes 2-11: PCR products indicating the activities of active site preselection clones 1-10, respectively. 195

Figure 5.19 Active site sequences from active clones isolated from CSR using natural nucleoside triphosphates. The base mutations are underlined, and the amino acid mutations are indicated below the nucleotide sequence. Clones marked with an asterisk have the same sequence as wild-type Taq gene. Bases outside of the mutated region are shown in blue. 195

Figure 5.20 Flowchart describing the building of the Generation 1 libraries. The active site residue 605-617 were partially mutagenized to create an active site library. After preselection of the active site library, it was fused to the O-helix library. The library resulting from the combination was used for CSR with modified nucleotides. 197

Figure 5.21 Revised CSR scheme to accommodate unnatural nucleic acids. Step 5 shows the introduction of modified nucleoside triphosphates into spCSR. 199

Figure 5.22 First amplification of the isolated selection product. Lane 1: negative control (no dNTPs added), Lane 2: amplification of 2'-F-araT amplicon, Lane 3: amplification of dC^{aa} amplicon, Lane 4: amplification of dA^{imm} amplicon, Lane 5: positive control (natural dNTPs) and Lane 6: NEB's low molecular weight ladder. 201

Figure 5.23 Library activities under standard PCR conditions. Lane 1 and 6: NEB's low molecular weight ladder, Lane 2: 2'-F-araT library, Lane 3 dC^{aa} library, Lane 4 dA^{imm} library, Lane 5 dNTP library..... 202

Figure 5.24 First amplification of the isolated selection product from Round 2. Lane 1 + 6: NEB's Low Molecular Weight Ladder. Lane 2: amplification of 2'-F-araT amplicon. Lane 3: amplification of dC^{aa} amplicon. Lane 4: amplification of dA^{imm} amplicon. Lane 5: positive control (natural dNTPs)..... 202

List of Abbreviations

Abbreviations

APS	ammonium persulfate
Bz	benzoyl
bp	base pairs
DNA	deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
ssDNA	single-stranded deoxyribonucleic acid
DNAzyme	DNA enzyme
DEPC	diethyl pyrocarbonate
Dz	DNAzyme, DNA enzyme
DDT	dithiothreitol
DMSO	dimethyl sulfoxide
DMT	4, 4'-dimethoxytrityl
DPO4	DNA polymerase 4
dNTPs	deoxynucleoside triphosphates
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
FACS	fluorescence-activated cell sorting
FANA	2'-deoxy-2'-fluoro- β -D-arabinonucleic acid
HPLC	high performance liquid chromatography
IDT	Integrated DNA Technologies
IPTG	Isopropyl β -D-thiogalactoside
mRNA	messenger ribonucleic acid
NAPS Unit	Nucleic Acid Protein Services Unit
O.D. ₆₀₀	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis

PCR	polymerase chain reaction
PDB	protein data bank
Pfu	Family B DNA polymerase from <i>Pyrococcus furiosus</i>
PICS	propynylisocarbostyrl
PMSF	phenylmethysulfonyl fluoride
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAzyme	RNA enzyme
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELEX	<u>S</u> ystematic <u>E</u> volution of <u>L</u> igands by <u>E</u> xponential enrichment
Taq	<i>Thermus aquaticus</i> DNA polymerase I
Tris	tris(hydroxymethyl)methylamine
tRNA	transfer ribonucleic acid
U	units
UV	ultraviolet
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Modified and Unmodified Nucleotide Abbreviations

A	adenine
T	thymine
G	guanine
C	cytosine
dATP	2'-deoxyadenosine triphosphate
dTTP	2'-thymidine triphosphate
dGTP	2'-deoxyguanosine triphosphate
dCTP	2'-deoxycytidine triphosphate
dA ^{imm} TP	8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine triphosphate
dA ^{ime} TP	8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate

dA ^{imp} TP	8-(3-(4-imidazolyl)aminopropyl)-2'-deoxyadenosine triphosphate
dC ^{aa} TP	5-aminoallyl-2'-deoxycytidine triphosphate
dU ^{aa} TP	5-aminoallyl-2'-deoxyuridine triphosphate
dU ^{ga} TP	5-guanidiniumallyl-2'-deoxyuridine triphosphate
dU ^{ph} TP	5-(para-hydroxybenzamido)methyl-2'-deoxyuridine triphosphate
2'F-araTTP	1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)thymine triphosphate

Amino Acid Abbreviations

A	Ala	alanine
C	Cys	cystine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histamine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Acknowledgements

I thank my supervisor Dr. David Perrin for his help and for the freedom to explore the science that lies at the interface of chemistry and biology. I thank the lab members of the Perrin Lab past and present who made the working environment light-hearted and informative. In particular, Curtis Lam's chemistry was integral to many parts of this dissertation, and Dr. Marcel Hollenstein's coaching on DNAzyme selections was most efficient. I wish to thank Dr. Leonerd Lermer for guiding me in my early days in the lab and ensuring that my nucleic acid chemistry was carried out meticulously. The work in the past and the current successes of my former students Dr. David Kwan, Mr. Jefferson Chan and Mr. Sherman Farahani are a constant motivation for me. Dr. Martin Tanner, Mr. Curtis Lam and Ms. Lainie Senger provided excellent feedback on my dissertation.

I also would like to thank the current and former members of UBC's Bioservices Laboratory, Dr. Elena Polishchuk, Mrs. Jessie Cheng and Mrs. Candice Martin for maintaining Bioservices and providing fruitful discussions in microbiology. I would like to extend my appreciation to the Department of Chemistry's IT and mechanical shop, in particular Mr. Milan Coshizza, for the maintenance of the laboratory equipment. Thanks must be extended to Mr. John Ellis and the staff of Chemistry Stores for ensuring our radiation and other supplies arrived promptly. The funding from Canadian Initiative for Health Research operating funds, National Science and Engineering Research Council and Protein Engineering Network Centre of Excellence has enhanced my education by allowing me freedom in my research and providing me the means to partake in conferences.

Outside of the laboratory scene, the fine company of Dr. Jon May, Dr. "Deetch" Dietrich, Dr. Marcel Hollenstein, Dr. Curtis Harwig, and the staff of the Pendulum always made for a pleasant morning coffee break. Michelle Tran was very diligent in reminding me to get some sun and fresh air during our afternoon walks.

Last, but not least, I thank my family who provided me a life in Canada.

Dedicated to my family

**Chapter 1: Nucleic Acids, Nucleic Acid Enzymes and *Thermus aquaticus* DNA
Polymerase I**

1.1 Nucleic acids

Nucleic acids are one of the four major molecule classes found in organisms, the others being: protein, glycosides and lipids. As such, they encode the genetic information that ultimately provides for an organism's physical characteristics. While nature seems to have reserved information storage as the primary role for nucleic acid in the modern world, researchers are discovering that nucleic acids have a far more diverse multifunctional role. In addition to roles in long-term storage of genetic information, they have been used for short-term amplification of genetic information, substrate recognition, binding, and catalysis.

The monomeric unit of nucleic acid is composed of a phosphate group, a nitrogenous nucleobase, and a ribose sugar (Figure 1.1). Natural nucleic acids fall into one of two major categories: the first category is deoxyribonucleic acid (DNA) and the second is ribonucleic acid (RNA). The main chemical difference between the monomeric units of the two biopolymers is the presence or absence of a single hydroxyl group at the 2' position of the ribose sugar. The presence or absence of this single hydroxyl group affects the chemistries of the two biopolymers, giving them different roles and stabilities in the cell.

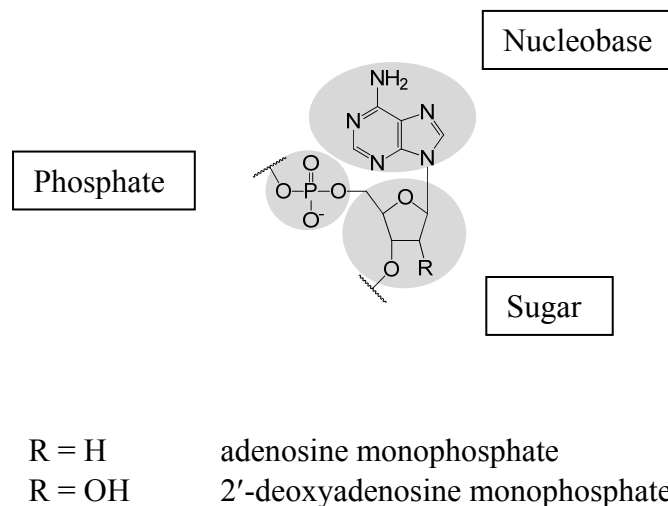


Figure 1.1 Diagram of a nucleotide monomeric unit. Adenosine monophosphate for R = OH or 2'-deoxyadenosine monophosphate for R = H. The subunits phosphate, sugar, and nucleobase groups are marked with grey ovals.

1.1.1 DNA

DNA is the biopolymer responsible for the transference and long-term storage of genetic information in living organisms. The stability of DNA and its role as long-term information storage has often led to DNA being referred to as the blueprints of the physical characteristics of the organism. The genetic information of these blueprints is encoded within the sequence of the nucleobases. Nucleobases break down into two groups: the monocyclic pyrimidines and the bicyclic purines. The pyrimidine thymine (T) pairs with the purine adenosine (A), and the pyrimidine cytosine (C) pairs with the purine guanosine (G) (Figure 1.2). Using these four nucleobases, 4^n combinations can be made for a sequence containing n nucleotides. These combinations can be used to create a set of instructions to a specific physical trait called a gene. Decoding of a gene's information often results in the expression of a protein. In protein synthesis, a single amino acid, the monomeric unit for peptides and proteins, is encoded by a short sequence of three nucleobases called a codon. The list of nucleotide sequences and their associated amino acids is referred to as the genetic code. (Crick, Brenner, Watstobi.Rj, & Barnett, 1961)

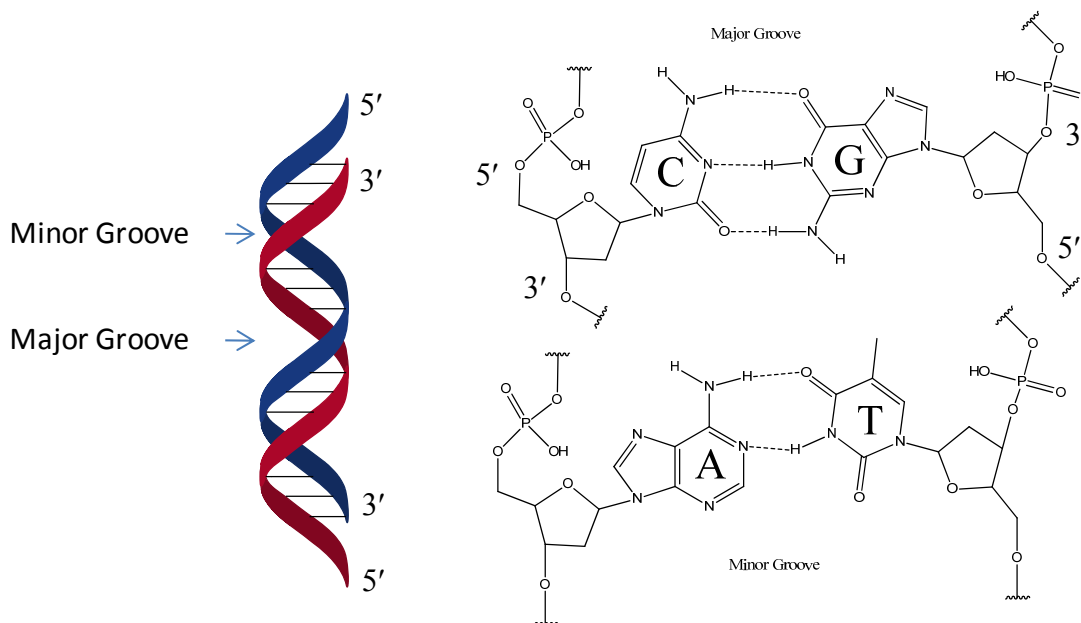


Figure 1.2 ChemBioDraw 12's cartoon representation of B-form double helix DNA illustrating the major groove and minor groove (left) and a close-up of the nucleobase pairings, the 5' and 3' oxygens, and the positions of the bases that face each groove (right).

The monomeric unit of DNA lacks a 2' hydroxyl, making it more resistant to hydrolysis compared to RNA, and is the superior biopolymer for the long-term storage of genetic information. This characteristic of the monomeric unit is complemented by features of the polymer that result in high stability and generally low reactivity. In the cell, DNA is commonly found as antiparallel double-stranded material, which helps in preventing non-specific H-bonding by the nucleobases to intracellular biomolecules. Besides H-bonding between the nucleobases of the two strands, the duplex structure is further stabilized by the π -stacking and hydrophobic interactions of the planar aromatic bases. The two DNA strands are intertwined to form a double helix (Figure 1.2 left). There are three forms of the double helix, which include the right-handed helical twisted A- and B-forms. A-form contains about 11 base pairs per twist and B-form contains about 10 base pairs per twist. Z-form is a left-handed helical structure made up of about 2.4 base pairs per twist. B-form DNA is the most common of the three forms. Notable architectural features of the A- and B-forms of the double helix are the major and minor grooves. These locations play an important role in the engineering of functionalized nucleobases.

1.1.2 RNA

In the cell, RNA commonly perform duties such as the relaying of genetic information as messenger RNA (mRNA), facilitating recognition of amino acids as transfer RNA (tRNA), and synthesizing peptides and proteins as ribosomal RNA (rRNA). In the biosynthesis of proteins, DNA is used as a template to first transcribe RNA which is used in the translation of peptides and proteins. By using messenger RNA, genetic information is quickly amplified by producing multiple strands of RNA from a single strand of DNA. In a similar fashion, many ribosomes can be affixed to a single strand of mRNA to maximize peptide synthesis. With this efficient system for the production of protein, a mechanism must be in place to quickly down regulate protein production. Unlike DNA, RNA can be easily degraded due to the presence of a hydroxyl group at the 2' position. Degradation of RNA can be readily accomplished chemically using strong base or enzymically using RNaseA, ribozymes or DNazymes, a process which will be discussed in Section 1.1.4 – 1.2.5.

Another chemical distinction between DNA and RNA is the use of different nucleobases to pair with adenine. RNA uses uracil, not thymine, to base pair with adenine. The two bases are very similar in composition and chemical functionality. Unlike thymine, however, uracil lacks the methyl group in the 5 position (Figure 1.3, right nucleobase). Due to favorable interactions with commercial polymerases, chemical modifications are often grafted at the 5 position of uracil (Dewey, Zyzniewski, & Eaton, 1996; Tarasow, Tarasow, & Eaton, 1997; Vaish, Fraley, Szostak, & McLaughlin, 2000) and 2'-deoxyuracil (Held & Benner, 2002; Held, Roychowdhury, & Benner, 2003; Jäger & Famulok, 2004; Jäger *et al.*, 2005; Sakthivel & Barbas, 1998; Sawai *et al.*, 2007; Thum, Jager, & Famulok, 2001).

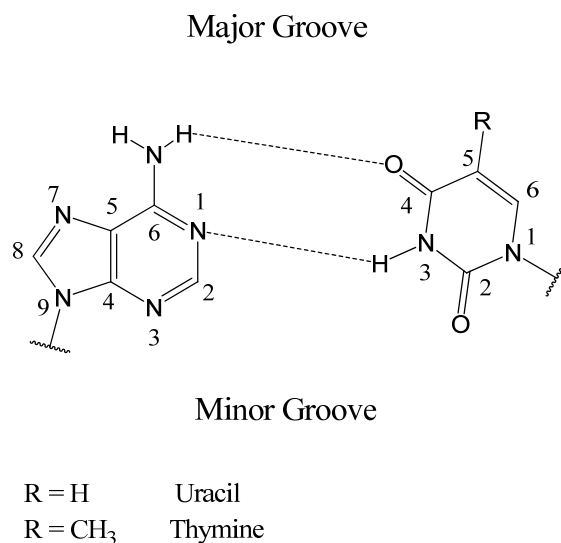


Figure 1.3 Nucleobase numbering – shown is adenosine paired with either uridine or thymidine.

Nominally, RNA in the cell is commonly considered to be single-stranded. However, this does not imply that RNA lacks 3D structure or that it is uniformly single-stranded. The nucleobases of RNA still retain their H-bonding potential and can stabilize small motifs consisting of folded RNA. A common secondary structure is the short stem loop which would be analogous to the α -helices and β -sheets found in proteins. Other possible structural motifs are double-stranded regions, junction regions, bulge regions, internal loop regions, hairpin loop regions, and single-stranded regions (Figure 1.4). Double-stranded RNA commonly adopts the right-handed helical A-form. An example

of RNA with higher order structure is *Schistosoma mansoni* hammerhead ribozyme (Figure 1.5).(Canny *et al.*, 2004) Clearly recognizable is the hairpin loop, the junction loop, the internal loop and double-stranded secondary structures. Additional H-bonding between the hairpin loop and the internal loop stabilize the three-dimensional structure of the ribozyme. The importance of the structure in the function of the hammerhead ribozyme will be further discussed in section 1.1.4.1.

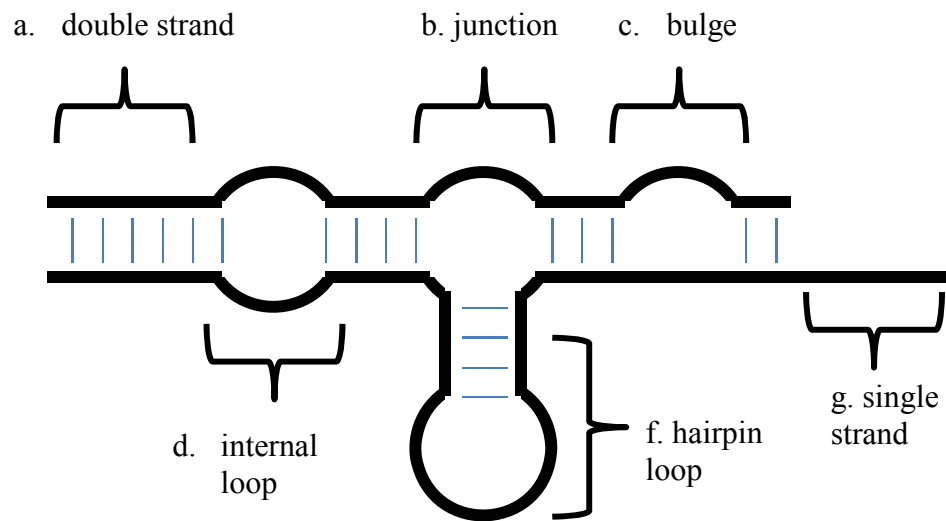


Figure 1.4 Secondary RNA structures: a. double-stranded; b. single-stranded; c. hairpin loop; d. bulge; e. internal loop; f. junction.

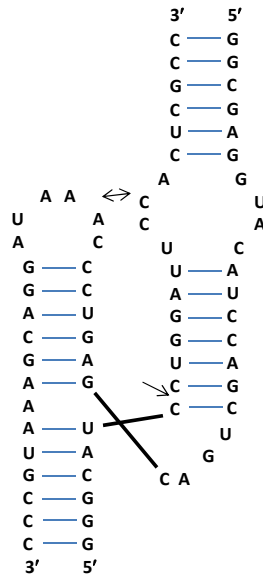


Figure 1.5 Representation of the secondary structure of *Schistosoma mansoni* hammerhead ribozyme. Shown above is the *trans*-cleaving ribozyme. The loop-loop interaction is indicated by the double-headed arrow and the cleavage site is indicated by the single-headed arrow.

Not all 3D structures of RNA are formed spontaneously. Formation of a well-defined structure may be initiated upon binding of a specific effector molecule. Nature uses this structural change as a signal to regulate the translation of mRNA in conditions containing a certain concentration of effector molecule. These sections of RNA that are often found at the 5' end of the mRNA are referred to as riboswitches.(Blouin, Mulhbach, Penedo, & Lafontaine, 2009; Winkler & Breaker, 2003) The specificity of substrate binding and RNA's ability to fold up into complex tertiary structure led researchers to speculate that RNA can function beyond substrate binding.

1.1.3 Ribonucleic acid enzymes and the RNA world

RNA has the ability to form complex 3D structures capable of substrate specific binding, which led researchers to believe that there exists higher-order structured RNA capable of performing catalysis. In 1982, the reported discovery of RNA catalysis of the *Tetrahymena* Intron I followed by the discovery of RNA catalysis of ribonucleoprotein RNase P in 1983 eradicated the concepts of catalysis being performed exclusively by

proteins and nucleic acids being used solely as information media or carrier biomolecules.(Guerrier-Takada, Gardiner, Marsh, Pace, & Altman, 1983; Kruger *et al.*, 1982) Natural RNA enzymes (ribozymes) perform functions in one of the following categories: 1) nucleolytic cleavage,(Doherty & Doudna, 2001) 2) splicing, or 3) peptidyl bond formation.(Cech, 2000) The most elaborate catalysis is the peptidyl bond formation carried out by the ribosome. While the ribosome is two-thirds RNA and one-third protein, the catalytic core of the ribosome was determined to be composed entirely of RNA.(Cech, 2000) Researchers concluded that it is the RNA and not the protein component that is responsible for catalysis. As this introduction is meant to orient the reader on selected types of catalysis, it will not include an extensive description of all naturally occurring ribozymes which can be found in reviews.(Butcher, 2001; Wu, Huang, & Zhang, 2009)

The catalytic complexities of RNA are evidence that RNA alone can play dual roles of genetic information storage and catalysis. In the absence of both DNA and protein, RNA might hypothetically possess functional competence to sustain a metabolic system on its own. Thus, it is possible then that RNA predates both DNA and protein and was at one time the dominant biomolecule in an era referred to as the RNA World.(Gilbert, 1986) Chemically rich proteins may have supplanted the majority of the ribozymes of the RNA world, but ribozyme relics of that proposed era remain. Nevertheless, only a handful of reactions catalyzed by natural ribozymes have been discovered and the chemical diversity of the reactions is limited.(Fedor & Williamson, 2005; Wu *et al.*, 2009) Except for the ribozyme component of the ribosome, natural ribozymes act upon RNA substrates. One of these reactions is the cleavage of the ribophosphodiester bond in RNA. The nucleolytic properties of ribozymes will be focused on in section 1.1.4.1.

1.1.4 Cleavage of RNA: applications and optimizations

Rapid and sequence specific ribophosphodiester bond cleavage is of importance in both the contexts of therapeutics and biotechnology. With regards to therapeutics, regulation of RNA is an important part of an organism's viability. Inappropriately high levels of mRNA expression due to malfunctioning regulation are characteristic of

oncogenesis as well as viral pathologies; therefore, a process leading to the attenuation of protein expression can be achieved through the degradation of specific mRNA transcripts that encode for the protein target. This is also referred to as post-transcriptional gene silencing. Degradation of RNA is usually accomplished by a transphosphorylation that results in ribophosphodiester bond cleavage. This process cannot occur randomly otherwise the expression of essential proteins would also be attenuated. Instead, sequence-specific cleavage of mRNA is required for post-transcriptional gene silencing. Nature has a process in place called RNA interference (RNAi)(Fire *et al.*, 1998) that uses a proteinaceous RNA-induced silencing complex (RISC) to sequence-specifically cleave mRNA. This RISC is activated by a short piece of RNA referred to as small interfering RNA (siRNA) which directs cleavage to a very specific mRNA sequence through the use of Watson-Crick base pairing. Nucleic acid enzymes can also perform sequence-specific post-translational gene silencing and are being studied for therapeutic use.(Khan, 2006; Peracchi, 2004)

Ribophosphodiester bond cleavage has found its way into biotechnological applications. One prominent use of ribophosphodiester bond cleavage is the detection of metals. For example, Yi Lu and coworkers have developed DNAzymes with high sensitivity and selectivity to metals such as Pb^{2+} , Cu^{2+} , Hg^{2+} , and UO_2^{2+} .(J. Liu *et al.*, 2007; J. Liu & Lu, 2003, 2004, 2007a, 2007b, 2007c) The binding of the metal cation of interest effectively switches the DNAzyme to an “on” state, allowing for catalytic cleavage of a signal molecule. Perrin and coworkers have also selected for chemically-augmented DNAzyme mercury sensors that either turn “off” or turn “on” in the presence of mercury.(Hollenstein, Hipolito, Lam, Dietrich, & Perrin, 2008; Thomas, Ting, & Perrin, 2004)

Investigators have focused on two aspects of ribophosphodiester bond cleavage that must be optimized. These two aspects are the substrate specificity and cleavage rate constant. Two of nature’s enzymes are presented here to showcase nature’s optimization of these two aspects. The first example is the hammerhead ribozyme, which illustrates high substrate specificity; the second example is the protein enzyme RNase A which illustrates a high catalytic rate constant. Both the hammerhead ribozyme and RNaseA attain catalytic efficiency, $k_{\text{cat}}/K_{\text{M}}$, of $\sim 10^9 \text{ M}^{-1} \text{ min}^{-1}$.

1.1.4.1 RNA degradation by *S. mansoni* hammerhead ribozyme

Although devoid of the chemical functionalities present in proteins, ribozymes can be efficient biocatalysts in certain contexts and in certain conditions. For example, the hammerhead ribozyme (Figure 1.6 A) is a naturally occurring ribozyme that performs nucleolytic cleavage of a ribophosphodiester bond. (Doherty & Doudna, 2001) This ribozyme is utilized in the rolling circle replication mechanism of viruses and viroids by cleaving the transcribed products into single genome lengths.

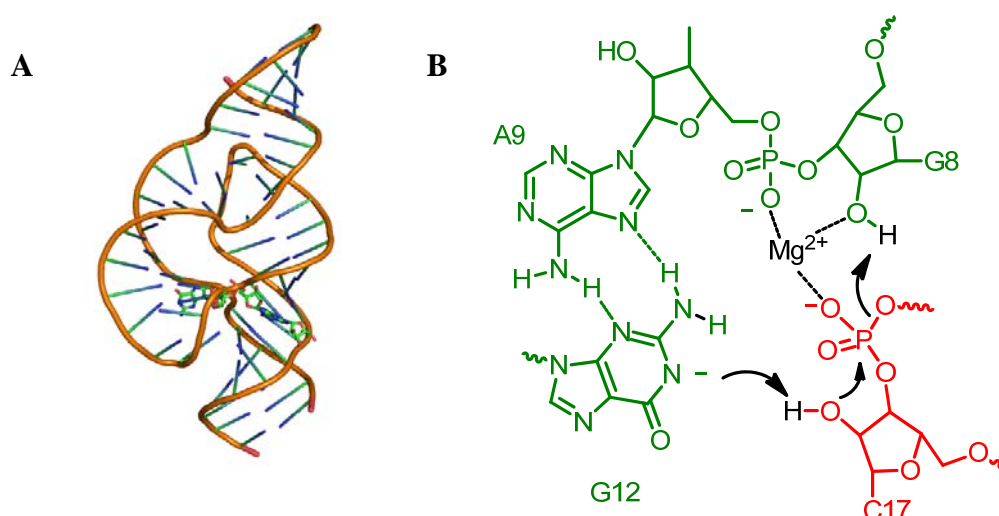


Figure 1.6 **A** Cartoon representation the crystal structure of the *S. mansoni* hammerhead ribozyme. PDB code 2GOZ build pymol 0.99rc6 using the cartoon setting. **B** Proposed Mg^{2+} -mediated cleavage mechanism of a ribophosphodiester bond by the hammerhead ribozyme. Ribozyme is shown in green and the substrate is shown in red.

Due to the minimal chemical functionalities available to RNA, ribophosphodiester bond cleavage is accomplished by employing divalent magnesium cations. One role of the divalent magnesium cation is the stabilization of the three-dimensional structure of the ribozyme which facilitates cleavage by promoting the in-line or S_N2 -like attack of the neighboring 2'-hydroxyl. It is noteworthy that sufficiently high concentrations of monovalent cations can also help stabilize three-dimensional structure allowing the hammerhead ribozyme to perform catalysis in the absence of divalent metal cations. (Murray, Seyhan, Walter, Burke, & Scott, 1998)

The divalent magnesium cations also aid in catalysis by inducing complex folds that perturb the pK_a values of existing functional groups toward neutral pH and by helping provide electrostatic stabilization. In this way, catalytic cleavage can be accomplished using an acid/base mechanism even at physiological pH. Although there is considerable debate surrounding the precise mechanism of ribozyme-mediated RNA cleavage, it appears that nucleobase G12 is pK_a perturbed, and initiates cleavage by deprotonating the 2' hydroxyl of the substrate (Figure 1.6 B). The role of the general acid in this mechanism is played by the 2' hydroxyl of nucleobase G8 that appears to necessarily be coordinated to a Mg^{2+} cation. This hydroxyl delivers a proton to the 5' hydroxyl of the 3' leaving group, which promotes the cleavage of the ribophosphodiester bond by completing the transphosphorylation.

While the catalytic rate constants of the minimal hammerhead ribozyme (HHR15, 1 min^{-1} at pH 7.5 in 10 mM Mg^{2+}) and the more efficient extended natural hammerhead ribozyme (1 min^{-1} at $\sim 0.07 \text{ mM } Mg^{2+}$) (Khvorova, Lescoute, Westhof, & Jayasena, 2003) are far lower than RNaseA's catalytic rate constant ($84\,000 \text{ min}^{-1}$) (Raines, 1998), the hammerhead ribozymes boasts the higher substrate specificity due to the Watson-Crick base pairing of the substrate to the enzyme. Nevertheless, the hammerhead ribozyme approaches catalytic perfection due to both a high k_{on} and a low k_{off} , in spite of the fact that the k_{cat} value fall far below that of RNaseA. When considering that high substrate specificity and the effects on the second order rate constant for multiple turnover, the minimal hammerhead displays a catalytic efficiency ($k_{cat}/K_M = 2.9 \cdot 10^7 \text{ M}^{-1} \text{ min}^{-1}$), which is just lower than the catalytic efficiency of RNaseA ($k_{cat}/K_M = 9.0 \cdot 10^8 \text{ M}^{-1} \text{ min}^{-1}$) for a poly C substrate. (Fedor & Uhlenbeck, 1992) What RNaseA lacks in substrate specificity, however, it makes up for in catalytic power (k_{cat}/k_{uncat}).

1.1.4.2 RNA degradation by RNaseA

RNase A is one of the most common proteins used as a model system for folding, stability and chemistry (Figure 1.7 A). (Raines, 1998) As discussed in section 1.1.4.1, the minimal hammerhead ribozyme and RNaseA have catalytic efficiencies close to the diffusion control limit. The k_{cat} contribution to the second order rate constant is an

impressive 84000 min^{-1} for the UpA dinucleotide substrate. (Delcardayre & Raines, 1994) Degradation of the RNA substrate is not sequence-specific, and cleavage can occur at any ribonucleotide containing a pyrimidine; cleavage can also occur at a purine, but the catalytic efficiency is much lower. For example, the $k_{\text{cat}}/K_{\text{M}}$ for a poly A substrate is $1.7 \cdot 10^4 \text{ min}^{-1}$. (Delcardayre & Raines, 1994) The protein's impressive stability and catalytic rate are due to the diverse chemical functionality available to proteins. Comprised of only 124 amino acid residues, RNaseA is a small enzyme, but shows incredible thermostability, attributed in part to four disulfide bonds, which involve all eight cysteines present in the protein. These four disulfide bonds help RNaseA retain its soluble, folded structure even in highly denaturing conditions. Purification of RNaseA typically involves a step in which the sample is incubated in denaturing conditions to disrupt the folding of any contaminating proteins. Such a simple method of purification led RNaseA to be the most easily produced and most studied protein of the 20th century.

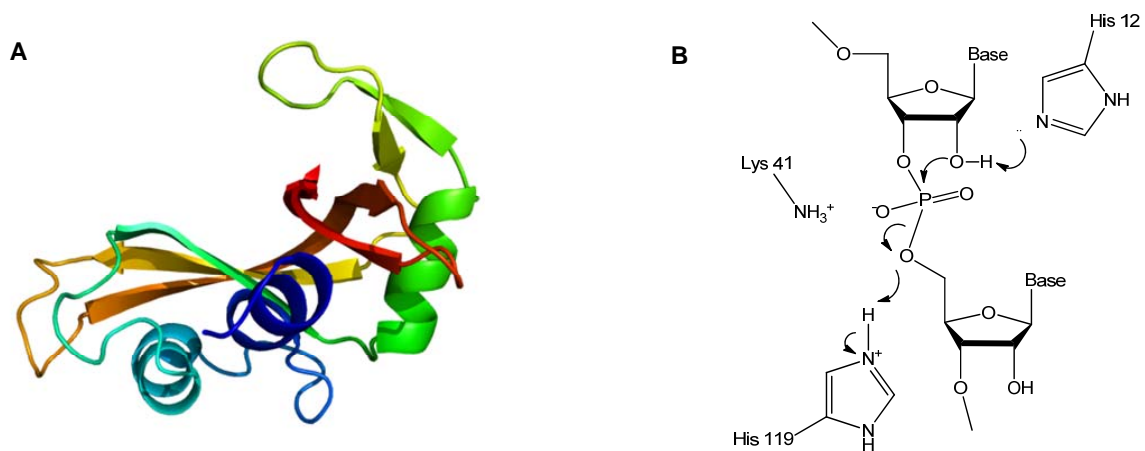


Figure 1.7 **A** Pymol's cartoon representation of the NMR Structure of RNase A. **B** Proposed mechanism of ribophosphodiester bond cleavage.

The catalytic mechanism of RNaseA involves a transphosphorylation reaction between the 2'-hydroxyl of a ribose and its neighboring phosphodiester group. (Raines, 1998) RNaseA utilizes a general acid/base mechanism mediated by two histidine residues; histidine 12 acts as a general base by deprotonating the 2' hydroxyl of the

substrate RNA; lysine 41 provides electrostatic stabilization to stabilize the buildup of negative charge of the pentacoordinate phosphorane intermediate; histidine 119 delivers a proton to the 5' hydroxyl of the 3' nucleoside. Protonation promotes the elimination of the 5' hydroxyl leaving group and the S_N2-like or inline attack by the 2' oxygen (Figure 1.7 B). Transphosphorylation effectively destroys the ribophosphodiester bond and results in strand cleavage. While there are other residues involved in the stabilization of the pentacoordinate phosphorane intermediate, the key steps of this mechanism are executed by only two types of functional groups – the imidazole and the amino groups. Unfortunately for natural nucleic acids enzymes, similar functional groups are not available for use in catalysis. Although ribozymes have high sequence specificity and appreciable catalytic rate constants at high divalent metal cation concentrations, natural nucleic acid enzymes such as the hammerhead ribozyme cannot reach catalytic rate constants as high as RNaseA without divalent metal cations due to the lack of the chemistry of side chain functional groups found in RNaseA.

1.2 *In vitro* selection of nucleic acid enzymes

The first *in vitro* selection of a nucleic acid enzyme with improved function was from the lab of Gerald Joyce.(Robertson & Joyce, 1990) In this selection, variants of the *Tetrahymena* ribozyme were selected according to their ability to sequence specifically cleave DNA. Similar to the selection of protein enzymes, the ribozyme selection mentioned above utilized an existing wild-type structure. However, designer nucleic acid enzymes have made further progress in *de novo* enzyme creation than protein enzymes due to predictable secondary structures like hairpin loops and to technologies such as *in vitro* selection and the Systematic Evolution of Ligands by EXponential enrichment (SELEX).(Breaker & Joyce, 1994; Ellington & Szostak, 1990; Tuerk & Gold, 1990) This powerful technique has led to the discoveries of both novel ribozymes and a new class of biomolecule - the DNA enzyme.(Baum & Silverman, 2008; Hobartner & Silverman, 2007; Joyce, 2004; S.K. Silverman, 2004, 2005; S. K. Silverman, 2008)

1.2.1 SELEX and generalized combinatorial selection

Selection is a process that allows the preferential proliferation of a species with a certain phenotype that is used to overcome a selection pressure. A selection pressure is a condition that eliminates or reduces the number of a species not bearing this phenotype. Natural selection leading to the evolution of a species can take up to several millions of years, so replicating parts of this process would not be feasible on an experimental time scale available to researchers. As such, *in vitro* (cell free) selection was developed to evolve species on a shorter time scale (Figure 1.8). The evolution of the species can be engineered by applying the appropriate selection pressure, and each generation may be produced in the time span of hours or days.

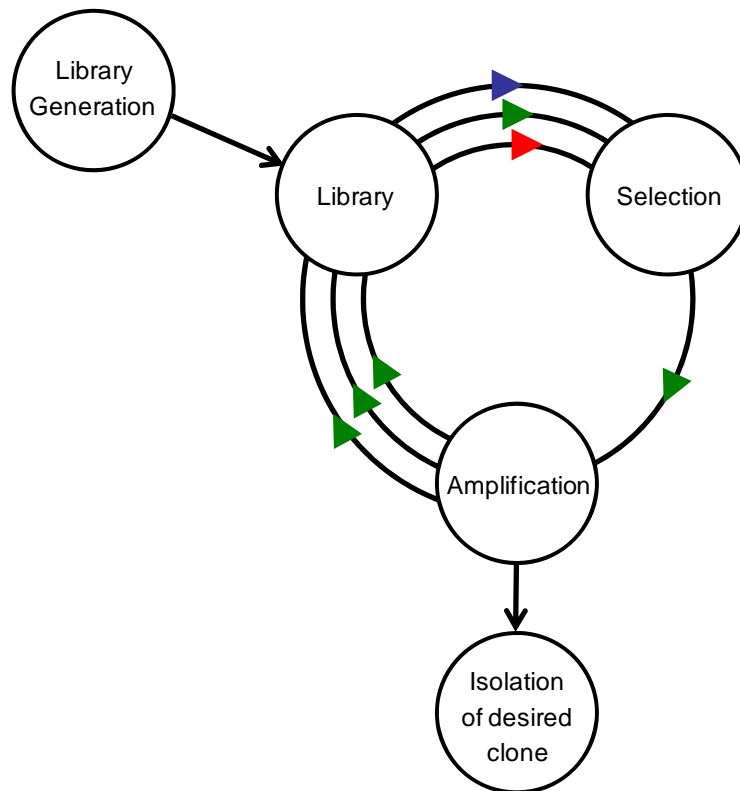


Figure 1.8 General *in vitro* selection scheme. Colored triangles represent individual clones.

A selection process that isolates novel functional nucleic acids of a specific sequence that tightly binds to a target molecule was independently developed in the separate labs of Gold, Joyce and Szostak.(Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990) The selected nucleic acid ligands are referred to as aptamers (derivative of Latin “to fit”). The value of this technique is due to the fact that both genotype and phenotype are expressed on the same biopolymer. This greatly simplifies selection of ligands, as there is no need for decoding and, in the case of DNA-based ligands, the sequence can be obtained by directly sequencing the ligand. These aptamers can also be amplified using PCR either directly, or following reverse transcription for RNA ligands that generally must be converted to DNA. This process of using nucleic acid for the selection of ligands was coined Systematic Evolution of Ligands by Exponential Enrichment (SELEX).

Every selection begins with the construction of a library. A library is a collection of biopolymers where each individual biopolymer sequence has one or more positions partially or totally randomized. Therefore, any one molecule in the library has a high probability of being different from any other molecule. For a DNA-based library, there may be up to 10^{15} unique sequences. This number represents an upper limit as larger libraries would be more mass than what is practical to work with. For example, 10^{15} ssDNA molecules 80 nucleotides in length would be about 50 μg of material. Protein libraries, on the other hand, can be constructed through the mutation of the wild-type gene at the DNA level. This DNA would then have to be introduced or transfected into a cell for expression. Due to the limitations caused by transfection efficiencies of the cell, protein libraries reach an upper limit of about 10^9 . Therefore, functional nucleic acid selections have the advantage over protein selections in terms of library size as they are not limited by transfection efficiency. Advances in *in vitro* peptide library expression using mRNA display(Roberts & Szostak, 1997) and ribosome display(Hanes & Pluckthun, 1997) have allowed for peptide and antibody libraries to reach sizes on par with nucleic acid libraries by reducing the limitation caused by transfection efficiencies. However, in the selection of protein ligands using mRNA display, and ribosome display, additional steps must be incorporated to allow the genetic material to be attached to the ligand.

The overall change in the genetic makeup of a species through partial randomization of known structures followed by the isolation of the individuals with common traits is referred to as evolution. The process generally involves mutation of DNA that gives rise to altered RNA and ultimately to altered protein. Evolution is often used to enhance certain qualities of protein enzymes, and due to the limitations on the current knowledge in protein folding, protein mutants are often evolved from a wild-type structure and not selected from a fully random library. *In vitro* selection of nucleic acid ligands or enzymes allow for the isolation of functional biomolecules that have no precedence or basis of design on any existing biomolecule. Not limited to *in vitro* selection, functional nucleic acids can also be further enhanced by using evolution on a selected aptamer or nucleic acid enzyme.

1.2.1.1 *In vitro* selection of ribozymes

There is little diversity in the reactions catalyzed by natural ribozymes.(Fedor & Williamson, 2005) Many of the reactions are RNA-modifying in nature (eg. ligation, cleavage, and splicing), but the complexity of the ribosome and the existence of nucleotide-like co-factors (eg. nicotinamide adenine dinucleotide) used by proteins suggest that ribozymes have the potential to have catalyzed a wide range of reactions. Using SELEX to select for catalysis, researchers have discovered many ribozymes capable of novel reactions.(Franzen, 2010) To complement catalytic activities found in natural ribozymes such as ligation, cleavage and splicing, *in vitro* selected ribozymes cover a new range of catalysis such as ribonucleotide synthesis,(Lau, Cadieux, & Unrau, 2004; Unrau & Bartel, 1998) polymerization,(Zaher & Unrau, 2007) carbon-carbon bond formation,(Seelig & Jäschke, 1999; Tarasow *et al.*, 1997) alcohol dehydrogenation,(Tsukiji, Pattnaik, & Suga, 2003) and reduction,(Tsukiji, Pattnaik, & Suga, 2004) to name a few. Such a diverse range of reactions, with yet more to be found, strongly support the RNA World hypothesis. Regardless of the catalytic activity that is sought through selection techniques, the critical step to selecting clones capable of catalysis is the necessary and often facile differentiation between active and inactive clones. Strategies often include a selection pressure that demands a chemical bond

forming or bond breaking reaction that results in a physical change such as a change in mass through the conjugation or removal of a chemical group. The act of bond breaking is of considerable interest and utility in a field closely related to ribozymes - the field of DNazymes.

1.2.1.2 *In vitro* selection of DNazymes

The discovery of both natural ribozymes and the methodology used to select for novel ribozymes led to the hypothesis that nucleic acid enzymes can be engineered to be more amenable to chemical reactions and therapeutic applications. DNA, having the superior stability towards degradation, but otherwise many of the same properties governing synthesis and manipulation is a suitable alternative biopolymer for the discovery of nucleic acid enzymes, even though it bears one less functional group (absence of the 2' hydroxyl). Another advantage to using DNA is the ability to directly amplify the biopolymer using PCR without the need to first perform reverse transcription. Finally, solid phase chemical synthesis of DNA is far less expensive than the solid phase synthesis of RNA, which requires the use of additional protecting groups.

Unlike ribozymes, DNA enzymes to date are not known in nature. Nevertheless, researchers have successfully selected deoxyribonucleic acid enzymes (DNazymes) capable of performing novel catalytic function (Baum & Silverman, 2008; Franzen, 2010; Joyce, 2004; Schlosser & Li, 2009; S.K. Silverman, 2004, 2005; S. K. Silverman, 2008). Among the chemical reactions catalyzed are porphyrin metallation,(Y. F. Li & Sen, 1996), ribophosphodiester bond cleavage,(S.K. Silverman, 2005) metal sensing,(J. Liu *et al.*, 2007; J. Liu & Lu, 2003, 2004, 2007a, 2007b, 2007c; Z. Liu, Mei, Brennan, & Li, 2003; Schonbrunner *et al.*, 2006) ligation,(Joyce, 2004; Pradeepkumar, Höbartner, Baum, & Silverman, 2008; Purtha, Coppins, Smalley, & Silverman, 2005; Wang & Silverman, 2005a, 2005b) oxidation,(Travascio, Li, & Sen, 1998) and T-T dimer reversion to name a few.(Chinnapen & Sen, 2004) The following sections 1.2.2 – 1.2.4 will focus on strategies that involve nucleolytic cleavage of a ribophosphodiester bond by *in vitro* selected DNazymes.

1.2.2 Ribophosphodiester bond cleavage by DNazymes

In section 1.1.2, the application of mRNA cleavage in biotechnology and therapeutics was presented, while sections 1.1.2.1 and 1.1.2.2 described two important aspects, specificity and catalytic rate constants, respectively; of RNA cleavage that researchers seek to optimize. Whereas the hammerhead and other ribozymes illustrate the potential of attaining high substrate specificity, possibly at the expense of k_{cat} , RNaseA represents the benchmark for an extraordinarily high rate constant, k_{cat} , for catalytic cleavage. Section 1.2.1.2 discusses the benefits of using DNA over RNA in the discovery of a nucleic acid-based enzyme. Breaker and Joyce isolated the first DNA-based enzyme in 1994.(Breaker & Joyce, 1994) The selection strategy that they used involved an oligonucleotide bound at the 5' end using a biotin streptavidin system (Figure 1.9). The resulting oligonucleotide contained an embedded ribose followed by an engineered stem loop to place a region of random sequence close to the ribose cleavage site. Sequences capable of undergoing self-cleavage of the ribonucleotide bond will be able to liberate themselves from the solid support. The cleaved oligonucleotides can then be physically isolated from the bound inactive sequences.

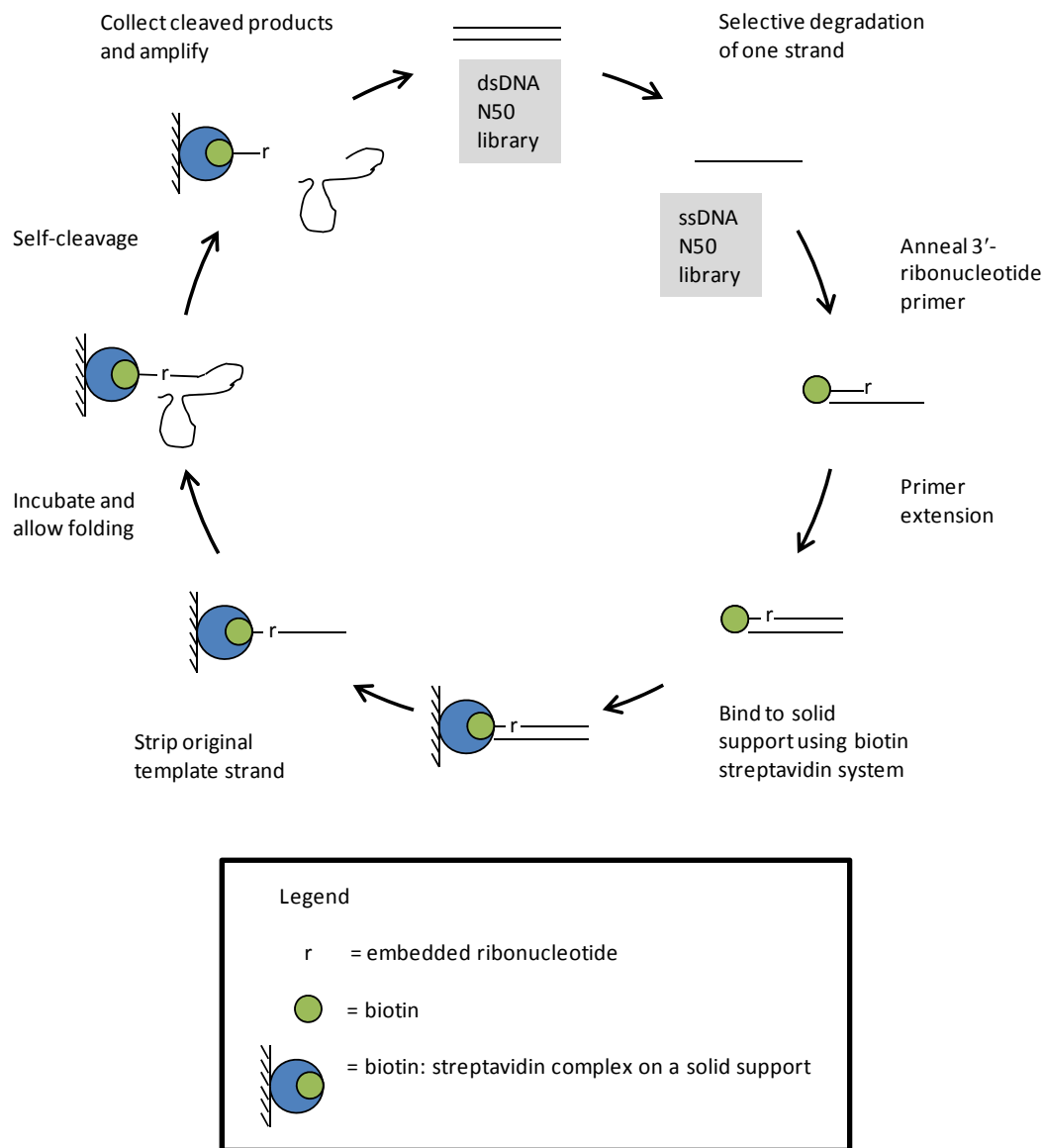


Figure 1.9 The method for selecting a ribophosphodiester bond-cleaving DNAzyme that was developed by Joyce and coworkers. This selection scheme became the general format for many subsequent self-cleaving DNAzyme selections.

Substrate specificity of DNAzymes that use Watson-Crick base pairing for target substrate recognition can be optimized by simply altering the guide sequence. A high catalytic rate constant, however, could not be achieved using the limited functional groups available to DNA. Therefore, divalent lead (Pb^{2+}), was included in the selection

buffer. The selected DNAzyme, much like the hammerhead ribozyme is dependent on metal cations to perform catalysis. While high concentrations of Pb^{2+} can be used in *in vitro* applications, Pb^{2+} is not present in the cell in significant concentrations. To better represent actual physiological conditions, magnesium cations were used in a subsequent selection (Breaker & Joyce, 1995). Magnesium cations are found in relatively low concentrations in the cell (~ 0.5 mM), but in comparison to lead cations, which are not bioavailable, magnesium cations represent cofactors that might prove useful. (Murphy, Freudenrich, Levy, London, & Lieberman, 1989) In a later selection, Joyce and coworkers addressed the higher availability of magnesium cations by selecting for divalent magnesium cation-dependent DNAzymes 8-17 and 10-23 that cleaved RNA. (Santoro & Joyce, 1997) In this thesis, DNAzymes will often be referred to with the prefix Dz. Under multiple turnover conditions, Dz10-23 was reported to have a k_{cat}/K_M of $4.5 \cdot 10^9 \text{ M}^{-1} \text{ min}^{-1}$ (50 mM MgCl_2 , 37 °C). Constructs based on Dz8-17 and Dz10-23 are currently being used for *in vivo* experiments and biomolecular engineering. For *in vivo* applications, one must consider the physiological availability of Mg^{2+} compared to the high concentrations used to reach high rate constants *in vitro*.

1.2.3 Divalent metal cation-independent cleavage

While Mg^{2+} -dependent DNAzyme catalytic rate constants reached $> 1 \text{ min}^{-1}$, the Mg^{2+} concentrations used to attain these rates were on the order of ~ 10 mM. These concentrations of magnesium cations far exceed the *in vivo* concentrations (~ 0.5 mM) found in the cell. (Murphy, Freudenrich, Levy, London, & Lieberman, 1989) Dependence on divalent metal cations obviously limit the efficacy of DNAzymes in physiological conditions. Therefore, a divalent metal cation-independent DNAzyme would be desirable. Appreciable k_{cat} values for DNAzymes in the absence of high concentrations of divalent metal cations are often difficult to obtain. However, nucleic acid enzymes are not completely dependent on divalent metal cations. For example, the hammerhead ribozyme was found to perform detectable divalent metal cation-independent catalysis in the presence of a high concentration of monovalent cation e.g. (Na^+ , Li^+ , or K^+) to promote proper folding. (Murray *et al.*, 1998) This is powerful evidence that the chemical

functionality of the hammerhead ribozyme is sufficient for divalent metal cation-independent catalysis while the role of the divalent M^{2+} is therefore restricted to folding on otherwise catalytically competent M^{2+} -independent species.

Seeking to avoid divalent metal cation dependence altogether, Faulhammer *et al.* attempted to select for a histidine dependent self-cleaving DNAzyme with minimal dependence on divalent metal cations.(Faulhammer & Famulok, 1996) Unfortunately, the isolated DNAzyme was discovered to be histidine-independent and dependent on either Mg^{2+} or Ca^{2+} . The authors surmised that the zwitterionic histidine may be poorly bound to the DNAzyme due to the lack of strong electrostatic attraction with the negatively charged DNAzyme. Three subsequent selections was used to assess the dependence of DNAzymes on divalent metal cations and one of the selections incorporated the use of a positively-charged cofactor spermine.(Faulhammer & Famulok, 1997) The second selection was carried out using Mg^{2+} and the third selection used no cofactors. All three led to the isolation of active DNAzymes.

In the same vein, in 1997, Geyer and Sen performed a selection leading to the isolation of a divalent metal cation-independent DNAzyme, G3, and the further evolved Na8.(Geyer & Sen, 1997) The selection of these divalent metal cation-independent DNAzymes is concrete evidence that DNA, like the hammerhead ribozyme, bears sufficient functionality for catalysis that does not absolutely necessitate a divalent metal cation. With this evidence, the development of divalent metal cation-independent DNAzymes turned towards the enhancement of catalytic rate. While the goal of selecting divalent metal cation-independent DNAzymes was to remove the dependence on cofactors that are not in abundance, it is noteworthy that Roth *et al.* successfully isolated a histidine-dependent DNAzyme capable of self-cleavage in the absence of divalent metal cation.(Roth & Breaker, 1998) Undoubtedly, the focus on histamine and amine derived inspiration from the M^{2+} -free mechanism of RNase A that employs the amino acids histidine and lysine.

1.2.4 Ribophosphodiester bond cleavage by modified DNazymes

Progress towards making nucleic acid enzymes more protein-like started with the assessment of the minimal functionality necessary for catalysis. That is, answering the question whether or not nucleic acid, without the assistance of any protein-like functionality in the absence of divalent metal cations, can support catalysis. Doing so has led to the discovery that DNA meets the minimum functionality requirements necessary for catalysis.(Faulhammer & Famulok, 1997; Geyer & Sen, 1997) Investigators sought to overcome DNA's chemical deficiencies by employing cofactors. The selections that introduced other non-metal cation small molecules such as histidine and spermine offered an improvement in catalysis.(Faulhammer & Famulok, 1997; Roth & Breaker, 1998) To ensure that the nucleic acid enzymes bear the necessary functional groups, chemical conjugation can be used to append the functional groups directly to nucleic acids.

The first experiment to utilize modified nucleotides for *in vitro* selection was performed by Latham *et al.* in 1994.(Latham, Johnson, & Toole, 1994) In this selection, a thrombin-binding aptamer was selected using 5-pentynyl-dUTP in the construction of the oligonucleotide. While this aptamer did not outperform existing unmodified aptamers and was, in fact, four to ten times weaker in binding than the unmodified aptamers, the modifications were shown to be essential for aptamer function. Nevertheless, this work proved researchers' ability to incorporate designer nucleotides into DNA and successfully perform a modification-dependent selection leading to the desired function. The first selection of a chemically modified nucleic acid enzyme was done using RNA in 1997. Tarasow *et al.* reported the *in vitro* selection of a pyridyl-modified ribozyme capable of Diels-Alder bond formation.(Tarasow *et al.*, 1997) While it was later found that this reaction can be catalyzed using unmodified nucleotides, (Seelig & Jäschke, 1999) Tarasow's Diels-Alder ribozyme (DA-ribozyme) was the first modification-dependent nucleic acid enzyme. Just as DNA was used as an alternative biopolymer for RNA in the selection of unmodified nucleic acid enzymes, modified deoxyribonucleotides can be used in place of natural deoxyribonucleotides in the selection for modified DNazymes.

Natural nucleotides do not have pK_a values near neutrality. This highly prized property is exemplified by the imidazole functionality whose pK_a value is near neutrality (~ 6.95). Santoro *et al.* first introduced protein-like functionality to DNazymes with the

introduction of a chemically appended 4-imidazoleacrylic acid group to the primary amine of 5-(aminoallyl)-2'-dUTP using an amide bond.(Santoro, Joyce, Sakthivel, Gramatikova, & Barbas, 2000) Modified oligonucleotides were synthesized by transcribing the random region using the natural nucleotides of dA, dC, dG and the modified dUTP. This DNAzyme was selected in the presence of Zn^{2+} and was found to be Zn^{2+} -dependent. With the addition of Zn^{2+} , this DNAzyme proved capable of ribophosphodiester bond cleavage through an undetermined mechanism. While this selection successfully integrated modifications onto functional DNAzymes, the contribution of the modification in the catalytic mechanism was not well defined as the DNAzyme was zinc dependent and zinc-dependent unmodified DNAzymes can perform the same function.(J. Li, Zheng, Kwon, & Lu, 2000)

Introducing modified nucleotides into combinatorial selections for a divalent metal cation-independent ribophosphodiester bond cleaving DNAzyme, Perrin *et al.* introduced two protein-like functional groups into a DNAzyme selection and succeeded in isolating the intramolecular or *cis*-cleaving divalent metal cation-independent DNAzyme, Dz9₂₅-11 ($k_{\text{obs}} = 0.044 \text{ min}^{-1}$) (Figure 1.10).(Perrin, Garestier, & Hélène, 2001) Dz9₂₅-11 is the first DNAzyme to utilize two protein-like functionalities to catalyze divalent metal cation-independent ribophosphodiester bond cleavage. The goal was to produce an RNase mimic that truly replicates the high efficiency transesterification reaction found in RNaseA. A study done by Thomas and Perrin assigned the putative roles of the general acid, general base, and charge stabilizing amine to the modified nucleobases A, A, and U, respectively, of Dz9₂₅-11.(Thomas, Yoon, & Perrin, 2009) This study used a combination of modification knockouts and affinity tagging to determine the critical nucleobases and confirmed that it is indeed the appended chemical functionality that is responsible for the catalysis. Solid phase synthesis of a Dz9₂₅-11 variant capable of intermolecular or *trans* cleavage was also successfully produced.(Lermer, Roupioz, Ting, & Perrin, 2002)

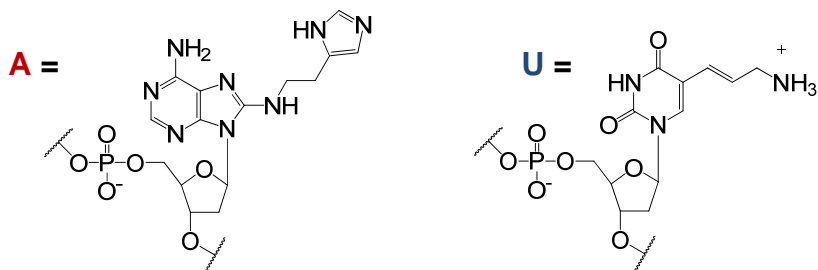
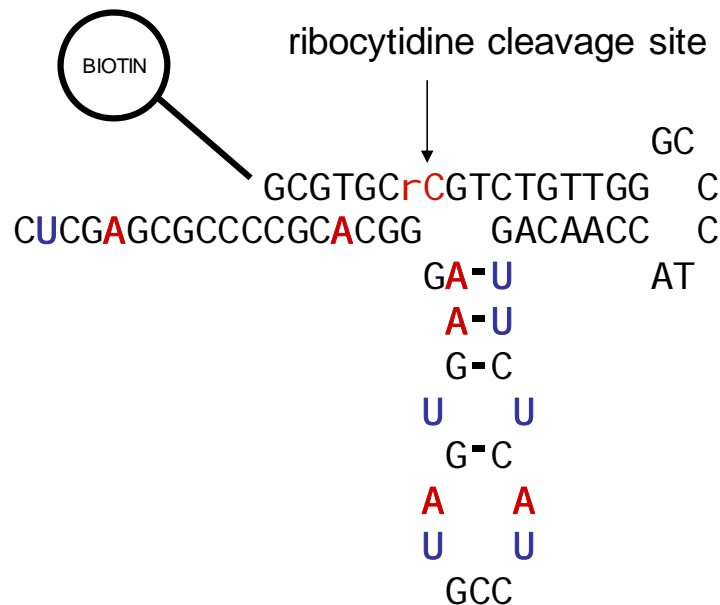


Figure 1.10 Primary sequence and secondary structure of *cis*-cleaving Dz9₂₅-11. 8-Modified-dA is represented by a red A. 5-modified dU is represented by a blue U. Embedded ribonucleotide is indicated by an rC.

Similarly, Sidorov *et al.* selected for a DNAzyme utilizing the amino and imidazole functionalities on dA and dU, respectively. (Sidorov, Grasby, & Williams, 2004) The imidazole functionality was introduced using the same 4-imidazolyl acrylamide-modified dUTP that Santoro used as opposed to the 8-modified dA used in 9₂₅-11. Unlike 9₂₅-11, which targets a single embedded ribophosphodiester bond, Sidorov's DNAzyme cleaves a 12 nucleotide (nt) section of RNA. Cleavage occurs at either one of two UpA positions within the RNA segment. Interestingly, synthesis of the

DNAzyme using only natural deoxynucleotides resulted in the retention of 7 % self-cleavage activity compared to the fully functionalized DNAzyme. The *trans* variant of Sidorov's DNAzyme has yet to be produced.

1.2.5 Expanding the chemical landscape

Hollenstein *et al.* continued the protein mimicry by introducing a third functional group, the guanidinium, into a DNAzyme selection. (M. Hollenstein, C. Hipolito, C. Lam, & D. M. Perrin, 2009a) This third functional group was chosen for its positive charge, which facilitates stabilization of nucleic acid duplex structures and could potentially offer additional electrostatic stabilization to negatively charged intermediates. The result of this selection was Dz9-86 (Figure 1.11). Dz9-86 conveyed a higher catalytic rate constant ($k_{\text{obs}} = 0.134 \pm 0.026 \text{ min}^{-1}$, 24 °C) and increased thermostability over Dz9₂₅-11. Recently, a subsequent selection was performed that utilized a region composed of 40 degenerate nucleobases – double the length of the degenerate region used in the Dz9-86 selection. (M. Hollenstein, C. J. Hipolito, C. H. Lam, & D. M. Perrin, 2009b) The DNAzyme isolated from this selection was named Dz10-66 ($k_{\text{obs}} = 0.50 \pm 0.05 \text{ min}^{-1}$, 24 °C). While the selection of Dz10-66 was not an attempt to increase the diversity of the chemical landscape, it does allow access to a larger quantity of functional groups and potentially larger catalytic motifs, and challenged the concept of the “tyranny of the small motif,” which states that small catalytic sequences are preferentially selected, thereby reducing the chance of discovering larger, potentially more efficient, catalysts. (Joyce, 2004) Successful selection of Dz10-66, the fastest divalent metal cation-independent DNAzyme utilizing 3 amino acid-like functionalities sets the current benchmark for modified DNAzymes catalyzing RNA cleavage. Dz10-66 was further studied by creating a derivative, named Dz10-66t that is capable of performing intermolecular cleavage of a DNA substrate with an embedded ribose. The second order rate constant, $k_{\text{cat}}/K_{\text{M}}$, of this construct capable of true catalytic turnover was determined to be $6 \cdot 10^5 \text{ M}^{-1} \text{ min}^{-1}$. This value falls short of the second order rate constants of the hammerhead ribozyme, RNaseA and 10-23 in their respective optimized conditions, but shows the field of modified DNAzymes for use in physiological conditions is progressing steadily and warrants further exploration.

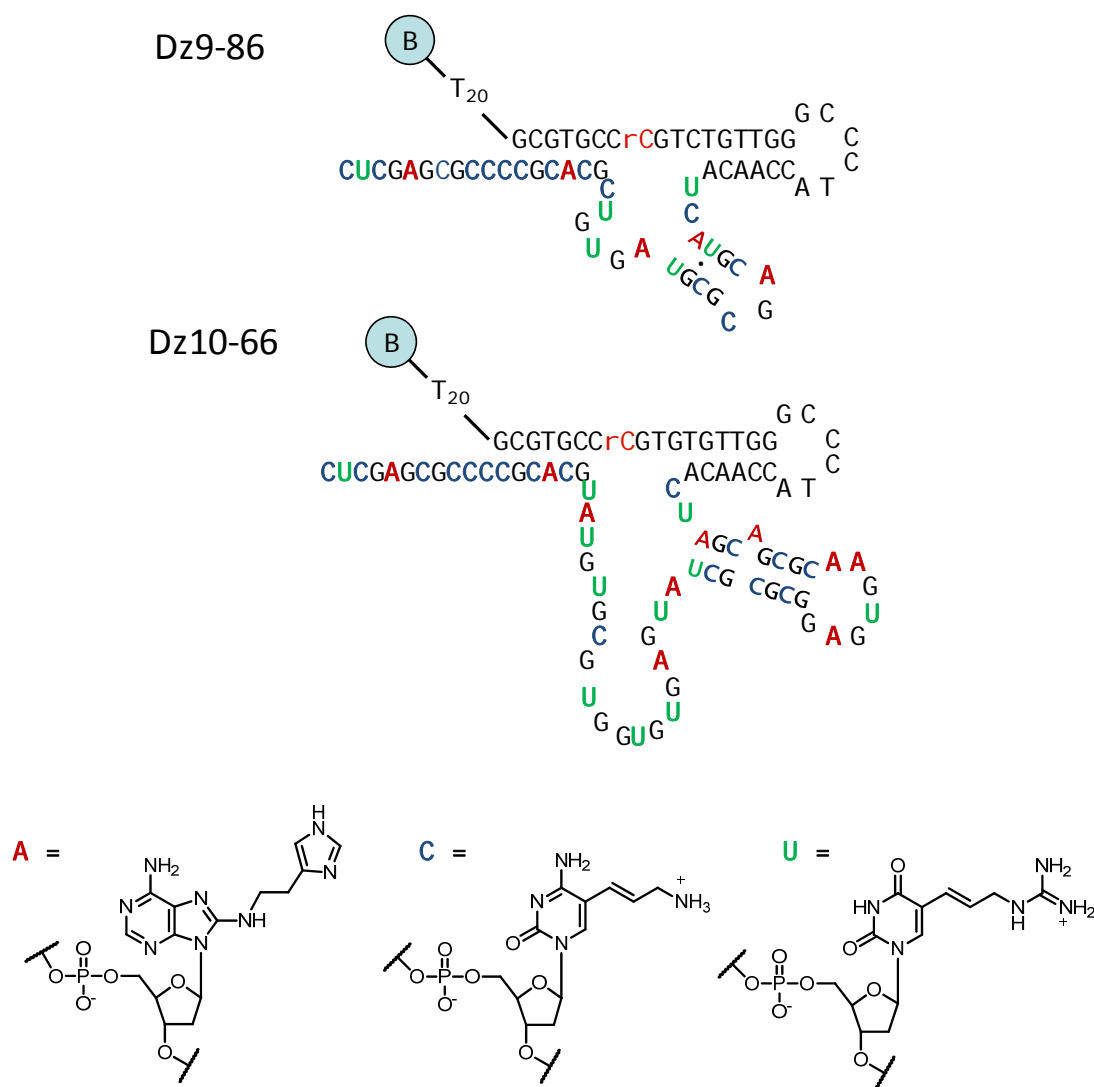


Figure 1.11 Primary sequence and secondary structure of *cis*-cleaving DNazymes 9-86 and 10-66. 8-Modified dA, 5-modified dC and 5-modified dU are represented by boldface A, C and U, respectively. Embedded ribonucleotide is indicated by a red rC. The modified U: G mismatch is indicated with a middle dot (•).

1.3 Enzymic recognition of modified DNA

The phenotype of a species is the total collection of the expressed traits and dictates the survival of a species through a selection. In the previous sections, *in vitro* selections were designed to isolate nucleic acid enzymes according to their ability to perform one specific function such as ribophosphodiester bond cleavage. However, this one function is one of many possible traits unique to that species. Other traits of the

species may come into play when considering the fitness of a species as it progresses through the generations of the selection. With regards to modified DNazymes, ease of synthesis and ease of replication are two traits that are sometimes overlooked due to the simplicity of these two procedures when working with unmodified DNazymes. Construction of modified DNA by polymerases plays a large role in the *in vitro* selection of modified DNazymes and negative effects caused by modified nucleotides cannot be overlooked. The activity of other nucleic acid-modifying protein enzymes besides polymerases can also be affected by modified nucleotides. Another class of protein enzyme, the restriction enzyme, also recognizes a specific sequence and cleaves dsDNA. While not used in our selection of DNazymes, selective destruction of modified and unmodified DNA has been the basis for other *in vitro* selections (Tawfik & Griffiths, 1998) and has potential application in the future selection of modified DNazymes.

1.3.1 Incorporation of modified nucleotides

The synthesis of a full-length unmodified DNazymes is generally a simple process as every step of the selection, including the primer extension to produce the full-length DNazyme, occurs at high efficiency. In the case of modified DNazymes, however, the efficiency of extension is lowered by the use of poorly incorporated modified nucleoside triphosphates. Therefore primer extension or elongation using modified nucleoside triphosphates introduces another selection pressure. The polymerases must be able to synthesize highly modified clones with efficiency. Otherwise, the highly modified clones will not be fully elongated. These poorly elongated clones will not be selected regardless of the potential activity of the unrealized fully elongated construct. Many studies using modified DNA have required the efficient incorporation of modified nucleotides. (Cahova, Havran *et al.*, 2008; Cahova, Pohl *et al.*, 2008; Hocek & Fojta, 2008; Lam, Hipolito, & Perrin, 2008; Macickova-Cahova & Hocek, 2009) To facilitate incorporation, investigators have engineered functionalized nucleotides for improved uptake. Famulok and coworkers have designed modified nucleoside triphosphates that are incorporated quite efficiently by commercial polymerases, and successfully produced the first functionalized DNA with all four

nucleotides replaced with modified nucleotides using primer-mediated incorporation.(Thum *et al.*, 2001) A notable characteristic of the linkers used in the modifications is the use of rigid olefins tethered to a position on the nucleobase that orients the modification into the major groove. Common attachment points that orient the modification towards the major groove are the 7 position of 7-deazapurines and the 5 position of pyrimidines (Figure 1.3). The ability to produce highly modified DNA allows investigators to study the elongation of DNA using a highly modified DNA template.(Capek *et al.*, 2007; Gurlain *et al.*, 2001; Hollenstein *et al.*, 2008; Hollenstein *et al.*, 2009a; Hollenstein *et al.*, 2009b; Jäger & Famulok, 2004; Jäger *et al.*, 2005; M. Kuwahara, Hanawa *et al.*, 2006; M. Kuwahara, Nagashima *et al.*, 2006; M. Kuwahara *et al.*, 2003; Latham *et al.*, 1994; Lee *et al.*, 2001; Masud, Kuwahara, Ozaki, & Sawai, 2004; Obayashi *et al.*, 2002; Ohbayashi *et al.*, 2005; Ohsawa *et al.*, 2008; Perrin, Garestier, & Hélène, 1999; Perrin *et al.*, 2001; Sakthivel & Barbas, 1998; Santoro *et al.*, 2000; Shoji, Kuwahara, Ozaki, & Sawai, 2007; Sidorov *et al.*, 2004; Thum *et al.*, 2001; Vaught *et al.*)

1.3.2 Transcription of a modified DNA template

To test whether or not there is an amplification bias for certain unmodified DNAzyme sequences, a study was done by Schossler *et al.* that measured the amplifiability of the unmodified DNAzyme gene pools based on their generation.(Schlosser, Lam, & Li, 2009) Their results suggest that there is little to no amplification bias for unmodified DNAzymes. A similar study has yet to be done on modified DNAzymes although work presented herein begins to address this issue. Nevertheless, amplification bias has been found in past selections and was in fact the basis of selection in the very first *in vitro* evolution, which involved the preferential amplification of viral genomic RNA that were easily copied.(Mills, Peterson, & Spiegelman.S, 1967) The final molecules of the evolution were only 15 % of the original length of the genome. The majority of the sequence has been removed leaving only the portions critical for RNA amplification. Speiglmán and coworker's *in vitro* evolution on the amplifiability of genomic RNA illustrated the overwhelming effect of how the

superior amplifiability of a species allows it to dominate the gene pool.(Mills *et al.*, 1967)
One of the best current examples of an amplification bias in the presence of non-canonical bases comes from the work of Kimoto *et al.*, which shows that a single non-canonical base pair has enough effect on its surrounding such that certain sequences flanking that base pair are preferentially amplified. Their work involves the PCR amplification of an unnatural base pair 7-(2-thienyl)-imidazo[4,5-b]pyridine (**Ds**) and 2-nitro-4-propynylpyrrole (**Px**).(Kimoto, Kawai, Mitsui, Yokoyama, & Hirao, 2009)
Preferred amplification of sequences resulting from the presence of modified nucleobases has yet to be explored.

The critical step in the unbiased amplification of modified DNA is the ability of a polymerase to efficiently use modified DNA as a template and synthesize a complementary strand with fidelity. The synthesis of the complementary strand by polymerase is referred to as read-through. After successful polymerization of four modified nucleotides, Famulok and coworkers investigated read-through of modified templates, which they then had access to.(Jäger & Famulok, 2004; Jäger *et al.*, 2005; Thum *et al.*, 2001) What was discovered was that Family A polymerases frequently failed to produce full length product while Family B polymerases were far more successful. Characteristics of these polymerase families will be discussed in section 1.4.1. The assessment of the commercial polymerases' abilities to incorporate modified nucleotides and read-through modified templates opens the potential for the generation of dsDNA that bears modifications on both strands referred to as doubly-modified dsDNA.

1.3.3 Amplification of doubly-modified dsDNA

If a modified nucleotide is a substrate for high-efficiency incorporation using a template strand bearing the same modification, the modified nucleotide could potentially be used in polymerase chain reaction (PCR) (Figure 1.12) to produce doubly-modified dsDNA. PCR is the exponential amplification of a nucleic acid template using multiple thermal cycles consisting of denaturation of a dsDNA template at about 94 °C, primer-template hybridization at about 55 – 65 °C and polymerase-mediated extension at about 72 °C. The two newly extended products can serve as templates for the next cycle. The

amount of dsDNA is doubled with every round; giving the theoretical 2^n times the amount of starting template where n is the number of cycles.

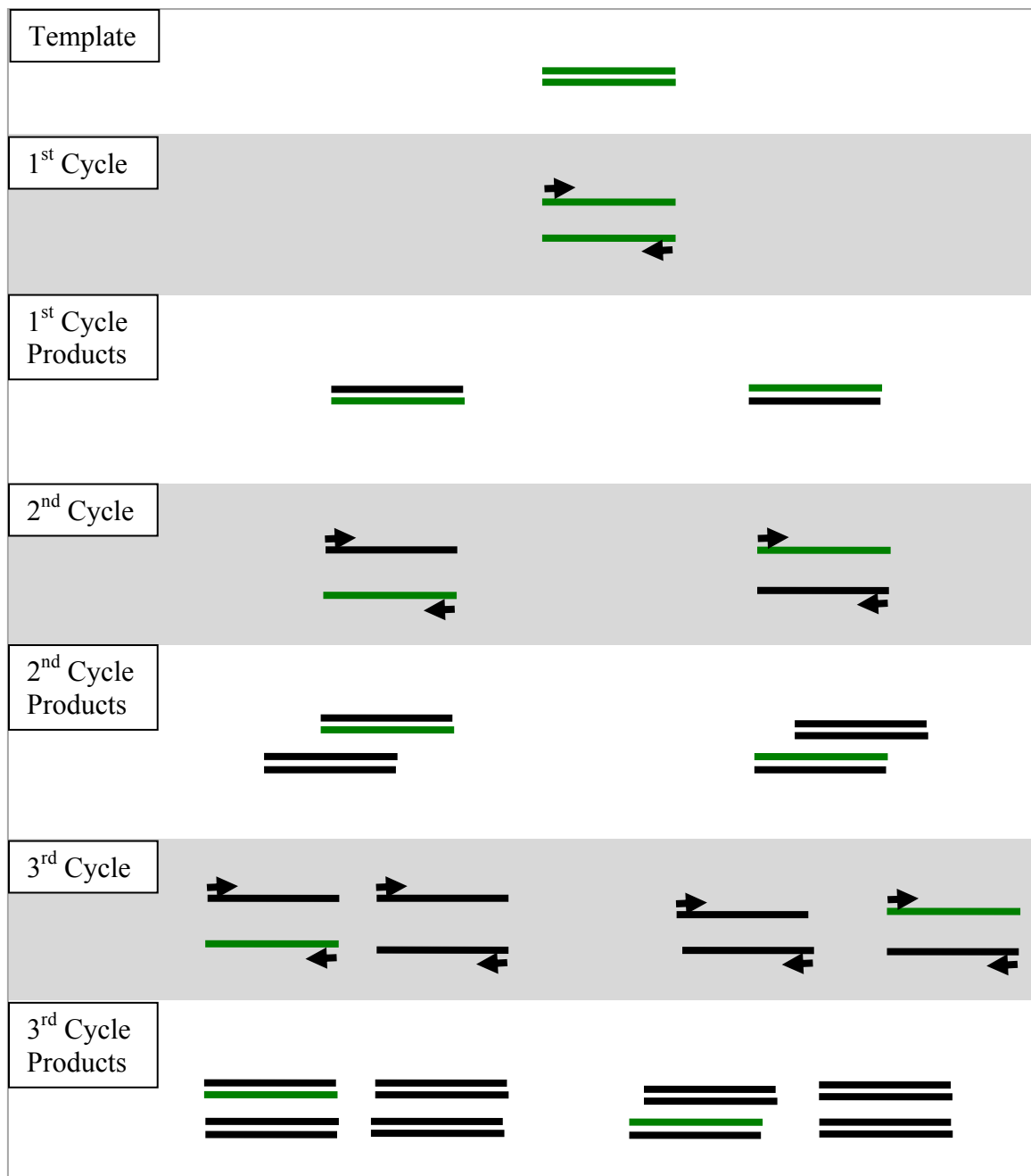


Figure 1.12 Polymerase chain reaction (PCR). Each cycle involves a step carried out at each of three different temperatures. Green lines represent the original DNA template. Black arrows represent primers. Black lines represent the newly synthesized DNA. 1. Template DNA is denatured at high temperature. 2. Temperature is lowered and primers anneal to the single-stranded template DNA. 3. Temperature is raised to the optimal extension temperature for the thermostable polymerase. Both the original DNA and the newly synthesized DNA are used as template for the subsequent cycle. The theoretical amount of DNA doubles with each cycle.

The amplification of doubly-modified dsDNA using a modified nucleoside triphosphate is not necessary for a DNAzyme selection as one polymerase can be used exclusively for incorporation of modified dNTPs and a different polymerase can be used exclusively for read-through and amplification. However, efficient amplification of doubly-modified dsDNA is a rapid, although not exclusive, method of simultaneously assessing incorporation and read-through efficiencies. Several groups have investigated the successful PCR amplification of DNA through the use of a chemically augmented nucleotide in place of its natural counterpart.(Gourlain *et al.*, 2001; M. Kuwahara, Hanawa *et al.*, 2006; Masayasu Kuwahara *et al.*, 2003; M. Kuwahara, Nagashima *et al.*, 2006; M. Kuwahara *et al.*, 2003; Masayasu Kuwahara, Tamura, Kitagata, Sawai, & Ozaki, 2005; Obayashi *et al.*, 2002; Ohbayashi *et al.*, 2005; Sakthivel & Barbas, 1998; Sawai *et al.*, 2001) Most notable is the successful amplification in which multiple unnatural nucleotides are used in lieu of their natural counterparts.(Jäger & Famulok, 2004; Jäger *et al.*, 2005; Thum *et al.*, 2001) The resulting doubly-modified dsDNA is functionalized with a high density of chemical augmentation.

1.3.4 Restriction enzyme digestion of modified dsDNA

Restriction enzymes are part of an organism's defense mechanism against foreign genetic material.(Arber & Linn, 1969) Chemical modification of the organism's own DNA helps protect the organism's genome from destruction. Restriction enzyme digestion has become a powerful tool in biotechnology and is frequently used in cloning and identification of DNA through site-specific cleavage. Minimal literature exists regarding the resistance of chemically-functionalized DNA to restriction enzyme digestion. The resistance to cleavage by restriction enzymes of 5'-modified dUTP containing dsDNA was briefly mentioned.(Perrin *et al.*, 1999; Sakthivel & Barbas, 1998) Hocek and coworkers recently reported the resistance to degradation bestowed upon heteroduplex dsDNA by 8'-modified 2'-deoxyadenosine. (Macickova-Cahova & Hocek, 2009) Even modified nucleotides residing outside of a restriction site consisting of unmodified nucleotides bestow resistance to restriction enzyme digestion.(Perrin *et al.*, 1999) Further investigations of this potentially useful resistance should be spearheaded.

1.4 Evolved polymerases

There are three main DNA polymerase families of interest to the modified nucleotide community: Family A, Family B, and Family Y. (Braithwaite & Ito, 1993; Ohmori *et al.*, 2001) DNA polymerases with homology to the product of the *polA* gene *Escherichia coli* (*E. coli*) specifying DNA polymerase I are categorized into Family A. Common polymerases that fall into the Family A category are Klenow, Taq and T7 polymerase. Likewise, DNA polymerases with homology to the product of the *polB* gene *E. coli* specifying DNA polymerase II are categorized into Family B. Common commercial polymerases like Vent (exo-) and Pfu fall into the Family B category. The Family Y polymerases are low fidelity polymerases that have gained attention due to their ability to bypass DNA lesions. (Ohmori *et al.*, 2001) One of the more notable commercial Family Y polymerases is DNA polymerase IV (Dpo4). (Boudsocq, Iwai, Hanaoka, & Woodgate, 2001) The majority of the above mentioned polymerases are used in PCR. The following sections will focus on the evolution of Taq polymerase.

1.4.1 *Thermus aquaticus* DNA polymerase I

The first experiments in the development of PCR used *E. coli* DNA polymerase I. (Saiki *et al.*, 1985) After the extension step, this polymerase would heat-deactivate during the denaturation step of the following cycle, and fresh enzyme would need to be added with each cycle during the annealing phase. Replacement of this mesophilic enzyme with a homologous thermostable polymerase isolated from *Thermus aquaticus* (*T. aquaticus*) improved on this by minimizing polymerase inactivation and allowing for the polymerase to extend at much higher temperatures. (Saiki *et al.*, 1988) Annealing at higher temperatures is also improved due to the reduction of primer-template mismatches, which are more prone to occur at *E. coli* DNA polymerase I's operating temperature of 37 °C. The polymerase chain reaction took a huge leap forward with the replacement of mesophilic *E. coli* DNA polymerase I with thermophilic polymerase *T. aquaticus* DNA polymerase I or simply Taq (Figure 1.13). (Eom, Wang, & Steitz, 1996)

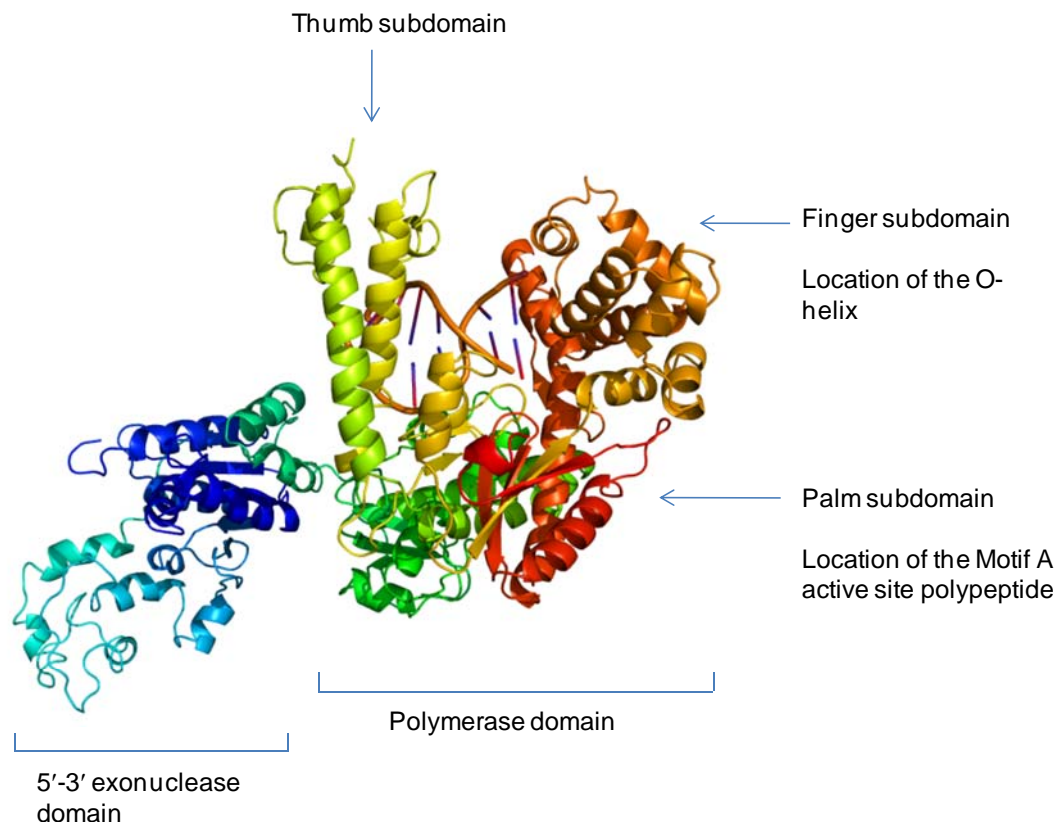


Figure 1.13 PyMol's cartoon representation of the crystal structure of *Thermus aquaticus* DNA polymerase I. PDB entry 1TAU.

Taq is an 832 amino acid nucleotidyl transferase enzyme and is composed of an N-terminal 5'-3' exonuclease domain linked to a C-terminal polymerase domain (Figure 1.13). The polymerases of Family A, Family B, and Family Y all have a characteristic right hand like structure consisting of a thumb domain, a palm domain and a finger domain.(Brautigam & Steitz, 1998) Each domain makes critical contact with its DNA substrate. The thumb domain makes contact with the double-stranded portion of the primer and template hybrid. After binding to the primer and template hybrid, Taq can accept an incoming nucleoside triphosphate and ensure proper Watson-Crick base pairing prior to initiating bond forming catalysis, pyrophosphate release and translocation. The finger domain, through a complex and dynamic process,(Patel, Suzuki, Adman, Shinkai, & Loeb, 2001) is responsible for ensuring that the incoming nucleoside triphosphate satisfies proper Watson-Crick base pairing. The palm domain contains the residues that

facilitate the attack by the primer's free 3' hydroxyl onto the α -phosphate of a bound nucleoside triphosphate. This attack is divalent metal cation-mediated and results in the elimination of a pyrophosphate group and a one base extension of the primer.

Translocation of the polymerase along the template strand allows access to the next unpaired template base. The process of incorporation can then be repeated until the polymerase reaches the 5' end of the template strand. Due to its "extendase" ability, Taq will add one additional dA at the 3' end in an untemplated fashion creating a one base 3' overhang. While critical to the biological functions of its organism, Taq's thermostability, processive speed and variable fidelity has additionally made it a powerful tool for *in vitro* studies particularly through its use in polymerase chain reaction (PCR)(Saiki *et al.*, 1988) and associated mutagenic techniques such as error prone PCR (epPCR).(Cadwell & Joyce, 1992)

Certain aspects of Taq continue to make it a very desirable enzyme to develop in terms of both protein engineering and evolution. Taq has a very well-studied active site; studies have shown that this active site is highly mutable and can be altered to produce novel catalytic activities.(Patel, Kawate, Adman, Ashbach, & Loeb, 2001; Patel & Loeb, 2000a) A domain, so called "Motif A" located at the palm region of the active site is mainly composed of a single polypeptide that is partially responsible for substrate recognition and catalysis. D610, located within Motif A, is a highly conserved residue that is responsible for chelating one of two critical magnesium cations.(Patel & Loeb, 2000a; Steitz, 1998) Mutation of this residue results in a total loss of polymerase activity. Every other residue in the active site polypeptide can be mutated to some extent without deactivating the enzyme.

Notable mutations to the active site are I614K and E615G. The I614K mutant retains polymerase activity but has low fidelity, allowing it to misincorporate bases.(Patel, Kawate *et al.*, 2001) E615 is referred to as the steric gate residue and is responsible for the discrimination of dNTPs from NTPs. Romesberg and coworkers have found in their evolution of the Stoffel Fragment that the E615G mutation was common to all successfully evolved polymerases.(Xia *et al.*, 2002) Holliger and coworkers found that their own isolated clones also contain the E615G mutation.(Ong, Loakes,

Jaroslowski, & Holliger, 2006) With such flexibility in the active site of the polymerase, Taq has great potential for customized activity.

As mentioned above, the full length Taq protein is composed of two domains. The N-terminal domain is a 5'-3' exonuclease attached to the C-terminal polymerase domain. This exonuclease domain catalyzes the degradation of the RNA primers used in the production of Okazaki fragments in discontinuous DNA synthesis. Removal of this domain does not ablate polymerase function. Two truncated versions of Taq polymerase are the KlenTaq,(Barnes, 1992) which is the M236-E832 amino acids, and Stoffel Fragment,(Lawyer *et al.*, 1993) which is the S290-E832 amino acids. While these truncated versions are slower at primer extension, they have increased thermostability and are often used for customized PCR using varying magnesium cation concentrations.

1.4.2 Evolution of Taq and Stoffel Fragment

Directed evolution is the change of the genetic makeup over several generations as a selection performed within an experimental timescale to reduce undesired mutants within a species and allows the selective proliferation of desired mutants. Three different techniques have been applied to Taq or Stoffel Fragment to isolate mutants with improved function. The first technique that will be discussed is a screening technique called complementation. Subsequently, two directed evolution techniques known as phage display and compartmentalized self-replication (CSR) will be described.

1.4.2.1 Complementation

Loeb and coworkers have probed the mutability of Taq's active site by partially randomizing the polypeptide located within the active site and screening for active mutants (Loeb 1996, 2000).(Patel & Loeb, 2000a; Suzuki, Baskin, Hood, & Loeb, 1996) The mutability of the polymerase finger subdomain was also determined in a similar fashion. The first region is a polypeptide located in Motif A in the palm subdomain of the polymerase domain (Figure 1.14).(Patel, Kawate *et al.*, 2001; Patel & Loeb, 2000a) The second region is the O-helix of the finger subdomain (Figure 1.15).(Suzuki, Avicola,

Hood, & Loeb, 1997) For both screens, the wild type sequences for those regions were partially degenerate.

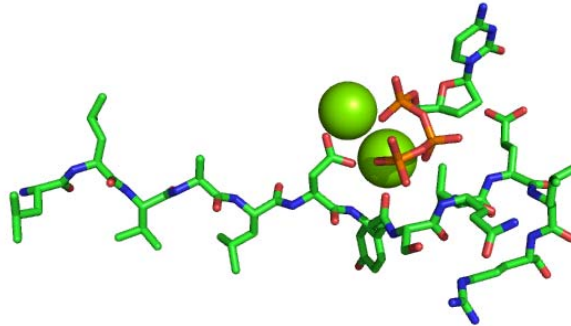


Figure 1.14 Line representation of the active site polypeptide region contained within the Motif A and the bound dideoxycytidine triphosphate from the crystal structure of Taq. PDB entry 1TAU. Green, blue, red, and orange lines represent carbon, nitrogen, oxygen and phosphorus, respectively. Amino acid residues 605 to 617 are shown. Asp610 and the triphosphate from the dideoxycytidine triphosphate are coordinated to magnesium cations, which are represented by green spheres.

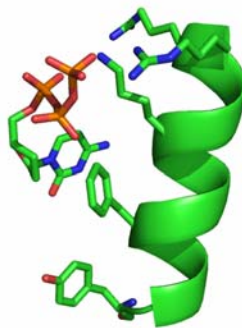


Figure 1.15 PyMol's cartoon and line representation of the O-helix and a bound dideoxycytidine triphosphate from the crystal structure of Taq. PDB entry 1TAU. Green, blue, red, and orange lines represent carbon, nitrogen, oxygen and phosphorus, respectively. Shown is the α -helix composed of amino acid residues 659-671. For clarity, only the side chains of residues Y659, F663, L667, R670 and R671 are shown. Residues Y659 and F663 are implicated in the base pairing process, and residues L667, R670, and R671 provide electrostatic stabilization for the negatively-charged triphosphate.

The host cell, the cell whose machinery is being used to express protein, is in this case is a thermosensitive strain of *E. coli*. The thermosensitive *E. coli* strain, recA718/polA12, was used in the genetic complementation to find Taq mutants capable of compensating for the host's own DNA polymerase I that is inactive at 37 °C. At 37 °C, the host cell does not have sufficient DNA polymerase I activity to remain viable (Figure 1.16). The screen identifies the mutant Taq polymerases with sufficient activity to complement the loss of the host cell's DNA polymerase I activity. Using this system, two libraries were screened for active polymerases containing mutations in two specific regions. Both the active site and O-helix were found to be highly tolerant to amino acid changes. However, there are some amino acids that are highly conserved. Aspartate 610, for example, binds to a critical magnesium cation and it was discovered that mutation of this amino acid to any other amino acid completely eliminates activity. All active mutants were found to have this residue unchanged. In the O-helix, the residues oriented towards the nucleotides within the active site were found to be resistant to mutation, but the residues oriented away from the substrate could tolerate mutation.

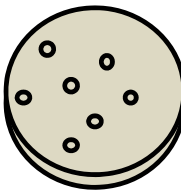
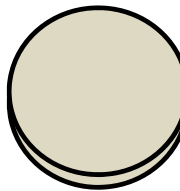
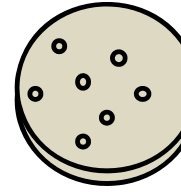
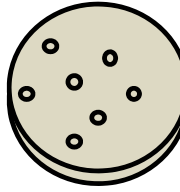
	30 °C	37 °C
recA718/polA12		
recA718/polA12 expressing Taq		

Figure 1.16 Genetic complementation by Taq polymerase to *E. coli* recA718/polA12 with a thermosensitive DNA polymerase I. *E. coli* recA718/polA12 is only capable of growth when supplemented with an active Taq polymerase gene.

1.4.2.2 Phage display

Directed evolution of Taq using phage display has been spearheaded by the groups of Winter,(Jestin, Volioti, & Winter, 2001) Jestin,(Vichier-Guerre, Ferris, Auburger, Mahiddine, & Jestin, 2006) and Romesberg.(Fa, Radeghieri, Henry, & Romesberg, 2004; Leconte, Chen, & Romesberg, 2005; Xia *et al.*, 2002) Phage display involves expressing a protein of interest on the coat proteins of a phage particle. Two coat proteins commonly used are pVIII and pIII. The pVIII protein is the major coat protein covering the largest surface area of the phage particle. The pIII coat protein is localized to one area and is limited to only a few copies per particle. Along with a few other coat proteins, these proteins assemble and encapsulate the phagemid DNA, providing a physical link between the expressed proteins and the DNA that encodes them.

Initial experiments involved Stoffel Fragment expressed as a fusion to the pIII coat protein for the facile control of copy number. (Jestin *et al.*, 2001) The polymerase substrate, a primer bearing a maleimide group, was then chemically linked to the phage particle via the pVIII coat protein. In later experiments, the polymerase was expressed as a fusion protein to the Fos leucine zipper peptide. The complementary Jun leucine zipper was expressed as a fusion protein to the pIII coat protein. The Jun/Fos dimerization promotes the conjugation of the polymerase to the pIII coat protein.

Romesberg and co-workers utilized X30 helper phagemid DNA to express a pIII protein fused to an acidic peptide. The Stoffel Fragment was then expressed as a fusion with pIII to create a mix of pIII proteins with a variety of adducts. Upon assembly of the phage particle, both pIII-acid peptide and pIII-Stoffel can be displayed on the same particle. A basic peptide fused to a DNA primer sequence was used to hybridize to the acidic peptide and irradiation with UV light promoted the photo deprotection of a blocked cysteine leading to a disulfide bridge between the two peptides. This disulfide bridge produces a covalent link between the phage particle and the DNA primer. Using phage display, Romesberg and coworkers successfully selected for Stoffel Fragments with improved uptake of ribonucleotides, and in a separate selection isolated polymerase P2, a Stoffel mutant capable of incorporating and extending a propynylisocarbostyryl (PICS) self-pair (Figure 1.17).

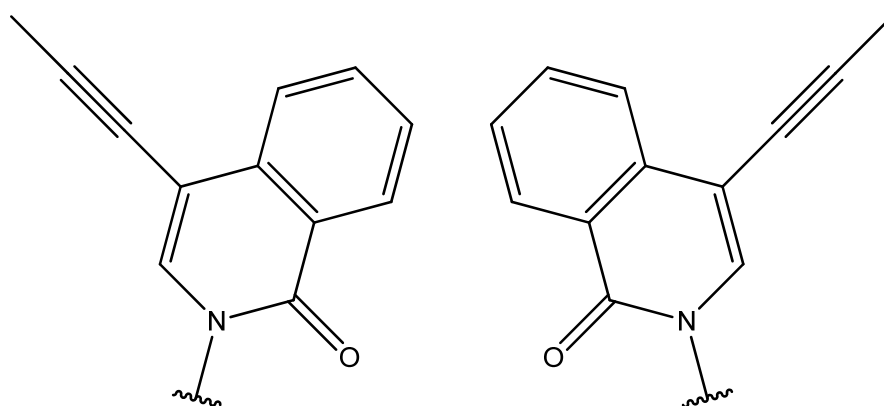
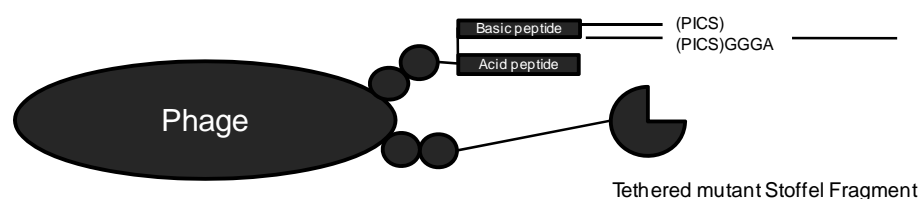
A**B**

Figure 1.17 **A** The synthetic propynylisocarbostyryl (PICS) nucleobase analog self-pair. **B** Cartoon representation of a phage particle displaying both pIII-Stoffel and pIII-Acid peptide.

1.4.2.3 Compartmentalized self-replication

A prominent method of evolving Taq is through the use of a technique called compartmentalized self-replication (CSR), which is based on Tawfik and Griffith's *In Vitro* Compartmentalization (IVC). (Ghadessy, Ong, & Holliger, 2001; Tawfik & Griffiths, 1998) IVC involves the formation of femtolitre volumes (~65 fl) of aqueous phase encapsulated by surfactants and oil phase in the form of a reverse micelle. Such emulsions can be formed through agitation of the two phases or extrusion through membranes containing an appropriate pore size. The process of encapsulation results in up to 10^{10} compartments in a milliliter of total emulsion volume. Encapsulation effectively isolates the contents within from neighboring compartments. IVC was first

validated using the selection of a Hae III methyltransferase to selectively modify and protect the DNA that encodes the Hae III methyltransferase gene isolated within the droplet. Unprotected DNA was preferentially destroyed with the addition of the restriction enzyme Hae III. Successful selection showed that compartmentalization through the use of an emulsion could be used to select for a task-directing enzyme and its encoding gene in the presence of a ten million-fold excess of another gene.

Since its inception, IVC has been used for the evolution and discovery of both protein and nucleic acid enzymes.(Agresti, Kelly, Jaschke, & Griffiths, 2005; d'Abbadie *et al.*, 2007; Ghadessy *et al.*, 2001; Ghadessy *et al.*, 2004; Loakes, Gallego, Pinheiro, Kool, & Holliger, 2009; Ong *et al.*, 2006; Ramsay *et al.*, 2010; Yonezawa, Doi, Kawahashi, Higashinakagawa, & Yanagawa, 2003) The original IVC selection process where unwanted material is reacted or destroyed is akin to kinetic resolution of stereoisomers in synthetic chemistry.(Tawfik & Griffiths, 1998) The utility of isolating and modifying genes that encode for fit enzymes extends beyond simply selective protection and destruction. Holliger and co-workers exploited the water and oil emulsion's thermostability in order to evolve Taq polymerase through the selective PCR amplification of desired genes rather than the destruction of undesired genes. *E. coli* cells expressing a library of Taq mutants were added to a PCR, and the cells were individually compartmentalized within reverse micelles (Figure 1.18). Statistically, a single compartment would contain one *E. coli* cell expressing a single type of mutant polymerase. Activity in the presence of an applied selection pressure allows the mutant polymerase to selectively amplify its own gene. With only the genes encoding active mutant polymerase being amplified, the gene pool will become enriched with these types of mutants bearing the desired traits. Holliger and coworkers named this technique compartmentalized self-replication (CSR) and used it to select for mutant Taq polymerases with either increased thermostability or 130-fold resistance to the inhibitor heparin. Both selections demand that the polymerase amplify its own gene in the presence of a new condition that applies selection pressure: i.e. incubation at high temperatures prior to thermocycling or amplification in the presence of heparin. Since polymerases with increased thermostability or increased heparin resistance were successfully selected, the use of CSR for the evolution of Taq was validated.

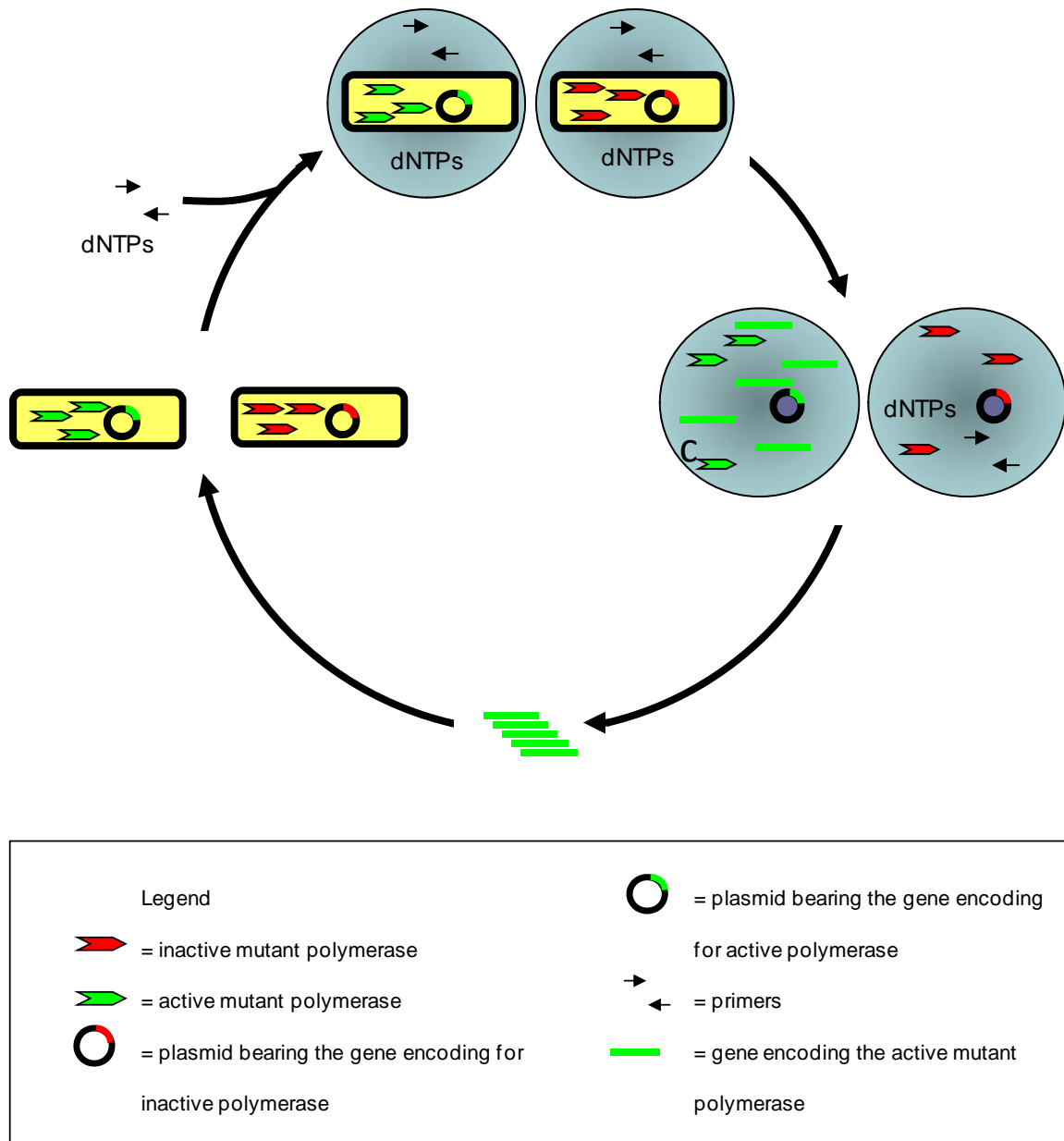


Figure 1.18 Compartmentalized self-replication. The polymerase gene library is cloned into an expression vector and subsequently transformed into an *E. coli* cell. After the polymerase library is expressed, the cells are isolated and encapsulated with PCR reagents using a water and oil emulsion. The emulsion is then thermocycled, destroying the *E. coli* cell wall and allowing amplification to occur. Post CSR, the amplicons enriched in genes encoding active mutants can be collected and purified for the next round of selection.

Holliger and co-workers continued evolving Taq polymerase for other traits using CSR. The two selection pressures of increase thermostability and heparin resistance were quite different from each other and a subsequent selection showed just how versatile the CSR process is. In a follow up experiment, the substrate specificity of Taq was altered by placing a mismatched base on the 3' end of the primers used in the PCR reactions.(Ghadessy *et al.*, 2004) For PCR, and ultimately CSR, to occur, mutants would necessarily be capable of extending a 3' mismatch. The dominant mutants from this selection not only had the ability to extend mismatches, but had also evolved the ability to incorporate a diverse spectrum of unnatural nucleoside triphosphate substrates. CSR is also being used to evolve polymerases capable of incorporating completely new nucleoside triphosphate analogs. In a recent publication, Holliger and coworkers evolved a polymerase capable of incorporating hydrophobic bases first reported by Kool.(Loakes *et al.*, 2009) In a similar technique to the 3' mismatch selection, the selection pressure was introduced by using primers that forced the polymerase to read-through non-canonical base pairs at specified positions. Instead of a 3' mismatch, a hydrophobic base was introduced in its place. Stringency was increased by introducing a hydrophobic base in the middle of the primer as well.

In the experiments mentioned above, the selection technique known as compartmentalized self-replication involved replication of the entire gene. Amplification of a large gene can be difficult and may impose additional selection pressure to the system. In the event that a defined region of the gene is known to be important to the evolution, the stringency of the selection can be reduced by demanding that the polymerase only amplify that defined region or “short patch” of the gene. Holliger and coworkers modified CSR to only involve the amplification of a defined region. They named this technique short patch compartmentalized self-replication or spCSR and used it to evolve Taq to incorporate ribonucleotides.(Ong *et al.*, 2006) The selection demanded that active mutant polymerases be able to amplify a portion of its own gene using any three of the four different deoxyribonucleotides and the ribonucleotide counterpart of the deoxyribonucleotide that was left out. Examination of the structure of Taq allowed them to focus on a single peptide of the active site. Decreasing the target amplicon size to decrease selection pressure is crucial as the selection required not only

the successful incorporation of multiple ribonucleotides, but also the efficient read-through of the resulting hybrid RNA/DNA product. Very recently, CSR has been used to evolve a Family B polymerase, *Pyrococcus furiosus* DNA polymerase I or Pfu, to accept nucleotides that have been base-functionalized with a fluorescent group and to produce dsDNA with a high density of functionalization.(Ramsay *et al.*, 2010) With the growing success and versatility of this technique to produce modified nucleotide compatible polymerases, an increasing repertoire of unnatural nucleotides will be available for *in vitro* selections.

1.5 Thesis objectives

The aim of this thesis is to further develop sequence specific RNA cleavage at high catalytic rates in physiological conditions using modified DNAzymes. *In vitro* selected DNAzymes are currently limited by their lack of functional diversity compared to protein enzymes. Recognition of this limitation has led to the development of biomolecule conjugates with the favorable attributes of both the sequence specific substrate recognition characteristics of nucleic acids and the high divalent metal cation-independent catalysis accomplished by proteins. Chapter 2 describes an improved synthesis of an 8-modified 2'-deoxyadenosine triphosphate that was based on a previous synthesis of the phosphoramidite congener. Updated synthesis affords substantial quantities of the triphosphate. Chapter 3 explores the effects of conformationally restraining the chemically appended functional group, imidazole, by shortening the functional group linker through the use of a yet untested modified nucleoside triphosphate, 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine triphosphate, in the discovery of a DNAzyme with enhanced catalysis.

Selections using different modified nucleotides introduce new selection pressures that have unforeseen results on the fitness of clones bearing a particular sequence. Chapter 4 examines the recognition of modified DNA by protein enzymes and how it affects DNAzyme fitness. Past selections using modified nucleotides showed a slight bias in the percentage of 8-modified dAs in the final gene pool. Bias against the critical 8-modified dA should be minimized to avoid the loss of sequence and chemical space. It

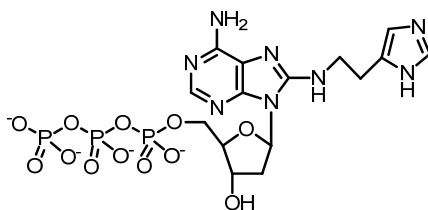
is hypothesized that two most notable selection pressures that could cause this bias stems from the polymerase-mediated synthesis of the fully modified DNAzyme clones and the read-through of successfully selected modified clones. Other enzymic activities related to the enzymic processing of modified DNA such as restriction digests and *in vivo* processing will be examined.

Inefficient processing by polymerases hinders the further discovery of modified DNAzymes and decreases the repertoire of functional groups available to researchers. Instead of sacrificing modified nucleotide design to accommodate commercial polymerases, the polymerases can be altered to accommodate modified nucleotides designed for enhanced catalytic activity. Chapter 5 describes progress towards the directed evolution of a DNA polymerase for improved uptake and read-through of functionalized nucleotides. Short patch compartmentalized self-replication, which was developed by Holliger and coworkers for the successful selection of several evolved polymerases with improved incorporation and extension of modified nucleotides, is the chosen method for the directed evolution of *Thermus aquaticus* DNA polymerase I. We hope to use this method of forming reverse micelles to isolate each mutant polymerase and its corresponding gene from neighboring mutant polymerases to enrich only those capable of efficiently incorporating modified nucleotides as well as synthesizing a new strand of DNA using a modified template.

Chapter 2: Chemically Synthesized Modified Nucleobase Analog

2.1 Introduction

The introduction of the modified nucleosides into the selection of self-cleaving DNazymes has begun to alleviate dependence on non-physiological conditions. Specifically, the need for high divalent metal cation concentrations is overcome by the use of nucleosides modified with chemical functionalities found in the active site of RNase A. (Perrin *et al.*, 2001; Sidorov *et al.*, 2004) The two modified nucleoside triphosphates used in the selection of the RNase mimic Dz9₂₅-11 were 5-aminoallyl dUTP (dU^{aa}TP) and 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA^{ime}TP) **2.1** (Figure 2.1). The former nucleoside is commercially available. The latter was originally synthesized using adenosine monophosphate as starting material.

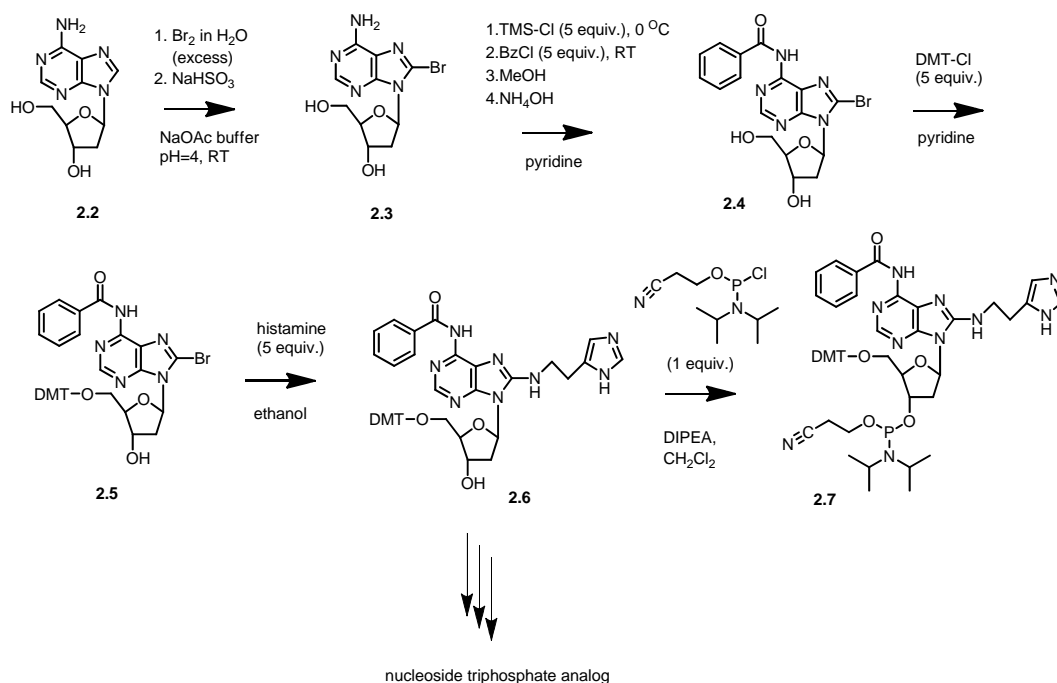


2.1

Figure 2.1 The structure of 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA^{ime}TP) **2.1**.

In subsequent experiments, the phosphoramidite version, 8-(4-imidazole)aminoethyl-2'-deoxyadenosine phosphoramidite **2.7** (Figure 2.2), was synthesized so that *trans*-cleaving Dz9₂₅-11, which was capable of multiple turnover, could be produced using solid phase DNA synthesis. (Lerner, Hobbs, & Perrin, 2002; Ting, Thomas, Lerner, & Perrin, 2004) As the production of related 8-modified adenosine triphosphate analogs was being explored, a major concern was the possibility of trace amounts of unmodified dATP could contaminate the modified nucleotide product. An example of this was an anomaly that arose during the synthesis and enzymic incorporation of an 8-(2-(4-imidazolyl)ethylthio)-modified 2'-deoxyadenosine triphosphate raised suspicions that contaminating unmodified dATP was being synthesized and preferentially incorporated by enzymes. Since the sulfur adduct followed the same original synthesis as **2.1** that started with 2'-deoxyadenosine monophosphate, the purification of the synthetic intermediates was made difficult by

their negative charge. As a result, an alternate method of producing 8-modified 2'-deoxyadenosine triphosphates was sought. A variation of the synthetic method used to synthesize **2.7** was chosen. The phosphoramidite method has the advantage of starting with the nucleoside, and the negatively-charged phosphate groups are added in the last steps. The phosphoramidite method facilitated purification by using uncharged intermediates that could be purified by flash chromatography. Furthermore, this method has been successfully employed to produce other 8-modified nucleosides and has the advantage of producing an intermediate **2.6** that could be used to synthesize either the nucleoside triphosphate for enzymic studies or the nucleoside phosphoramidite for chemical scale up. Branching off from compound **2.6**, the synthetic scheme was adapted with the addition of a few protecting group changes so that the triphosphate group can then be introduced in the last step.



Bz = benzoyl
DMT = dimethoxytrityl
DIPEA = diisopropylethylamine

Figure 2.2 Literature synthesis of (6*N*-benzoyl-8-(2-(4-imidazolyl)ethylamino)adenyl)-3'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxy-β-D-ribofuranose. The triple arrows branching off from **2.6** represent a synthetic route that affords the corresponding modified nucleoside triphosphate analog.

To facilitate the synthesis of **2.1**, a modified synthetic scheme that started with compound **2.6**, the precursor to the phosphoramidite, was used. Using this method, grams of the intermediates can be easily produced and converted to either the modified nucleoside phosphoramidite or the modified nucleoside triphosphate. The modified nucleoside phosphoramidite was used to produce modified oligonucleotides using solid phase synthesis. The isolated product dA^{ime}TP **2.1** was characterized using enzymic incorporation followed by MALDI analysis.

2.2 Objective of this work

An alternate method of producing 8-modified dATP was required to facilitate purification of 8-modified intermediates and ensure unmodified contaminations are kept to a minimum. In this chapter, the literature synthesis of **2.7** was performed and the resulting compound was used to chemically synthesize an oligonucleotide. Also, an alternate synthesis of dA^{ime}TP that did not start from an adenosine monophosphate, but from adenosine is also described. The enzymic incorporation reaction of dA^{ime} nucleoside monophosphate followed by MALDI analysis was used as characterization.

2.3 Materials and methods

2.3.1 General methods

Chemical reactions were monitored using aluminum-backed silica gel thin layer chromatography from Macherey-Nagel. Compounds were visualized by irradiating the silica gel plate with UV light at 254 nm. Acid treatment of the compounds on the silica gel plates was accomplished by briefly immersing the plate in 1 M HCl. Flash chromatography was performed using 230-400 Mesh Silicycle silica gel.

2.3.2 Nuclear magnetic resonance, mass spectrometry and matrix assisted laser desorption ionization instruments

¹H and ¹³C NMR spectra were obtained using a Bruker AV300 spectrometer at 300 MHz and 75 MHz, respectively. Mass spectrometry (MS) was performed on a Bruker

Esquire-LC, Micromass LCT or Waters LC/MS instrument in positive ion mode and matrix-assisted laser desorption ionization (MALDI) was performed Bruker Biflex IV at the Mass Spectrometry Centre at The University of British Columbia.

2.3.3 Oligonucleotides

The following oligonucleotides are mentioned in this chapter (5'-3'):

ODN 2.1 GTCGACTCTAGAAGATCTATCA^{ime} (CH-4-42-ss)

ODN 2.2 GTCGACTCTAGAAGATCTATCA (CH-6-42-ss)

ODN 2.3 GTCTGTTGGGCCCATCCAACA (7DMP4)

ODN 2.4 ACTCTAGAGGATCCCCGGGTA^{ime}CCGAGCTCGAATTCAC (CH-4-42-A)

ODN 2.5 TTCGGCGTCCCGCGGGAGGCCCTCCAGCCCCT (CH091A)

Oligonucleotides **ODN 2.1**, **ODN 2.2**, and **ODN 2.4** were synthesized at and purchased from the University Core DNA Services at the University of Calgary. Oligonucleotides **ODN 2.3** and **ODN 2.5** were synthesized at and purchased from the Nucleic Acids and Protein Services Unit at the University of British Columbia. **ODN 2.1** was gel purified. Oligonucleotide names as recorded in the notebook are in parentheses.

2.3.4 Synthesis of (6*N*-benzoyl-8-(2-(4-imidazolyl) aminoethyl)adenyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxy- β -D-ribofuranose and (6*N*-benzoyl-8-(2-(4-imidazolyl)ethylamino)adenyl) -3'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxy- β -D-ribofuranose
Compounds (6*N*-benzoyl-8-(2-(4-Imidazolyl)ethylamino)adenyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxy- β -D-ribofuranose **2.6** and (6*N*-benzoyl-8-(2-(4-Imidazolyl)ethylamino)adenyl) -3'-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxy- β -D-ribofuranose **2.7** were synthesized according to literature precedent.(Ikehara & Kaneko, 1970; Lerner, Hobbs *et al.*, 2002; Singh, Kumar, & Ganesh, 1990) Briefly, adenosine was brominated at the 8 position and the product **2.3** crystallized when the pH was brought up to 7 using 0.5 M NaOH. The N-6 position was

protected with a benzoyl group and subsequently the 5' hydroxyl group was protected with a dimethoxytrityl group. The bromide at the 8 position was displaced with free histamine in a nucleophilic aromatic substitution reaction to give compound **2.6**. Compound **2.6** was treated with 2-cyanoethyl-*N,N*-diisopropylphosphoramidyl chloride to afford **2.7**.

2.3.5 Synthesis of (6*N*-benzoyl-8-(2-(4-imidazolyl) aminoethyl)adenyl)-3'- *O*-(methoxyacetyl)-2'-deoxy- β -D-ribofuranose (**2.8**):

A solution of compound **2.6** (373.8 mg, 0.49 mmol) was dissolved in dry pyridine (1.92 ml). Methoxyacetic anhydride (0.128 ml, 0.94 mmol) was added dropwise with constant stirring. The reaction was stirred overnight at room temperature. By TLC, the reaction was complete (R_f = 0.50 in 10 % methanol in chloroform) and solvent was removed using a rotary evaporator to afford an oil. The oil was dissolved in 80 % acetic acid:water (3.2 ml) and was allowed to stir for three hours at room temperature to remove the acid labile dimethoxytrityl group from the 5' oxygen. The reaction was complete by TLC (R_f = 0.20 in 10 % methanol in chloroform). The product **2.8** (174.2 mg, 0.325 mmol, 67 %) was purified using flash chromatography and appeared as a white solid. ^1H NMR (300 MHz, CDCl_3 , 25 °C): δ = 8.50 (s, 1H, 2-H), 8.01 (d, J = 7.2 Hz, 2H, o-Bz-H), 7.69 (s, 1H, (C-2)-Imid), 7.59 (t, J = 7.41 Hz, 1H, p-Bz-H), 7.51 (t, J = 7.68 Hz, 2H, m-Bz-H), 6.78 (s, 1H, (C-5)-Imid), 6.63 (dd, J = 5.39, 10.04 Hz, 1H, 1'-H), 5.58 (d, J = 6.26 Hz, 1H, 3'-H), 4.18-4.14 (m, 1H, 4'-H), 4.06 (d, J = 11.44 Hz, 1H, 5'-H), 3.99 (s, 1H, CO-CH₂-O), 3.94 (dd, J = 1.21, 11.43 Hz, 1H, 5'-H), 3.73 (m, 2H, N-CH₂-), 3.51 (s, 3H, -O-CH₃), 2.99 (m, 2H, -CH₂-Imid), 2.78 (m, 1H, 2'-H), 2.27 (dd, J = 5.433, 14.085, 1H, 2'-H). ^{13}C NMR (75 Mhz, CD_3OD , 25 °C) δ = 170.5 (ester), 166.9 (amide), 153.5 (C-4), 153.4 (C-8), 147.8 (C-2), 143.2 (C-6) 134.6, 134.2, 134.0 (C-4 imid, C-2 imid, C-Bz), 132.4 (C-p-Bz), 128.4 (C-m-Bz), 128.0 (C-o-Bz), 123.8 (C-5), 116.9 (C-5-imid), 85.4 (C-4'), 83.7 (C-1'), 76.2 (C-3'), 69.2, 61.4, 58.3 (ester-CH₂-O-, C-5, -O-CH₃), 42.2 (N-CH₂-), 34.9 (C-2'), 25.81 (CH₂-imid). MS (ESI⁺): calcd. For $\text{C}_{25}\text{H}_{29}\text{N}_8\text{O}_6^+$: 537.6 Found: 537.5

2.3.6 Synthesis of 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA^{ime}TP) 2.1

8-(2-(4-Imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate (dA^{ime}TP) **2.1** was synthesized by triphosphorylating **2.8** according to Ludwig and Eckstein. (Ludwig & Eckstein, 1989) A solution containing **2.8** (51.6 mg, 96.3 μ mol) was dissolved in 400 μ l of a 1:3 pyridine: dioxane mixture. 2-chloro-4H-1, 3, 2-benzodioxaphosphorin-4-one (0.0457 mg, 225.6 μ mol) dissolved in 200 μ l dioxane was added. A gum formed upon addition of 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. Dry DMF was added until the gum dissolved. Tetrabutylammonium pyrophosphate (100 mg, 182.3 μ mol) dissolved in 300 μ l dry DMF was added to the reaction. The reaction was stirred for ten minutes. Reaction was quenched with the addition of 1 % iodine in 98:2 pyridine: water (v/v) (2 ml, 157 μ mol). The reaction was stirred for 15 minutes. To consume excess iodine, 5 % NaHSO₃ was added until the colour from the iodine disappeared. The reaction was evaporated to dryness and dissolved in water (10 ml). After 30 minutes, concentrated ammonium hydroxide (20 ml) was added. Deprotection was carried out for 5.5 hours. The reaction was evaporated to dryness. Products were purified on fourteen separate 10 cm by 20 cm (length by width) glass-backed TLC plates (R_f = 0.12 in 6:4:1 dioxane:H₂O:NH₄OH) and eluted from the silica gel using 3 % NH₃OH. About 0.4 mg of crude dA^{ime}TP was recovered. The product was HPLC-purified using a Waters 600 system and a Phenomenex Jupiter 10 μ C4 300A column giving 98 μ l of a 7.9 mM solution (Table 2.1, Retention time = 4.0 minutes). This solution was diluted and divided into 1 mM aliquots (50 μ l). The product was characterized by enzymic incorporation using terminal transferase and oligonucleotide **ODN 2.3** followed by analysis of the products using MALDI. Products found using MALDI were 6377.4 (primer), 6799.7 (primer **ODN 2.3** + dA^{ime}), and 7222.8 (primer **ODN 2.3** + (2 \times) dA^{ime}). Mass difference for the first addition dA^{ime} monophosphate = 422.3. Calculated mass of the incorporated dA^{ime} monophosphate (neutral) = 422.3.

Table 2.1 HPLC Gradients.

Time (minutes)	Flow rate (ml/min)	% Acetonitrile/H ₂ O (0.05 M ammonium acetate pH 7)
0	1	0
20	1	25
40	1	50
41	1	100
46	1	100
47	1	0
52	1	0

2.3.7 Chemical synthesis of an 8-modified dA containing oligonucleotide

Compound **2.7** was sent to University Core DNA Services at The University of Calgary to synthesize oligonucleotide **ODN 2.1** (5'-GTCGACTCTAGAAGATCTATCA^{ime}-3') bearing a dA^{ime} at the 3' end. The corresponding unmodified oligonucleotide **ODN 2.2** (5'-GTCGACTCTAGAAGATCTATCA-3') was also synthesized. MALDI (5'-GTCGACTCTAGAAGATCTATCA-3'): 6720.3 MALDI (5'-GTCGACTCTAGAAGATCTATCA^{ime}-3'): 6829.4 Calculated mass difference due to the 8-(2-(4-imidazolyl)aminoethyl) group: 109.1 Mass difference found: 109.1

2.3.8 PCR amplification using an 8-modified dA oligonucleotide

PCR amplification was performed using the modified primer **ODN 2.1** (5'-GTCGACTCTAGAAGATCTATCA^{ime}-3'). The reaction contained 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, 1 µM primer **ODN 2.5** (5'-TTCGGCGTCCCGCGGGAGGCCCTCCAGCCCCT-3') and 1 µM **ODN 2.1** (5'-GTCGACTCTAGAAGATCTATCA^{ime}-3'), and 25 ng pTaq plasmid and 2.5 U of Taq polymerase (NEB) in a final volume of 50 µl. The reaction was thermocycled 20 × (94

°C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). Products were visualized on a 2 % agarose gel.

2.4 Results

2.4.1 MALDI spectra of unmodified and modified oligonucleotide primer

The phosphoramidite **2.7** was synthesized according to literature precedent (Ikehara & Kaneko, 1970; Lermer, Hobbs *et al.*, 2002; Singh *et al.*, 1990) to confirm the presence of the (4-imidazolyl)aminomethyl group in the intermediate compound **2.6** and used to produce oligonucleotides intended for use for a polymerase selection. One oligonucleotide, **ODN 2.1**, in particular was synthesized with the dA^{ime} at the 3' position. The inspiration for this was motivated by the selections done designing the primers to introduce selection pressure in compartmentalized self-replication (CSR). (Ghadessy *et al.*, 2004; Loakes *et al.*, 2009; Loakes & Holliger, 2009) It was our intention to use this same technique of using modified primers in CSR to evolve a polymerase that has improved tolerance for our 8-modified adenosines.

Only two nanomoles of the modified oligonucleotide **ODN 2.1** were recovered from solid phase synthesis. MALDI analysis of this modified oligonucleotide and the unmodified version was performed by Dr. Leonerd Lermer. The unmodified oligonucleotide gave a peak with a mass to charge ratio of 6720.3 while the modified oligonucleotide gave a mass to charge ratio of 6829.4 (Figure 2.3). The difference in mass to charge corresponds to a group matching the mass of an 2-(4-imidazolyl)aminoethyl group (109.1). With the presence of the functional group confirmed, primer **ODN 2.1** could be used in PCRs to assess its compatibility with polymerases.

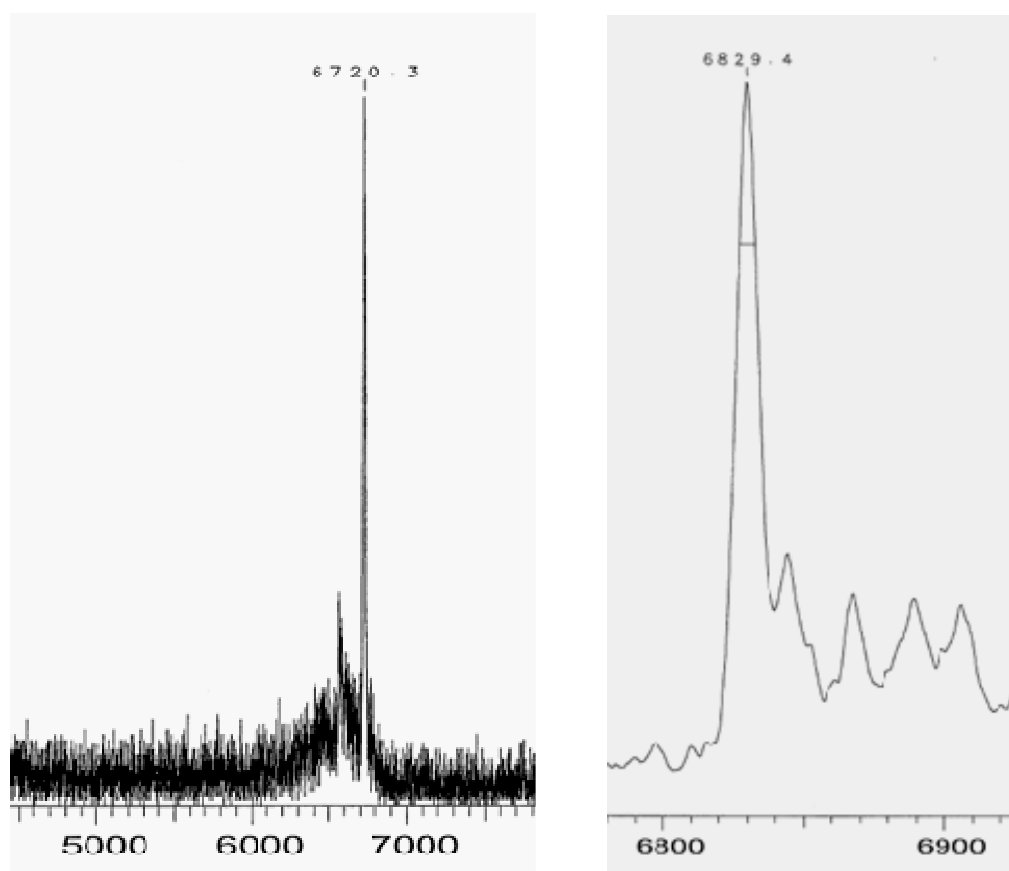


Figure 2.3 MALDI spectrum of unmodified primer oligonucleotides **ODN 2.2** showing a mass to charge ratio of 6720.3 (left) and 3'-dA^{ime}-modified primer oligonucleotide **ODN 2.1** showing a mass to charge ratio of 6829.4 (right). Values differ by a mass corresponding to a 2-(4-imidazolyl)aminoethyl group.

2.4.2 PCR using a modified oligonucleotide primer

Once oligonucleotide primer **ODN 2.1** was synthesized and the presence of the modification was confirmed by MALDI, the first test was to assess its compatibility with Taq in a standard PCR. The amount of amplicon produced in this PCR will give an indication of the difficulty Taq will have extending a 3' dA^{ime} primer. Commercial Taq from New England Biolabs (NEB) was used in a PCR reaction to test Taq's ability to incorporate nucleotides past a 3' dA^{ime} primer. A band corresponding to the correct amplicon size of 600 bp was evident in the 2 % agarose gel (Figure 2.4). These results confirm that Taq can polymerize beyond a 3'-dA^{ime} in the oligonucleotide primer **ODN 2.1**. Therefore, using 3'-dA^{ime} primers in CSR is a promising method of introducing a

selection pressure into a directed evolution and evolving a polymerase's ability to accept a base-modified nucleoside triphosphate substrate.

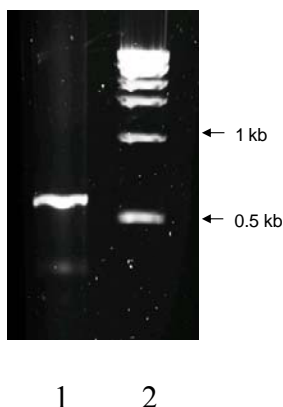


Figure 2.4 PCR amplification using Taq polymerase and **ODN 2.1**, a primer bearing an 8-modified dA at the 3' end. Lane 1: PCR amplification product. Lane 2: Invitrogen 1kb Molecular Weight Ladder.

2.4.3 Synthesis of dA^{ime}TP

As discussed in the introduction of Chapter 2, we sought to produce **2.1** in a way that we could make gram-scale amounts of the intermediate compound where every intermediate could be easily purified using flash chromatography to eliminate contamination from unmodified adenosine. The synthetic scheme shown in Figure 2.5 describes an improved synthetic approach to producing dA^{ime}TP **2.1**. This new synthetic scheme incorporates elements used in the literature synthesis of **2.7** (Lerner, Hobbs *et al.*, 2002) with a literature method of producing nucleoside triphosphates (Ludwig & Eckstein, 1989)

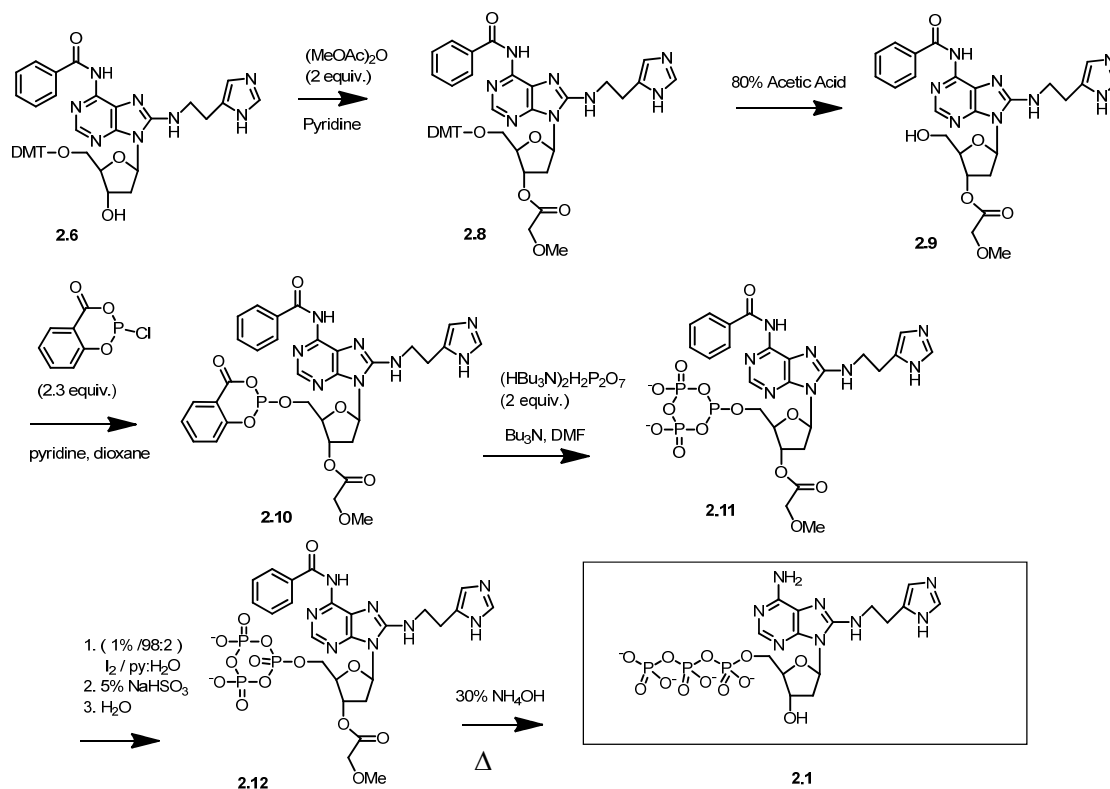


Figure 2.5 Synthesis of (8-(2-(4-Imidazolyl)ethylamino)adenyl)- 5'-triphosphate-2'-deoxy-β-D-ribofuranose.

Compound **2.8** was easily obtained by 3'-protecting compound **2.6** with a methoxyacetyl group. Deprotection of the 5' dimethoxytrityl group was accomplished using 80 % acetic acid: water (v/v). After the protecting group swap, the nucleoside was then subjected to a multistep-one-pot triphosphorylation starting with phosphitylation of the free 5' hydroxyl intermediate **2.8**. The salicylyl group protecting the phosphite was displaced with inorganic pyrophosphate and subsequent oxidation using iodine oxidized the phosphite to a phosphate affording the cyclic triphosphate intermediate **2.12**. Ammonium hydroxide treatment with heat deprotected the N-6 amine and 3' hydroxyl group. The base treatment simultaneously promoted the ring-opening hydrolysis of the cyclic triphosphate to the triphosphate, which affords compound **2.1**. To purify the product away from the salts and other byproducts, the product of the crude reaction mixture was first purified using thin layer silica gel chromatography followed by HPLC. HPLC-purified product **2.1** was enzymically incorporated into an oligonucleotide using terminal transferase.

2.4.4 Enzymic incorporation of dA^{ime} monophosphate

Richard Ting performed the enzymic incorporation of the dA^{ime} nucleotide. Chemically synthesized compound **2.1** was used to extend an oligonucleotide **ODN 2.3** using terminal transferase (Figure 2.6 left). The products of the extension reaction were prepared for MALDI analysis, which was performed by Dr. Leonerd Lermer (Figure 2.6 right). The difference in mass to charge ratio (422.3) matches a value that corresponds to the addition of uncharged dA^{ime} nucleotide (422.3). Two incorporations can be seen on both the polyacrylamide gel and the MALDI spectrum. However, the second addition is faint and the corresponding peak is not as well resolved as the peak of the first incorporation. Therefore, the mass to charge value for the peak corresponding to two incorporations appears one mass to charge unit too high. Since the resolution was unacceptable for the second peak, the mass difference from the first incorporation was considered to be the true mass difference and positively supports the incorporation of a dA^{ime} nucleotide.

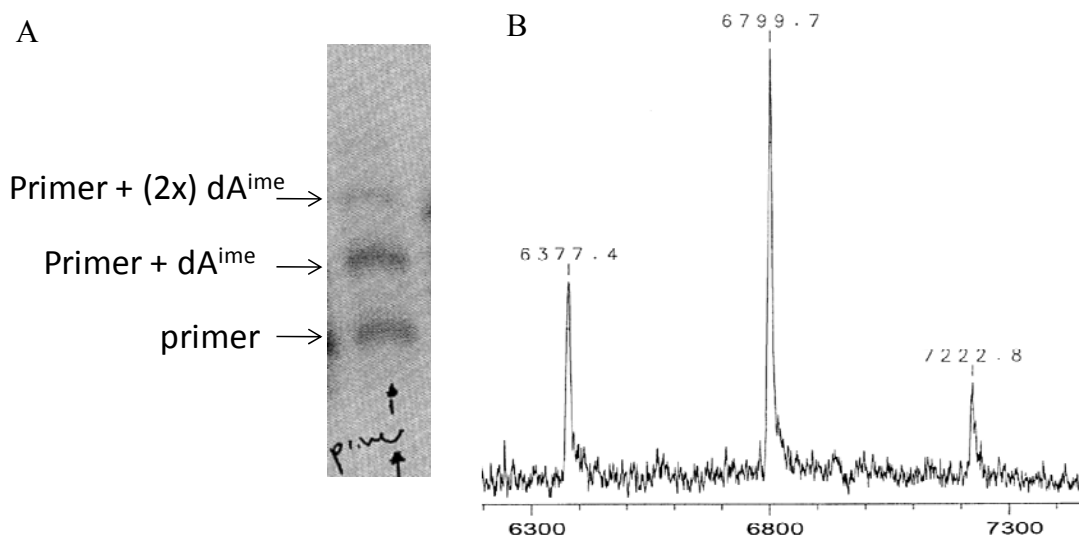


Figure 2.6 A) Polyacrylamide gel showing a single and a double incorporation of the dA^{ime} nucleoside monophosphate by terminal transferase (left). B) Representative MALDI spectrum of the incorporation reaction (right). Mass to charge ratios found are 6377.4, 6799.7, 7222.8. Peak values differ by a mass corresponding to a single neutrally charged dA^{ime} monophosphate.

2.5 Discussion

2.5.1 Modified oligonucleotide synthesis

The phosphoramidite was synthesized for use in the chemical synthesis of modified oligonucleotides and primers. The short oligonucleotide primer **ODN 2.1** was synthesized with the intended use in a polymerase selection described in Chapter 5. The presence of the 2-(4-imidazolyl)aminoethyl functional group attached to oligonucleotide primer **ODN 2.1** was verified according to a mass difference calculated from the MALDI analyses of a modified oligonucleotide and an unmodified oligonucleotide of the same sequence. **ODN 2.1** is also shown to support PCR using commercial Taq, which shows us that any fraying of the 3' end that might result from weakened Watson-Crick base pairing due to the modification does not inhibit the PCR reaction. Despite these successes, there were a few problems with this oligonucleotide. First, the yield was very low. Two nanomoles were recovered after purification from a two hundred nanomole scale synthesis. We attribute this to introducing the modified base at the 3' end of the

oligonucleotide and that commercial solid phase synthesis normally begins with the solid support preloaded with the first base. One other oligonucleotide, **ODN 2.4** was synthesized with the modification in the middle of the oligonucleotide, but was not characterized by MALDI. Oligonucleotide **ODN 2.4** was synthesized in much higher yield. Two-hundred and fifty nanomoles of oligonucleotide **ODN 2.4** was received by the author indicating that **2.7** is suitable for solid phase synthesis, but yields will be low if dA^{ime} is required at the 3' end. As stated earlier, the primer **ODN 2.1** with the modification at the 3' end was meant to be used as a selection primer. However, low synthetic yields may have been a problem if a repertoire of such primers needs to be made. Also, recent changes to the way that the polymerase selection was performed have made the use of this particular primer design obsolete. Nevertheless, valuable information was attained through the synthesis of the modified primer and the future syntheses of new selection primers will be designed accordingly.

The versatility of the synthetic method of producing the phosphoramidite analog was exploited by other members of the lab. One other 8-modified adenosine analog was produced by Richard Ting who was given 4 grams of **2.4**. Dr. Ting successfully functionalized and phosphitylated **2.4** to produce an 8-sulfur analog of **2.7**. The photolabile sulfur analog was used to produce a light-activated DNAzyme.(Ting, Lermer, & Perrin, 2004) After this successful synthesis, compound **2.5** became the key compound in the synthesis of a repertoire of 8-modified analogs of **2.1** and **2.7**. (Lam *et al.*, 2008)

2.5.2 Triphosphate synthesis

The final steps of the modified nucleoside triphosphate synthesis involve a multistep-one-pot reaction that, unfortunately, had a low yield of about 1 %. After work up and evaporation of the solvents, the triphosphate compound is amidst many chemical byproducts in the round bottom flask. It is quite possible that some of compound **2.1** remained in the round bottom flask trapped within the salt byproducts. Even with the low yield, however, we managed to isolate more than enough of the final product for many enzymic incorporation and the yield was not further optimized by the author. The yield could be improved by dissolving all of the salts and applying that solution onto silica gel

plates as well. Future syntheses should integrate dissolving the salt byproducts and purifying any leftover modified nucleoside triphosphate as a final step.

This method of synthesis for dA^{ime}TP **2.1** facilitates the purification of chemically-augmented nucleoside intermediates. One of the benefits of this method is that compound **2.5** can be derivatized with an assortment of functional groups. Using this method, Curtis Lam proceeded to synthesize various 8-modified adenosine triphosphates using both nucleophilic aromatic substitution displacement as well as Sonagashira coupling.(Lam *et al.*, 2008) One of his new nucleotides, 8-(2-(4-imidazolyl)aminomethyl)-2'-deoxyadenosine triphosphate (dA^{imm}TP), became the focus of a selection described in Chapter 3.

**Chapter 3: A DNzyme Selection used to Evaluate Effects of Shortening Linker
Length between Imidazole and 8-Modified 2'-Deoxyadenosine**

3.1 Introduction

The discovery of divalent metal cation-independent DzG3 by Geyer and Sen proved that unmodified DNA has sufficient functionality for detectable catalytic self-cleavage at a ribophosphodiester bond.(Geyer & Sen, 1997) Much like the hammerhead ribozyme mentioned in Chapter 1, DzG3 achieves high substrate specificity through the use of Watson-Crick base pairing. Although the rate enhancement of DzG3 over spontaneous self-cleavage reached 10^7 -fold improvement, the k_{obs} of DzG3 falls short of the first order rate constant of divalent metal cation-dependent nucleic acid enzymes and is several orders of magnitude below the first order rate constant of the non-specific RNA cleaving RNaseA.

The addition of exogenous non-metal cation cofactors e.g. histamine and spermine has been used in the past to assist the DNAzyme with catalysis.(Faulhammer & Famulok, 1997; Roth & Breaker, 1998) Two limitations of this approach are that the DNAzyme is then required to evolve cofactor binding as well as catalysis and that the selection pressure increases with the number of cofactor binding sites required. These pressures could potentially limit the number of cofactors and therefore limit the chemical space of the DNAzyme. The chemical functionalization of a given nucleoside triphosphate used in the selection of DNAzymes removes the need of selecting for the binding of any exogenous metal cation or non-metal cation cofactor. The first divalent metal cation-independent DNAzyme functionalized with two chemical groups, Dz9₂₅-11 ($k_{\text{obs}} = 0.044 \pm 0.01 \text{ min}^{-1}$),(Perrin *et al.*, 2001) achieved an improvement in rate constant over unmodified DNAzymes like DzG3 ($k_{\text{obs}} = 2.8 \pm 0.4 \cdot 10^{-3} \text{ min}^{-1}$).(Geyer & Sen, 1997) Dz9₂₅-11 utilized amino and imidazole functional groups to mimic lysine and histidine, respectively, which are found to be the key catalytic residues in the active site of RNaseA.(Barnard, 1969) Sidorov *et al.* later selected for a DNAzyme that similarly uses amino and imidazole functionalities by way of introducing an imidazole-modified dU and a propargylamino-modified dA into the selection to cleave a short 12 nt section of RNA (Figure 3.1).(Sidorov *et al.*, 2004) The cleavage was not site-specific, however, and cleaved at either of two 5'-UA-3' sites ($k_{\text{obs}} = 0.06 \text{ min}^{-1}$ for the major site and $k_{\text{obs}} = 0.07 \text{ min}^{-1}$ for the minor site) in the RNA region. While these nucleotides were chosen for their high compatibility with commercial polymerases compared to those used by Perrin

et al., the activity of the DNAzyme constructed using unmodified counterparts shows that the DNAzyme is still 7 % active and that the catalysis is not completely dependent on the chemical augmentations. This finding raises the question of whether or not the modifications are absolutely necessary for catalysis. Mechanistic studies on the functional groups in the Dz9₂₅-11, on the other hand, proved that they are directly involved in catalysis and with the exception of one modified dA removal of a single key modification results in at least a ten-fold decrease in activity. (Thomas *et al.*, 2009) Of the three above mentioned DNAzymes: Sidorov's DNAzyme, DzG3, and Dz9₂₅-11, only Dz9₂₅-11 was chemically synthesized to perform intermolecular cleavage with multiple turnover to prove its catalytic function.

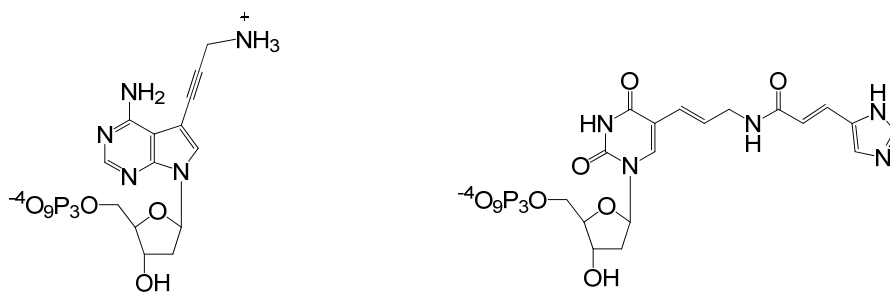


Figure 3.1 The two modified nucleoside triphosphates propargylamino-modified deaza-dATP (left) and imidazolyl-dUTP (right) used by Sidorov *et al.*

The next leap in catalytic rate enhancement was the addition of a third functional group. This third functional group, the guanidinium, is not a chemical group of the critical catalytic residues His12, His119, or Lys41 of RNase A, but is known to stabilize duplex DNA (Prakash, Puschl, & Manoharan, 2007; Roig & Asseline, 2003) and can potentially provide additional electrostatic stabilization to the negatively-charged pentacoordinate phosphorus intermediate. (Trautwein, Holliger, Stackhouse, & Benner, 1991) With a third modified nucleoside triphosphate, the guanidinium-modified dU^{ga}TP, introduced in a selection with both the amino-modified and imidazole-modified nucleoside triphosphates, a new benchmark for the rate constant for divalent metal cation-independent cleavage was attained. The DNAzyme isolated from this selection, Dz9-86 ($k_{\text{obs}} = 0.134 \pm 0.026 \text{ min}^{-1}$, 24 °C) shown in Figure 1.11, with this third functional group

did indeed have a rate enhancement over Dz9₂₅₋₁₁ ($k_{\text{obs}} = 0.044 \pm 0.01 \text{ min}^{-1}$, 37 °C). (Hollenstein *et al.*, 2009a) An additional advantage bestowed is increased thermostability of Dz9-86 ($k_{\text{obs}} \sim 0.15 \text{ min}^{-1}$, 37 °C) over Dz9₂₅₋₁₁. Although selected at room temperature, Dz9-86 was reported to have a rate maximum between 35 - 40 °C. Dz10-66 ($k_{\text{obs}} = 0.63 \pm 0.04 \text{ min}^{-1}$, 37 °C), the DNAzyme subsequently selected using a 40 nt degenerate region, was also selected using the dU^{ga}TP and while it likewise showed similar thermostability as Dz9-86, it presented yet another increase for the benchmark minimum rate constant for divalent metal cation-independent self-cleavage. (Hollenstein *et al.*, 2009b)

In terms of intermolecular cleavage and true catalytic turnover, the catalytic efficiency, $k_{\text{cat}}/K_{\text{M}}$, is the value of interest. The maximum value for catalytic efficiency is set by the diffusion of the substrate to the enzyme and has a value of $\sim 10^9 \text{ M}^{-1} \text{ sec}^{-1}$. Recall, Dz10-23 has a $k_{\text{cat}}/K_{\text{M}}$ of about $4.5 \cdot 10^9 \text{ M}^{-1} \text{ min}^{-1}$. Currently, our best divalent metal cation-independent DNAzyme, Dz10-66t, has a $k_{\text{cat}}/K_{\text{M}}$ of $\sim 6 \cdot 10^5 \text{ M}^{-1} \text{ min}^{-1}$ which is well below the $k_{\text{cat}}/K_{\text{M}}$ of Dz10-23. Since the binding of the substrates for Dz10-23 and Dz10-66t is mediated by Watson-Crick base pairing and increasing substrate affinity by elongating the binding regions will result in slow product release, improving the k_{cat} of our divalent metal cation-independent DNAzymes by optimizing the functional group-dependent catalytic mechanism is explored in this chapter.

The addition of functional groups has improved DNAzyme catalysis, but the method of chemical attachment, i.e. position and composition of the linker of the functional group, has not been thoroughly explored. The positional attachment has varied from sugar to base with much emphasis focused on the ease of incorporation. While its incorporation is limited to a specific polymerase, the 8-(4-imidazolyl)aminoethyl-2'-deoxyadenosine triphosphate has been successfully employed in several DNAzymes. (Hollenstein *et al.*, 2008; Hollenstein *et al.*, 2009a; Hollenstein *et al.*, 2009b; Perrin *et al.*, 2001) The 8 position has shown much success and will be retained as a key attachment point.

The composition of the functional group linker is another variable that must be explored. The functional group linker is crucial for precise placement of the imidazole. We hypothesized that by limiting the free movement of the functional group by using a

rigid or short linker at the 8 position and constraining it into its active conformation within the active site; we can obtain higher catalytic rate constants. Although one may be tempted to use the protein analogy of site-directed mutagenesis where one would simply replace the modified adenosines with modified nucleotides bearing new linkers, doing so may not accurately validate the short linker as such replacements could cause large structural changes in the relatively small DNAzyme. Therefore, a new selection utilizing these constrained modified adenosines must be performed. A modified 2'-deoxyadenosine triphosphate **3.4** using a rigid unsaturated alkynyl linker at the 8 position of 2'-deoxyadenosine has been tested for incorporation and was shown to be a poor substrate for DNA polymerases (Figure 3.2). (Lam *et al.*, 2008) Poor incorporation immediately eliminates its use in a DNAzyme selection. Modified 2'-deoxyadenosine triphosphates bearing a short methyl linker **3.2** (dA^{imm}TP) or a longer propyl linker **3.3** (dA^{imp}TP) were also synthesized. The longer propyl linker variant was found to be a poor substrate for Sequenase V2.0. The short methyl linker consists of an aminomethyl group, which is one methylene group shorter than the linker found in 8-(2-(4-imidazolyl))aminoethyl-2'-deoxyadenosine triphosphate, **2.1**. The incorporation efficiency of modified nucleotide **3.2** was found to be adequate for a selection.

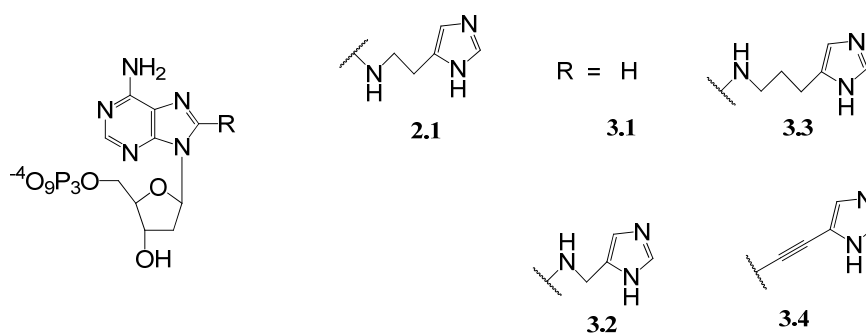


Figure 3.2 Chemical structure of **2.1** (dA^{ime}TP), **3.1** (dATP), **3.2** (dA^{imm}TP), **3.3** (dA^{imp}TP), **3.4** (alkynyl-linked imidazole dATP).

3.2 Objective of this work

The previous work on DNAzymes mentioned above shows the incremental improvements made to the general selection scheme for DNAzymes that perform catalysis in specific conditions. In particular, high activity in physiological conditions (low divalent metal cation concentrations, near neutral pH) is sought. Through the use of chemical modifications, modified DNAzymes that achieve catalytic rate constants in these conditions and are orders of magnitude higher than unmodified DNAzymes have been discovered. Increasing the diversity of chemical functionalities and potentially enlarging the catalytic motifs by switching from a 20 base degenerate region to a 40 base degenerate region have beneficial outcomes on the selections of modified DNAzymes. In this study, the fine tuning of the modifications' spatial orientation was investigated.

The goal of this chapter is to further our knowledge of modified nucleotides in selections and through this understanding isolate DNAzymes with enhanced properties. To acquire this knowledge, we needed to determine whether or not using the chosen 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine bearing a constrained linker will lead to a selected DNAzyme with a catalytic rate constant higher than that of the previously selected Dz10-66. To determine this, we selected for a catalytically active DNAzyme that is dependent on the 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine and compared its rate constant to that of Dz10-66. The temperature and pH dependence of the isolated DNAzyme was also assessed.

3.3 Materials and methods

3.3.1 Chemicals and reagents

The nucleoside triphosphates dU^{ga}-TP, dA^{imm}-TP and dA^{imp} were synthesized according to literature protocol (Hollenstein *et al.*, 2009a; Lam *et al.*, 2008) by Curtis Lam. The protected dA^{ime} nucleoside was synthesized according to literature protocol (Lerner, Hobbs *et al.*, 2002; Prakash, Krishna Kumar, & Ganesh, 1993) and further converted to the triphosphate using the Ludwig and Eckstein method. (Ludwig & Eckstein, 1989) dC^{aa}-TP was purchased from Trilink. α -P³²-dGTP was purchased from Perkin Elmer. Chemicals used were purchased from Sigma. X-Gal and SOC media was

purchased from Invitrogen. Ampicillin disodium salt was purchased from Sigma. Water used in all experiments was first diethyl pyrocarbonate (DEPC)-treated and autoclaved.

3.3.2 Cells, plasmids, enzymes, and protein

Sequenase Version 2.0 was purchased from USB Corporation. Lambda exonuclease, Taq polymerase and Vent (exo-) were purchased from New England Biolabs. Single-stranded Binding Protein was purchased from Epicentre. SUPERase-In was purchased from Ambion. RNase A was purchased from Fermentas. pGEM-T Easy was purchased from Promega. *E. coli* DH10B Electromax cells, SOC Media, LB Broth Base (Lennox L Broth Base), and Terrific Broth (TB) were purchased from Invitrogen.

3.3.3 Oligonucleotides

All oligonucleotides were purchased from Integrated DNA Technologies. Biotinylated primer **ODN 3.2** was gel purified. All other oligonucleotides were extracted using phenol:chloroform:isoamyl alcohol 25:24:1 followed by precipitation with 3 % LiClO₄ in acetone. The precipitate was centrifuged and the pellet was washed with ethanol (500 µl), dried and dissolved in water. The oligonucleotides were then passed through a G-25 column for desalting. The numbers in bold following the sequence are the oligonucleotide names used in a previous publication.(Hollenstein *et al.*, 2009b) The following oligonucleotides were used (5' to 3'):

ODN 3.1 GAGCTCGCGGGGCGTGCN₄₀CTGTTGGTAGGGCCCAACAGACG **1**

ODN 3.2 biotin-T₂₀GCGTGCCrCGTCTGTTGGGCCC **2**

ODN 3.3 phosphate-CGTCTGTTGGGCCCCTACCA **3**

ODN 3.4 GAGCTCGCGGGGCGTGC **4**

ODN 3.5 phosphate-ACGACACAGAGCGTGCCCGTCTGTTGGGCCCCTACCA **5**

ODN 3.6 TTTTTTTTTTTTTTTTTTTTGGAGCTCGCGGGGCGTGC **6**

ODN 3.7

GAGCTCGCGGGGCGTGCAACGACCCACACGACCTGCGAACCCTAGAGAGC

ODN 3.8 ATGACTTGTGGTAGGGCCCAACAGACGGGCACGCTCGTGTGT **7**

rC indicates a cytidine

3.3.4 Buffers and cocktails

1 (cleavage buffer, no divalent metal cations) 50 mM sodium cacodylate, 200 mM NaCl, 1 mM EDTA, pH 7.4

2 (wash buffer, TEN) 50 mM Tris-HCl, 1 mM EDTA, 200 mM NaCl, pH 7.5

3 (template stripping buffer) 0.1 M NaOH, EDTA 1 mM

4 (NaOH neutralization buffer) 25 mM sodium cacodylate pH 6

5 (elution buffer) 1 % LiClO₄ in 10 mM Tris-HCl pH 8

6 (pH variance buffers) mM Tris-HCl, 1 mM EDTA, 200 mM NaCl pH 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0

Cocktail for the First amplification of selection products (5x First Amp Cocktail): 32.6 µM primer **ODN 3.3**, 39.1 µM primer **ODN 3.4**, 1.5 mM dNTPs, 5 mM MgSO₄, and 5x Thermopol Buffer

Cocktail for the second amplification for the amplification of gel purified amplicons from the amplification of the selection products (5x Second Amp Cocktail): 32.6 µM **ODN 3.5**, 43.5 µM **ODN 3.6**, 1.5 mM dNTPs, 5 mM MgSO₄, and 5x Thermopol Buffer.

Formamide loading buffer: 10 ml formamide (Sigma), 0.2 ml 0.5 M EDTA, pH 8.0, 10 mg bromophenol blue (0.1 % w/v), 10 mg xylene cyanol (0.1 % w/v)

3.3.5 Detection of radioactive DNA

Bands of radioactive DNA, resolved on denaturing polyacrylamide gels, were visualized by first exposing the gels to storage phosphor screens. Low activity gels were exposed overnight. Imaging of the screen was done using a GE Typhoon 9200 Phosphorimager. For data manipulation, the program Imagequant Version 5.2 was used to determine the number of counts associated with a particular band by encompassing the band and representing the counts as a function of image intensity.

3.3.6 *In vitro* selection

The first round of selection follows a scheme shown in Figure 3.3. Subsequent rounds closely follow the same scheme with minor variations to increase stringency. Details of each step are described in the following paragraphs.

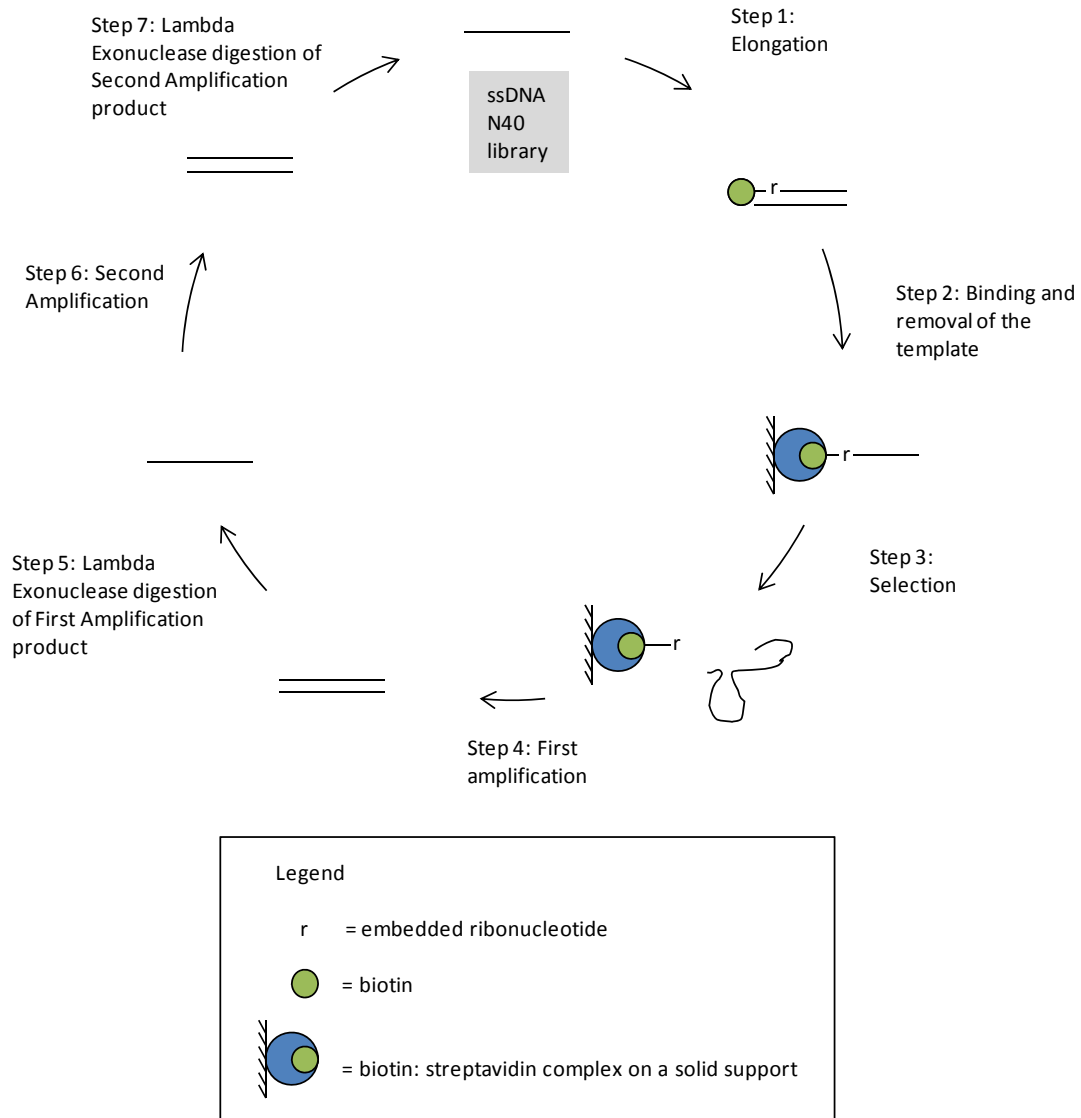


Figure 3.3 Modified DNAzyme selection scheme.

Step 1: Elongation

For round 1, template **ODN 3.1** (15 pmol) was annealed to ribose embedded primer **ODN 3.2** (15 pmol). The primer was extended using 4.6 U of Sequenase Version 2.0 (0.35 μ l of a 13 U/ μ l stock) in the following conditions: 1x Sequenase buffer, 5 mM dithiothreitol, Single-stranded Binding Protein (5 U), SUPERase In (5 U), 50 μ M dA^{imm}TP, 10 μ M dU^{ga}TP, 10 μ M dGTP, 10 μ M dC^{aa}TP and 5-15 μ Ci α -P³²-dGTP in a final volume of 20 μ l. The reaction was overlaid with mineral oil and incubated (37 °C, five hours). After incubation, a 0.5 M solution of EDTA (0.5 μ l) was added to quench the reaction.

Step 2: Binding and removal of the template

The elongation reaction mixture was added to prewashed streptavidin magnetic particles and incubated at room temperature for 30 minutes. The streptavidin magnetic particles with bound modified DNA were washed with TEN (2 \times 50 μ l), (0.1M NaOH, 1 mM EDTA) (5 \times 50 μ l), (25 mM cacodylate, pH 6 buffer) (1 \times 100 μ l), DEPC H₂O (1 \times 100 μ l).

Step 3: Selection

Cleavage buffer (50 μ l) was added to the bound modified single-stranded oligonucleotides and left to incubate for one hour at room temperature. After one hour of incubation the beads were magnetized and the supernatant was removed. 5x First Amp Cocktail (2 μ l) and 1 % LiClO₄ in acetone (1 ml) were added to the supernatant to precipitate cleaved product. The cleaved products were centrifuged for 15 minutes and the supernatant was separated from the pellet. The pellet was washed with ethanol (500 μ l), agitated and spun in the centrifuge for 15 minutes. The ethanol was decanted and any residual ethanol was removed in the speedvac. Formamide loading buffer (20 μ l) was added to the pellet. For size controls, water (10 μ l) was added to the beads and the beads were split equally. To one set of beads, RNaseA (1 μ l of 10 mg/ml) was added and incubated at room temperature for 20 minutes. After twenty minutes of incubation, 99:1

formamide loading buffer:biotin (100 mM in DMF) (200 μ l) was added to the beads. To the other set of bead, 99:1 formamide loading buffer:biotin (100 mM in DMF) (200 μ l) was added directly. Products were resolved by 7 % denaturing PAGE. Polyacrylamide gels were cast using 42 cm by 32 cm glass plates separated by a 0.4 mm spacer. Power was applied for 2 hours at 44 watts. In the early rounds of selection, products are not detectable and gel was excised at the height where the cleaved product control appears (Figure 3.4).

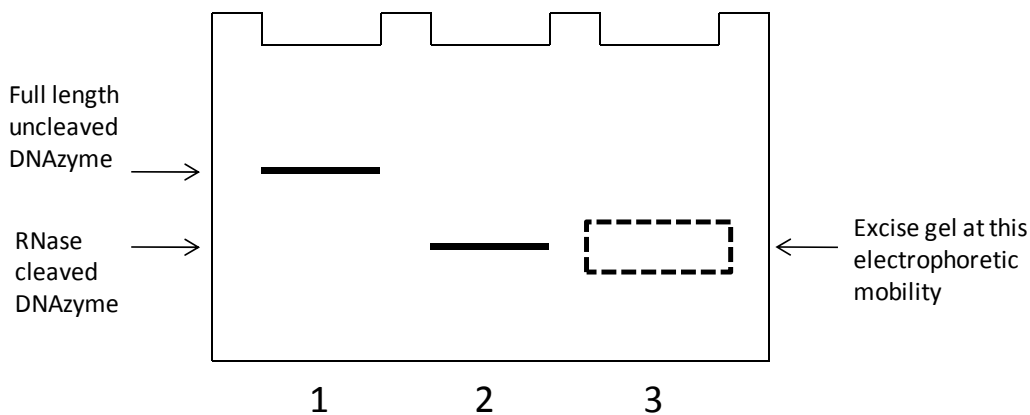


Figure 3.4 Schematic representation of a polyacrylamide self-cleaving DNase selection gel. Lane 1: Full length uncleaved DNase control. Lane 2: RNaseA-cleaved DNase control. Lane 3: Decants from DNase selections. In early rounds, radioactivity from selected DNase is too faint to be detected. Gel is excised at the same height as the RNaseA-cleaved DNA control.

Step 4: First amplification

The portion of excised gel was crushed using a flamed sealed pipette tip and the DNA was eluted by freezing and thawing (2×500 μ l elution buffer). Elutions were combined in a single eppendorf tube and dried in a speedvac. The dried sample was dissolved in water (100 μ l) and 5x First Amp Cocktail (2 μ l). Ethanol (1 ml) was added to precipitate the DNA. The sample was agitated briefly on a vortex and centrifuged for 15 minutes. The supernatant was removed and the pellet was dried using a speedvac to evaporate residual ethanol. After dissolving in water (30 μ l) the DNA was desalted using a short spin column. PCR amplification of the purified products was carried out with the addition of 5x First Amp Cocktail (8 μ l), 3.3-10 μ Ci α -dGTP and 1.5 μ l of 2 U/ μ l Vent

(exo-) in a total volume of 40 μ l. The reaction was thermocycled $30 \times (54/15s, 75/40sec, 95/15s)$. Amplicons were extracted with phenol:chloroform:isoamyl alcohol 25:24:1 (40 μ l) and precipitated with the addition of ethanol (400 μ l). After agitating and 15 minutes of centrifuging the ethanol was decanted and residual amounts were evaporated.

Step 5: Lambda Exonuclease digestion of First Amplification product

The amplicons were dissolved in water (35 μ l), 10x Lambda exonuclease buffer (4 μ l) and, added last, 5U lambda exonuclease (1 μ l). Degradation of the phosphorylated strand was carried out for 1.5 hours at room temperature. The reaction was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 (40 μ l). To precipitate the single-stranded DNA, 3 % LiClO₄ in acetone (300 μ l) was added. The sample was agitated and centrifuged for 15 minutes. The supernatant was decanted and the pellet was washed using ethanol (400 μ l). The sample was again agitated and centrifuged for 15 minutes. The ethanol was decanted and any residual was evaporated by heating the sample to 65 °C. To load the sample onto a 10 % denaturing polyacrylamide gel, it was first dissolved in water (15 μ l) and formamide loading buffer (15 μ l). The single-stranded product was visualized and excised. The gel slices were crushed, eluted, precipitated and desalted as before.

Step 6: Second Amplification

A second amplification of the products involved combining 2nd Amp Cocktail (40 μ l), water (150 μ l), Vent (exo-) (9 μ l of 2 U/ μ l) and DNA product (4 μ l) obtained in the First amplification. The second amplification reaction was thermocycled using the same program as in the First amplification, and subsequently extracted using phenol:chloroform:isoamyl alcohol 25:24:1 (200 μ l). The amplicon product from the second amplification was precipitated with the addition of ethanol (2 ml), agitated briefly and centrifuged for 15 minutes. The supernatant was decanted away from the solid pellet containing the second amplification product and residual liquid was evaporated by heating the sample to 65°C and allowing the pellet to air dry.

Step 7: Lambda Exonuclease digestion of Second Amplification product

The pellet containing the second amplification product was dissolved in water (90 μ l) and 10x lambda exonuclease buffer (10 μ l). The phosphorylated strand was degraded with the addition of lambda exonuclease (1 μ l of 5 U/ μ l stock). The degradation reaction was incubated at room temperature overnight. The reaction was extracted using phenol:chloroform:isoamyl alcohol 25:24:1 (100 μ l). The single-stranded DNA was precipitated with the addition of ethanol (1 ml), agitated briefly and centrifuged for 15 minutes. The supernatant was decanted, and residual liquid was evaporated by heating the sample to 65 °C and allowing it to air dry. The single-stranded DNA was dissolved in formamide loading buffer (40 μ l) and 1 M NaOH (1 μ l) prior to loading on a 10 % denaturing polyacrylamide gel. The polyacrylamide gel was cast using 17 cm by 16.5 cm glass plates separated by 1 mm spacers. Power was applied for one hour at 11 watts. The single-stranded product was visualized by UV-shadowing and excised. The gel slice was crushed and eluted by freezing and thawing (3 \times 500 μ l elution buffer). The sample was dried using a speedvac and dissolved in water (100 μ l) and precipitated with the addition of 3 % LiClO₄ in acetone (300 μ l). The sample was agitated briefly and centrifuged for 15 minutes. The supernatant was decanted and the pellet was washed with ethanol (400 μ l). The sample was agitated briefly and centrifuged for 15 minutes. The supernatant was decanted and residual ethanol was evaporated. The single-stranded template was dissolved in water (50 μ l) and desalted using a long G-25 column. The template was quantified using UV absorption.

The remaining rounds of the selection were carried out as above with round specific modifications. Synthesis of the modified pools for rounds 2-20 were done on a 30 pmol scale. Decants (supernatant) from each of the incubated selection were collected and replaced with fresh cleavage buffer at the water wash step or incubation time indicated in Table 3.1. Decants from the reaction mixture were taken at 60 minutes for rounds 1 to 3. From rounds 4 to 21, decants were taken at 1 minute, 5 minutes, and 60 minutes with fresh cleavage buffer being added after the 1 minute and 5 minute decants. From round 9 onwards, the water used to wash the beads before the addition of cleavage

buffer was also collected and resolved on the selection gels. Products from each collected decant were resolved by 7 % PAGE (Figure 3.4) and excised. Each gel slice was processed in the same manner as the 60 minutes decant described above in round 1 to produce polyacrylamide gel-purified First amplification product.

Table 3.1 Decant (supernatant) collected after the indicated duration of incubation from each round of selection. Fresh cleavage buffer was added after each collection.

	Water wash	1 minute	5 minutes	60 minutes
Round 1-3				x
Round 4-8		x	X	x
Round 9-21	x	x	X	x

In order to increase the stringency as the rounds progressed, varying volumes of First amplification product corresponding to different decants were combined and used as templates to produce the second amplification product as shown in Table 3.2. Rounds 1 to 3 used 4 µl of gel purified First amplification product. Rounds 4 to 7 used the following amounts of First amplification product: 1 µl of 1 minute, 1 µl of 5 minutes, and 4 µl of 60 minutes. Rounds 8 to 10 used the following amounts of First amplification product: 1 µl of water wash, 1 µl of 1 minute, 1 µl of 5 minutes, and 1 µl of 60 minutes. Rounds 11 to 17 used the following amounts of First amplification product: 1 µl of water wash, 1 µl of 1 minute and 1 µl of 5 minutes Round 18 to 20 used only 1 µl of First amplification product from the water wash and 1 minute decant.

Table 3.2 Amount of purified First amplification product added to the second amplification reaction to increase stringency according to round.

	Water wash (1 µl)	1 minute (1 µl)	5 minute (1 µl)	60 minute (1 µl)	60 minute (4 µl)
Rounds 1-3					x
Rounds 4-7		x	x		x
Rounds 8-10	x	x	x	X	
Rounds 11-17	x	x	x		
Rounds 18-20	x	x			

3.3.7 General TA cloning method

The final generation of the selection was cloned using Promega's pGEM-T Easy. This cloning kit uses a technique called TA cloning and does not require the use of restriction enzymes to create ssDNA overhangs to facilitate ligation of the insert to the vector. Instead, inserts have an extra dA added to the 3'-end of dsDNA in a non-templated fashion by a polymerase conveying extendase activity (Figure 3.5). This non-templated addition of a dA is also referred to as A-tailing. The cloning vector (pGEM-T Easy) included with the cloning kit has been prepared by cleaving with a restriction enzyme that leaves two blunt ends each bearing 5'-phosphates. The vector is further processed by T-tailing (extending with a dT) the 3' ends using a polymerase with extendase activity. Inserts are added in a 3:1 insert to plasmid ratio. With the overhanging 3'-dT, the linearized vector can pair with the overhanging 3'-dA of the insert. The 3'-dA of the insert can be ligated to the phosphorylated 5' end of the cloning vector. Unless the PCR primers used to produce the PCR product were 5'-phosphorylated, the end product of the ligation is a double nicked circularized plasmid.

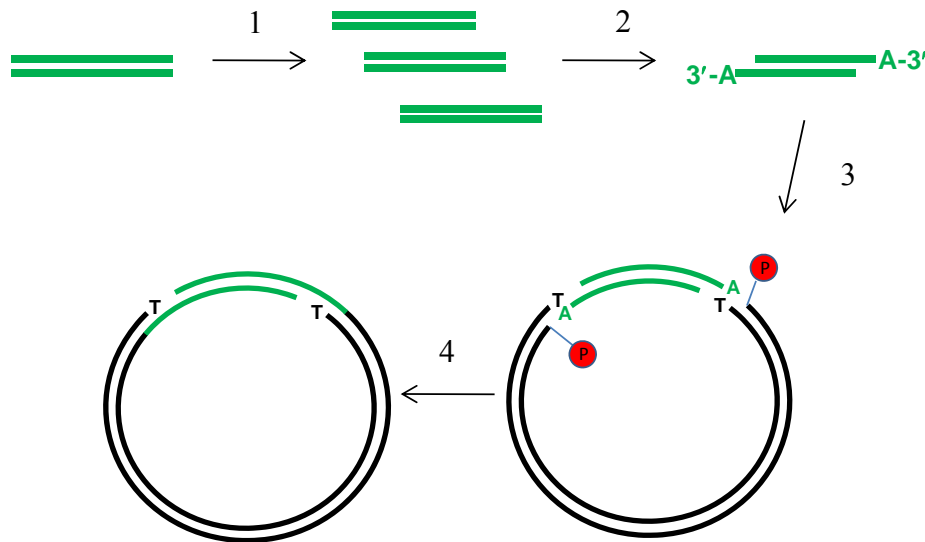


Figure 3.5 TA cloning. 1. A target DNA is amplified using PCR. 2. If the PCR did not employ the use of a polymerase with extendase ability, the PCR product is further extended with a single dA in an untemplated fashion by a different polymerase with extendase ability. 3. The A-tailed insert is incubated with T-tailed linearized plasmid. 4. The insert is ligated to a T-tailed plasmid and the newly constructed doubly-nicked circular plasmid is ready for transfection into the cell. 5'-Phosphate is represented by a P overlaid upon a red circle.

3.3.8 Cloning and sequencing of generation 20

Using primers **ODN 3.5** and **ODN 3.6**, generation 20 was amplified with Taq polymerase $20 \times (94\text{ }^{\circ}\text{C}/30\text{ seconds}, 55\text{ }^{\circ}\text{C}/30\text{ seconds}, 72\text{ }^{\circ}\text{C}/30\text{ seconds})$ to produce dsDNA with 3' A overhangs. The amplicon product was purified on 2 % agarose and extracted using Qiagen's QIAQuick Gel Extraction kit. The purified amplicon was TA cloned using Promega's TA cloning pGEM-T Easy cloning kit. The ligation was dialyzed on 1 % agarose for at least one hour followed by electroporation using Triton's Bactozapper and Invitrogen's *E. coli* DH10B Electromax cells (20 μl). The transfected cells were added to SOC media (980 μl) and incubated (37 $^{\circ}\text{C}$) with shaking (225 rpm) for one hour. Fifty microlitres of the transfected cells in SOC media were spread on LB agar (100 mg/l ampicillin, 1 mg X-gal). *E. coli* containing plasmids without insert DNA will produce blue colonies. Successful insertion of DNA into the vector disrupts the *lacZ* gene and resulting colonies appear white. White colonies were chosen at random and screened for single inserts of correct size by restriction enzyme digestion using Eco RI. Plasmids bearing single inserts were sent to UBC's Nucleic Acids and Protein Services Unit for sequencing.

3.3.9 Screening of the clones

Single-stranded templates for individual sequences obtained from the sequencing of the individual clones originating from the final generation were ordered from Integrated DNA Technologies. To screen the clones for self cleavage activity, 6 pmol of primer **2** were annealed to 5 pmol of template and extended using dC^{aa}TP, dU^{ga}TP, dA^{imm}TP, natural dGTP and $\alpha\text{-P}^{32}\text{-dGTP}$. Extension reactions were halted after 3 hours with the addition of 0.5 M EDTA (0.7 μl). Streptavidin (10 μl) was washed with water and magnetized. The water wash was decanted. Extension reactions were incubated with the prewashed streptavidin for 30 minutes at room temperature. After the 30 minute incubation, the beads were washed with TEN (2 \times 100 μl), (0.1M NaOH, 1 mM EDTA) (5 \times 100 μl), (25 mM cacodylate, pH 6 buffer) (1 \times 200 μl), DEPC-treated H₂O (1 \times 100 μl). Prior to magnetizing the final water wash two aliquots (2 μl) of beads in water were removed for controls. To one control sample, RNaseA (0.5 μl of 10 mg/ml) was added

and incubated at room temperature for 20 minutes prior to the addition of 99:1 formamide loading buffer:biotin (100 mM in DMF) (10 µl). Clones were incubated using the selection conditions (40 µl cleavage buffer, 24 °C). At time points 1, 5, 30, 60, 120 and 1320 minutes an aliquot (5 µl) was removed and quenched by adding 99:1 formamide:biotin loading buffer (15 µl). Cleaved and uncleaved oligonucleotides were resolved on a 7 % denaturing polyacrylamide gel, visualized using a phosphorimager (Amersham Typhoon 9200) and polygons were drawn around the bands corresponding to cleaved and uncleaved products. The data were fitted to a single-exponential equation (1) using Sigmaplot 2001 for an initial estimate of the observable rate constant, k_{obs} .

$$P_t = P_{\infty} \times (1 - e^{-kt}) \quad 1$$

where:

P_t = amount of cleaved product
 P_{∞} = total amount of DNAzyme
 k = observed rate constant
 t = time

3.3.10 Kinetics of native Dz20-49 and modified dA^X replacements

For four individual extension reactions, the template for Dz20-49 **ODN 3.7** (15 pmol) was annealed to 5'-biotin-modified ribose imbedded primer **ODN 3.2** (15 pmol), and the primer was extended under the same conditions as the selection rounds described in section 3.3.6 with the exception of using dNTPs of varying compositions. The nucleoside triphosphate cocktails used contained dCTP, dGTP, dTTP and one of either dA^{imm}TP, dATP, dA^{ime}TP, or dA^{imp}TP. An aliquot (6.67 µl, 5 pmol) of each of the four extension reaction was bound to streptavidin magnetic beads and washed with TEN (2 × 100 µl), (0.1M NaOH, 1 mM EDTA) (5 × 100 µl), (25 mM cacodylate, pH 6 buffer) (1 × 200 µl), DEPC-treated H₂O (1 × 100 µl). Prior to magnetizing the final water wash two aliquots (2 µl) of beads in water were removed for controls. To one control sample, RNaseA (0.5 µl of 10 mg/ml) was added and incubated at room temperature for 20 minutes prior to the addition of 99:1 formamide loading buffer:biotin (100 mM in DMF)

(10 µl). The beads with the elongated Dz20-49 constructs were incubated in cleavage buffer (100 µl) at room temperature (24 °C) and aliquots (5 µl) were removed and quenched at the following time points 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes. Cleaved and uncleaved oligonucleotides were resolved on a 7 % denaturing polyacrylamide gel, visualized with a phosphorimager (Amersham Typhoon 9200) and polygons were drawn around the bands corresponding to cleaved and uncleaved products. To explore the possibility that more than one folded states is catalytically active, equation (2) that would account for a folded state A and a folded state B was used to test for a better fit for the acquired data points.

$$P_t = P_{\infty} \times (1 - e^{-kt}) + P_{\infty}' \times (1 - e^{-k't}) \quad 2$$

where:

- P_t = total amount of cleaved product at time t
- P_{∞} = total amount of DNAzyme in folded state A
- P_{∞}' = total amount of DNAzyme in folded state B
- k = observed rate constant for folded state A
- k' = observed rate constant for folded state B
- t = time

3.3.11 37 °C kinetics of Dz20-49

The 37 °C kinetics of Dz20-49 were carried out using 5 pmol of the Dz20-49 extension reaction described in section 3.3.10 except that the incubation temperature was maintained at 37 °C using a VWR temperature-controlled water bath. The cleavage buffer was warmed to 37 °C prior to its addition to the bead-bound oligonucleotides. Collection of time point samples, resolution by PAGE and calculation of the rate constant was performed as described in section 3.3.10.

3.3.12 pH dependence

The k_{obs} values for Dz20-49 was determined at various pH values ranging from 6.0 to 9.0. Cleavage buffers used were composed of 50 mM Tris-HCl, 1 mM EDTA and 200 μ M NaCl. Cleavage buffers pH values used were 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. About 1 pmol of extension product was bound to streptavidin beads and washed. After being washed, the beads were treated with one of the cleavage buffers (40 μ l) mentioned above and incubated at room temperature (24 °C). Aliquots (5 μ l) were removed at times 5, 60, 120, 330, 540, 1350 and 1800 minutes followed by quenching with the addition of 99:1 formamide loading buffer: biotin (100 mM in DMF) (15 μ l). The data were processed as above to determine cleavage rates and fitted to equation (3). (Roychowdhury-Saha & Burke, 2006)

$$k_{obs} = \frac{k_{max}}{(1+10^{(pK_a-pH)}+10^{(pH-pK_a)}+10^{(pK_a-pK_a')})} \quad 3$$

3.4 Results

3.4.1 Progress of the selection

A selection was performed to determine the effects of making discrete changes to the linker length of the tether joining an imidazole to the 8 position of adenine. The protocol used to select for Dz10-66(Hollenstein *et al.*, 2009b) was used in this selection with the exception of using 8-(4-imidazolyl)aminomethyl)-2'-deoxyadenosine triphosphate **3.2** in place of 8-(2-(4-imidazolyl)aminoethyl)-2'-deoxyadenosine triphosphate **2.1** that was used in the original Dz10-66 selection. As discussed in the introduction of Chapter 3, the value of the rate constants of the clones resulting from this selection compared to the value of the rate constant of Dz10-66 will help determine whether or short linkers are better than longer linkers when selecting functionalized DNazymes. The selection described in this chapter did result in the isolation of active mutants. However, not only was the rate constants of the isolated clones informative, but the progress of the selection gave us insight on the performance of the clones and the value of the dA^{imm} nucleotide in a modified DNzyme.

The first sign of activity was detected at round 9. Round 10 clearly showed activity with a faint band appearing from the 60 minute decant (Figure 3.6). Activity increased slowly through the subsequent rounds, and while total cleavage after 60 minutes of incubation in cleavage buffer seemed to continue increasing over rounds 10 to 20, it did not go beyond 2 - 3 %. The selection was stopped after 20 rounds to give the 20th generation, which is also referred to as the final generation. The start of one additional round (round 21) was used to assess the activity of the 20th generation as shown using 7 % denaturing PAGE with narrower lanes (Figure 3.7). The products were not excised for amplification. While the resolution is diminished, the bands are easier to visualize giving higher contrast and seemingly higher activity (5 % total cleavage) in that final generation (Figure 3.8). Nevertheless, this low level of activity at this high number of rounds was a clear indication that it would be more informative to stop the selection and isolate the available clones rather than to continue a possibly futile attempt at finding a DNzyme with a rate constant higher than Dz10-66.

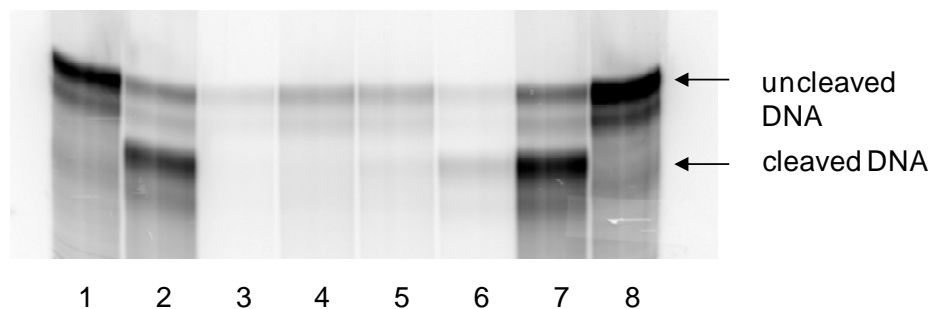


Figure 3.6 Polyacrylamide gel depicting round 10 self cleaved DNA in the reaction buffer. Lanes 1 and 8: Full length DNA, lanes 2 and 7: RNase treated DNA. Lanes 3-6: Decanted product from water wash, 1 min, 5 min, and 60 min.

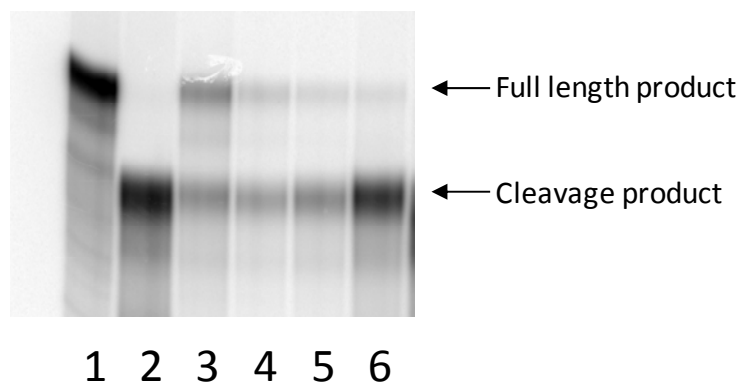


Figure 3.7 Polyacrylamide gel depicting round 21 self-cleaved DNA. Lane 1: full length DNA, Lane 2: RNase treated DNA. Lanes 3-6: Decanted product from water wash, 1 min, 5 min, 60 min.

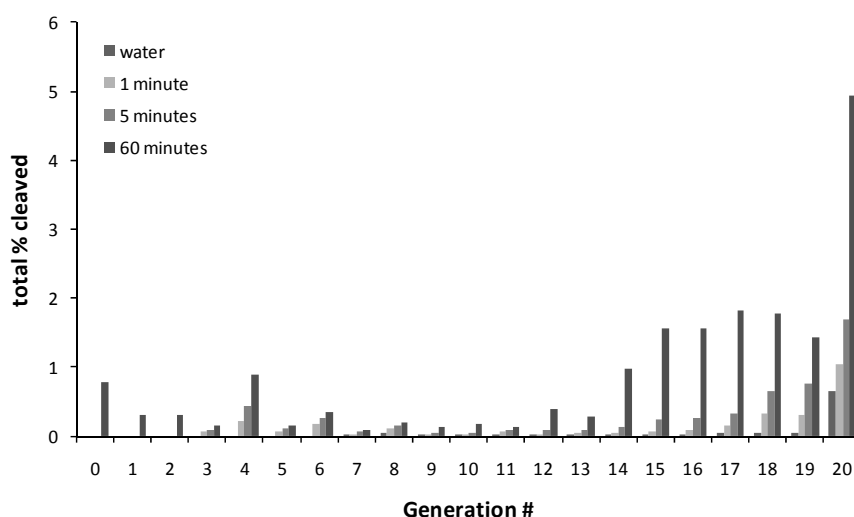


Figure 3.8 Bar graph depicting the fraction of self-cleaved DNAzyme clones as a function of generation.

3.4.2 Activity of the clones

The amplified product of the final generation was ligated into the *lacZ* gene of the vector pGEM-T Easy. A blue/white screen was used to identify *E. coli* colonies containing vectors with *lacZ* genes disrupted by the insertion of foreign DNA. From the cloning of the final generation of the selection, sixty-one white colonies were picked at random and each was used to inoculate TB media (1 ml). The plasmids were harvested and treated with restriction digest enzyme Eco RI to determine the size the inserted DNA.

From these sixty-one white colonies, forty-two contained inserted DNA of the correct size. The other plasmids contained DNA inserts of the incorrect size or no DNA insert at all. The concentrations of thirty-six of these plasmids were standardized to 50 ng/μl and the plasmids were sent to UBC NAPS for sequencing. Once the sequences were obtained, template oligonucleotides for thirty-three unique sequences were synthesized by Integrated DNA Technologies (IDT). These thirty-three template oligonucleotides were used to enzymically synthesize individual clones and screen the clones for activity. A Dz10-66 variant which was enzymically synthesized with dA^{imm} in place of the dA^{ime} of the native DNazyme was also tested for activity along with the clones from this selection. Although the Dz10-66 variant was efficiently synthesized to full length with minimal truncation products, it showed no activity. The majority of the clones showed very little activity (<5 % total cleavage) even after 22 hours of incubation. The sequences of the clones, their rate constants referred to as k_{obs} , and the total percentage cleaved after 22 hours (1320 minutes) is summarized in Table 3.3. Clones with appreciable observed rate constants, but low yields (<20 %) were not pursued for further studies. Clone 13 showed a high amount of cleavage after 22 hours, but several truncation products, which also appeared active, were present and could cause problems with future kinetic analysis. Other notable clones are clone numbers 7, 44 and 49 with k_{obs} calculated to be $4.73 \cdot 10^{-3} \text{ min}^{-1}$, $3.49 \cdot 10^{-3} \text{ min}^{-1}$ and $3.33 \cdot 10^{-3} \text{ min}^{-1}$, respectively (Figure 3.9). Clone 7 showed slight smearing. Due to clone 49's observed rate constant, high yield, minimal truncation products and minimal smearing, this clone was chosen for further kinetic analysis.

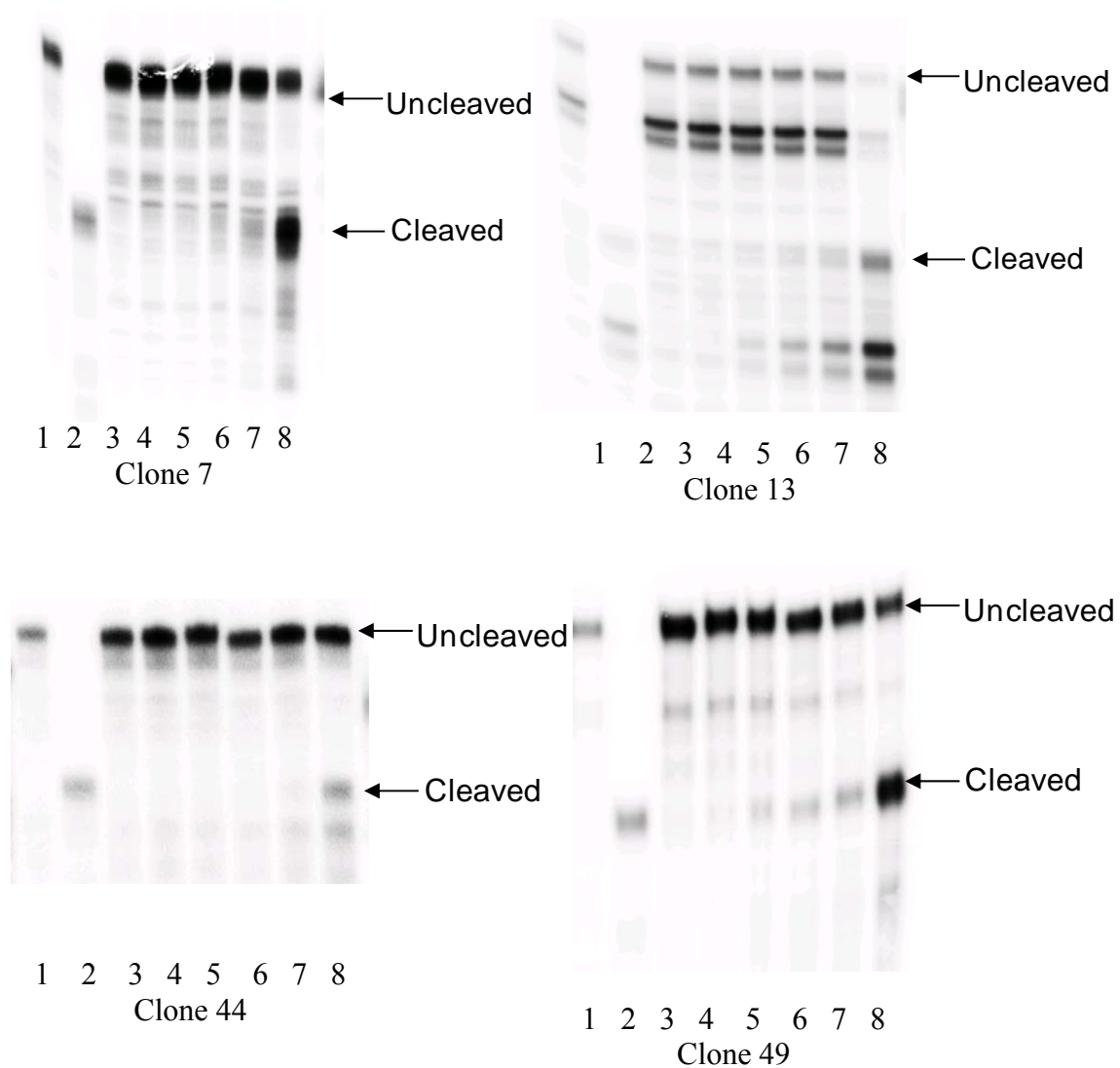


Figure 3.9 Screening of clone 7, 13, 44, and 49. Lane 1: Full-length DNazyme control. Lane 2: RNaseA-cleaved DNazyme control. Lanes 3-8: self-cleavage reactions at times 1, 5, 30, 60, 120 and 1320 minutes, respectively.

Table 3.3 Generation 20 sequences and k_{obs} .

Clone #	Degenerate Region Sequence (N40)	% yield	Truncates	k_{obs} ($\cdot 10^{-3} \text{ min}^{-1}$)
1	AUGCAUGGUU <u>AUUGUAGCAUGUGCUGUGUAGCAGCAGCGUUU</u>	<5	+++	n.d.
2	AUGCAUGGUU <u>AUUGAGUCGAGGCAUGUUAGUGAGUGUGUGCUU</u>	<5	++	n.d.
4	CUGCAUGGUU <u>AUUGAGGCGAGGCAUGUGAGGGAUUGGCUG</u>	<5	+	n.d.
5	UCAUAGUCUCGGUGGCACGUUCGUAGGUGUGAUUGUGUGU	<5	++	n.d.
6	AGUUAGUCUCUCCAGUGGCUCGCAUGAUUGUGUAGUGUGUG	6	+	2.9
7*	GU <u>AUGAGCAGUGUGUGGGAGGCGCGCUUGUGCUUGCGUUAGU</u>	64	+	4.7
8	AGUCAUGUAGUCAGUCUGCGGCACGCCGUGGUGAGGGAUGUGC	14	+	7.9
9	AUGCAUGCUU <u>AUUGAGGCGAGGCAUGCGUCGAGUGUGUGUGCGGU</u>	<5	+	n.d.
10	AGUCAUGU <u>AU</u> UCCGUUGCUAGCGCAGCAUGUGCUGUGUGUUG	<5	++	n.d.
11	GUGUUUGCUCGGCUGUGGUGCGCAGUGUGGUGCAAGUGUGU	<5	+	n.d.
13	GU <u>AUGAGUGGAGUGUGGGAGGCAUGCUUGUGGUGAGGUGGCUUU</u>	80	+++	2.4
14	ACUGUUGAGCACUAGUGAGGUGUGCACGAGUGGUGUCGGUCU	<5	++	n.d.
18	AUGCAUGCUU <u>AUUGAGGCGUGGCACAGU</u> AUGUGUGUGAGU	<5	+++	n.d.
19	AUGCAUGGUU <u>AUUGAGGCGUGGCCAGGGUGGCAGU</u> AGUGUU	<5	+	n.d.
21	AUGCAUGGUU <u>AUUGAGGCGUGGUACGCUUGUGCUUGU</u> AGUGAGU	<5	++	n.d.
22	CUGCAUGGUU <u>AUUGAGUCGAGGCAUGUGAGGGAUUGGCUG</u>	<5	-	n.d.
24	AUGCAUGGUU <u>AUUGAGUCUAGGCACGUGAUGAGUGUGAGUGCG</u>	<5	-	n.d.
25	AUGCAUGGUU <u>AUUGAGUCGAGGCUUGUGAGGGAUUGGCUG</u>	<5	-	n.d.
31	UUGCAUGGUU <u>AUUGAGGCGUGGUCGAGUGGU</u> AGUGAGU	<5	+++	n.d.
35	ACUACCAUGUGGUCUACA <u>AUGGCGGAGCACCAGUU</u> AUGUUU	<5	++++	n.d.
38	UGUCAUGUUCUCCGUGGCUCGUACGCCGUGUGUGUGUGUUA	<5	+	n.d.
41	UUCGCAUCGUGAGUGAGGCACGGUGGCGACGUGUUGUGGUGCA	<5	+	n.d.
43	AUGCAUGGUU <u>AUUGAGGUGUGGCCAGGUGGCAGU</u> AGUGUU	<5	++	n.d.
44	AGUCAUGCUCUCCAGUGGUUCGCAUGUAUGUGAGUGGAGUGU	20	+	3.5
45	AGCAGCUCGAGUCAGUUUGCGGCACGCAUGGUGGUUCGCGUGU	<5	+	n.d.
46	AUGCAUGGUU <u>AUUGAGGCGUGGCACAGU</u> AGGGUGUGAGU	<5	+	n.d.
48	AUGCAUGGUU <u>AUUGAGUCGAGUGUAGUGUAGCAGCAGCGUUG</u>	<5	+++	n.d.
49	AGUCAUGCUCUCUAGUGGUUCGCAGGUCGUGUGGGUCGUU	80	+	3.3
50	AUGCAUGGUU <u>AUUGAGUCGAGUGUAGUGUAGCAGU</u> AGCGUUG	<5	+++	n.d.
52	GCAUAGUCUCGGUGGCACACUCGUAGAGGUGGUAGUGUCA	<5	++++	n.d.
54	AGUCAUGUAGUCCGUUGCUAGCGCGCCAUGUGCUGUGUUG	<5	++	n.d.
55	AUGCAUGGUU <u>AUUUAGUCGAGGCAUGUGAGGGAUUGGGUG</u>	<5	-	n.d.
57	CUGUUUGCUCGACU <u>AUGGCGCGCAGUGUGGUCUU</u> AGUGUUU	<5	++	n.d.
<i>For reference: Dz10-66 constructed using $da^{imm}TP$</i>				
10-66	CUAGCAGCGCAAGUGAGGCGCGCUAUGAGUGUGGUGCGUGUAU	<5	-	n.d.

n.d. = none detected

- = no significant truncates

+ = slight truncates

++ = truncates

+++ = significant truncates

++++ = almost no fully extended products

* = slight smearing

3.4.3 Predicted secondary structure of Dz20-49

The sequence of the self-cleaving species 20-49, here on referred to as Dz20-49, was input into mFOLD(SantaLucia, 1998; Zuker, 2003) along with the following parameters: forced base pairs 1 to 6, 200 mM NaCl and 24 °C (Figure 3.10). As calculated by mFOLD, the secondary contains three stem loops – one engineered (I) and two arising from the degenerate region (II and III).

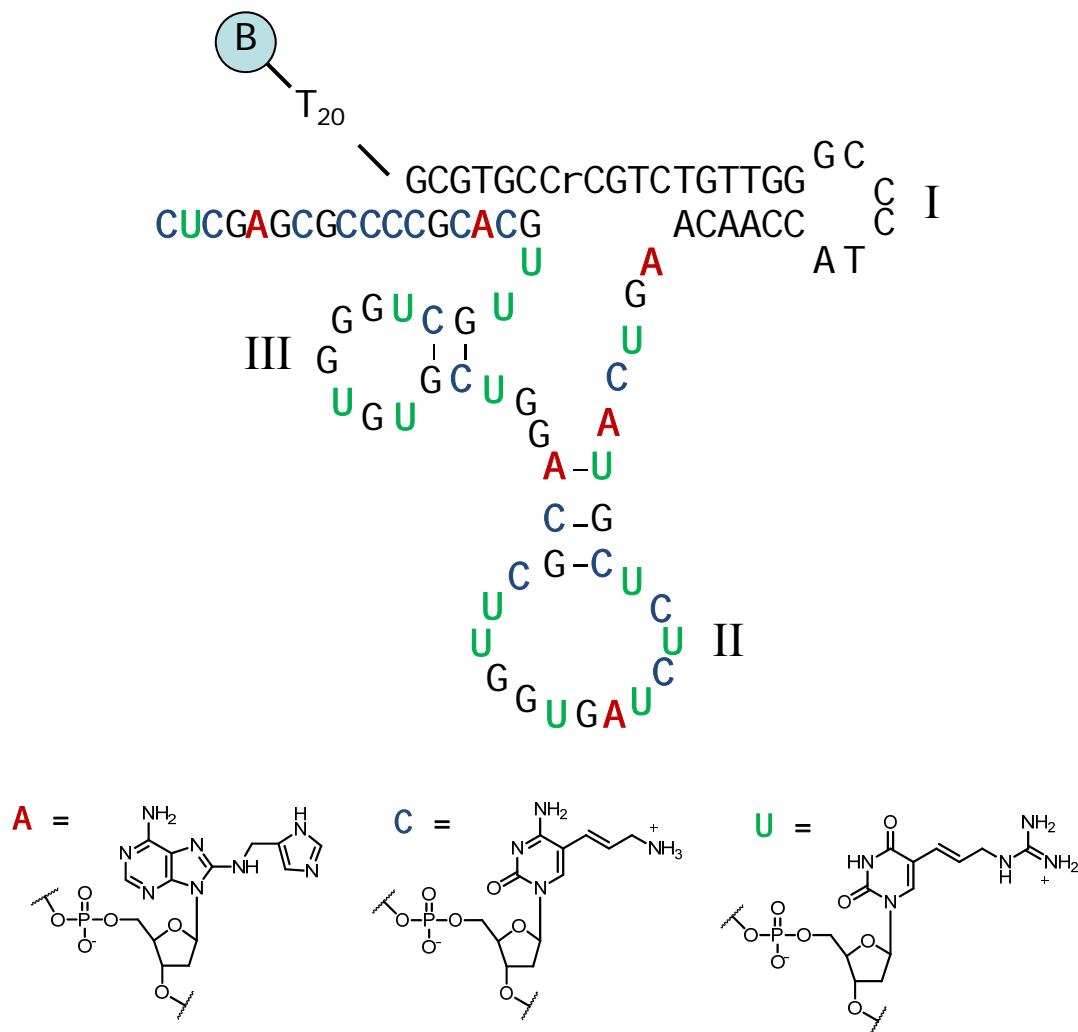


Figure 3.10 Predicted secondary structure according to mFOLD(SantaLucia, 1998; Zuker, 2003) with the parameters 200 mM NaCl and 24 °C. Engineered stem loop is indicated by a I. Indicated loops II and III were predicted by mFOLD. Modified bases are indicated by boldface A, C, and U. Embedded ribonucleotide is indicated by an rC.

3.4.4 Dz20-49 kinetics

3.4.4.1 Room temperature (24 °C) kinetics

The rate of self cleavage of Dz20-49 at room temperature (24 °C) was determined by sampling the reaction at several points over the course of 96 hours (Figure 3.11) using a stop-quench kinetic assay. A notable feature of the data set when fitted to either a single-exponential and double-exponential equations is that the pre-exponential term, which represents the fraction of DNazymes cleaved at time equal to infinity has an apparent value that is significantly lower than unity. This indicated that a significant fraction of the DNzyme is inactive, which may be due to misfolding although misincorporation of a non-cognate base cannot be entirely excluded. Correction for these inactive DNazymes increases the amplitude of the graph and raises the value of the pre-exponential term to a value of one. However, changing the amplitude does not affect the value of the calculated rate constant. Rate constants were calculated using either a single- or double-exponential equation without correcting for the inactive DNazymes. While the data appears to fit a double-exponential equation, the calculated values of the fast phase and the slow phase rate constants are subject to debate. The majority of the active DNazymes (~88%) are slow-cleaving species. The small contribution of the fast-cleaving species and the low number of data points leads to a calculated rate constant for the fast phase with an unacceptably high associated error; when the data were fit to the double-exponential equation (2), the high error ($\pm \sim 50 \%$) associated with the second exponential term, a minor contributor (of <13%) that appeared to be a fast-cleaving species, could not be accurately measured with these sets of data points. Calculation of the second minor term could be accomplished with more thorough kinetic data.(Ting, Thomas, & Perrin, 2007) However, studying biphasic kinetics of the Dz20-49 constructs is not the goal of this chapter. Comparing the magnitude of the rate constant of the DNzyme that composes the majority of the population to the magnitude of the rate constant of Dz10-66 is the goal. Calculated rate constants from the single-exponential equation were found to be close in value to the rate constants of the slow-cleaving phase derived from the double-exponential equation and are good approximations for our purposes. Therefore, all calculated rate constants reported were determined using a

single-exponential equation. Three separate trials gave P_{∞} and k_{obs} values of 0.75 ± 0.03 and $3.2 \pm 0.3 \cdot 10^{-3} \text{ min}^{-1}$, 0.68 ± 0.02 and $3.4 \pm 0.3 \cdot 10^{-3} \text{ min}^{-1}$, and 0.69 ± 0.02 and $3.4 \pm 0.3 \cdot 10^{-3} \text{ min}^{-1}$, respectively. The calculated average rate constant, k_{avg} , of Dz20-49 at 24 °C is $3.5 \pm 0.4 \cdot 10^{-3} \text{ min}^{-1}$ averaged from the three separate trials.

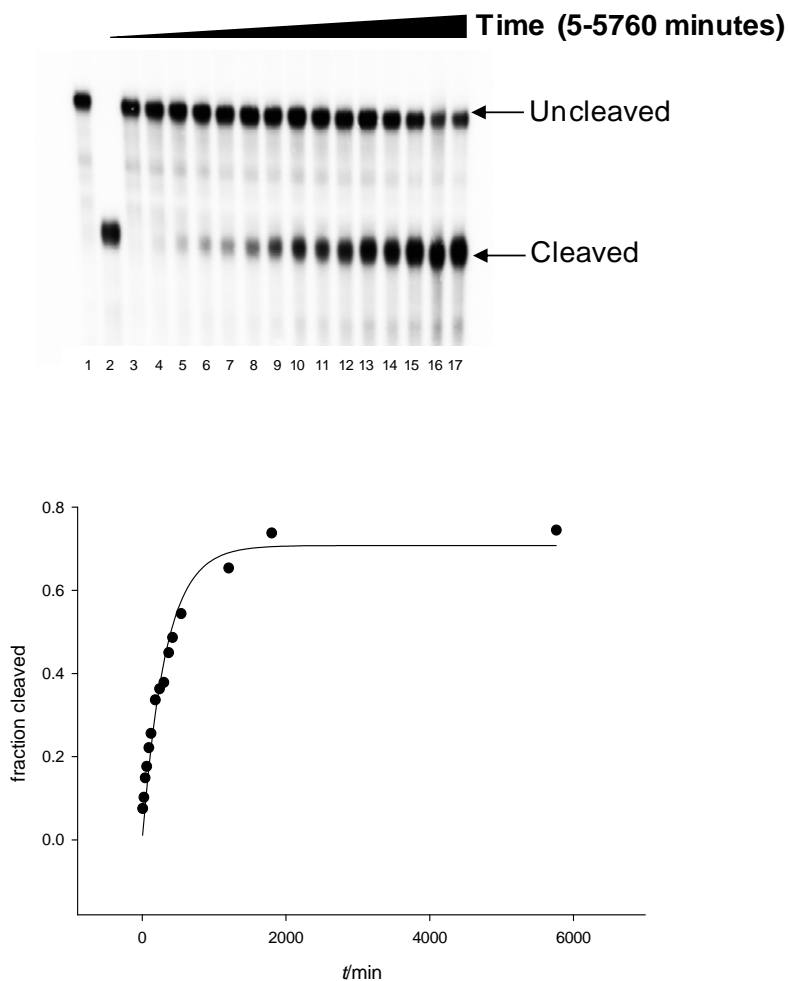


Figure 3.11 Top: Autoradiograph of the time-dependent self-cleavage of Dz20-49. Lane 1: Full-length DNAzyme. Lane 2: RNaseA-cleaved DNAzyme. Lanes 3-17: Time-dependent self-cleavage reaction sampled at times 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively. Bottom: The fraction of DNAzyme cleaved was graphed as a function of time. These data point were fitted to a single exponential equation. The self-cleaving rate was determined to be $3.1 \cdot 10^{-3} \text{ min}^{-1}$ for this particular run.

3.4.4.2 Modified dA requirement

To determine whether or not the imidazole function was playing a critical role in the DNAzyme, a Dz20-49 variant was synthesized by replacing dA^{imm}TP with dATP in the enzymic synthesis reaction. The resulting construct contains only the guanidinium and aminoallyl functionalities. The ablation of the imidazole functionality assesses its importance in either direct catalysis or indirectly in terms of folding of a catalytically competent structure. The imidazole ablated construct was unable to perform any significant self-cleavage (Figure 3.12). Although the exact role of any one imidazole cannot be ascertained without a more systematic knockout of individual imidazoles, this study does show that the imidazole functionality is absolutely critical for self-cleavage.

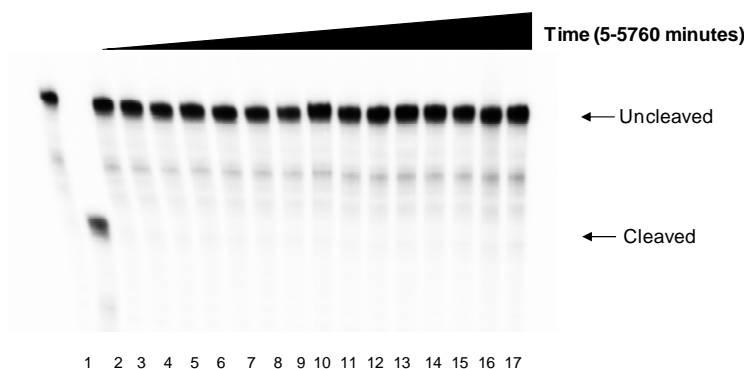


Figure 3.12 Autoradiograph showing the lack of activity from a Dz20-49 variant bearing no imidazoles. Lane 1: Full-length DNAzyme analog. Lane 2: RNaseA-cleaved DNAzyme analog. Lanes 3-17: Time-dependant self-cleavage reaction sampled at times 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively.

3.4.4.3 Aminoethyl and aminopropyl replacement

As further proof of the importance of the imidazole in this construct, Dz20-49 was constructed using dA^{ime}TP. This construct showed significant self-cleavage matching the native construct. Although the Dz20-49's active site was putatively optimized for the dA^{imm} modification, the Dz20-49 variant displays significant self-cleavage activity showing that the catalytic potential of dA^{ime} compensates for the dA^{imm}-optimized

environment of the active site of Dz20-49 (Figure 3.13). The averaged k_{obs} from three different trials performed at 24 °C is $4.1 \pm 0.5 \cdot 10^{-3} \text{ min}^{-1}$. In a subsequent experiment, the dA^{imm} modification of Dz20-49 was replaced using dA^{imp} . This construct was only able to perform a trace amount of self-cleavage (Figure 3.14).

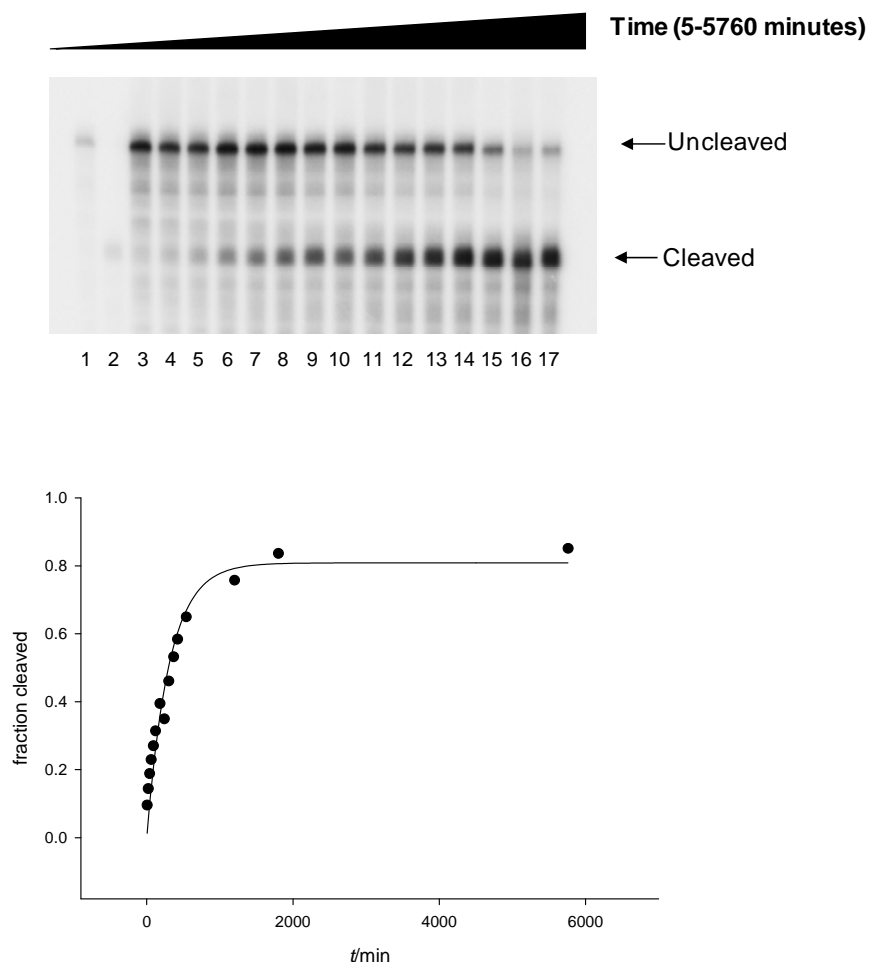


Figure 3.13 Self-cleavage kinetics of the aminoethyl linker analog of Dz20-49. Top: Autoradiograph of the self-cleavage reaction. Lane 1: Full-length DNAzyme analog. Lane 2: RNaseA-cleaved DNAzyme analog. Lanes 3-17: Time-dependent self-cleavage reaction sampled at times 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively. Bottom: Graph showing the fraction of DNAzyme cleaved as a function of time. The rate constant was calculated to be $3.2 \cdot 10^{-3} \text{ min}^{-1}$ for this particular run.

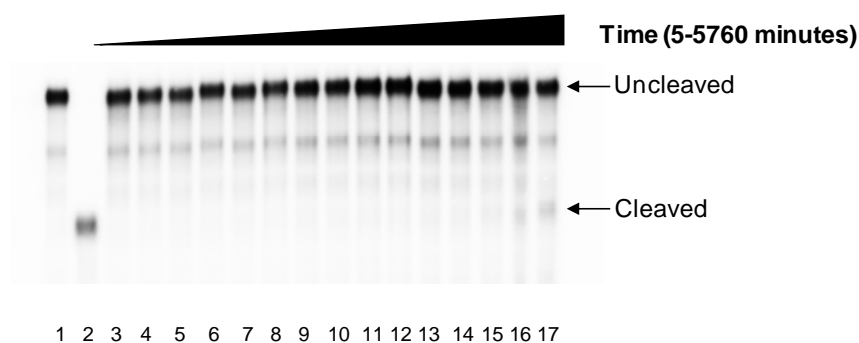


Figure 3.14 Autoradiograph of the self-cleavage kinetics of the aminopropyl linked analog of Dz20-49. Lane 1: Full-length DNAzyme analog. Lane 2: RNaseA-cleaved DNAzyme analog. Lanes 3-17: Time-dependant self-cleavage reaction sampled at times 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively.

3.4.4.4 37 °C kinetics

The divalent metal cation-independent DNAzymes DzG3 and Dz9₂₅-11 were both found to have optimal operating temperature at the temperatures that they were selected at room temperature, ~24 °C. The k_{obs} values of these two DNAzymes were lowered at the physiologically relevant 37 °C. The k_{obs} values for Dz9-86 and Dz-10-66 increased when self-cleavage is performed at 37 °C indicating that the catalytic rate enhancement of the elevated temperature outweighs any distortion of the active structure or that the active structure has minimal, if any, distortion. To probe which trend Dz20-49 follows, kinetic experiments were performed at 37 °C.

Dz20-49 was tested for increased catalysis at an elevated temperature. The temperature 37 °C was chosen as it is a physiologically relevant temperature and previously selected DNAzyme showed a rate maximum around this temperature. Unlike the previously selected DNAzymes, however, raising the temperature for self-cleavage to 37 °C was found to be detrimental to Dz20-49 (Figure 3.15). Lack of a sigmoidal curve in the graph suggests that there is no step involving a competitive folding step resulting in an active conformation. To determine if the graph is monophasic or biphasic, the data points were fitted to both a single-exponential equation and a double-exponential

equation, and they appear to fit the single-exponential equation the best. The average k_{obs} of three individual trials was determined to be $2.0 \pm 0.5 \cdot 10^{-3} \text{ min}^{-1}$. This drop in k_{obs} suggests that the Dz20-49 has a less stable three-dimensional structure compared to the closely related Dz9-86 and Dz10-66.

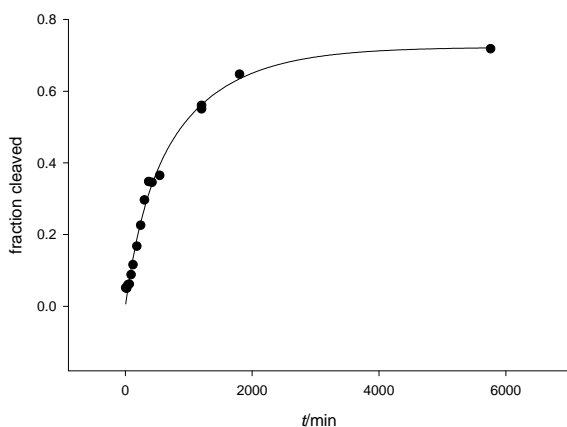


Figure 3.15 Kinetics of Dz20-49 at 37 °C. Temperature was maintained using a water bath at 37°C. The self-cleaving rate constant was determined to be $1.61 \cdot 10^{-3} \text{ min}^{-1}$ for this particular experiment.

3.4.4.5 pH rate profile

The pH dependence of the DNAzyme will give insight into the mechanism of catalysis and what roles the imidazoles play in that mechanism. Two possible outcomes are that the pH rate profile appears log-linear or bell-shaped with respect to pH. Two nucleic acid enzymes that show a log-linear profile are the hammerhead ribozyme and DzG3. The log-linearity indicates that the rate limiting step for these two nucleic acid enzymes is a single proton transfer. The second possibility is that the pH rate profile is bell-shaped. This is an indication that the mechanism undergoes a two-step general acid/general base mechanism. Imidazole-dependent Dz9₂₅-11, Dz9-86, and Dz10-66 all show a bell-shaped pH rate profile.

Dz20-49's cleavage rate constants were determined at the following pH values: 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. Using the computer program SigmaPlot 2001, the rate constants were fitted into equation 1 to obtain the graph shown in Figure 3.16. The value of the rate constants diminishes at high and low pH. This observation is characteristic of a general acid / general base mechanism. The rate maximum appeared to be between 7.0 and 7.5. The pK_a and $pK_{a'}$ were determined to be 6.1 and 8.1, respectively.

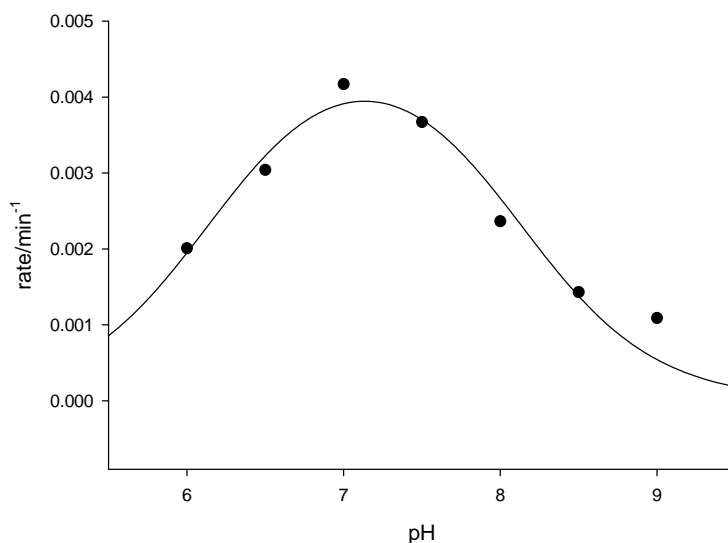


Figure 3.16 pH profile of Dz20-49. Rates were determined at the following pH: 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0.

3.5 Discussion

3.5.1 General discussion

In the past, the quest for higher catalytic rate constants for functionalized DNazymes has been achieved by adding additional functional groups and increasing the size of the degenerate region.(Hollenstein *et al.*, 2009a; Hollenstein *et al.*, 2009b) In this chapter, the selection of a DNzyme with a higher rate of cleavage than Dz10-66 produced an active DNzyme, Dz20-49. Unfortunately, the overall goal of attaining a higher catalytic rate constant than the rate constant of Dz10-66 was not achieved. Despite Dz20-49 being heavily functionalized, Dz20-49's catalytic rate constant was

nearly three orders of magnitude lower than Dz10-66. Post selection analysis involving replacement studies determined if the imidazoles are critical to the function of the DNAzyme. Determining temperature and pH dependence was also used to help elucidate the roles of the imidazoles.

3.5.2 Selection for ribophosphodiester bond cleaving activity

Progress of the selection described in this chapter was sluggish compared to the selection of Dz10-66. In the selection of Dz10-66, the activity was so high by the third round, over 30 % of the total pool was self-cleaved in one hour.(Hollenstein *et al.*, 2009b) The rapid progress of the Dz10-66 selection is an indicator of the presence of DNAzymes with high rate constants or a high content of sequences capable of self-cleavage in the gene pool. In this selection, activity remained undetected until faint activity appeared in round nine. In contrast to the Dz10-66 selection, this absence of activity until such a late round suggested the absence of DNAzymes with high rate constants. In attempts to enrich the pool in fast cleaving DNAzymes once the activity was detectable, the stringency was increased by amplifying the DNA in the 0 (water wash), 1, and 5 minute decants instead of only the 60 minute decant. By round 18 only the 0 and 1 minute decants were used.

The overall percent yield of the generations after 20 rounds failed to go above 2 % after 60 minutes. The selection was called to a halt at Round 20. In this chapter, the end product of round 20 is the gene pool referred to as generation 20 or the final generation. Although no further rounds were done, the activity of generation 20 was examined. The polyacrylamide gel used to observe activity of round 20 was cast with narrower lanes, which produced clearer bands and made the overall % cleavage appear much higher (~5 %) than that of the immediately previous rounds (~2 %). Regardless of a possibly misleading gel showing high activity, the performance of the final generation of this selection is far below the performance of the very early generations (generation 2, 30 % yield) of the Dz10-66 selection. This observation, the lack of high activity, established the primary goal of this current selection, the evaluation of the performance of the methyl linker vs. the ethyl linker. Therefore, no further rounds were performed.

The low activity of the generations makes competing reactions a threat to the success of the selection. One known competing reaction, spontaneous hydrolysis, is a concern only in the early rounds. Preferential amplifiability due to the modifications is another possible competing reaction. The modifications in previous selections were known to be important (but not of them critical) for catalysis.(Thomas *et al.*, 2009) If the dA^{imm} is critical for catalysis then it would be present, perhaps in high number, in the active clones. The benefits of the modification are assumed to be enhanced acid/base catalysis, but the negative impacts of the dA^{imm} modification are still unknown. In the later rounds, a competing reaction involving the preferential amplifiability of certain clones, due to the modifications and regardless of their activities, becomes a concern. Preferred amplifiability will be further discussed in Chapter 4. The purpose of this current chapter is the identification and characterization of the most superior dA^{imm}-dependent DNAzyme and comparing it to Dz10-66.

3.5.3 Dz20-49's secondary structure

DNAzyme Dz20-49 contains three stem loop structures according to mFOLD (Figure 3.10).(SantaLucia, 1998; Zuker, 2003) Loop I is engineered to position the N40 region close to the ribophosphodiester bond. Loop II and Loop III are part of the N40 region and constitute the active site motif. That said, however, the program mFOLD does not take into account the effects of the modifications and one would expect that the sheer number of the modifications would have a great influence on the three-dimensional structure. Therefore, while a predicted structure is presented, the actual structure is possibly very different and will not be known until an NMR or X-ray crystal structure is elucidated. The structure predicted by mFOLD doesn't take into account many structural features of our modified DNAzyme and is most likely inaccurate, but it does provide a starting point for a possible three-dimensional structure.

3.5.4 dA^x replacement studies

By replacing dATP for dA^{imm}TP in the synthesis of Dz20-49, we effectively produced the DNAzyme construct without the 8-(4-imidazolyl)aminomethyl side chain

modification. The other two modifications, the allylamino and guanidinium, were not systematically replaced in this study as those substitutions were thoroughly covered in the literature for other DNAzymes (Hollenstein *et al.*, 2009a; Thomas *et al.*, 2009) and the focus of this chapter was the tolerance of the imidazole functionality to discrete changes in linker length. Consistent with the characterization of Dz9-86 and Dz10-66, the construct with the ablated imidazole functional group was inactive. Much more interesting, however, is that the DNAzyme constructed using dA^{ime}TP in place of dA^{imm}TP was active with a catalytic rate nearly identical to the native Dz20-49. This interesting finding shows us that the imidazole functionalities are not necessarily finely tuned into position and in this case the DNAzyme activity is not diminished by increasing the entropy of the imidazoles groups by using a longer linker. This finding is contrary to what was first hypothesized. Indeed, the catalytic value of the dA^{ime} has been proven in many preceding selections and again showed its value here by matching the catalytic rate constant of Dz20-49 in a DNAzyme scaffold that was not optimized for its use. Any detrimental effects on the catalytic rate due to using a scaffold not optimized for the functional group of dA^{ime} is compensated by its superior rate enhancing ability.

To examine the potential of extending the linker, dA^{imp}TP **3.3** (Figure 3.2) was used in place of dA^{imm}TP to create the Dz20-49 construct with the imidazole linker extended by two additional methylene groups. The construct showed only trace activity (Figure 3.14). Caution should be used in the interpretation of this finding. This finding does not indicate that the aminopropyl linker is unsuitable for selection, only that the scaffold produced by the primer extension using the Dz20-49 template, may be optimized for shorter linkers. Also, as mentioned in the discussion about the three-dimensional structural changes in this chapter's introduction, these chemical changes must have a large effect on the overall scaffold structure. Indeed, dA^{imp}TP still has much potential to construct active DNAzyme catalysts even though no selection using this particular nucleoside triphosphate has been performed.

3.5.5 Temperature dependence

The room temperature rate constants of Dz20-49 and Dz10-66 were compared. At room temperature (24 °C), Dz20-49 has a k_{obs} of $3.5 \pm 0.4 \cdot 10^{-3} \text{ min}^{-1}$. This self-

cleavage rate constant is lower than the self-cleavage rate constants of Dz10-66 (k_{obs} of $0.50 \pm 0.05 \text{ min}^{-1}$, 24 °C). This is strong evidence of the catalytic inferiority of the dA^{imm} used by Dz20-49 vs. the dA^{ime} used by Dz10-66 and other DNazymes.(Hollenstein *et al.*, 2008; Hollenstein *et al.*, 2009a; Hollenstein *et al.*, 2009b; Perrin *et al.*, 2001) The reasons for the lack of catalytic rate enhancement by the dA^{imm} are unknown. It was hypothesized that shortening the linker by one methylene group would not drastically change the pK_a of the imidazole group. However, we did not verify this with a titration prior to the selection. If the pK_a has been shifted away from near neutrality, the DNzyme does not benefit from the imidazole functional group any more than it would a pK_a -perturbed nucleobase like G12 of the hammerhead ribozyme.(Doherty & Doudna, 2001) It was also hypothesized that the shorter linker would help produce a well-defined active site. The reduced degrees of freedom of the imidazole were introduced so that the putative catalytic imidazoles would be better positioned into their active conformation. Imidazoles are also thought to play a role in folding. Once again, it was hypothesized that constraining the imidazoles would promote proper folding of a well-defined three-dimensional structure. The lack of thermostability suggests that the constrained imidazoles do not stabilize three-dimensional structure the same way that the imidazoles do in Dz10-66. This lack of stabilization of the three-dimensional structure may persist at room temperature. The aminoethyl linker could possibly have the minimum linker length to promote imidazole-mediated interactions between neighbouring DNA strands. These are the proposed reasons for the lack of catalytic rate enhancement.

The self-cleavage rate constant was also determined at physiologically relevant temperature of 37 °C. The change in rate could follow one of two possibilities: 1) the rate could increase like Dz9-86 and Dz10-66 or 2) the rate could decrease like DzG3. In the first situation, the rate does increase, is indicative of the effects of a temperature-dependent rate constant dominating the reaction over any negative disruption of the active structure at a given temperature. On the other hand, if the rate decreases, this indicates that the possible disruption to the three-dimensional structure of the DNzyme is the dominant factor. We have hypothesized that the thermostability of Dz9-86 and Dz10-66 is attributable to the guanidinium groups. However, Dz20-49 also has numerous (fourteen) guanidinium groups and yet does not show the same thermostability

as Dz9-86 and Dz10-66. The imidazole is hypothesized to play a role in the folding of the three-dimensional structure. This role may have been undermined in the characterization of Dz9-86 and Dz10-66 as the increased thermostability was immediately credited to the guanidinium functionality. In fact, it could be that the contribution of the guanidinium to the thermostability is enhanced in the presence of the dA^{ime} modification. This brings to the forefront the possibility that the dA^{ime}, perhaps in concert with the dU^{ga}, aids in the stabilization of the three-dimensional structure whereas the dA^{imm} does not. Since the results of the 37 °C kinetics show a drop in k_{obs} ($2.0 \pm 0.5 \cdot 10^{-3} \text{ min}^{-1}$), this indicates that the DNAzyme has less thermostability than Dz10-66 even though it bears more dU^{ga} modifications.

3.5.6 pH dependence of Dz20-49

Although the bell-shape of the pH rate profile implies a two-step deprotonation and protonation event, it is difficult to assign the calculated $\text{p}K_a$ values to the general acid or general base as assigning the values to either would both give a bell-shaped curve. Dz20-49 has a maximum rate between 7.0 and 7.5 and the calculated $\text{p}K_a$ and $\text{p}K_a'$ were determined to be 6.1 and 8.1, respectively. For Dz10-66, the maximum rate was found to be about pH 7. The $\text{p}K_a$ and $\text{p}K_a'$ values were determined to be 6.6 and 7.5, respectively. Dz20-49's rate constant is in the same order of magnitude as non-modified DzG3.(Geyer & Sen, 1997) DzG3 performs catalysis without the use of an imidazole group. This implies that the functional groups present in Dz20-49 offer no rate enhancement over those non-modified DNAzyme, which are not dependent on imidazoles for deprotonation and/or protonation of the substrate. DzG3 did not show significant pH dependence between pH 5-9 and the authors suggest that the rate-limiting step is not a proton transfer, but perhaps a conformational change. Dz20-49's bell-shaped pH rate profile may suggest a two-step proton transfer mechanism mediated by the imidazoles, and the replacement studies discussed in section 5.3.5 do show that the imidazoles are critical for catalysis. However, the roles of the imidazoles in the catalytic mechanism have not been defined as they have been with 9₂₅₋₁₁.(Thomas *et al.*, 2009) Indeed, the imidazoles may play a large role in the folding of the active structure, and their pH-sensitive nature may cause

the rapid disruption of the active structure at high and low pH values. Disruption of the active structure due to the protonation or deprotonation of imidazoles at low and high pH values, respectively, would also cause a drop in k_{obs} at high and low pH values. This could possibly appear as a bell-shaped curve if the rate-limiting step depends on a well-defined conformation at high and low pH values. Although not part of this study, unambiguous identification of the general acid and general base of a modified DNAzyme is possible through the use of solvent kinetic isotope effects, tagging, and rescue studies.(Thomas *et al.*, 2009)

Chapter 4: Enzymic Recognition of Modified DNA

4.1 Introduction

As seen in Chapter 3, the synthesis of modified DNA and its use in a selection can lead to the isolation of a functional DNAzyme. Modifications to DNA can have other effects besides bestowing catalytic function. Enzymic manipulation of the functional or non-functional modified DNA should be characterized as this could play an important role in selections. The manipulation of single-stranded modified DNA in the context of PCR, the production of doubly-modified DNA, and the resistance of doubly-modified DNA to digestion by restriction endonuclease will be explored.

4.1.1 Modified DNA as a template

Spearheaded by Dr. Marcel Hollenstein and inspired by the works of Dr. David Perrin, several selections successfully resulted in the isolation of functionalized DNAzymes. The final generation of each selection was cloned and sequenced to identify the most efficient individual clone in the pool. Amongst the selected DNAzymes are ones that sense mercury (Dz10-33),(Hollenstein *et al.*, 2008) cleave a short section of RNA (Dz12-57), increase the diversity of functional groups (Dz9-86)(Hollenstein *et al.*, 2009a) and combine both increased modification diversity and increased sequence space (Dz10-66).(Hollenstein *et al.*, 2009b) The imidazoles of these modified DNAzymes appear to be critical to the catalytic phenotype of the DNAzyme. Research done on the role of the functional groups in Dz9₂₅-11 has shown us that a minimal number of critical imidazoles are necessary for acid/base catalysis while the rest appear to be needed for folding.(Thomas *et al.*, 2009) Since this is the primary desired trait for the ribophosphodiester bond-cleaving DNAzymes, selection schemes have been designed to optimize this one particular self-cleaving task.(Hobartner & Silverman, 2007; S.K. Silverman, 2005) Catalysis is the end activity desired from the DNAzyme and is the selection pressure purposefully introduced in a selection, but is only one of the traits that affect the success of the selection.

The design of the nucleotide dA^{imm}TP used in Chapter 3 was meant to test the effects of varying linker arm length with the hope of enhancing acid/base catalysis.

Previous incorporation studies showed that modified dA^{imm}TP is accepted as a substrate in primer extensions only slightly less efficiently than dA^{ime}TP.(Lam *et al.*, 2008) While the incorporation and polymerizability of modified nucleotides were considered and addressed, efficiency of read-through of modified DNA and its repercussions in the context of a selection has not been explored. More specifically, the efficiency of the interaction of the polymerase and the modified nucleotide present in a modified DNA template strand, such as Dz20-49, imposes another selection pressure that may not have been accounted for, and could result in the elimination of active clones regardless of their catalytic activity (Figure 4.1). One example of the selection of amplifiability of a molecule comes from the first ever *in vitro* evolution involved in the selective amplification of RNA, and this experiment gave us an insight on one of the major factors to be considered as a selection pressure.(Mills *et al.*, 1967) Amplifiability of a modified or unmodified nucleic acid is the determining factor that governs the survival of a species through a round of selection; as such this property cannot be undermined or compromised.

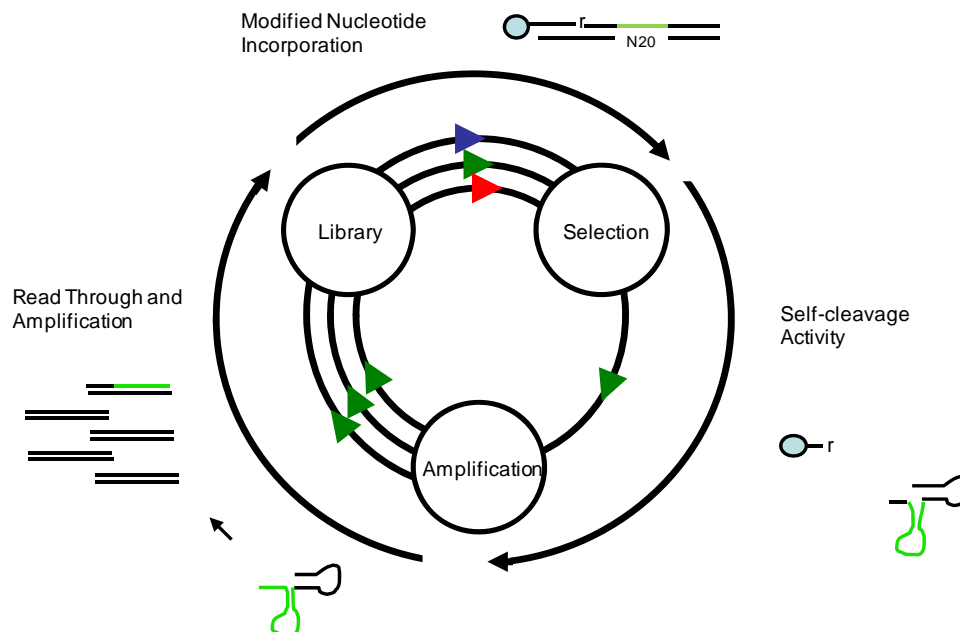


Figure 4.1 DNAzyme selection processes that affect fitness of a sequence in a self-cleaving DNAzyme selection include both incorporatability and read-through for amplification.

Efficient read-through of 8-substituted adenosines is critical for unbiased amplification. In this chapter, we noticed a subtle trend of low 8-modified dA content in the final generations from different selections provided by Dr. Marcel Hollenstein's five selections (mercury sensor Dz10-13 selection, RNA-cleaving Dz10-91 selection, Dz10-96 selection, Dz10-66 selection and the two modifications, N40 selection). Mr. Curtis Lam provided the final generations for an ATP sensor selection and a stachyose sensor selection. These two selections used DNA bearing the phenol functional group that was introduced by elongating the DNAzymes using the modified nucleotide 5-(para-hydroxybenzamido)methyl-2'-deoxyuridine-5'-triphosphate (dU^{ph}TP) (Figure 4.2 center). The stachyose sensor selection resulted in the isolation of a modification-dependent DNAzyme, DzSTA17, which was found to operate independently of stachyose. The biased nucleotide content seen in Dr. Hollenstein's selections did not observed in the dT content of the clones of Curtis Lam's two selections. In Mr. Lam's ATP sensor and stachyose sensor DNAzyme selections, the lone modified nucleoside triphosphate used was the dU^{ph}TP, which is well tolerated by Vent (exo-) polymerase. Although Mr. Lam's selections did not successfully isolate DNAzymes activated by binding either ATP or stachyose, the isolated DNAzymes were active Mg^{2+}/Zn^{2+} -dependent self-cleaving DNAzymes that also depend on the phenol modification.

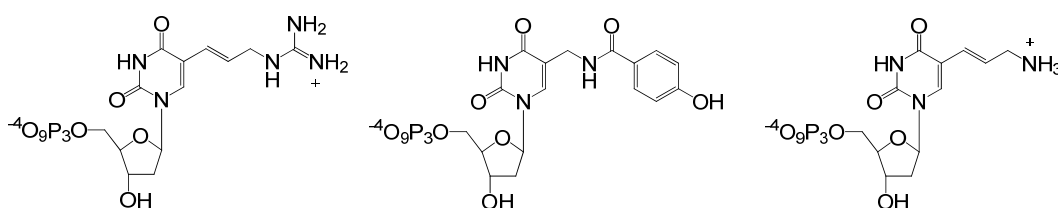


Figure 4.2 Modified dUTP derivatives dU^{ga}TP (left), dU^{ph}TP (center) and dU^{aa}TP (right) used in PCR, restriction digests and transfection.

To begin analyzing this low 8-modified dA sequence content phenomenon, the number of clones is analyzed according to their 8-modified dA sequence content for Dr. Hollenstein's selection and the Dz20-49 selection described in Chapter 3 or their 5-

modified dU sequence content for Curtis Lam's selections. Analyzing this relation provides insight into the inherent pressure by the unnatural nucleotides that are orthogonal to the beneficial catalytic rate enhancement. This relationship is further explored by comparing the amounts of amplicon produced by amplification using a modified template containing dA, dA^{imm}, dA^{ime}, or dA^{imp}. The relative quantities of amplicons produced helps explain why certain sequences encoding for fewer dA^{imm} appear in the final round of the selection described in Chapter 3. Our findings show that the native DNAzyme construct Dz20-49 is very poorly amplified compared to the fully unmodified version of the DNAzyme and even the dA^{ime} version, referred to as Dz20-49H, or the dA^{imp} version. In terms of amplifiability as well as overall fitness, sequences that contain dA^{imm} bear disadvantages in an *in vitro* selection and this modification should be avoided in future selections.

4.1.2 Restriction enzyme digestion of modified DNA

Although not a part of our DNAzyme selection scheme, resistance to enzymic degradation is an important characteristic of modified and unmodified DNAzymes, both of which for potential clinical applications should exhibit higher stability and resistance to degradation than RNA. If the modifications somehow bestow additional stability against nucleases, this would be beneficial for future selections and therapeutic applications. An endonuclease of interest to us is the type II restriction enzyme. Type II restriction enzymes are dimeric and cleave double-stranded DNA (dsDNA) at palindromic sites 4-8 bases in length. Therefore, when evaluating the efficiency of cleavage of modified dsDNA by type II restriction enzymes, both strands of dsDNA should be modified.

For this chapter, three modified versions of dUTP will be used to produce doubly-modified dsDNA (Figure 4.2). These are the 5-guanidiniumallyl-2'-deoxyuridine-5'-triphosphate (dU^{ga}TP), the 5-(para-hydroxybenzamido)methyl-2'-deoxyuridine-5'-triphosphate (dU^{ph}TP) and the 5-aminoallyl-2'-deoxyuridine-5'-triphosphate (dU^{aa}TP). All three have been successfully used in DNAzyme selections and are amenable to PCR using customized conditions. Once the PCR products have been produced, they can be

subjected to treatment with restriction enzymes. One must be careful assigning resistance to digestion to the modifications when testing PCR amplicon(s). The potentially mutagenic nature of base modifications in PCR may cause the amplicon(s) to contain mutations that change the sequence of the restriction site. To ensure that no mutagenesis has occurred, the amplicon can be cloned and sequenced. To facilitate the cloning process, the modified amplicons can be amplified into unmodified dsDNA the same method that was done for the modified DNAzyme selections. However, in addition to the learning about the potentially mutagenic nature of the modified bases, the direct cloning of doubly-modified dsDNA and the introduction into the cell is an opportunity to observe how the cell processes a plasmid bearing these modifications.

Production of doubly-modified dsDNA was used to investigate the resistance to restriction digest as well as the ability to introduce modifications into the cell using electroporation. The modified dsDNA was subjected to restriction endonuclease treatment to examine the susceptibility and the resistance to degradation bestowed upon the dsDNA. Subsequent sequencing of the plasmid produced indicates the cell's ability to either read through the modifications or excise the modified base through proofreading mechanisms and replace it with an unmodified analog such that the transfected DNA can be replicated without loss of sequence information.

4.2 Objective of this work

The goal of this chapter is to assess the enzymic recognition of modified DNA and how it affects certain reactions or processes both in a selection context and outside of a selection context. Specifically, the final generations of the seven selections from Dr. Marcel Hollenstein and Curtis Lam were cloned and sequenced. The final generations of five different selections, all of which used dA^{ime}TP, were obtained from Dr. Marcel Hollenstein. The final generations of two different selections both of which used dU^{ph}TP, were obtained from Curtis Lam. Selections using 8-modified dAs have shown a slight bias against its representation in the clones of the final sequence. The origins of this bias lie in both the read-through of DNA bearing 8-modified dA as well as the incorporation of 8-modified dA to produce modified DNA. The number of sequences

encoding for the crucial 8-modified dA from Dr. Marcel Hollenstein's selection as well as the selection described in Chapter 3 will be determined and graphed according to their dA content. The number of sequences encoding for the well-tolerated dU^{ph} from Curtis Lam's selections will be determined and graphed according to dT content as well. The amplifiability of DNA bearing the dA^{imm} modification used in Chapter 3 will be assessed by performing a PCR against DNA templates of the Dz20-49 sequence, but bearing slightly longer linker arms at the dA modifications.

Well-tolerated modified nucleoside triphosphates dU^{ga}TP, dU^{ph}TP, and dU^{aa}TP, all of which have been successfully utilized in DNAzyme selections, will be used in PCR to produce doubly-modified dsDNA. With doubly-modified dsDNA in hand, the resistance to restriction enzyme digestions will be assessed. To ensure that resistance is not due to mutation as the mutation potential of dU^{ga} and dU^{ph} has not yet been determined, the doubly modified dsDNA will be ligated directed into a plasmid and transfected into *E. coli*. Depending on the sequence attained from the harvested plasmid, we will be able to determine if there was any change to the sequence, specifically the restriction sites, stemming from mutation in the PCR or processing within the cell. This will be the first experiment using DNAzyme-relevant nucleotides dU^{ga} and dU^{ph} within a living system.

4.3 Materials and methods

4.3.1 Chemicals and reagents

Modified nucleoside triphosphates 5-guanidiniumallyl-2'-deoxyuridine-5'-triphosphate (dU^{ga}TP) and 5-(para-hydroxybenzamido)methyl-2'-deoxyuridine-5'-triphosphate (dU^{ph}TP) were synthesized by Curtis Lam. Modified nucleoside triphosphate 5-aminoallyl-2'-deoxyuridine-5'-triphosphate (dU^{aa}TP) was purchased from Trilink. Plasmid pUC18 was purchased from Fermentas.

4.3.2 Enzymes

All restriction enzymes were purchased from Invitrogen.

4.3.3 Oligonucleotides

The following oligonucleotides are mentioned in this chapter (5'-3'):

ODN 4.1 TAATACGACTCACTATAGGGTAACGCCAGGGTTTTCC (Mod PCR-F)

ODN 4.2 GCTAGTTATTGCTCAGCGGGGAATTGTGAGCGGATAACA

(Mod PCR-R)

Oligonucleotides **ODN 4.1** and **ODN 4.2** were synthesized at and purchased from the Integrated DNA Technologies (IDT). Oligonucleotide names used in notebooks follow the sequence in parentheses.

4.3.3 TA cloning and sequencing of the final DNzyme generations

The mercury sensor Dz10-13 selection (two modifications, N40 selection), Dz9-86 selection (3 modifications, N20 selection), Dz10-66 selection (3 modifications, N40 selection), and Dz12-91 selection (12 nt RNA primer, 3 modifications, N20 selection) were carried out by Dr. Marcel Hollenstein. The ATP sensor selection and stachyose sensor DzSTA-17 selection were carried out by Curtis Lam. The Dz20-49 selection (3 modifications, one being dA^{imm}, N40) was carried out by the author and was described in Chapter 3.

Dr. Marcel Hollenstein and Curtis Lam provided Taq-amplified dsDNA of each of the final generation from each of their respective selections. Purification of the PCR product was done using Qiagen's PCR Clean-up kit. Once purified, the DNA was TA cloned using Promega's pGEM-T Easy kit following the kit protocol. Ligation reactions were incubated at 4°C overnight. An aliquot (5 µL) was dialyzed by placing the sample on 1 % agarose for 1 hour. The dialyzed ligation (0.5 µL) was added to of *E. coli* DH10B ElectroMax cells (20 µL, Invitrogen) and incubated on ice for 2 minutes. The cells were electroporated using Triton's Bactozapper. White colonies were used to inoculate TB (1 ml) containing 100 mg/ml ampicillin. After incubating overnight at 37 °C with shaking at 255 rpm, plasmids were isolated using Invitrogen's Purelink Plasmid

Purification Kit. Eluted plasmids (9 μ l) were treated with 1x Invitrogen Buffer 3 (1 μ l) and 1 U Eco RI (0.1 μ l of a 10 U/ μ l stock). The concentration of plasmids containing cloned insert of correct size were determined by UV spectroscopy and an aliquot from each was standardized to 50 ng/ μ l. Standardized plasmids were submitted to UBC's Nucleic Acid Protein Services Unit for sequencing using primer SP6.

4.3.4 Modified template synthesis for PCR, purification and standardization

Modified templates, Dz20-49 and variants, were synthesized on a 15 pmol scale according to protocol described in Chapter 3. Nucleosides triphosphates used for extensions were one of the following cocktails: 1) natural dNTPs, 2) dGTP, dC^{aa}TP, dU^{ga}TP and dA^{imm}TP, 3) dGTP, dC^{aa}TP, dU^{ga}TP and dA^{ime}TP and 4) dGTP, dC^{aa}TP, dU^{ga}TP and dA^{imp}TP. Synthesized templates were bound on streptavidin, washed TEN (2 \times 50 μ l), (0.1M NaOH, 1 mM EDTA) (5 \times 50 μ l), (25 mM cacodylate, pH 6 buffer) (1 \times 100 μ l), DEPC H₂O (1 \times 100 μ l), and resuspended in DEPC H₂O (20 μ l). The beads were split into two equal portions. To one portion, RNaseA (1 μ l of a 10 mg/ml stock solution) was added and the sample was left to incubate at room temperature for 20 minutes. Uncleaved DNA was used as a control to ensure full extension. 99:1 formamide loading buffer:biotin (100 mM in DMF) (10 μ l) was added to each sample. The samples were heated at 95 °C for 5 minutes and resolved using 7 % PAGE. Bands corresponding to RNaseA-cleaved material were excised. The gels containing cleaved products were crushed and the DNA was eluted by freezing and thawing (2 \times 500 μ l elution buffer). Templates were dried in a speedvac. Pellets were dissolved in H₂O (100 μ l), 5x First Amp buffer (1 μ l) and ethanol (1 ml). Samples were agitated on a vortexer briefly and centrifuged for 15 minutes. The supernatant was decanted and the pellet was air dried on a 65 °C heat block. The pellet was dissolved in DEPC H₂O (50 μ l) and desalted by passing the sample through a G25 spin column.

A calibration curve relating the radioactivity to the dilution of unlabeled dGTP in a solution was produced. A solution containing the same amount of radiation that was obtained from the same stock of dGTP α ³²P was produced. Radioactivity was measured

using Imagequant will be quantitatively described in terms of autoradiographic “volume.” Dilutions used were 30 ×, 90 ×, 900 ×, 1800 ×, 18000 ×. The natural log of the volumes were plotted against the natural log of the dilution and fitted to the equation (3):

$$x = \frac{y - b}{a} \quad 3$$

where:

$x = \ln$ (dilution factor)

$y = \ln$ (volume)

$a = \text{slope}$

$b = \text{intercept}$

The calibration curve provides a relation between the radioactivity and dilution factor. Using the calibration curve, one can calculate a dilution factor, x , for each of the modified templates. Purified DNA templates were spotted along with the calibration standards in 1 µl and 10 µl volumes. The radioactive volumes for the 10 µl samples were used for determination of template concentration. The dilution factor, x , of deoxyguanosines can then be used to determine the concentration of the DNA based on the 19 deoxyguanosines that get incorporated during the enzymic synthesis of Dz20-49 and Dz20-49 dA^X variants according to the equation (4):

$$[\text{template}] = \frac{[dGTP] \times 0.1}{19 \times e^x} \quad 4$$

where:

$x = \ln$ (dilution factor)

[dGTP] = the concentration used in the extension = 10 µM

The factor of 19 corrects for the 19 deoxyguanosines and the factor of 0.1 corrects for the 10 µl sample volume.

4.3.5 PCR using modified templates

Once the templates concentrations were determined, templates were standardized to 10 fM and used in PCR. PCR reactions contained 1 × Thermopol buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 μM dNTPs, 1 fM DNA template, 1 μM of each primers **ODN 3.5** and **ODN3.6** and 5 U Vent (exo-) or 5 U Taq polymerase in final volume of 50 μl.

The Vent (exo-) polymerase catalyzed reactions were thermocycled 45 times (94 °C/30 sec, 57 °C/ 30 sec, 72 °C/ 30 sec). Ten microlitre aliquots were removed after cycles 37, 39, 41, 43 and 45. The Taq polymerase catalyzed reactions were thermocycled 35 times (94 °C/30 sec, 57 °C/ 30 sec, 72 °C/ 30 sec). Ten microlitre aliquots were removed after cycles 27, 29, 31, 33 and 35. For comparison, the Vent (exo-)-produced amplicons from an aliquot of the 43rd cycle of the PCR (2 ul) were resolved on 2 % agarose gel. Amplicons produced from the 33rd cycle of the Taq PCR were likewise resolved.

4.3.6 Production of doubly-modified dsDNA

The multiple cloning site of pUC18 was amplified using modified dU^xTPs to produce doubly modified dsDNA. Three separate PCRs contained 1 × Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 1 μM primer **ODN 4.1** (5'-*TAATACGACTCACTATAGGGTAACGCCAGGGTTTTCC*-3'), 1 μM primer **ODN 4.2** (5'-*GCTAGTTATTGCTCAGCGGGGAATTGTGAGCGGATAACA*-3'), 200 μM dNTPs (dATP, dCTP, dGTP, and one of dU^{ga}TP, dU^{ph}TP or dU^{aa}TP), 250 ng pUC18, 5 U Vent (exo-), 5 mM MgSO₄ and DEPC treated water to a total of volume 50 μl. Non-hybridizing regions composed of sequences foreign to pUC18 are shown in italics. These three reactions were thermocycled 20 × (94 °C/1 min, 55 °C/1 min, 72 °C/5 minutes) with a final extension 72 °C/5 minutes.

For an unmodified control product reaction, the following reaction mixture was made:

1 × Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 1 μM primer **ODN 4.1** (5'-*TAATACGACTCACTATAGGGTAACGCCAGGGTTTCC*-3'), 1 μM primer **ODN 4.2** (5'-*GCTAGTTATTGCTCAGCGGGGAATTGTGAGCGGATAACA*-3'), 200 μM dNTPs, 50 pg pUC18, 2 U Vent (exo-), 5 mM MgSO₄ and DEPC treated water to a total of volume 50 μl. Non-hybridizing regions composed of sequences foreign to pUC18 are shown in italics. Sequence foreign to the template plasmid is indicated in italics. The reaction was thermocycled 15 × (94 °C/20 sec, 55 °C/20 sec, 72 °C/20 sec). PCR products were purified using QIAGEN's PCR Purification Kit. Product DNA was eluted from the spin column using 30 μl of elution buffer. One microlitre of purified product was run on a 2 % agarose gel.

4.3.7 Restriction digests of doubly-modified dsDNA

To 50 ng of modified or unmodified amplicon (5 μl from a 10 ng/μl stock) and water (3 μl), one of the following combination of 10x buffers and 10 U enzymes was added with the restriction enzyme was added last:

- 1) React 2 buffer (1 μl), Hind III (1μl from a 10 U/μl stock)
- 2) React 10 buffer (1 μl), Sal I (1μl from a 10 U/μl stock)
- 3) React 2 buffer (1 μl), Xba I (1μl from a 10 U/μl stock)
- 4) React 3 buffer (1 μl), Bam HI (1μl from a 10 U/μl stock)
- 5) React 3 buffer (1 μl), Eco RI (1μl from a 10 U/μl stock)
- 6) React 4 buffer (1 μl), Sma I (1μl from a 10 U/μl stock)
- 7) React 8 buffer (1 μl), Kpn I (1μl from a 10 U/μl stock)

DNA substrates were digested for 2 hours at 37 °C. After 2 hours of incubation, 5 μl of each reaction was loaded on 2 % agarose gels.

4.3.8 Transfection of doubly-modified dsDNA

Doubly-modified dsDNA amplicons were A-tailed using reactions containing 1 × Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 μM dATP, 50 ng doubly-modified dsDNA, 5 U Taq (1 μl from a 5 U/μl stock) in a 10 μl final volume. The reactions were incubated at 72 °C for 20 minutes. TA cloning was performed using reactions containing an aliquot of the A-tailing reaction (2 μl). Ligations were incubated at 4 °C overnight and subsequently dialyzed on 1 % agarose for 1 hour. Electroporation, plating, harvesting and sequencing were carried according to protocol out as described in section 4.3.3.

4.4 Results

4.4.1 Sequences of the final DNAzyme generations

Sequences were obtained for the final generations of Dr. Hollenstein's selections (mercury sensor Dz10-13 selection, RNA-cleaving Dz10-91 selection, Dz10-96 selection, Dz10-66 selection and the two modification, N40 selection) and Curtis Lam's selections (ATP sensor selection and stachyose sensor selection) mentioned in this chapter and the Dz20-49 selection mentioned in Chapter 3. Sequences were obtained according to the protocol outlined in section 3.4.2 and are summarized in Table 4.1-4.8. All selections contained notable self-cleaving DNAzymes which, except for the ATP selection, are indicated by an asterisk. These clones were further studied with the exception of clone N40-12 (Table 4.1) which contained the sequence for Dz9₂₅-11.

An interesting similarity between the most active clone of the mercury selection, Dz10-13, and the most active clone of the dA^{imm}-dependent selection of Chapter 3 is that there are only four modified dA in the approximately forty base sequences. This represents a non-random ten representation of a base that affords critical and catalysis-enhancing properties. This observation prompted closer examination of the individual sequences of the final gene pools. The number of sequences is graphed according to the indicated modified base content in Figure 4.3-4.10. The gene pool of the mercury sensor selection showed an average dA content of about 12.5 % (5 out of ~40). While the content of the other 8-modified dA selections are not nearly as striking, they all show a

dA content lower than the statistical average content of 25 %. For the purposes of selection, a bias against dA content in the sequences is a concern to investigators who seek to expand the chemical space of DNAzyme, but not at the cost of sequence space. More concerning is the process by which this bias comes about. In a DNAzyme selection where expected reactions result in the exponential amplification of a desired sequence, unexpected reactions and side reactions may also result in exponential amplifications of undesired sequences. The result could be the failure to enrich the gene pool in the desired clones ultimately causing the entire selection to fail.

Table 4.1 Individual clone sequences from the two modifications, N40 selection.

Cl one	Sequence	#of A	Acti vi ty
N40-2	GTG AGTGT ACCGTGGTGGGTTGTCGCGCTG ACCA TCCTG	5	n. d.
N40-12*	GTTCTC AT CCGT AGTGA AGGCACGG AGCC ACCCTCCCGC	5	n. d.
N40-17	AGCCA ACCAGCGGT AGTGA AGGCATGCTCTCCTCCCCTCGT	7	n. d.
N40-19	GTTGTTG AGTGGCGAT GA CTCTTTCC CCTTGCA AT GGTCGT	5	n. d.
N40-21	GCCTGC ACCCTAG CTTG ACGATGGA ACCGC AGCCT CGGCCC	7	n. d.
N40-33	GCGC AGAGCTAGTGA GTGCACGCACGCCGGTGCGCGTCA	7	n. d.
N40-50	GGTGTCC TTTCG ACCGTGTCT AGCTGT CCGGTCTGTGTGCT	2	n. d.

n. d. = not determined or data not available

* = indicates that this sequence bears the Dz9₂₅-11 catalytic sequence and most likely has similar activity

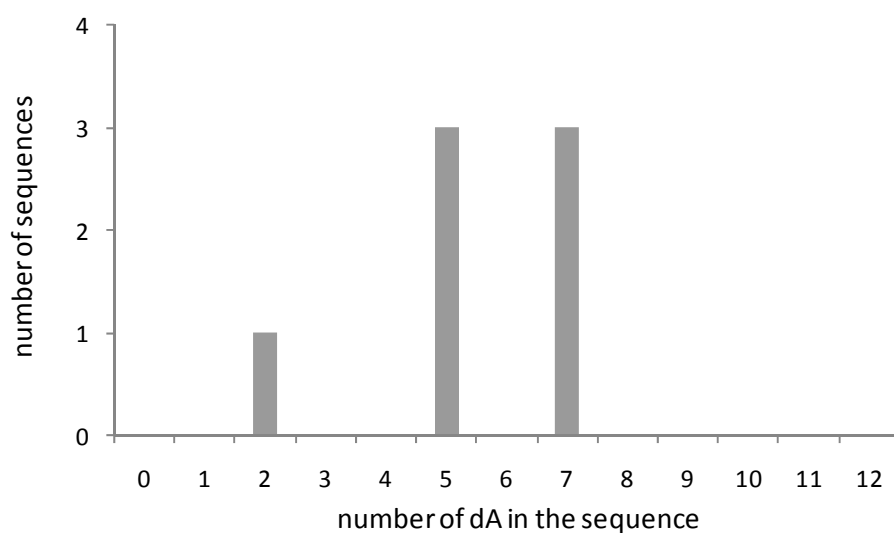


Figure 4.3 Bar graph depicting the number of sequences as a function of the number of dA^{ime} modifications present in the N40 region, which after the base deletions and additions during the selection consists of a range of 39 - 41 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA^{ime} and dU^{raa}.

Table 4.2 Individual clone sequences from the Dz10-13 (mercury sensor) selection.

Clone	Sequence	#of A	Activity
3	CACAGTGTGTGAGGC ACTGTACGGTGAGTGGTGCTT	6	n. d.
5	TTCTC ATCCGTAGTGAGGGACGCGGCGCTCCCCGTT	4	n. d.
8	C ATAGTGCGTGAGGCCGCCC ACTCCCCGTCGTGGTA	5	n. d.
9	TTCTC ATCCGTAGTGAGGGACGCGGCGCTCCCCGTT	4	n. d.
10	GCACAGTGTGTGAGGC ATGTGCGAGTGTGCTGTCTCG	5	n. d.
13*	CACACGTGTGTGACGGCCTCGCGGCCCGCCCTGCCGTA	4	++
17	C ATAGTGCGTGAGGCGCGTCTGCAGCGTGGTGGTTT	4	n. d.
21	CACAGTTGTGTGATGGCTGGGCAGCCAGCGGTGGTCA	6	n. d.
23	CACACGTGTGTGATGGCTCTGCTTCCCACCGGTCCTCA	5	n. d.
24	C ATAGTGCGTGAGGCTTGC GCTGTCA GTCA GCAGGTCG	5	n. d.
31	CACAGTTGTGTGAGGCGCGTGTACAGTGCAGGCGGGTCT	5	n. d.
34	CACAGTTGTGTGATGGCTGGCTCTGCAGCCCGCTGGGT	4	n. d.
40	ACGTGTGTGAGGCATGCGCTCACTGCCGTGGGTCT	4	n. d.
41	CACAGTGTGTGAGGTT CGCTCTGCCTCCCTGCTACA	5	n. d.
42	C ATAGTGCGTGAGGCACGGTTACG CCGTGGTGTGT	5	n. d.
49	CACAGTGTGTGAGGCGTGCGC AGTGTTAGT GCCTG	5	n. d.

n. d. = not determined or data not available

* = indicates the sequence most characterized

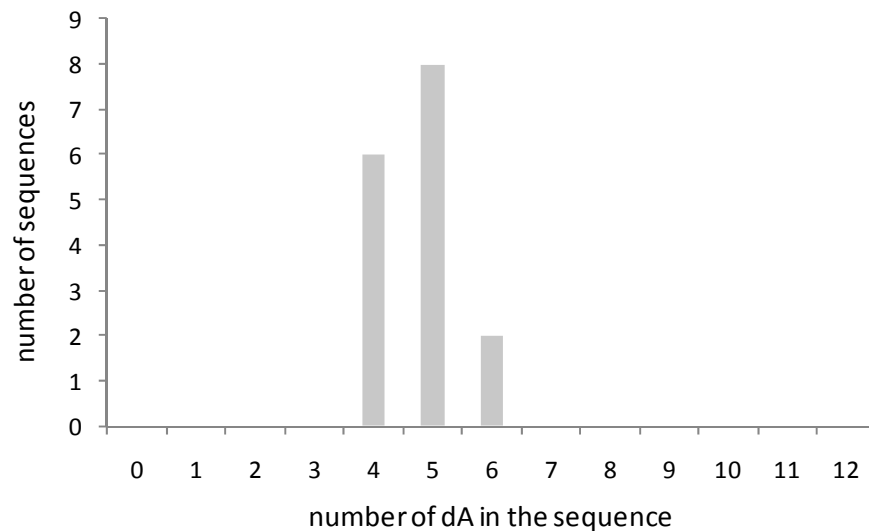


Figure 4.4 Bar graph depicting the number of sequences as a function of the number of dA^{ime} modifications present in a N40 region, which after the base deletions during the selection consists of a range of 35 - 39 bases, for a selection for a mercury sensor DNAzyme modified with dA^{ime} and dU^{aa}. (Hollenstein *et al.*, 2008)

Table 4.3 Individual clone sequences from the Dz9-86 (3modifications, N20) selection.

Clone	Sequence	#of A	Activity
3	CTGCGTGTGTCTTGTGTGCG	0	n. d.
4	CT AGCAGCGCCAGTGAGGCTCGCGTT	4	++
5	GAGAGTGTACTGTGGGTGTA	4	+
9	TTGTTGCATCGCATGTGATG	3	n. d.
11	GGT AGTGTGGACGGT TTT	2	n. d.
12	CCGTGTGTGTGTCGTGT GTA	1	n. d.
13	GAGAGTGTATCGTGGTGGGTC	3	+
17	GAGAGTGTACCATGTGTGTA	5	++
18	ATGGAGCGCTAGTGATGTTTC	4	n. d.
28	TTGCC AGCGGTAGTGTCGCTCGCT	2	+
29	TTGCC AGCGGCAGTGAGGCTTC	3	++
33	TCGCC AGCGGCAGTGAGGTTTCGA	4	++
36	TTCCC AGCGGAGTGTCGCC	2	++
40	GG ATTGCAGTAGGTTGTCCG	3	n. d.
43	GTGTTTGCTTGCT ATGGCTC	1	+
52	CC ATACAGCGCTAGTGTC	4	n. d.
53	ACGCATCGTTAGTGAGGGTGC	4	n. d.
54	TGGTGTGTCTGGGTGGGTGT	0	n. d.
55	GGC ACAGGGGAGTGGT GTTG	3	n. d.
56	TTGCC AGCGGCAGTAGTGAC	3	n. d.
57	TGTGC ACAGTGTTCCGTG	2	-
58	TTGCC AGCGGTAGTGATG	3	n. d.
59	AAGCAGCGTTAGTGAGCGC	5	++
60	GCCT ACTAGCGGAGTGAG	3	n. d.
61	GAGAGTGTTTGATGGGTGTG	2	+
63	GAGAGTGTTTTATGGGTGTC	2	-
64	TT AGCAGCGCATGTGGTGGCTTG	3	+
67	CAGTGTGTCACAGCGGT GTA	3	n. d.
77	CTGGAGTGGACTATGTGG GTA	4	n. d.
81	TGTCGCGT ACGGTTGGT GTTG	1	n. d.
82	T AGAGCTTCGTACGGGAGGT	4	n. d.
83	GC ACGTATCGGGTTTTGTT	2	+
86	GTGTTTGCTCGCT ATGGTTC	1	n. d.
87*	TC ATGCAGCGGTAGTGTC	3	+++
88	CC ANGTATCGGGTTTTGTT	2	n. d.
90	GTGTTTGCTTGCT ATGGCTC	1	n. d.
91	GC ACGTATCGGGTTTTGTT	2	n. d.
92	TT AGCAGTGCA TGTGATGGCTGC	4	n. d.
98	CCGT ACGTGTGTTGTGCGTT	1	n. d.

n. d. = not determined or data not available

* = indicates the sequence most characterized. Clone #87 was renamed Dz9-86

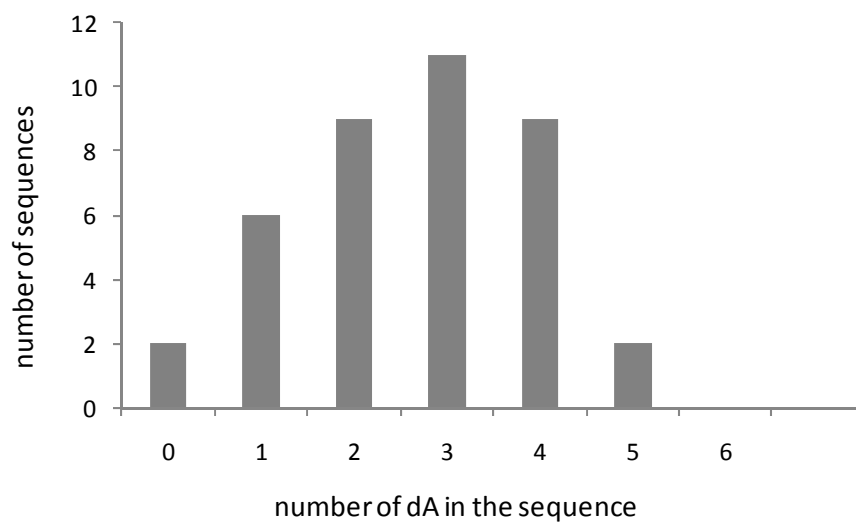


Figure 4.5 Bar graph depicting the number of sequences as a function of the number of dA^{ime} modifications present in a N20 region, which after the base deletions and additions during the selection consists of a range of 18 - 23 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA^{ime}, dU^{ga} and dC^{aa}. (Hollenstein *et al.*, 2009a)

Table 4.4 Individual clone sequences from the all RNA-cleaving Dz12-91(3 modifications, N20) selection.

Clone	Sequence	#of A	Activity
9	TGCAGCGCTAGTGTGTGGAATGCGT	3	n. d.
19	CTCAGCGCTAGTGTGAGGCTCGCTT	3	n. d.
24	GGCAGCTCGGTCCCGTTTGGC	1	n. d.
33	CGTCATGGTGGTGTGTGGTT	1	n. d.
34	TGCAGCGCTAGTGTGAGGCATT	3	n. d.
40	TACCGCAGGGTGATTAGTGTC	4	n. d.
42	TACCGCAGCGTGATTAGTGTC	4	n. d.
44	TTAGCAGTGTACTATGTAGTGCC	4	n. d.
45	TTAGTGTACAGTGGGTGTA	4	n. d.
47	CGCAGCGCTAGTGTGAGGCTTC	3	+
47b	CGCAGCGGTAGTGTGAGGCTTC	3	n. d.
51	TGCAGCGCTAGTGTGAGGCATGTGT	4	+
53	GCACGCACAGGTTTTTGT	4	n. d.
55	CATGCAGCCCTAGTGTGTGGCATT	3	n. d.
57*	ATGATGCAGCGCATGTGTC	4	++
58	TGCAGCGCTAGTGTGGGCTCGTT	2	+
59	TGCAGCGCTAGTGTGCGGCACGCGC	3	+
74	TGCAGCGCTAGTGTGGGTGTA	3	+
79	AATTTGCAGCGCGTAGTGTC	4	n. d.
80	AGTTACAGTGGTAGCGGTTG	4	+
85	TCTTATGTAGCGCCAGTGTC	3	n. d.
86	TGCAGCGCTAGTGTGAGGCATTGT	4	+
87	AATGTCACAGAGGTCCGGTG	5	n. d.
R2	TGCAGCGTTAGTGTGGGTACGCGT	4	-
R10	AGTTATAGTGGTAAGCGGTTG	5	++
R11	GCACGTACCGGGTTTTTGT	2	++
R11b	TTGATGCAGCGCATGTGTC	3	++
R12	GCTAAGTTCGCCATTTGGTGG	3	+
R12b	AGTTACAGTGGTAGCGGTG	4	+
R13	ATGCCAGCGGCAGTGTGCTCGCTT	3	-
R13b	GTTACAGTGGTAGCGGTTG	3	+
R14	TACCGCACCGTGATTAGTGTC	4	n. d.
R17	TAGCCAGTGGTAGTGTGCTCGCTTT	3	n. d.
R18	TGCAGGACTAGTGTGAGGCTCGTCT	4	n. d.
R3	TGCAGCGCTAGTGTGAGGCATGCTT	4	++

n. d. = not determined or data not available

* = indicates the sequence most characterized

Table 4.4 Individual clone sequences from the all RNA-cleaving Dz12-91(3 modifications, N20) selection continued.

Clone	Sequence	# of A	Activity
R22	CTCC AG CTGT AG TGTGAGGCTCGCTT	3	n. d.
R23	TACCGCAGCGTGATT AG TGTC	4	+
R24	TACCGCAGCGTGATT AG TGTC	4	+
R26	GGC AG CTCGGCCCCGTTTGCG	1	+
R26b	TATGC AC AGAGTTTCGGGT	3	n. d.
R28	TGC AG CACT AG TGTCTGGC AT GTTG	4	-
R30	CTG AGG AGTGGTTGGTCTT	2	n. d.
R32	T AC AGCGCT AG TGTG AGGC ATT	5	n. d.
R33	GGTTTG AC GC CGT AGTGTC	2	+
R34	TC AT G TAGT ACGT AG TGTCGCC	4	+
R34b	GCT AA GTTTGCGTGTGGGTTTC	2	-
R4	TACCGCAGCGTGATT AG TGTC	4	n. d.
R5	TGC AG CGCT AA TGTGTGGCTCGTTT	3	n. d.
R6	GCT AA GTCCGCC ATT GGTGG	3	+
R6b	CTCC AG CGGT AA TGTG AGG CTCGCT A	5	n. d.
R8	TACCGCAGCGTGATT AG TGTC	4	n. d.
R9	TGC AG CGCT AG TGTG AGGC ATT	4	n. d.
R37	TGC AG CGCT AG TGTG AGGC ATT	4	n. d.
R44	AGTTAT AGTGGT AG CGGTTG	4	++
R45	TGC AG CGCT AG TGTG AGGCAC GTGT	4	++

n. d. = not determined or data not available

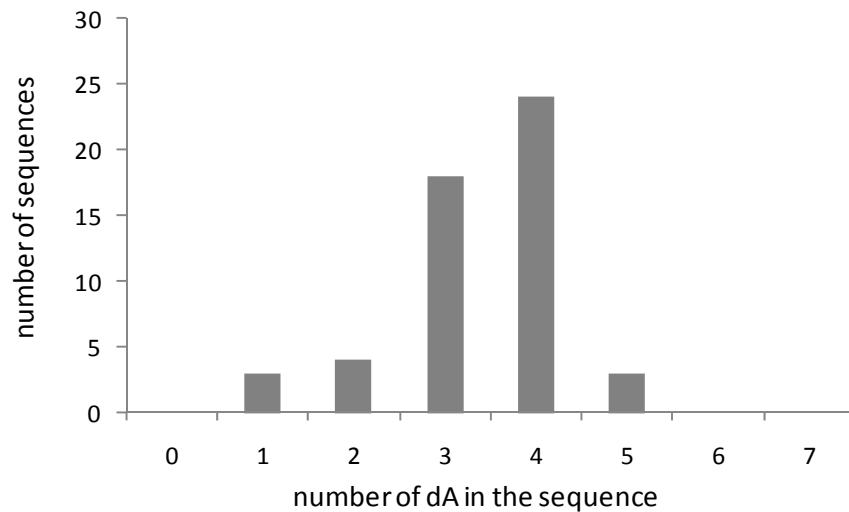


Figure 4.6 Bar graph depicting the number of sequences as a function of the number of dA^{ime} modifications present in a N20 region, which after the base deletions and additions during the selection consists of a range of 19 - 26 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA^{ime}, dU^{ga} and dC^{aa} capable of cleaving at a region composed of twelve ribonucleotides.

Table 4.5 Individual clone sequences from the stachyose sensor selection.

Clone	Sequence	# of T	Activity
1	GAAACGAGCTCGAACGAAGGCACGATGGCCAATCACCTATTG	6	+++
4	GAAATTAGTTGACGCAGGCATTGCACGTGTGAATGGTGTTA	12	-
17*	GAAACTAGTTGACGAAGGCACGTGCTGACTATCGAGTGCAAA	8	++++
22	GATACTAGTTGACGCAGGCACGATTGAGTATCCCCTTCACAA	10	+++
46	GATACTAGTTGACGCAGGCACAAATGACCATCCCGTGCATA	9	+++
48	GATACGAGCCGTACGAAGGCACGATGCGCAATCGCCTATTG	7	++
53	TCACATGAGTGTGAGACACGCAGTGGTCCGGATTGGGCAA	8	-
54	GAACTGAGCTCAGACGCAGGCTAAACGCTGAAATAGCGCTA	6	++
107	GATACTAGTTGACGAAGACACTAATGACTATCTAGTGCATA	11	-
109	GATACTAGTTGATGATGGCACTCCCTAGCAGTCTACTACT	12	-
112	GCATCCTGCTCGCTGTGGAGGCCACCAAATCGCTATGGAAT	9	-
113	GAAACTAGTTGACGAGGCACCTTCTGAGGATCGTGTCCACA	9	+++
120	GATACTAGTTGACGAAGGCACATATCGCGTTTTACGTGCAAA	11	+++
124	GAAACTAGTTGACGCAGGCACCTTCTGAGCATCTCCTTCACAA	10	+++

* = indicates the sequence most characterized

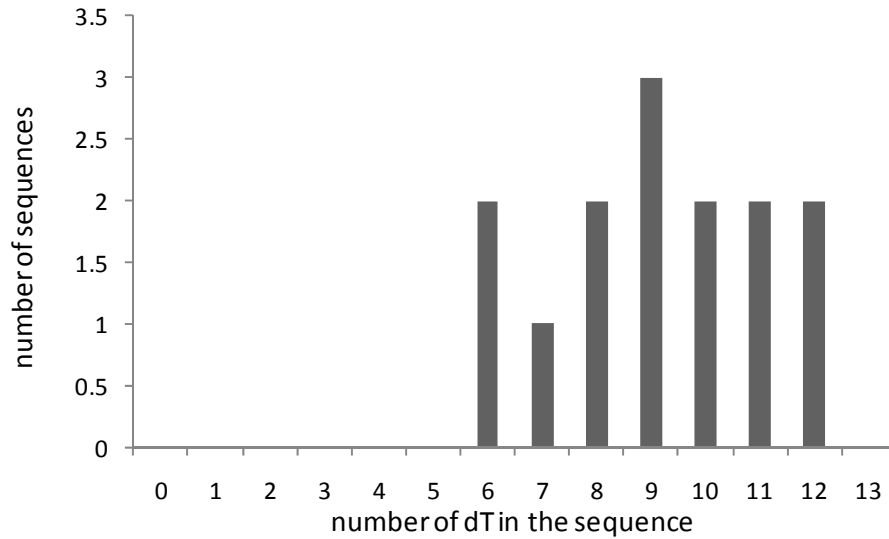


Figure 4.7 Bar graph depicting the number of sequences as a function of the number of dU^{ph} modifications present in a N40 region, which after the base additions during the selection consists of a range of 40 - 42 bases, for a selection for a stachyose sensor DNAzyme modified with dU^{ph} only.

Table 4.6 Individual clone sequences from the ATP sensor selection.

Clone	Sequence	# of T	Activity
2	GAACTAGTTGACGGAGGCACTGCATGGTGAGTGCTGTTT	10	+++
3	GATACCAGCCGTACGCAGGCACGTTACCCACAAAGCAAAA	5	-
13	GAACTAGTTGACGACGACACGCCAACGCCCTGTCGTGTATA	8	+
14	GAACTAGTTGACGGAGGTACGGTCATGAGGCGGTGTGGTCTA	10	+++
16	GATAGTAGTTCACGGAGGCACTACTAAGTATGGTGTGCAAA	10	++
18	GAAGCTAGTTGACGCAGGCACGGAGTGGGTGAGCGTGGTTA	8	+++
20	GAACTAGTTGACGCAGGTACCCCTACACAGTGTAAACATA	8	-
22	GATACTAGTTGACGAGGGCATGCCCTCCCTCCCGTAGTGTA	10	-
37	GAACTAGTTGACGGGGGCATAGGTCAGTATGCTTTGTTT	13	-
47	GAAACAAGCTGCACGAAGGCACGCACGGTCCGTATCGGCCTA	5	+
101	AAC TGCCATATTACCTACAATGGCGGAGTGCGCCCTATGATC	10	-
102	GATAGCAGCGTCTACGAAGGCACCATGGTCCTAGCATGTTA	9	+
105	GAAGCAAGCTACACGAAGGCACGCACACTGGTGCCTGGTTA	6	-
108	GAACTAGTTGACGCAGACACTGCTAAGTATCGTGAGCAAG	8	-
110	GAAC TGAGCCCAGACGCCGGCTAAAAGTCGAAATGGTGTTA	7	-
114	GATATCAGCCGATACGAAGGCAACAATATGCCTTTCATTG	10	-
115	GAACTAGTTGACGCAGGCATTATGCAGTATCTCATAGAAA	10	+

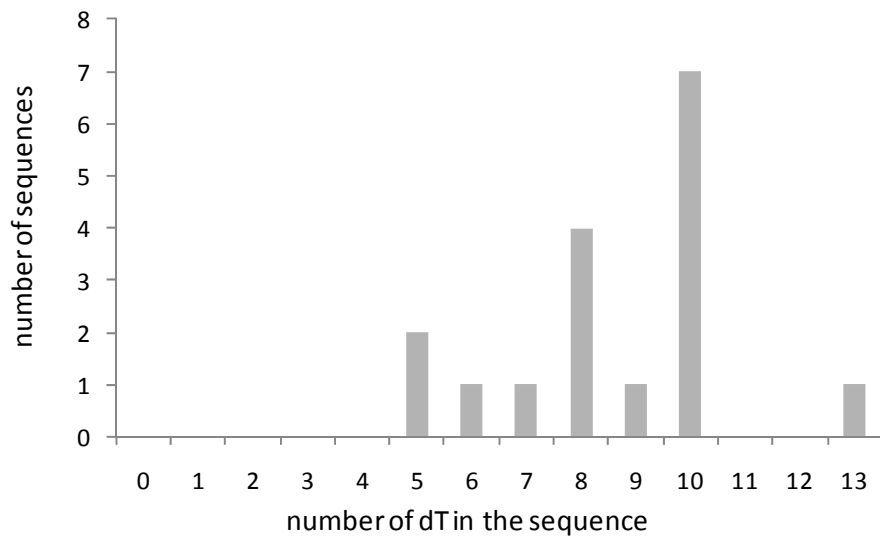


Figure 4.8 Bar graph depicting the number of sequences as a function of the number of dU^{ph} modifications present in a N40 region, which after the base additions during the selection consists of a range of 40 - 43 bases, for a selection for a ATP sensor DNAzyme modified with dU^{ph} only.

Table 4.7 Individual clone sequences from the Dz10-66 (3modifications, N40) selection.

Clone	Sequence	#of A	Activity
N2	TTAGCAGCATGAGTGAGAGCGCGCACTGTGAGTGGCGTTT	8	-
N5	TTCGCAGCGTGAGTGAGGTGCGCTCTGTAGGGTGAGTGGTGCT	5	+++
N8	TTAGCAGCGCAATGAGGTGCGCATGTGAGTGTGTGCTA	8	-
N12	CTAGCAGCGCATGTGAGGCATGCATGGTTGAGTGTGGCC	6	++
N16	TCCGCAGCGTGAGTGAGGTGCGCGTGAGTGGGATTGCTGTT	5	++
N17	CTAGCAGCGCATGTGAGTGGACGCATTGAGGCCAGGGTTA	9	++
N20	TTCGCAGCGTGAGTGTGCTCGCTTTAGTCTTGTTGCT	3	++
N22	TTGTTGCAGCGCTAGTGACGCACGCATGCAGTGATGTGGT	6	++
N23	TTAGCAACGCCAGTGCCGCATGCGTGAGTGAGCTGTTCTT	7	-
N26	CTAGCAGCGCATGTGTCGCTCGCATGCGAGTTCGTGCAT	6	++
N30	TTCGCAGCGTGAGTGAGTGATGCATGAGTGGGAGTGGTGTA	8	+++
N32	CTAGCAGCGCACGTGATGGCTCAGCATGTGAGGGGGGTGC	6	++
N33	TGTTGCAGCGCTAGTGAGGCGCGTACGAGTGGTCGATGTCTT	5	++
N37	TTAGCAGCGCATGTGAGGCATACACGAGAGGGTGTGTTGTA	7	+++
N38	TTAGCAGCCCCAGTGAGGCACAGTAATGTGAGCGTTTCGTT	9	-
N40	GTTGCAGCCCTAGTGATGGTGCGCCATGCTGGCTGTGGTGGA	5	-
N41	GTTGCAGCGCATGTGATGGTCTCGCATGCTTGTGTGTGAGT	5	+++
N43	TTAGCAGCGCATGTGATGGCATGCGCACGGGGGTTTCGGCTG	6	++
N46	TTCGCAGCGTGAGTGAGAGCATGCATGGAGTGGGTGCGTGT	7	+++
N66*	CTAGCAGCGCAAGTGAGGCGCGCTATGAGTGTGCGTGCGTGTAT	8	+++

* = indicates the sequence most characterized

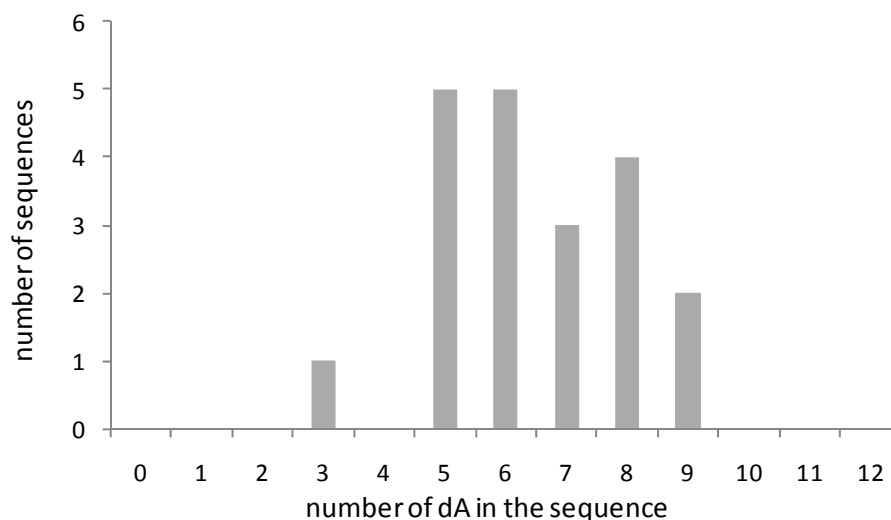


Figure 4.9 Bar graph depicting the number of sequences as a function of the number of dA^{ime} modifications present in a N40 region, which after the base deletions and additions during the selection consists of a range of 38 - 43 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA^{ime}, dU^{ga} and dC^{aa}. (Hollenstein *et al.*, 2009b)

Table 4.8 Individual clone sequences from the Dz20-49 (3 modifications, N40) selection.

Clone	Sequence	# of A	Activity
1	AUGCAUGGUU <u>UUUG</u> AGCAUGUGCUGUGUAGCAGCAGCGUUU	8	-
2	AUGCAUGGUU <u>UUUG</u> AGUCGAGGCAUGUUAGUGAGUGUGUCUU	8	-
4	CUGCAUGGUU <u>UUUG</u> AGGCGAGGCAUGUGAGGGAUUGGCUG	7	-
5	UCAUAGUCUCGGUGGCACGUUCGUAGGUGUGAUUGUGUGU	5	-
6	AGUU <u>AUGCUCUCC</u> AGUGGCUCGCAUGAUGUGUAGUGUGUG	6	+
7	GU <u>AUGAGC</u> AGUGUGGUGGGAGGCGCGCUUGUCUUGCGUUAGU	5	+
8	AGUCAUGUAGUCAGUCUGCGGCACGCCGUGGUGAGGGAUUGUGC	7	+
9	AUGCAUGCUU <u>UUUG</u> AGGCGAGGCAUGCGUCGAGUGUGUGCGGU	7	-
10	AGUCAUGUAUUCCGUUGCUAGCGCAGCAUGUGCUGUGUGUUG	6	-
11	GUGUUUGCUCGGCUGUGUGCGCAGUGUGGUCGAAGUGUGU	3	-
13	GU <u>AUGAGGG</u> AGUGGUGGGAGGCAUGCUUGUGUGAGGUGGCUUU	6	+
14	ACUGUUGAGCACUAGUGAGGUGUGCACGAGUGGUGUCGGUCU	7	-
18	AUGCAUGCUU <u>UUUG</u> AGGCGUGGCACAGUAUGUGUGAGU	8	-
19	AUGCAUGGUU <u>UUUG</u> AGGCGUGGCCAGGUGGCAGUAGUGUU	7	-
21	AUGCAUGGUU <u>UUUG</u> AGGCGUGGUACGCUUGUGCUUGUAGUGAGU	7	-
22	CUGCAUGGUU <u>UUUG</u> AGUCGAGGCAUGUGAGGGAUUGGCUG	7	-
24	AUGCAUGGUU <u>UUUG</u> AGUCUAGGCACGUGAUGAGUGUGAGUGCG	9	-
25	AUGCAUGGUU <u>UUUG</u> AGUCGAGGCUUGUGAGGGAUUGGCUG	7	-
31	UUGCAUGGUU <u>UUUG</u> AGGCGUGGUCGAGUGGUAGUGAGU	6	-
35	ACUACCAUGUGGUCUACAAUGGCGGAGCACCAGUUAGUUU	10	-
38	UGUCAUGUUCUCCGUGGCUCGUACGCCGUGUGUGUGUUUA	3	-
41	UUCGCAUCGUGAGUGAGGCACGGUGGCGACGUGUUGUGGUGCA	6	-
43	AUGCAUGGUU <u>UUUG</u> AGGUGUGGCCAGGUGGCAGUAGUGUU	7	-
44	AGUCAUGCUCUCCAGUGGUUCGCAUGUAUGUGAGUGGAGUGU	7	+
45	AGCAGCUCGAGUCAGUUUGCGGCACGCAUGGUGGUUCGCGUGU	6	-
46	AUGCAUGGUU <u>UUUG</u> AGGCGUGGCACAGUAUGGGUGUGAGU	8	-
48	AUGCAUGGUU <u>UUUG</u> AGUCGAGUGUAGUGUAGCAGCAGCGUUG	9	-
49*	AGUCAUGCUCUCUAGUGGUUCGACGGUCGUGUGGGUCGUU	4	+
50	AUGCAUGGUU <u>UUUG</u> AGUCGAGUGUAGUGUAGCAGUAGCGUUG	9	-
52	GCAUAGUCUCGGUGGCACACUCGUAGAGGUGGUAGUGUCA	8	-
54	AGUCAUGUAGUCCGUUGCUAGCGGCCAUGUGCUGUGUUG	5	-
55	AUGCAUGGUU <u>UUUU</u> AGUCGAGGCAUGUGAGGGAUUGGGUG	8	-
57	CUGUUUGCUCGACU <u>AUGGCGCGC</u> AGUGUGGUCUUAGUGUUU	4	-

* = indicates the sequence most characterized

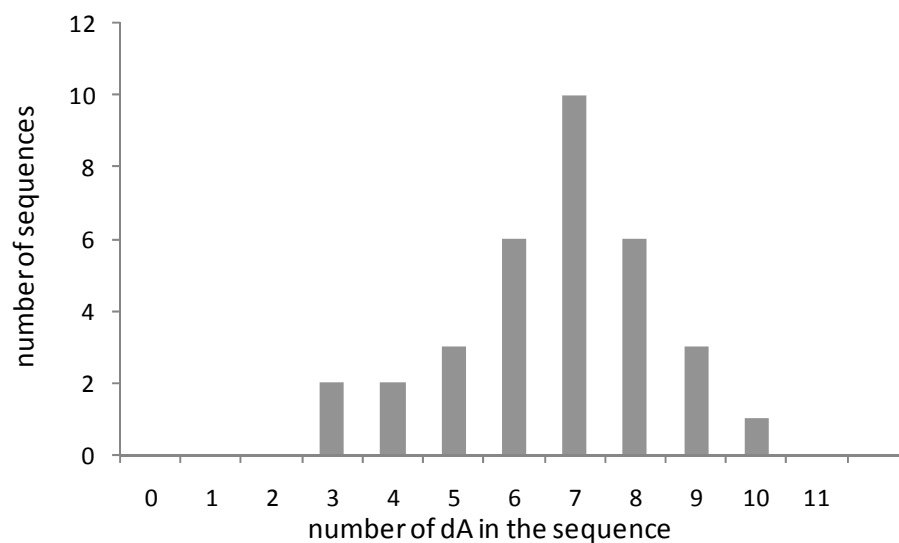


Figure 4.10 Bar graph depicting the number of sequences as a function of the number of dA^{imm} modifications present in a N40 region, which after the base deletions and additions during the selection consists of a range of 39 – 45 bases, for a selection for a divalent metal cation-independent DNAzyme modified with dA^{imm}, dU^{ga} and dC^{aa}.

4.4.2 Quantification of various dA-modified templates based on Dz20-49

As the efficiency of modified DNA template read-through was suspected to be lower than that of unmodified DNA templates, modified templates were synthesized, RNaseA-cleaved, and gel purified. These modified DNA templates were used to compare the efficiency of read-through by comparing the amounts of amplicon produced from PCR. Quantification of the modified DNA templates was done using a calibration curve (Figure 4.11). A sample calculation using equations (3) and (4) is shown on page 123. A standard stock solution was made by adding the same volume (0.75 μ l) of the same radioactive stock solution as the volume and stock solution used in the original primer extension reactions to water to a final volume (20 μ l) that matches the extension reactions. The solution used to produce the calibration curve was made from the same stock solution of dGTP α ³²P used to produce the templates to ensure the concentration of radioactivity source was consistent. The actual radioactivity of the stock solution does not need to be known exactly as the radioactivity to dG ratio changes with time due to the dGTP α ³²P constantly decaying. However, it is imperative that the calibration standard be made from the same stock solution so that both calibration standard and templates will have decayed by the same amount at any given time.

Because the concentration of dG is related to the radioactivity, the concentration of dG in a sample containing purified template can be calculated by the radiation emitted. First, the calibration curve was produced by loading various dilutions (1 μ l) on a glass-backed silica gel plate and relating the natural log of the radioactivity to the natural log of the dilution factor. Second, the radioactivity for a given volume of template solution is quantified and a value for the “dilution factor” for that particular template is calculated using equation (3). Third, the template’s “dilution factor” can be used to calculate the concentration of template by accounting for the nineteen dGs in the Dz20-49 sequence by dividing by nineteen as shown in equation (4). Templates were diluted to make 10 fM stock solutions for PCR.

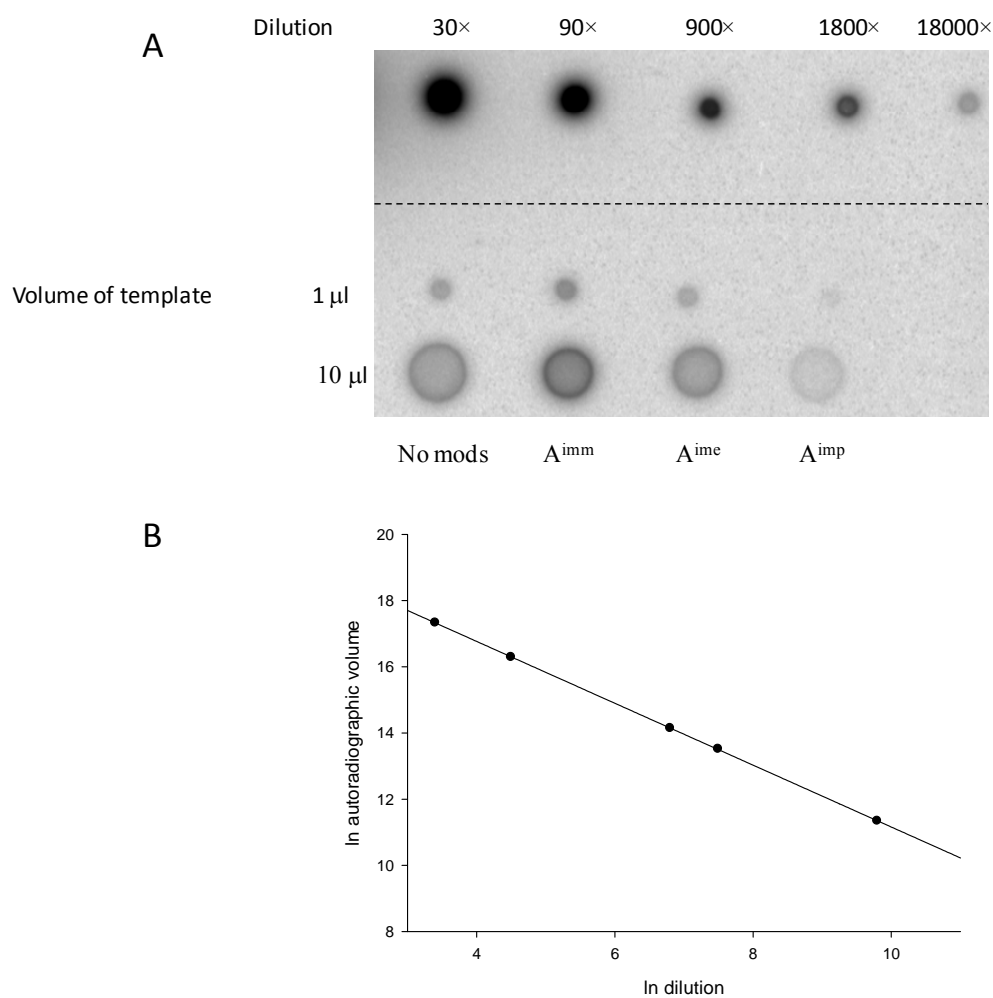


Figure 4.11 Calibration curve used to determine modified DNA concentration. **A** Exposed screen showing the radioactive intensity of calibration standards at various dilutions and 1 μl and 10 μl of synthesized modified and unmodified templates. **B** Calibration curve produced by autoradiography for determining modified template concentration. Graph fitted to equation $y = -0.9344x + 20.5035$, $R^2 > 0.99$.

Example calculation of dA^{imm} -modified template concentration:

Using the acquired data points (Figure 4.11), to produce the calibration curve gives:

$$x = \frac{y - b}{a}$$

$$a = -0.9344$$

$$b = 20.50$$

Calculating associated DF = “dilution factor:”

$$\ln DF = \frac{\ln volume - b}{a}$$

$$\ln DF = \frac{\ln 1.34 \times 10^6 - 20.50}{-0.9344}$$

$$\ln DF = 6.84$$

Calculating dA^{imm} -modified template concentration:

$$[template] = \frac{[dGTP] \times 0.1}{19 \times e^{\ln DF}}$$

$$[template] = \frac{[10^{-6}] \times 0.1}{19 \times e^{6.84}}$$

$$[template] = 5.61 \text{ pM}$$

4.4.3 Amplicons from modified template PCR

To simulate and examine the relative amplifiability of the trace amounts of modified DNA recovered from selection, the modified templates whose sequences are based on Dz20-49 and contain various combinations of relevant modifications were amplified using PCR. In this case, the main determinant of amplifiability is the initial read-through of the template strand into unmodified DNA. Once the first strand of unmodified DNA is produced, the PCR will proceed normally using the newly synthesized unmodified DNA. Using conditions similar to those found in the selection of Chapter 3 (trace amount of template, use of Vent (exo-)), we see that given the same conditions and number of cycles the dA^{imm} -modified template is amplified with reduced efficiency compared to dA^{ime} -modified or dA^{imp} -modified templates. Note that in Figures 4.12 and 4.14 the ethidium bromide appears in greater quantity on the right sides of the gels as seen in the background fluorescence. Therefore, side-by-side comparisons of amplicons produced using different templates, but that same polymerase and number of cycles were done (Figure 4.13 and Figure 4.15).

For PCRs using Vent (exo-), the dA^{ime} and dA^{imp} templates are amplified with almost the same efficiency (Figure 4.12). The dA^{imm} -modified template is amplified far less efficiently than the other two dA^{x} -modified templates. All dA^{x} -modified templates produce less amplicon than the fully unmodified template (Figure 4.13). For PCRs using Taq polymerase, some mispriming is also seen as suggested by the appearance of artifacts (Figure 4.14). A trace amount of amplification artifact can be seen in the unmodified template control. The amount of artifact produced in the modified template PCR is much more pronounced. In the PCR containing the dA^{imm} -modified template, the major product is artifact. The relative ratio of the amounts of amplicon, both product and artifact combined, is similar to what is seen for Vent (exo-) (Figure 4.15).

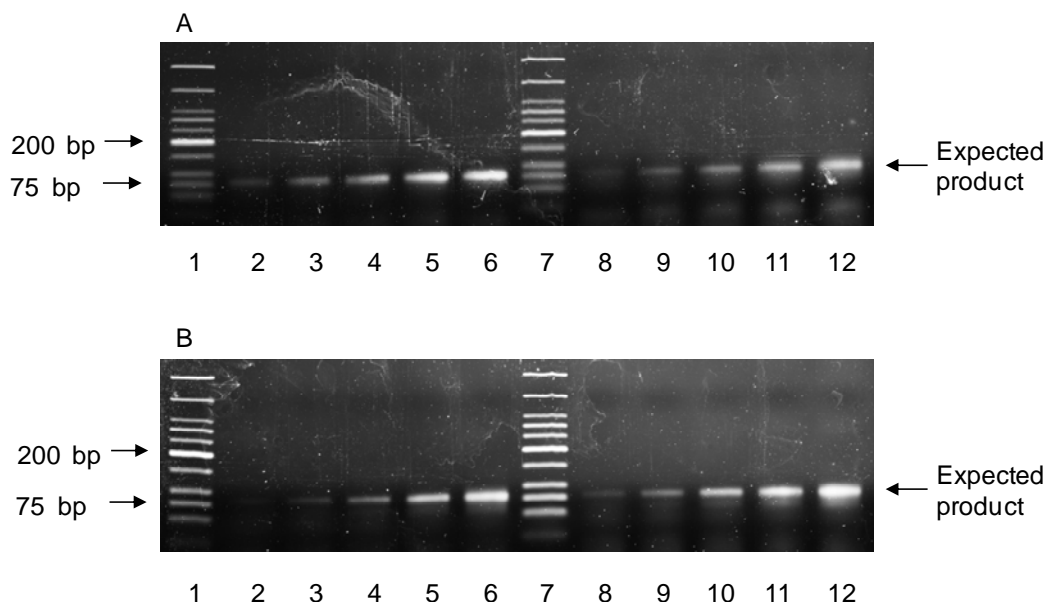


Figure 4.12 Amplification of Dz20-49 templates with varying modifications using Vent (exo-) polymerase. **A** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 37, 39, 41, 43, 45 cycles of PCR, respectively, using unmodified template. Lanes 8-12: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imm}-modified template. **B** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{im}-modified template. Lanes 8-12: 37, 39, 41, 43, 45 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imp}-modified template.

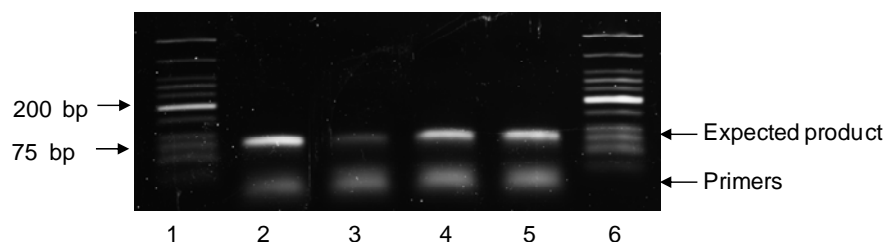


Figure 4.13 Comparison of the amplicons produced by Vent (exo-) using modified templates. Lanes 1 + 6: NEB Low Molecular Weight Ladder. Lane 2: Amplicon produced after 43 cycles of PCR using template containing no modifications. Lanes 3-5: Amplicons produced after 43 cycles of PCR using templates containing modified dC, modified dU and one of dA^{imm}, dA^{ime}, dA^{imp}, respectively.

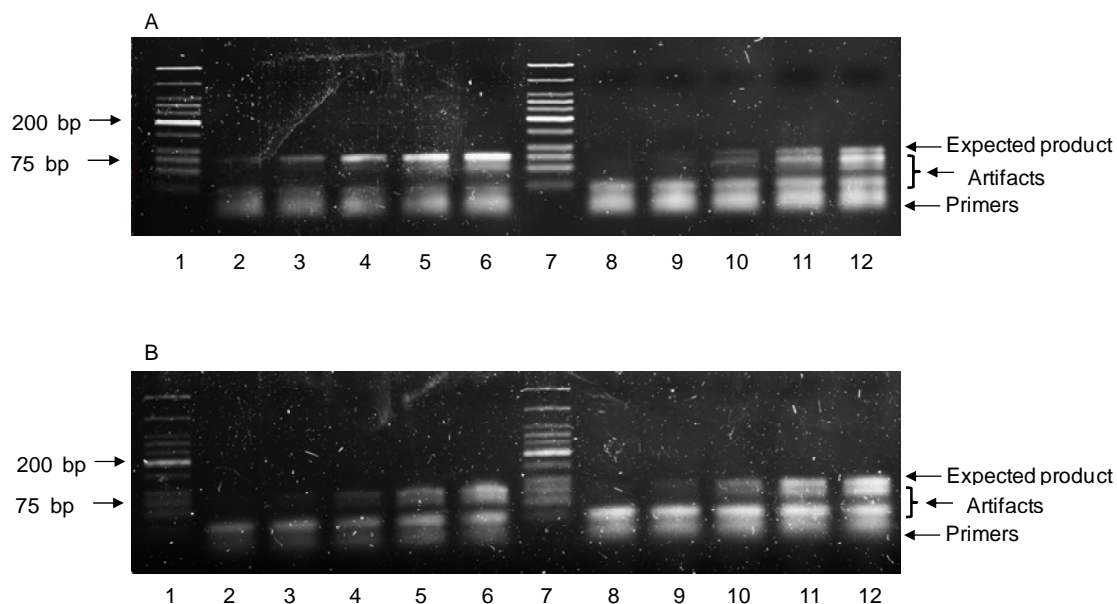


Figure 4.14 Amplification of Dz20-49 templates with varying modifications using Taq polymerase. **A** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 27, 29, 31, 33, 35 cycles of PCR, respectively, using unmodified template. Lanes 8-12: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imm}-modified template. **B** Lanes 1 and 7: NEB Low Molecular Weight Ladder, Lanes 2-6: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{im}-modified template. Lanes 8-12: 27, 29, 31, 33, 35 cycles of PCR, respectively, using dU^{ga}, dC^{aa} and dA^{imp}-modified template.

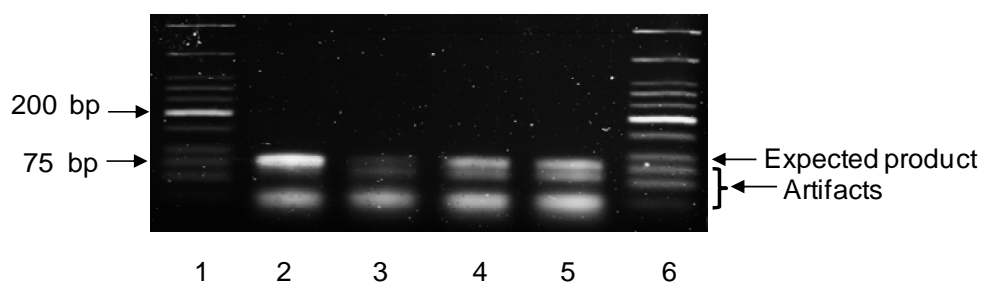


Figure 4.15 Comparison of the amplicons produced by Taq using modified templates. Lanes 1 + 6: NEB Low Molecular Weight Ladder. Lane 2: Amplicon produced after 33 cycles of PCR using template containing no modifications. Lanes 3-5: Amplicons produced after 33 cycles of PCR using templates containing modified dC, modified dU and one of dA^{imm}, dA^{ime}, dA^{imp}, respectively.

4.4.4 Production of doubly-modified dsDNA

The advantages of ssDNA with a high density chemical functionalization were described in section 1.2.5. It would then be of interest to study the properties of doubly-modified double-stranded DNA as our folded DNAzyme could have many double-stranded motifs. Doubly-modified dsDNA was produced by performing PCR using modified nucleoside triphosphates dU^{ga}TP, dU^{ph}TP or dU^{aa}TP. Amplicons were produced cleanly and were agarose gel purified (Figure 4.16).

A

5' - *TAA TAC GACT CACTA TAggggt*aacgccaggggttttccCAGTCACGACGTTGTAAAACG
3' - *ATTATGCTGAGTGTTAT*cccatgtgcggtcccaaaaggGTCAGTGCTGCAACATTTTGC

HindIII SalI XbaI BamHI KpnI

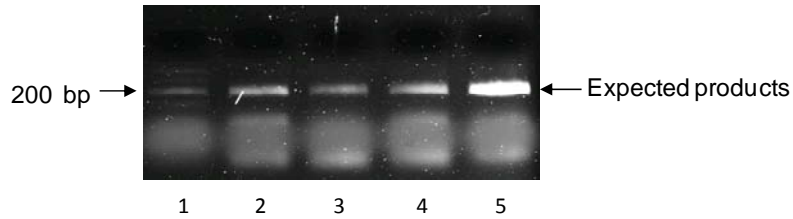
ACGCCAGTGCC**AAGCTT**GCATGCCTGCAG**GTCGACTCTAGAGGATCCCGGGTACC**GAGC
TGCCGGTCACGG**TTCGAAC**GTACGGACGTC**CAGCTGAGATCTCCTAGGGGCCCA**TGGCTCG

SmaI

Eco RI
 TCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATtgttatccgctcacaattc
 AGCTTAAGCATTAGTACCAGTATCGACAAAGGACACACTTTAacaataggcgagtgttaag

cCCGCTGAGCAATAACTAGC- 3'
gGGCGACTCGTTATTGATCG- 5'

B



C

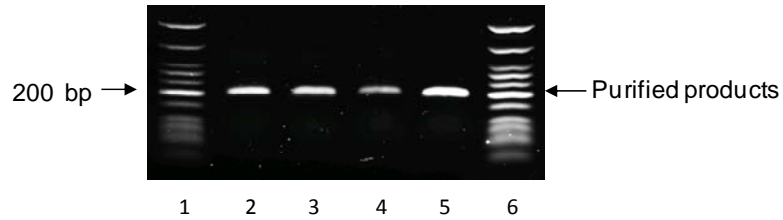


Figure 4.16 Amplicon sequence and modified PCR products. **A** Amplicon sequence showing foreign sequence introduced by PCR primers in italics, primer binding regions in lowercase, and restriction digest sites shown in boldface. **B** Lane 1: NEB Low Molecular Weight DNA Marker. Lanes 2-5: dU^{ga}-modified amplicon, dU^{ph}-modified amplicon, dU^{aa}-modified amplicon and unmodified amplicon, respectively. **C** Gel purified PCR products. Lane 1 and 6: NEB Low Molecular Weight DNA Marker. Lanes 2-5: dU^{ga}-modified amplicon, dU^{ph}-modified amplicon, dU^{aa}-modified amplicon and unmodified amplicon, respectively.

4.4.5 Restriction digests of doubly-modified dsDNA

Section 4.4.3 describes reduced activity between densely functionalized DNA templates and two commercial polymerases. Another type of commercial enzyme commonly used in molecular biology and in selections is the restriction enzyme or restriction endonuclease. Determination of efficiency of restriction enzyme digestion of densely functionalized dsDNA would be of interest. Doubly-modified dsDNA amplicons were also subjected to restriction digestion. For size control, unmodified DNA was produced and subjected to digestion for the same duration of time. The dU^{ph}-modified and dU^{ga}-modified amplicons showed resistance to cleavage by Hind III, Bam HI, Kpn I, Eco RI, Sal I and Xba I (Figure 4.17). With the exception of Sma I, all restriction enzymes cleave at a site containing a dT (Table 4.9). Compared to the cleavage of unmodified amplicon, the modifications provide no resistance to cleavage against Sma I. The dU^{aa}-modified amplicon showed resistance to cleavage by all of the tested restriction enzymes. Percent cleavage by the various restriction enzymes is summarized in Table 4.9.

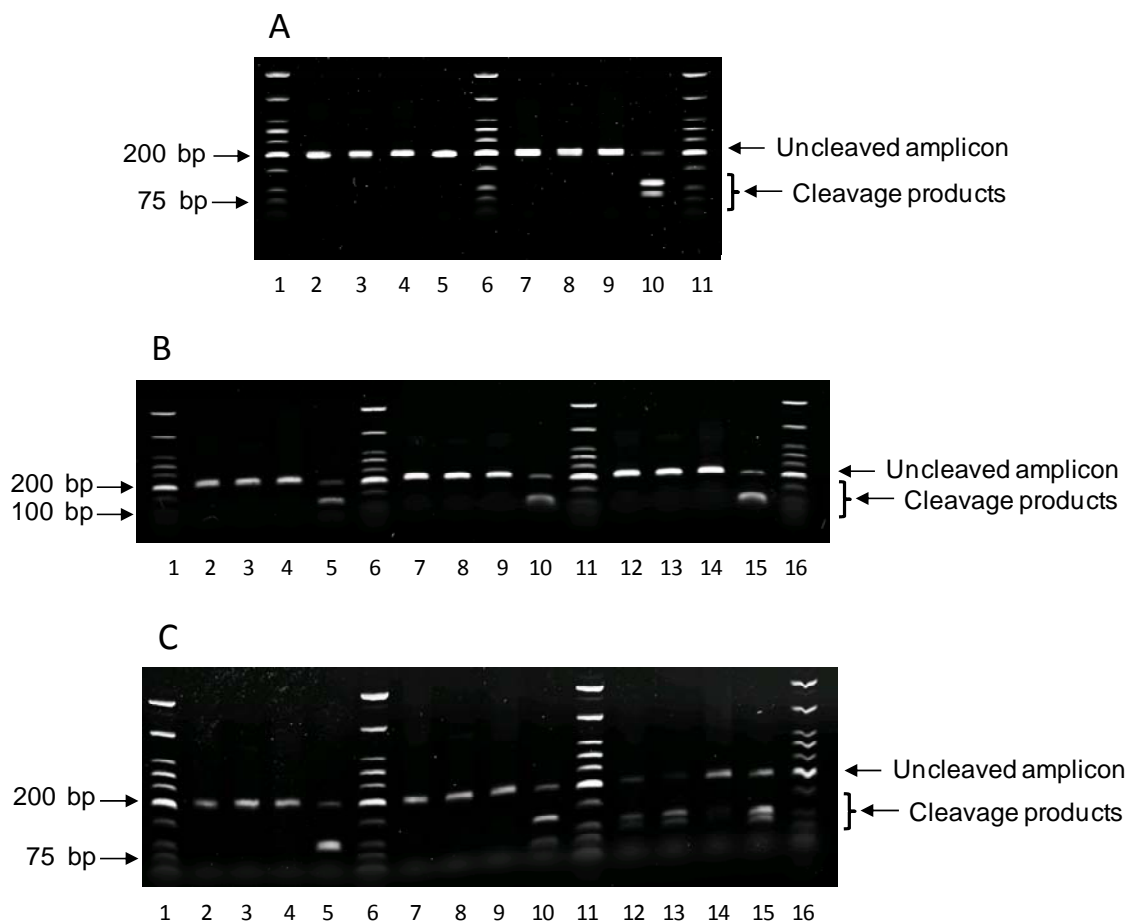


Figure 4.17 Restriction digests of modified dsDNA. **A** Lanes 1, 6 + 11: NEB Low Molecular Weight DNA Marker. Lanes 2 -5: Untreated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 -10: Kpn I treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. **B** Lanes 1, 6, 11 + 16: NEB Low Molecular Weight DNA Marker. Lanes 2 – 5: Hind III-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 – 10: Sal I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 12 – 15: Xba I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. **C** Lane 1, 6, 11 + 16: NEB Low Molecular Weight DNA Marker. Lanes 2 – 5: Bam HI-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 7 – 10: Eco RI-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively. Lanes 12 – 15: Sma I-treated dU^{ga}-, dU^{ph}-, dU^{aa}-modified DNA and unmodified DNA, respectively.

Table 4.9 Summary of the percentage of dU-modified dsDNA cleaved by the indicated restriction enzyme after two hours of incubation at 37 °C.

restriction enzyme	Restriction site shown 5' to 3'	dT	dU ^{ga}	dU ^{ph}	dU ^{aa}
Sal I	GTCGAC	80	<5	<5	<5
Xba I	TCTAGA	81	<5	<5	<5
Bam HI	GGATCC	86	<5	<5	<5
Eco RI	GAATTC	81	<5	<5	<5
Sma I	CCCGGG	75	74	86	43
Kpn I	GGTACC	87	<5	<5	<5
Hind III	AAGCTT	79	<5	<5	<5

4.4.6 Sequences originating from doubly-modified dsDNA

The possibility of a modified nucleotide causing detrimental mutations could greatly affect the outcome of a DNAzyme selection. The mutagenic potential of the modified dU described in this chapter was examined by the introduction of doubly-modified dsDNA into bacterial cells followed by the sequencing of the bacteria-produced plasmids. Sequences were obtained from plasmids originating from dU^{ph}-modified and dU^{ga}-modified plasmids (Figure 4.18). Few white colonies were obtained possibly as a result of inefficient production of ligated product. A transformation efficiency of only $\sim 2 \times 10^5$ cfu/ μ g was attained. This could be due to inefficient A-tailing by Taq. The few sequences of the multiple cloning sites obtained were virtually identical to the original multiple cloning site sequence. One sequence, however, did have a T to A mutation in the Bam HI site. Transfection using the dU^{aa}-modified plasmid failed to yield plasmids containing correctly sized inserts.

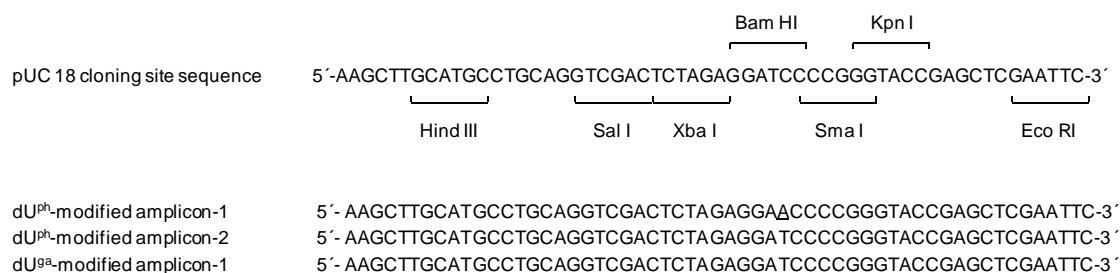


Figure 4.18 Multiple cloning site sequences obtained from plasmids originating from dU^{ga}-modified or dU^{ph}-modified DNA. The dT to dA base mutation is indicated with an underlined A.

4.5 Discussion

Arising from an observation that active DNazymes Dz10-13 and Dz20-49 only contain four modified dA's in their N40 regions, the possible origin for such a low percentage of these critical modified nucleotides was explored. Gathering more information from other similar selections led us to suspect the possibility of polymerase-mediated bias. Other enzymic interactions, resistance to restriction enzyme digestion and the processing of a plasmid bearing modified DNA by *E. coli*, were assayed.

4.5.1 General analysis of the sequences of the final DNazyme generations

The concept of fitness involves the survival of a certain species from one generation to another. While catalytic cleavage plays a large role in this, one of several selection pressures is unintentionally applied even before the catalyst is produced. The incorporation of modified nucleotides depends on the efficiency of incorporation and the ability to incorporate multiple modifications sequentially. After self-cleavage, the clones undergo another selection pressure. This pressure also involves the interaction of polymerases and the modified DNA catalyst during the transcription of modified DNA to unmodified DNA.

The numbers of clones were graphed according to their dA content or their dT content as these bases were modified and in the case of dA, this base was found to be the most refractory to incorporation and yet the most integral to M²⁺-independent activity.

Clones from 8 different selections were graphed relating population to modification content for a particular modification. Two selections used a library with a 20 base degenerate region. Statistically, the average number of dA's in any given sequence of the selections using twenty degenerate positions should be five (20 positions/ 4 possible bases). For the library with a forty degenerate positions, the average number of dAs in a given sequence should be ten (40 positions/ 4 bases). Deviation from this average number could be indicative of a biased incorporation of modified dA or it could also indicate a biased replication of the modified DNA catalyst to unmodified DNA. From the data, one cannot determine whether the bias was caused by poor incorporation or if it was caused by poor read-through. For the 8-modified dA, poor incorporation has already been established (Lam *et al.*, 2008) and likely eliminates sequences containing a high percentage of dA's or sequential dA's in early rounds.

The average number of modified dAs in all of the selections is lower than the expected statistical average. The selection containing the greatest shift in distribution is the mercury sensor Dz10-13 selection (Figure 4.11). Graphical representation of the distribution of modified dA shows that there is a bias against sequences containing a large number of dA. This bias can come from two components of fitness which are the low efficiency of incorporation of the modified nucleotide in the primer extension synthesis of the modified catalyst or the low amplifiability of a sequence derived from a modified strand of DNA. In the case of the mercury sensor Dz10-13 selection, however, it can be hypothesized that the selection promotes species to jettison all non-essential imidazole modifications to avoid non-essential mercury cation binding. The selection, perhaps, favors clones with a minimal number of critical imidazoles. This is not unexpected as one can assume that imidazoles facilitate the binding of unnecessary mercury cations, which could be detrimental in general acid/general base chemistry.(Thomas *et al.*, 2004)

Another effect that incorporation and read-through bias can have on the fitness of a sequence is on the replication of sequences with sequential modified dA's. Interestingly, clone #66 (also known as Dz10-66) from the Dz10-66 selection is one of the very few sequenced and highly active clones that possess two sequential dAs. If there is a bias against sequential 8-modified dA, not only are we reducing the total number of

available sequences using a defined number of positions, also referred to as sequence space, but more importantly we are losing chemical space, which is contrary to our overall goal of enriching functional DNA selections. Perhaps the second imidazole provides some structural stabilization or perhaps the sequential modifications are not critical and the ablation of one of those modifications may have a minimal effect on catalysis. Assessing the need for these two sequential dA would require a systematic knockout of each followed by a kinetic analysis.

The average number of A's (~7) in the sequences of Dz20-49 selection may be a little surprising at first. While it fits the shift to a lower-than-expected number of modified A's, one would expect that there would be a definitive shift similar to that of the mercury selection due to the difficulty in the amplification of the dA^{imm} as seen in section 4.4.3. That is, if the 8-(2-(4-imidazolyl))aminomethyl functional group is critical for catalysis, but highly detrimental to amplification, we would expect a mean distribution around four modifications so that only the critical imidazoles are present. Instead, the mean appears to be around 6 to 7. This was probably because the pool of the 20th generation had not reached convergence and sequences originating from uncatalyzed transphosphorylation represent a significant percentage of the gene pool as indicated by the high percentage of inactive clones.

Upon closer examination of the sequences of the inactive clones, a recurring short sequence UUAUU stands out amongst many of the clones. Interestingly, this UUAUU is absent in the sequences of the active clones in the final generation suggesting that it does not enhance the catalytic trait of the clones, but enhances some other trait that allows sequences containing UUAUU to be enriched. The noticeable UUAUU appears to be at the center of a larger sequence AUGCAUGGUUAUUUGAGUCUAGG. There are slight variations amongst the sequences, but there is a striking homology between inactive clones possibly due to this larger recurring sequence. We hypothesize that there exists sequences which allow for preferential amplification of modified DNA bearing dA^{imm}. Such contextual amplification was elegantly described by Kimoto *et al.* who identified certain nearest-neighbor sequences that preferentially co-amplify with an unnatural base pair 7-(2-thienyl)-imidazo[4,5-b]pyridine (**Ds**) and 2-nitro-4-propynylpyrrole (**Px**). (Kimoto *et al.*, 2009) While our modifications do not reflect an unnatural base

pairing motif, it is conceivable that the nearest-neighbour sequences that dG, dU^{ga}, and dC^{aa} may greatly influence the amplifiability of the dA^{imm} similar to the unnatural nucleosides of the **Ds:Px** system. Evidence of this effect is suggested by the sequences of the final generation of our selection; the core sequence AUGCAUGGUUAUUGAGUCUAGG and slight variations thereof appear in ~ 30 % of the clones, none of which is active. Notably this sequence is absent in all of the active clones and this observation supports our hypothesis that an over-represented sequence does appear for reasons other than its contribution to catalytic activity.

Ideally, where a selection demands only catalytic activity, one would expect that inactive sequences would be eliminated from the gene pool, rather than overpopulate it. Yet in this case, motifs that are barely active are over-represented, even though some contain a greater percentage of modified dA's than Dz20-49. When certain sequences are minimally active but preferentially amplified, these too will likely prevail because such a selection selects for both catalytic activity and amplifiability. Although sequence dependent amplifiability has not yet been observed in selections of unmodified catalysts, (Schlosser *et al.*, 2009) it is well known that unmodified DNA sequences e.g. triplet repeats are poorly amplified in a PCR, and thus subtle amplification differences may also determine the outcome of unmodified selections, only to a lesser extent. Tyranny of the small motif, whereby certain small catalytic sequences are found embedded in larger libraries, may indeed reflect such preferences. (Joyce, 2004)

Curtis Lam's ATP sensor and stachyose sensor selections did not show the same bias in amplification as the aforementioned modified dA-dependent selections. In the ATP sensor and stachyose sensor selections, the only modified nucleoside triphosphate used was dU^{ph}TP. The distribution of the clones does lie with a mean around the average of ten modifications per clone. The modifications used in the ATP and stachyose selections are very well tolerated and do not show any detrimental effects to the fitness on the selected clones. That is, the mean number of modifications was not shifted lower than the expected ten. This is an ideal system of a selection using a modified nucleotide. These observations are by no means replacements for proper exhaustive analysis of the populations and sequence-dependent amplifiability. However, as we strive to expand the chemical space of DNAzymes, we should be vigilant and make note of all decreases in

the sequence space. Observations like these are important to better understand selections where the smallest side reaction can be exponentially amplified and be highly detrimental to the selection.

4.5.2 Evaluation of modified DNA as a template in PCR

The efficiency of amplification of the active modified DNAs is extremely important in a DNAzyme selection as it is an aspect of fitness of a DNAzyme bearing modifications. The fitness of a DNAzyme determines its continuation of its genetic information onto the final generation. In our amplification tests, conditions were made to simulate the first amplification condition of our DNAzyme selections. Therefore, the tests included low template concentration conditions and high number of cycles of PCR. The trace amount of template was meant to represent the few active clones in the large starting library. The primers and the use of Vent (exo-) were also meant to simulate our first amplification conditions. Of the modified templates described in sections 4.3.4, 4.4.2, and 4.4.3, the native Dz20-49 construct that contains dA^{imm} appeared to be the most difficult template to amplify as the least amount of amplicon was produced in PCRs using this template. Amplification of Dz20-49 constructed with dA^{ime} and dA^{imp} in place of dA^{imm} served as better templates resulting in higher amounts of amplicon. The dA^{ime} and dA^{imp} constructs amplified roughly the same, but both produced less amplicon than the fully unmodified control as expected.

PCR amplification of the modified templates using Taq clearly indicates that the polymerase is having difficulty with read-through. The product ratios were similar to the ratios of the amplicon amounts produced by Vent (exo-). Amplification of the three modified templates by Taq resulted in decreased amounts of amplicon compared to that of the unmodified template and all three produced an artifact. In the amplification of Dz20-49, artifacts dominated the PCR and very low amounts of amplicon of the correct size was actually being produced. The appearance of artifact in the PCR using Taq is strong justification for our choice of Vent (exo-) for the first amplification in our selections. Now knowing that the dA^{imm} modification causes difficulty in amplification and, more importantly, in selection, one can make the assumption that read-through bias

as well as incorporation bias plays a role in the selection of Dz20-49. If one were to continue researching the origins of bias, one experiment would be to see if there is a sequence bias between templates heavily modified with dA^{imm} versus templates containing very few dA^{imm}. The extension times used in the amplification tests were also meant to simulate conditions used in the amplification of the *in vitro* selection product described in Chapter 3. A follow up experiment that could be done is to see if the large difference in amplicon production due to differing modifications can be reduced by lengthening the extension time of the cycles to promote production of full-length unmodified strand of DNA. If found to be beneficial i.e. increased production of amplicon from the dA^{imm}-modified template, lengthening extension time be implemented in any future *in vitro* DNAzyme selections using difficult-to-amplify templates.

4.5.3 Production of doubly-modified dsDNA

The production of doubly-modified dsDNA is one of the final tests to demonstrate the fitness of a system made up of a thermophilic polymerase and a set of nucleoside triphosphates where at least one nucleoside triphosphate is modified. Most noteworthy successes have been done by Famulok and coworkers who produced PCR product with modifications on every nucleobase.(Jäger & Famulok, 2004; Jäger *et al.*, 2005) Other groups have also made contributions, but the most applicable studies have come from Sawai and coworkers who have done exceptional research on 5-modified pyrimidines and the thermophilic polymerases that accept them as substrates.(M. Kuwahara, Nagashima *et al.*, 2006; Sawai *et al.*, 2007) Through Famulok's and Sawai's work, we get a better grasp on the best method of linkage of functional groups and the polymerases to use according to their family. Famulok's method of attaching functional groups through the use of alkynyl groups has probably found the most success for immediate incorporation with a commercial polymerase. Both Famulok and Sawai have both found that the Family B polymerases are exceptional for the incorporation of 5-modified pyrimidines. In particular, KOD Dash is ideal for an acetamide linkage and Vent (exo-) and Pfu excel in the incorporation of nucleotides bearing an alkynyl linkage. Vent (exo-) appears to be the best Family B polymerase for general use for the 5-modified pyrimidines and was an

obvious choice for the generation of doubly-modified dsDNA using our DNAzyme-relevant nucleotides.

Since the incorporation efficiency of the two modified nucleoside triphosphates dU^{ga}TP and dU^{ph}TP had not yet been determined, extension times for the construction of the two modified dsDNA during PCR were increased to compensate for any poor incorporation. Also, these modifications can be detrimental to the read-through component of PCR. From the successful selections of modification-dependent DNAzymes, we know that modified ssDNA containing one of these two modifications can be amplified into unmodified dsDNA. However, the efficiency of that initial read-through is unknown. In the case of using modified nucleoside triphosphates in a PCR to produce modified dsDNA, the polymerase would necessarily have to use the modified extension products from the previous round as templates. In addition to this, the polymerase would be producing a modified strand while using a modified strand as a template. If the polymerase is unable to use modified template to synthesize modified DNA, the polymerase would only be able to utilize the original unmodified template for synthesis and produce, at maximum, one copy of modified DNA for each of the original unmodified strands of DNA per cycle. This type of amplification is referred to as linear amplification and while it is not nearly as efficient as exponential amplification, the amplicon products from either type of amplification will be doubly-modified dsDNA. Since exponential amplification of doubly-modified dsDNA is not required to produce doubly-modified dsDNA as linear amplification will suffice, a high amount of unmodified template was used to facilitate and maximize linear amplification to complement exponential amplification in the production of doubly-modified dsDNA. For the PCR, 250 ng of template was used in the reaction. While it is presumed that linear amplification contributes greatly to the production of amplicon, we know that from our *in vitro* selections, Vent (exo-) can read-through DNA modified with the dU^x nucleotides discussed in this chapter. However, only about 300 ng of doubly-modified dsDNA was recovered per modified nucleotide PCR, which is consistent with 20 cycles of linear amplification, and indicates that any modified DNA produced from exponential amplification (i.e. modified nucleotides incorporated across a modified DNA strand) does not contribute significantly to the overall total. Since the template is a plasmid, it greatly

differs in size from the 200 bp target and can be easily removed by gel purification. The amplicons of expected size were produced with little to no artifacts. Subsequent gel purification gave a pure product by agarose gel. These products were suitable for restriction enzyme digestion tests.

4.5.4 Restriction digest of modified dsDNA

For the restriction digest analysis, all three doubly-modified dsDNA were tested with seven different restriction enzymes. Six of the seven restriction enzymes used recognize 6 bp palindromic sites that contain dT. The remaining restriction enzyme Sma I recognizes the palindromic sequence CCCGGG, which contains no dT.

The first observation is that, in general, all of the doubly-modified dsDNA resisted restriction enzyme digestion from the six restriction enzymes containing dT in their respective restriction sites. The interesting observation is the cleavage by Sma I. Consistent to what has been mentioned in the literature before, (Perrin *et al.*, 1999) dU^{aa}-modified dsDNA, where the labeled strand is modified, is resistant to cleavage at a site not containing dT, but is flanked by dT. It was suggested that the cation helps neutralize the phosphate backbone which in turn causes distortions in the DNA. These distortions then impair the function of the restriction enzyme even though the modification is not within the actual restriction site. One would imagine the same trend would follow for the dU^{ga}, but interestingly dU^{ga}-modified and dU^{ph}-modified dsDNA was cleaved at roughly the same efficiency as unmodified DNA. Caution should be taken as the mutational properties of the modified nucleotides are not known at this point and that resistance to restriction enzyme digestion can arise from mutations in the restriction site of the target dsDNA.

In general, resistance to cleavage by restriction enzyme digestion can easily find its way in to modified DNase selections. In our modified nucleotide selections, we want to ensure that the DNases are dependent on the modifications. If the modifications have been removed intentionally or unintentionally, being able to identify the occurrence through the simple treatment of a restriction enzyme would be useful.

Also, the use of the restriction enzymes can be used to help remove unmodified constructs that may have arisen from contaminating amounts of unmodified nucleotides.

4.5.5 Processing of dU^{ph} and dU^{ga} *in vivo*

Sakthival and Barbas showed that dU^{aa}TP is a very poor substrate for PCR (Sakthivel & Barbas, 1998) whereas Lee *et al.* managed to produce substantial amounts of amplicon when amplifying short target sequence (<98 nt). (Lee *et al.*, 2001) Despite the low amount of amplification product, Sakthival and Barbas managed to ligate their doubly-modified dsDNA product, transfect the modified plasmid and harvest replicated plasmid from the host cells. They found that fidelity of the original sequence was retained throughout the PCR and cloning process. In our study, plasmids from a few white colonies for the dU^{ga}-modified and dU^{ph}-modified plasmid transfections were obtained. To facilitate detection of the amplicon product in the isolated plasmids, the primers used in the amplification contained sequences foreign to the template plasmid. This way the amplicon sequence can easily be detected using a second set of primers in PCR where the isolated plasmids are used as template. Doing so ensures that the multiple cloning site sequence originates from an amplicon and not from the original template, pUC18.

The sequencing of the plasmids revealed one mutation at the Bam HI restriction site of one of the dU^{ph}-modified DNA sequences. However, since the resistance to cleavage of the dU^{ph}-modified amplicon is practically complete for all restriction enzymes, we hypothesize that the resistance to restriction enzyme digestion of the dU^{ph}-modified amplicon is due to the modifications located in the restriction sites. Any mutation to the restriction sites will, of course, contribute to the appearance of resistance, but the primary cause of the resistance is credited to the inability of the restriction enzyme to recognize the dU^{ph} nucleotide. Sakthival and Barbas also attributed the resistance to restriction enzyme digestion to their 5-modifications on dU, and their subsequent cloning confirmed the fidelity of the replication of the restriction sites. In our experiments, the frequency of mutation required to cause mutations at six out of six dT-containing restriction sites would have to be extremely high to produce the pattern of resistance that was observed in Figure 4.17. If one would wish to pursue characterizing

the mutational properties of the modified dU nucleotides, a more thorough assay using the DNA mismatch binding protein, MutS, would be more appropriate.(Stanislawska-Sachadyn & Sachadyn, 2005)

A potential use for our polyguanidinium DNA is enhanced delivery; the polyanionic nature of oligonucleotides makes delivery through the cell membrane difficult. One class of cationic biopolymers that is internalized readily into the cell through yet undefined mechanisms is the cell penetrating peptide (CPP). The mechanism of internalization of polyarginine CPPs is currently under investigation.(Schmidt, Mishra, Lai, & Wong, 2010) Investigators have attached these polycationic peptides to oligonucleotides for increased cellular uptake. As the theme of this thesis is empowering the catalytic and/or binding potential of oligonucleotides through the use of chemical conjugation of amino acid side chains, it is imperative to mention that through the use of guanidinium-modified nucleotides, researchers hope to replicate the internalization using oligonucleotides as the biopolymer. After showing that the guanidinium group promotes uptake more than that of the amino group, Sawai and coworkers used guanidinium-coupled DNA to translocate dye-conjugated DNA into HeLa cells.(Ohmichi *et al.*, 2005) By modifying the oligonucleotides with positively charged functional groups, investigators hope to activate the mechanism that facilitates internalization of CPPs through recognition of the polycationic oligonucleotide biopolymer.

The finding that our modifications are replicated with fidelity leads us to the possibility of using modified DNA to facilitate transfection followed by replication into unmodified DNA. The process would produce a “traceless” piece of DNA with no evidence of prior modification. One possible use for highly charged DNA is in site-directed mutagenesis. Mutations are introduced using complementary primers that bear the desired mutation. After priming the template plasmid, a polymerase is used to replicate the entire plasmid. If this replication process is used to introduce a high number of cationic nucleotides, the resulting polycationic plasmid product may have cell penetrating properties. Since the product will also be resistant to cleavage by a variety of restriction enzymes, they can be added in place or to complement Dpn I in the destruction of unmodified template DNA. Once the polycationic plasmid enters the cell, the cell’s mechanism can replicate the plasmid into unmodified DNA with fidelity. This method

could remove the limitations of library size set by the transfection efficiency of the host cell. This is an ambitious goal, but not unreasonable considering our ability to create doubly-modified dsDNA with relative ease and pending a full assessment of the potential mutational properties of the dU^{ga} nucleotide.

Chapter 5: Towards the Evolution of *Thermus aquaticus* DNA Polymerase I

5.1 Introduction

As discussed in the previous chapters, we have focused efforts on engineering modified nucleobases towards enhanced function in a DNAzyme selection rather than on efficient incorporation. These efforts, however, may not produce the desired results due to poor incorporation by polymerases. For example, our imidazole modification of 2'-deoxyadenosine triphosphate tethered at the 8 position using relatively short linkers severely limits our choices of polymerases that we can use in a selection, and yet, generally speaking, it is this type of modification that has led to very successful DNAzyme selections.(Hollenstein *et al.*, 2008; Hollenstein *et al.*, 2009a; Hollenstein *et al.*, 2009b; Lerner, Roupiaz *et al.*, 2002; Perrin *et al.*, 2001) Sidorov *et al.* have also introduced the imidazole functionality into a DNAzyme in the form of a 5'-modified 2'-deoxyuracil.(Sidorov *et al.*, 2004) Not only was the imidazole attached to a position that oriented the modification towards the major groove, a long linker was also used. These two structural features facilitated incorporation by polymerase. While Sidorov's modified DNAzyme was found to be active, the unmodified variant of Sidorov's DNAzyme still retained 7 % activity causing the necessity of the modification to be subject to debate. In the end, it should be the function of the modified dXTP (and its appended function) within the modified nucleic acid enzyme that dictates the value of the modified dXTP, and not its ease or efficiency of polymerization in the process of finding a functional nucleic acid enzyme. Notwithstanding the need for high incorporation efficiency, however, we recognize that poor incorporation may in some cases afford reasonably good activity while in other cases it may be very detrimental to the selection process. Curtis Lam synthesized dA^{ime}-TP linker variants that were meant to improve upon enhanced catalytic function and improved incorporation.(Lam *et al.*, 2008) The crucial elements retained in the variants' design were the imidazole attachment through the 8 position of 2'-deoxyadenosine and use of a short linker. Unfortunately, closely related analogs of dA^{ime}-TP proved to be inferior to the original dA^{ime}-TP in terms of either enhanced catalysis, incorporation, or both as discussed in Chapter 3. While our endeavors to improve modified nucleotide properties through chemical synthesis continue, other non-chemical methods were also explored and described in this chapter.

Valuable modified nucleotides are not limited to nucleobase modifications. An unnatural sugar-modified nucleotide that has shown promise in the field of the gene silencing is the 2'-deoxy-2'-fluoroarabinonucleotide, which Damha and coworkers refer to as the FANA monomer.(Dowler, 2006 #464} The structure of 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)thymine 5'-triphosphate (2'F-araTTP) is shown in Figure 5.1. While the FANA monomers are not base-modified nucleotides, these modified nucleotides are important for the overall goal of specific gene silencing and could potentially be used in modified DNzyme selections as well. The current value of FANA monomers is their use in RNA interference (RNAi)(Aigner, 2007; Fire *et al.*, 1998) a process used to down regulate protein synthesis at the mRNA level, which is also known as post-transcriptional gene-silencing. When compared to small interfering RNA (siRNA), FANA-modified RNA was found to be four times more potent and had a half-life of 6 hours versus RNA's half-life of <15 minutes in serum.(Dowler *et al.*, 2006)

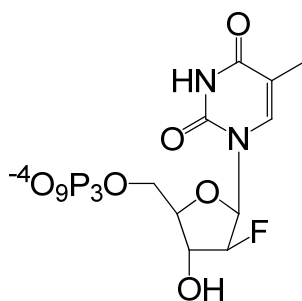


Figure 5.1 Structure of 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)thymine triphosphate (2'F-araTTP).

DNA:RNA hybrids are also capable of inducing RNaseH activity. When the phosphate group of the DNA backbone is replaced with a phosphorothioate group where the sulfur is introduced in place of one of the non-bridging oxygen atoms, DNA is bestowed with increased nuclease resistance. This phosphorothioate DNA (PS-DNA) can be combined with phosphorothioate FANA monomers. Much like FANA-modified siRNA, phosphorothioate FANA-DNA (PS-FANA-DNA) can induce RNaseH activity and raises the melting temperature, T_m , of the PS-FANA-DNA: mRNA heteroduplex.(Kalota *et al.*, 2006) One benefit of PS-DNA-FANA hybrids over PS-DNA is that it has been shown to be more potent, requiring only 20 % of the dose to get the same effect as PS-DNA. Also, PS-FANA-DNA is more persistent as 70 % remained in

the cell after 96 hours and none of the PS-DNA could be detected. Unfortunately, enzymic synthesis of FANA-DNA hybrids is limited by the low number of polymerases that efficiently incorporate or read-through the FANA monomer.(Peng & Damha, 2007) Similar to our nucleobase-modified nucleotides, the versatility and value of the FANA monomer would vastly increase if incorporation and read-through by polymerase were improved.

Instead of relying on solving the problem of poor incorporation through systematic modified nucleotide synthesis and testing incorporation, we sought to improve upon the incorporation of the modified nucleotides by modifying the polymerase used for incorporation. The aforementioned 8-modified 2'-deoxyadenosine triphosphates and 2'-F-araNTPs are normally poorly recognized by most polymerases, but to date have provided for extraordinary function. We attribute the extraordinary function of our modified nucleotide to key structure design features such as functional group attachment to the 8 position of 2'-deoxyadenosine via a short linker. To retain key elements of our nucleotide design and improve incorporation by polymerases, directed evolution of DNA polymerases will be used to improve nucleotide incorporation. A variation on short patch compartmentalized self-replication (spCSR) was used to attempt to evolve Taq polymerase's ability to incorporate modified nucleotides. The goal of the directed evolution was the improved incorporation of two nucleobase-modified nucleotides that were successfully used in a DNAzyme selection described in Chapter 3, dC^{aa} and dA^{imm}. To probe the versatility of spCSR for the directed evolution of polymerase, we also included 2'-F-araTTP as a substrate in spCSR to improve the incorporation of the FANA monomer.

At the onset of this undertaking, Taq was considered the best candidate for directed evolution. From the work seen in previous chapters, one may ask, "Why use one of the least competent polymerases for the incorporation of functionalized nucleotides for evolution?" A polymerase used extensively throughout the work in this dissertation due to its compatibility with many modified nucleotides is Vent (exo-). It is true that Vent (exo-) immediately shows its aptitude for incorporating modified nucleotides as seen in Chapter 4. Not only is Vent (exo-) capable of incorporating many of the modifications used in our lab, but also shows high efficiency for 5-modified dUTP and 5-modified

dCTP in general.(Sawai *et al.*, 2007) Recall from Chapter 1 that Vent (exo-) is a Family B polymerase and many of the other Family B polymerases also follow this trend of efficient 5'-modified dUTP and 5'-modified dCTP incorporation.(Sawai *et al.*, 2007) Taq is a Family A polymerase. That said, Family A and Family B operate using similar structural motifs and mechanisms.(Patel & Loeb, 2001) It is feasible that a Family A polymerase like Taq can be evolved to possess some characteristics of the Family B polymerases.

There are two main reasons for choosing Taq for directed evolution. First, Taq is a commonly used and well-studied enzyme. Studies show that Taq is highly mutable and activities can be altered with agreement to the crystal structure.(Suzuki *et al.*, 1996) From a researcher's point of view, this facilitates engineering of the enzyme both on paper and in practice. Second, one of the major themes of this thesis is the expansion of the utility of 8'-modified dA. The primary polymerase used for incorporation in DNazyme selections in our lab is Sequenase Version 2.0. Sequenase Version 2.0 is a variant of Family A polymerase T7.(Tabor & Richardson, 1989) Taq is also a Family A polymerase and most likely evolved from the same ancestral gene as T7 polymerase.(Braithwaite & Ito, 1993) It would make sense to attempt to evolve the pliable Taq for our 8'-modified dA nucleotides and other unnatural nucleotides as well.

Taq's recognition of nucleobase uses a very dynamic and complex process.(Patel & Loeb, 2001) The nucleobases, paired or unpaired, interact with the thumb, palm and finger subdomains of Taq and a conceived selection would most likely require mutations in multiple regions in the Taq gene. Altering this very precise process is a daunting task. Aside from incorporation of modified bases, two labs succeeded in evolving Taq to accept ribose-modified substrates.(Ong *et al.*, 2006; Xia *et al.*, 2002) Our lab focuses on nucleobase modifications; however, improving polymerase function to include incorporation of sugar-modified nucleotides, like the FANA monomer, would also be a sensible goal due to the simplicity of the recognition of the 2' substituent, which seems to be the responsibility of a single amino acid, E615.(Ong *et al.*, 2006; Xia *et al.*, 2002) Both groups recognize that a single polypeptide and in particular residue E615 seems responsible for Taq's ability to discriminate 2'-deoxyribose from ribose. Loeb and coworkers identified 23 mutant polymerases, isolated from a previous non-related

selection, that have the ability to synthesize RNA.(Patel & Loeb, 2000b) The most common mutation seems to be E615D, which appears in ten of the twenty three clones and is the only mutation appearing at that position. Romesberg and coworkers evolved the Stoffel Fragment into an RNA polymerase and found four of their best mutants contained mutations only at positions 614 and 615, which are located in Motif A of the active site.(Xia *et al.*, 2002) They claim that while creating space with the E615G or E615A mutation allows for ribose to be accepted, a mutation at 614 usually occurs to compensate for the loss of any secondary function the glutamate carboxylic acid might have. Holliger and coworkers used a modified version of compartmentalized self-replication (CSR) called short patch CSR or spCSR.(Ong *et al.*, 2006) This technique focuses on the incorporation of multiple ribonucleotides and attenuates the stringency by making the CSR target amplicon a small portion of the gene that encodes a short polypeptide as opposed to the whole gene that encodes the entire polymerase. This selection required the stringency to be as low as possible because the mutant DNA polymerases were required to incorporate multiple ribonucleotides. In this selection, every successful clone had the E615G mutation. Since Loeb's selection was not designed specifically for the incorporation of ribonucleotides, it was not surprising that the E615G mutation is absent and the conservative mutation E615D is present. The findings of these investigators lay the groundwork for the evolution of polymerases for the incorporation of sugar-modified nucleotides.

5.2 Objective of this work

As seen in the previous chapters, the biggest setback in the selection of nucleobase-modified DNAzymes has been polymerase recognition of modified nucleoside triphosphates. With regards to the synthesis of modified DNA through the incorporation of modified nucleotides, inefficient polymerization affects the synthesis of the DNAzyme libraries by favoring sequences bearing fewer 8-modified dA. Post-selection processing of the successful clones is also affected by polymerases due to inefficient read-through. In Chapter 4, we noted that the sequences of the final generations of 8-modified dA-dependent DNAzymes on average have low dA content

and this hints at preferential amplification of sequences of sequences lacking the crucial 8-modified nucleotide. Therefore, to avoid sacrificing our key structural elements of modified nucleotide design such as functional group attachment to the 8 position of 2'-deoxyadenosine using short linkers, a polymerase will be evolved to accommodate our 8-modified dATPs and other modified nucleotides for both incorporation and for read-through. Taq is the prime candidate for directed evolution using a customized version of spCSR. (Ong *et al.*, 2006) Our *in vitro* selection technique described in this chapter is a derivation of the spCSR protocol used by Ong *et al.* where a modified nucleoside triphosphate was used in place of its natural counterpart in spCSR. The two base-modified unnatural nucleoside triphosphates chosen, dC^{aa}TP and dA^{imm}TP, have found utility in DNAzyme selections. The last unnatural nucleoside triphosphate used is 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)thymine 5'-triphosphate (2'F-araTTP) (Figure 5.1). Once incorporated into DNA, the FANA monomer bestows upon DNA increased duplex stability with an RNA target and initiates RNase H activity. This spCSR technique of evolving Taq polymerase could be used as a universal method of introducing other unnatural nucleotides into DNA and, ultimately, into selections.

The goal of the work presented in this chapter is to isolate polymerase mutants capable of incorporating, elongating and amplifying modified DNA. However, this goal must be preceded by several smaller goals. First, we needed to show that in our hands, stable reverse micelle compartments can be produced and that cross reactions between compartments is minimized. Second, we needed to express a library of Taq mutants and show that they are functional in the compartments. Third, without the high selection pressure of incorporating the modified nucleotides, we needed to show that enrichment is possible using spCSR with a very low selection pressure i.e. mutants capable of simple self-replication using natural nucleotides. After these goals have been achieved, unnatural nucleoside triphosphates dA^{imm}TP, dC^{aa}TP and 2'F-araTTP was used in a selection for the directed evolution of Taq.

5.3 Materials and methods

5.3.1 Chemicals and reagents

1-(2-Deoxy-2-fluoro- β -D-arabinofuranosyl)thymine 5'-triphosphate (2'F-araTTP) was kindly donated by Dr. Masad Damha. CSR oil phase reagent Span 80 was purchased from Fluka. Tween 80, Triton-X 100 and mineral oil were purchased from Sigma. CSR oil phase is a combination of Span 80 (4.5 % v/v), Tween 80 (0.4 %, v/v) and Triton-X 100 (0.05 %, v/v) all in mineral oil. Bacteriological agar was purchased from Marine BioProducts. Pfx polymerase, LB Broth Base (Lennox L Broth Base), and Terrific Broth (TB) were purchased from Invitrogen. Bst polymerase (large fragment) was purchased from New England Biolabs.

5.3.2 Oligonucleotides

The following oligonucleotides were synthesized and purchased from Integrated DNA Technologies (IDT). Oligonucleotides are shown in the 5'-3' direction. Oligonucleotide names used in the notebooks are shown in parentheses.

ODN 5.1 Biotin-ATGACCATGATTACGAATTCGG (CH067D)

ODN 5.2 AGAAGATCTATCACTCCTTGCGGAGAGCCA (CH001C)

ODN 5.3 ACGAATTCGCCCCAAGGCCCTGGAGGAGGCC-3' (CH001B)

ODN 5.4 GTCGACTCTAGAAGATCTATCA (CH067B)

ODN 5.5 GGACTATAGCCAGATAGGGCTCAGGGTGCTGGCC (E615G-A)

ODN 5.6 GGCCAGCACCCCTGAGCCCTATCTGGCTATAGTCC (E615G-B)

ODN 5.7

GCTAGTTATTGCTCAGCGGTTTCGGCGTCCCGCGGGAGGCCCTCCAGCCCCCT
(CH091A-T7)

ODN 5.8 GCTAGTTATTGCTCAGCGGTAAGGGATGGCTAGCTCCTGGGA
(CH092D-T7)

ODN 5.9 TAATACGACTCACTATAGGGCGCAAACCGCCTCTCCCCGCGCGTTGG
(DK051A-T7pro)

ODN 5.10 GCTAGTTATTGCTCAGCGGAGAACATCCCCGTACGCAC (CH192A-T7)

ODN 5.11 GCTAGTTATTGCTACGCGGGTCGACTCTAGAAGATCTATCA
(CH067B-T7)

ODN 5.12 GCTAGTTATTGCTACGCGGCGCAAACCGCCTCTCCCCGCGCGTTGG
(DK051A-T7)

ODN 5.13 CGCAAACCGCCTCTCCCCGCGCGTTGG (DK051A)

ODN 5.14 GCTAGTTATTGCTCAGCGGAGATCTATCACTCCTTGGC (CH-6-163B)

ODN 5.15

GCCGAGGAGGGGTGGctattggtggccctggactatagccagatagagctcaggGTGCTGGCCCACC
TCTC (CH081B)

lowercase letters represent the wild-type base and the level of mutation is 88 % wild-type, 4 % each of the other 3 bases

ODN 5.16

ATCCAGCTGGCGGTCTCCGTGTGGATGTCCCGCCCCTCCTGGAAGACCCGGA
TCAGGTTCTCGTCGCCGGAGAGGTGGGCCAGCAC (CH145C)

ODN 5.17

GAACATCCCCGTACGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATC
GCCGAGGAGGGGTGG (CH145A)

ODN 5.18

GGGGTCCACGGCCTCCCGCGGGACGCCGAACATCCAGCTGGCGGTC
(CH145D)

ODN 5.19 CGCAAACCGCCTCTCCCCGCGCGTTGG (DK051A)

ODN 5.20 TAATACGACTCACTATAGGG (T7_{pro})

ODN 5.21 GCTAGTTATTGCTCAGCGG (T7_{term})

ODN 5.22

CGGGAGGCCGTGGACCCCCTGATGcgccgggcgcggaagaccatcaacttcggggtcctctacGGC
ATGTCGGCCCACCG (CH091B)

lowercase letters represent wild-type base and the level of mutation is 91 % wild-type , 3 % each of the other 3 bases

ODN 5.23 (5'-TGGCTAGCTCCTGGGAGAGGCGGTGGGCCGACATGCC-3')
(CH091C)

ODN 5.24 (5'-GTAAGGGATGGCTAGCTCCTGGGA-3') (CH091D)

5.3.3 Recombinant Taq

Plasmid pTaqacsac was kindly donated by Dr. Lawrence Loeb. The Taq gene was amplified out of the plasmid pTaqacsac using primers **ODN 5.1** (5'-Biotin-ATGACCATGATTAC**GAATT**CGG-3') and **ODN 5.2** (5'-AGA**AGATCT**ATCACTCCTTGGCGGAGAGCCA -3'). The amplicon containing the gene was purified using QIAGEN's Nucleotide Clean-Up Kit. The gene was then digested using Eco RI and Bgl II (indicated in boldface) and cloned into the Eco RI and Bam HI sites of pUC 18 (Fermentas). The resulting plasmid is referred to as pTaq.

5.3.4 Stoffel Fragment

Stoffel Fragment was amplified out of the plasmid pTaqtaqsac using the primers **ODN 5.3** (5'-AC**GAATT**CGCCCAAGGCCCTGGAGGAGGCC-3') and **ODN 5.4** (5'-GTCGACTCTAGA**AGATCT**ATCA-3'). Restriction sites Eco RI and Bgl II are shown in boldface. The Stoffel Fragment gene was digested with Eco RI and Bgl II and ligated into the Eco RI and Bgl II sites of the expression vector pUC18. The resulting plasmid was named pStoff.

5.3.5 E615G site-directed mutagenesis

Site directed mutagenesis to obtain the E615G single point mutant using the following reaction 1 × Pfu Ultra Buffer (composition not disclosed by Invitrogen), 200 μM dNTP, 2.5 ng / μl pTaq plasmid, 1 μM primer **ODN 5.5** (5'-GGACTATAGCCAGATAGGGCTCAGGGTGCTGGCC-3'), 1 μM primer **ODN 5.6** (5'-GGCCAGCACCCCTGAGCCCTATCTGGCTATAGTCC-3'), and 2.5 U of Pfu Ultra in a total volume of 50 μl. The mismatch bases that introduce the E615G mutation are underlined. The reaction was thermocycled 20 × (94 °C/30 sec, 57 °C/30 sec, 72 °C/5 min). After thermocycling, 10 μl of the reaction was treated with 4 U of Dpn I for 1 hour at 37 °C to digest template plasmid. After digestion, an aliquot of the reaction (5 μl) was dialyzed on 1 % agarose for 1 hour. An aliquot of dialyzed reaction (0.5 μl) was used to

transform Invitrogen *E. coli* DH10B Electromax cells (20 µl). Colonies were chosen at random and sequenced for the identification of successful mutagenesis. Expression was carried out for 12 hours without the use of an inducer.

5.3.6 Expression

Expression of all polymerases including libraries was carried out based on the protocol described by Desai and Pfaffle.(Desai & Pfaffle, 1995) Plasmids pTaq, pStoff, and ligated libraries were dialyzed on 1 % agarose for one hour to reduce the salt contents of the preceding ligation reaction mixture prior to electroporation. Plasmids were electroporated into Invitrogen *E. coli* DH10B Electromax cells (20 µl). After one hour incubation at 37 °C with shaking, the transformed *E. coli* (0.5 µl) was spread on LB (Invitrogen) Agar (Marine BioProducts) 100 mg / l ampicillin and the remainder was added to 100 ml of Terrific Broth (Invitrogen). Incubation was carried out at 37 °C with shaking @ 225 rpm overnight. Proteins were visualized on a 12 % sodium dodecyl sulphate polyacrylamide gel and stained with Coomassie Brilliant Blue (Bio-Rad).

5.3.7 Control experiments

5.3.7.1 Visualization of single cell encapsulation

pEGFP (Clontech) was used to transform DH5α *E. coli* cells (Invitrogen) and the cells were plated out. Cells from brightly fluorescing colonies were picked and added to water (200µl). The suspension was briefly agitated using a vortexer to separate the individual cells. A CSR oil phase was contained in a Corning Cryovial (2051) and rapidly stirred on a VWR stir plate turned on to setting number 10. The suspension of cells was added in portions (40 µl) every 30 seconds and the mixture was allowed to continue stirring for an additional 18 minutes after the addition of the aqueous phase. A glass pipette was used to smear a small amount of the emulsion through mineral oil located at the center of a microscope slide to dilute to reverse micelles. Dilution of the highly concentrated reverse micelles allows for single droplet visualization. The reverse micelles disperse upon placement on the cover slip. Visualization was done using an Olympus IX70 microscope. Cells were first visualized using UV light. Once the cells

were in focus, white light provided by the microscope was slowly added in until the border surrounding the cell could be distinctly seen. Images were captured using the program Image Pro Express.

5.3.7.2 Chemical detection of cross-reaction

Two separate PCRs named Reaction A and Reaction B were prepared. Reaction A contains a pair of primers that, when thermocycled, results in the amplification of a 160 bp target amplicon. Reaction A contains 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, primers **ODN 5.7** (5'-GCTAGTTATTGCTCAGCGGTTTCGGCGTCCCGCGGGAGGCCCTCCAGCCCCT-3') and **ODN 5.8** (5'-GCTAGTTATTGCTCAGCGGTAAGGGATGGCTAGCTCCTGGGA-3'), and pStoff plasmid and 5 U of Taq polymerase (NEB) in a total volume of 600 µl. Reaction B contains a pair of primers that, when thermocycled, produce a 0.8 kb target amplicon. This 0.8 kb amplicon also contains the target sequence for Reaction A. Reaction B contains 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, primers **ODN 5.9** (5'-TAATACGACTCACTATAGGGCGCAAACCGCCTCTCCCCGCGCGTTGG-3') and **ODN 5.4** (5'-GTCGACTCTAGAAGATCTATCA-3'), and 50 pg pStoff plasmid and 5 U of Taq polymerase (NEB) in a total volume of 600 µl. Solution phase controls were produced using Reaction A, Reaction B and a mixture of Reaction A and Reaction B. Volumes of each reaction used are shown in Table 5.1. Briefly, emulsified Reaction A (100 µl) and Reaction B (100 µl) were each placed into their own PCR tubes. Fifty microlitres of each of the solution phase Reaction A and Reaction B were mixed and placed into a PCR tube.

Table 5.1 Solution phase cross reaction controls.

	Reaction A	Reaction B	Mixture of solution phases
Reaction A (μl)	100	0	50
Reaction B (μl)	0	100	50
Total volume (μl)	100	100	100

An aliquot of each Reaction A (300 μl) and Reaction B (300 μl) was emulsified with CSR oil phase (600 μl). Total volume for each emulsion is 900 μl. Emulsified Reaction A (300 μl) was thoroughly mixed with emulsified Reaction B (300 μl). Emulsion PCRs of Reaction A, Reaction B and the mixture of emulsions are summarized in Table 5.2. The mixture of the two reactions (600 μl) was split amongst 6 PCR tubes. The remainders of each of the emulsified Reaction A (600μl) and emulsified Reaction B (600 μl) were each split amongst 6 PCR tubes.

Table 5.2 Emulsion phase cross reaction controls

	Reaction A	Reaction B	Mixture of emulsion phases
Emulsified Reaction A (μl)	600	0	300
Emulsified Reaction B (μl)	0	600	300
Total volume (μl)	600	600	600

All reaction mixtures listed in Tables 5.1 and 5.2 were thermocycled 20 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). After thermocycling, PCRs with emulsions were centrifuged for 15 minutes and the oil phase was decanted off leaving the concentrated reverse micelles. The emulsion was broken down with an extraction using water-saturated diethyl ether (1 ml). The aqueous phase was washed with one additional volume of water-saturated diethyl ether. The diethyl ether was removed and the aqueous layer was incubated without a lid in a 65 °C heat block until the diethyl ethyl stopped boiling off. The remaining aqueous phase was purified using Qiagen's Nucleotide Removal Kit. Products from the solution phase PCRs and the emulsion phase PCRs were analyzed on a 2 % agarose gel.

5.3.7.3 Chemical evidence of whole cell isolation

E. coli expressing Taq and *E. coli* expressing Stoffel Fragment were both incubated for 12 hours at 37 °C with 225 rpm shaking. The cell cultures were then cooled on ice. One milliliter from each culture was removed. The cells were spun down in a centrifuge (13000 rpm, 1 minute), and the media decanted. The cell pellets were resuspended in water (500 µl) and standardized to OD₆₀₀ = 1.0. An aliquot of each cell suspension (100 µl) was removed, placed in its own Eppendorf tube and spun in a centrifuge (13 000 rpm, 1 minute). The supernatant was removed. For expression controls, each pellet of Taq-expressing cells or Stoffel Fragment-expressing cells was resuspended using a PCR mixture containing 1 × Thermopol (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTP, 1 µM primers **ODN 5.10** (5'-GCTAGTTATTGCTCAGCGGAGAACATCCCCGTACGCAC-3') and **ODN 5.11** (5'-GCTAGTTATTGCTACGCGGGTCGACTCTAGAAGATCTATCA-3') in a final volume of 100 µl.

Table 5.3 Taq and Stoffel Fragment expression control reactions.

	Stoffel Fragment	Taq
Stoffel Fragment	100	0
Taq	0	100
Total volume	100	100

For the actual test for whole cell isolation by emulsion, six hundred microlitres of each suspension of cells in water were centrifuged and the supernatant was removed. Each pellet of Taq-expressing cells or Stoffel Fragment-expressing cells was resuspended using a PCR mixture containing 1 × Thermopol (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTP, 1 µM primer **ODN 5.12**(5'-GCTAGTTATTGCTACGCGGCGCAAACCGCCTCTCCCCGCGCGTTGG-3') and 1 µM primer **ODN 5.11** (5'-

GCTAGTTATTGCTACGCGGGTCGACTCTAGAAGATCTATCA -3') in a final volume of 600 µl. These above mentioned whole cell PCR mixtures were named Solution Phase Stoffel Fragment (SP Stoffel) and Solution Phase Taq (SP Taq). For controls of each solution phase reaction, an aliquot of each SP Stoffel (100 µl) and SP Taq (100 µl) were each placed in their own PCR tube. A mixture of the two solution phases was made by thoroughly mixing SP Stoffel (100 µl) and SP Taq (100 µl). The mixture of the two cell types (200 µl) was divided equally into two PCR tubes. Table 5.4 summarizes the solution phase whole cell controls.

Table 5.4 Solution phase whole cell PCR controls.

	Stoffel Fragment	Taq	Mixture of cells expressing Taq or Stoffel Fragment
SP Stoffel (µl)	100	0	100
SP Taq (µl)	0	100	100
Total volume	100	100	200

SP Stoffel (200 µl) and SP Taq (200 µl) were individually emulsified by making five additions (40 µl each) to CSR oil phase (400 µl) over 2 minutes followed by an additional 18 minutes of stirring following the protocol described in section 3.5.7.1. To produce a reaction mixture that tests the efficiency of whole cell isolation, one hundred microlitres from each SP Stoffel and SP Taq mixture was thoroughly mixed together. The intimate mixture of the two solution phases was added in 40 µl aliquots into stirring CSR oil phase over 2 minutes. Stirring was allowed to continue for another 18 minutes. The compositions of the emulsion reactions are shown in Table 5.5.

Table 5.5 Compositions of the emulsion phase whole cell isolation controls.

	Stoffel Fragment	Taq	Mixture of cells expressing Taq or Stoffel Fragment
SP Stoffel (μl)	200	0	100
SP Taq (μl)	0	200	100
CSR oil phase (μl)	400	400	400
Total emulsion (μl)	600	600	600

All PCR mixtures listed in Tables 5.3 – 5.4 were thermocycled 20 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/45 seconds). Amplicons from the emulsions were isolated and purified using the same method of breaking down the emulsion described in 5.3.7.2. Products were analyzed on a 2 % agarose gel.

5.3.8 Shuffled Taq library

Plasmid pTaq was used for shuffling. Five individual digestion reactions contained Invitrogen's 1x PCR Buffer minus Mg^{2+} (20 mM Tris-HCl (pH 8.4), 50 mM KCl), 10 mM $MnCl_2$ (USB) and ~4 μg plasmid pTaq in a total volume of 50 μl. The mixture was cooled to 15 °C in a water bath and 0.23 U of DNase I was added. Aliquots of digestion reaction (10 μl) were stopped after 1, 2, 3, 4, and 5 minutes with the addition of 0.5 μM EDTA (1 μl). The reaction progress was visualized on a 2 % agarose gel stained with ethidium bromide. From the 5 minute reaction, the fragments were run on a 2 % agarose gel and fragments under 100 bp were excised and purified using Qiagen's Gel Purification kit. Fragments were eluted from the spin column of the kit using 50 μl of the included EB solution. The reassembly reaction contained 1x Pfx polymerase Buffer (composition was not released by Invitrogen), 200 μM dNTPs, plasmid fragments (10 μl), and 2.5 U Pfx polymerase in a total volume of 20 μl. The reassembly was thermocycled using the program: 94 °C/ 3 minutes, 20 × (94 °C/ 1 minute, 55 °C/ 1 minute, 72 °C/ 1 minute), 72 °C/ 10 minutes. The Pfx-mediated reassembly reaction was used to provide a template for the amplification of the shuffled Taq gene. The

amplification reaction of the reassembly product contains 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, Pfx reassembly reaction (1µl), 1 µM primer **ODN 5.13** (5'-CGCAAACCGCCTCTCCCCGCGCGTTGG-3'), 1 µM primer **ODN 5.14** (5'-GCTAGTTATTGCTCAGCGGAGATCTATCACTCCTTGGC-3') and 5 U Taq in a final volume of 50 µl. The reassembly was thermocycled using the program: 94 °C/ 3 minutes, 20 × (94 °C/ 1 minute, 55 °C/ 1 minute, 72 °C/ 1 minute), 72 °C/ 10 minutes. The shuffled gene product was digested with EcoRI and Bgl II. The digested gene was then ligated to the pUC18 vector.

5.3.9 Construction of the Taq library via cassette mutagenesis

The active site of Taq contains a single polypeptide region referred to as Motif A (Figure 1.14). (Patel, Kawate *et al.*, 2001; Patel & Loeb, 2000a) The amino acids at positions 605-617 were chosen for mutagenesis. The active site library was constructed according to protocol.(Patel & Loeb, 2000a) Briefly, oligonucleotide **ODN 5.15** was chemically synthesized with the bases encoding for residues 605-617 partially mutagenized using cassette mutagenesis. Active site library was constructed by annealing 50 pmol **ODN 5.15** (5'-GCCGAGGAGGGGTGGctattggtggccctggactatagccagatagagctcaggGTGCTGGCCCACC TCTC-3') (where lowercase letters represent the wild-type base and the level of mutation is 88 % wild-type, 4 % each of the other 3 bases) which contains a partially degenerate region encoding for the active site of the polymerase to 50 pmol **ODN 5.16** (5'-ATCCAGCTGGCGGTCTCCGTGTGGATGTCCCGCCCCTCCTGGAAGACCCGGA TCAGGTTCTCGTCGCCGGAGAGGTGGGCCAGCAC-3'). The last 17 bases at the 3' ends of the two oligonucleotides are complementary. Oligonucleotides **ODN 5.15** and **ODN 5.16** were added to a reaction mixture containing 1x Thermopol (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, and 2.5 U Taq in a final volume of 100 µl and thermocycled 5 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds) to extend each strand and produce dsDNA.

Ten microlitres of the extension reaction was then PCR amplified using 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, 1 µM primers **ODN 5.17** (5'-GAACATCCCCGTACGCACCCCGCTTGGGCAGAGGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGG-3') and **ODN 5.18** (5'-GGGGTCCACGGCCTCCCGCGGGACGCCGAACATCCAGCTGGCGGTC-3') in a total volume of 100 µl to introduce restriction sites Bam HI and Sac II (shown underlined). This product was digested using Bam HI and Sac II and cloned into the Bam HI and Sac II of pTaq. The library was introduced into *E. coli* cells in preparation for expression following the protocol described in section 5.3.6.

5.3.10 Active site library preselection using unmodified nucleoside triphosphates

After 12 hours of incubation at 37 °C with shaking at 225 rpm, the cells were spun down (13 000 rpm, 1 minute), the media decanted and cell pellet resuspended in water. After resuspension in water, the OD₆₀₀ was measured and the suspension of cells was diluted using water to 1.0 OD. Two hundred microlitres of this diluted suspension was centrifuged and the supernatant was decanted. At a average density of 10⁸ cells for every 1 ml of a 0.1 OD suspension, 200 µl of a 1.0 OD suspension will contain 2 × 10⁸ cells. Limitations from cloning limit library sizes to a maximum of 10⁶. Therefore, each clone should be represented at least a hundred times. The cell pellet containing the library of mutant polymerases was resuspended in a PCR mixture containing 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8) 1 µM **ODN 5.19** (5'-CGCAAACCGCCTCTCCCCGCGCGTTGG-3'), 1 µM **ODN 5.4** (5'-GTCGACTCTAGAAGATCTATCA-3'), 200 µM dNTPs, RNaseA (5 ng) and water to a total volume of 200 µl to afford the CSR reaction mixture. Oligonucleotides **ODN 5.19** and **ODN 5.4** hybridize to regions on the polymerase gene that flank the mutagenized region.

CSR oil phase (4.5 % v/v Span 80 (Fluka), 0.4 % Tween 80 v/v (Sigma), 0.05 % v/v Triton X-100 (Sigma) in light mineral oil (Sigma)) (400 µl) was added to a 5 ml

Corning Cryogenic Vial (2051) along with an 8 × 3 mm stir bar with a pivot ring. The oil phase was stirred on a VWR stir plate at the maximum setting (speed setting 10) and the CSR reaction mixture was added in 40 µl aliquots every 30 seconds. After the last addition, the emulsion was allowed to stir for an additional 18 minutes. The emulsion (600 µl) was equally divided into six thin-walled PCR tubes (Axygen). The emulsions were subjected to thermocycling 30 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds).

After thermocycling, the six tubes were centrifuged (13 000 rpm, 15 minutes) to concentrate the droplets. The clear oil phase was decanted and the remaining concentrated droplets were combined into one Eppendorf tube using water-saturated diethyl ether (1 ml) to break the emulsion and facilitate the transfer. The aqueous layer was washed with another portion of water-saturated diethyl ether (1 ml). The extraction was centrifuged for 2 minutes and the organic layer was decanted. The aqueous layer was incubated on a 65 °C heat block to evaporate the residual diethyl ether. The aqueous layer was split into two equal portions and the DNA was precipitated with the addition of ethanol (1 ml) to each portion. The two samples were agitated on a vortexer briefly and centrifuged for 15 minutes. The ethanol was decanted leaving behind a solid pellet. Any residual liquid was evaporated by incubating the pellet in a 65 °C heat block until the pellet appeared dry. The DNA was recombined by dissolving both pellets in water (50 µl total volume). Qiagen's Nucleotide Removal Kit was used according to instruction to remove impurities from the DNA.

5.3.11 First amplification of preselection library products

Degradation of the selection primers was critical. To do this ExoSAP-IT (USB) was used. ExoSAP-IT is a combination of Exonuclease I and Shrimp Alkaline Phosphatase that degrades both single-stranded linear DNA (primers) and excess dNTPs. The byproducts of the degradation are nucleosides and phosphate. A reaction mixture was setup with the following components: 4.5 µl of CSR product, 0.5 µl of 5x Thermopol Buffer and 2 µl of ExoSAP-IT. The reaction was incubated in the thermocycler 1 × (37

°C/ 15 minutes, 80 °C/ 15 minutes). After allowing the reaction to cool to room temperature, 0.5 µl of 5x Thermopol Buffer (100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 0.1 % Triton X-100, 200 mM Tris-HCl pH 8.8), 0.1 µl dNTPs, primer **ODN 5.20** (5'- TAATACGACTCACTATAGGG -3') and primer **ODN 5.21** (5'- GCTAGTTATTGCTCAGCGG-3'), and of 0.4 µl Vent (exo-) (0.8 U) was added. Oligonucleotides **ODN 5.20** and **ODN 5.21** hybridize to regions containing sequences exclusive to the selection product so that only selection product and not plasmid DNA will be amplified. The reaction was thermocycled 30 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). The amplified preselection product was cloned into the Bam HI and Sac II site of pTaq.

5.3.12 Preselected library activity

The plasmids containing the preselected library resulting from the ligation described at the end of section 5.3.10 were dialyzed on 1 % agarose for at least one hour followed by electroporation using Tritech's Bactozapper and Invitrogen's *E. coli* DH10B Electromax cells (20 µl). The transfected cells were added to SOC media (980 µl) and incubated (37 °C) with shaking (225 rpm) for one hour. Fifty microlitres of the transfected cells in SOC media were spread on LB agar (100 mg/l ampicillin). The agar plate was incubated overnight at 37 °C. Ten colonies, each putatively containing a unique mutant polymerase, were picked at random and used to inoculate TB media (2 ml). Cells were grown at 37 °C overnight (12 hours) with shaking @ 225 rpm. Each of the cell cultures (50 µl) was centrifuged for 1 minute (13 000rpm) and the media was decanted. The cell pellets were resuspended in the following reaction mixture: 1x Thermopol, 5 mM MgSO₄, 1 µM **ODN 5.10** (5'- GCTAGTTATTGCTCAGCGGAGAACATCCCCGTACGCAC -3'), 1 µM **ODN 5.8** (5'- GCTAGTTATTGCTCAGCGGTAAGGGATGGCTAGCTCCTGGGA-3'), 200 µM dNTPs, RNaseA (5 ng). The reaction mixtures were thermocycled 20 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). All PCR products from the activity test of the ten mutant polymerases were visualized on a 2 % agarose gel.

5.3.13 O-helix library construction

The O-helix library was constructed on a protocol based on Suzuki *et al.* (Suzuki *et al.*, 1996). The O-helix library was constructed by first annealing primer 50 pmol of **ODN 5.22** (5'-CGGGAGGCCGTGGACCCCCTGATGcgccgggcgcggaagaccatcaacttcggggtcctctacGGCATGTCGGCCCACCG-3') (where lowercase letters represent wild-type base and the level of mutation is 91 % wild-type, 3 % each of the other 3 bases) to 50 pmol of primer **ODN 5.23** (5'-TGGCTAGCTCCTGGGAGAGGCGGTGGGCCGACATGCC-3') in a reaction mixture that contained 1x Thermopol (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, and 2.5 U Taq in a final volume of 100 µl and thermocycled 5 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). A portion (10 µl) of this reaction mixture was then used to provide a template for an additional amplification using 1 µM **ODN 2.5** (5'-TTCGGCGTCCCGCGGGAGGCGGTGCACCCCCT-3') and 1 µM **ODN 5.24** (5'-GTAAGGGATGGCTAGCTCCTGGGA-3'), 1x Thermopol (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 200 µM dNTPs, and 2.5 U Taq in a final volume of 100 µl and thermocycled. The PCR product was purified using Qiagen's PCR Purification Kit. The product was restriction enzyme digested using Sac II and Nhe I and cloned into the plasmids containing the preselected active site library.

5.3.14 CSR with modified nucleoside triphosphates

The preselected active site library combined with the O-helix library (OApre) was expressed for 12 hours and cooled on ice. An aliquot of cells (1 ml) was centrifuged at 13000 rpm for 1 minute and the media was decanted. The cells were resuspended in 500 µl water. The OD₆₀₀ of the suspension was measured and the suspension was diluted to an OD₆₀₀ of 1.0. Five aliquots (200 µl) were each placed in a fresh tube, centrifuged at 13000 rpm for 1 minute and the water was decanted. PCR mixtures that contained modified nucleotides and primers that hybridize to the regions that flank the section of DNA that encodes for the combined active site and O-helix regions were prepared. Each

of the pellets was resuspended using one of the following CSR reactions: 1x Thermopol Buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM MgSO₄, 0.1 % Triton X-100, 20 mM Tris-HCl pH 8.8), 5 mM MgSO₄, 1 µM **ODN 5.10** (5'-GCTAGTTATTGCTCAGCGGAGAACATCCCCGTACGCAC -3'), 1 µM **ODN 5.8** (5'-GCTAGTTATTGCTCAGCGGTAAGGGATGGCTAGCTCCTGGGA-3'), RNaseA (5 ng), and one of the following deoxyribonucleoside triphosphate combinations:

- (1) Negative control: no nucleoside triphosphates
- (2) FANA monomer: 200µM dATP, dCTP, dGTP, and 2'F-araTTP
- (3) dC^{aa}TP: 200 µM dATP, dGTP. dTTP and dC^{aa}TP
- (4) dA^{imm}TP: 200 µM dTTP, dGTP. dCTP and dA^{imm}TP
- (5) Positive control: 200 µM dNTPs

and water added up to a final volume of 200 µl.

The CSR reactions were emulsified individually by adding each of the five CSR reaction mixtures in five equal aliquots (40 µl) into of rapidly stirring CSR oil phase (400 µl) every 30 second according to the protocol described in section 5.3.7.1. After the final additions, the emulsions were allowed to continue stirring for 18 minutes. Each of the CSR emulsions (600 µl) was divided into aliquots (100 µl) in six PCR tubes. The reactions were thermocycled 94 °C/5 minutes, 30 × (94 °C/1 minute, 55 °C/1 minute, 72 °C/5 minute) followed by a 72 °C/5 minute final extension.

5.3.15 First amplification of CSR products

The products of the CSR were collected according to the protocol described in section 5.3.7. To amplify only the selection products and not the plasmid DNA, 10x Thermopol Buffer (100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 0.1 % Triton X-100, 200 mM Tris-HCl pH 8.8) (0.5 µl) and ExoSAP-It (2 µl) were added to the each of the products from the CSR reactions (4.5 µl). The reaction was incubated in a thermocycler at 37 °C for 15 minutes followed by 80 °C for 15 minutes to deactivate the enzymes. Additional 10x Thermopol Buffer (100 mM KCl, 100 mM (NH₄)₂SO₄, 20 mM MgSO₄, 0.1 % Triton X-100, 200 mM Tris-HCl pH 8.8) (0.5 µl), 10 mM dNTPs (0.2 µl), primer T7 term (5'-GCTAGTTATTGCTCAGCGG-3') (0.2 µl), 0.4 U Vent (exo-) and

DEPC-treated water (1.5 µl) was added. The reaction was incubated at 94 °C for 3 minutes then lowered to 65 °C. 4 U Bst polymerase (large fragment) was added to reaction which was then incubated for 30 minutes. The reactions were subsequently thermocycled 20 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). The products from each first amplification reaction were agarose gel purified and used as templates (1 µl) for a subsequent using Taq polymerase under standard PCR conditions. The amplified first amplification products were cloned into pTaq to produce libraries for the second round of selection.

5.3.16 Generation 1 activities

The new generation 1 libraries were expressed according to the protocol described in section 5.3.6. Before applying the polymerase libraries to the second round of selection, the successful expression and presence of activity of the libraries must be verified. Cells from each of the expressed libraries were collected from the media (50 µl) and centrifuged for 1 minute at 13000 rpm. The media was decanted and the cells were resuspended in the following reaction mixture: 1x Thermopol, 5 mM MgSO₄, 1 µM **ODN 5.10** (5'-GCTAGTTATTGCTCAGCGGAGAACATCCCCGTACGCAC -3'), 1 µM **ODN 5.8** (5'-GCTAGTTATTGCTCAGCGGTAAGGGATGGCTAGCTCCTGGGA-3'), 200 µM dNTPs, RNaseA (5 ng). The reaction mixtures amplify the region of DNA that encodes for both the active site and the O-helix (same PCR product as the selection) were thermocycled 20 × (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/30 seconds). PCR products were visualized on a 2 % agarose gel.

5.3.17 CSR using modified nucleoside triphosphates: Round 2

The library generated from the selection using 2'-F-araTTP in the first round, herein referred to as the FT-library, will be used in the second round of selection, which will exclusively use 2'-F-araTTP. The other two libraries selected using either dC^{aa}-TP (dC^{aa}-library) or dA^{imm}-TP (dA^{imm}-library) were also used for the second round of selection along with their respective modified nucleoside triphosphates exclusively as

well. Round 2 was carried out according to the protocols described in sections 5.3.10 and 5.3.11 except each of the three libraries only used the unnatural nucleoside triphosphate they were exclusively named after. The positive control library continued to use unmodified dNTPs. PCR products were visualized on a 2 % agarose gel.

5.4 Results

5.4.1 Expression of wild-type Taq and Stoffel Fragment

Successful evolutions of both the full Taq polymerase and the smaller 5'-3' exonuclease domain-deleted Stoffel Fragment have been reported.(Brakmann, 2005) It is beneficial then to have both the plasmid that expresses Taq and the plasmid that expresses Stoffel Fragment. Our initial experiments will be the directed evolution of Taq polymerase. However, it is a good idea to have as many different polymerases as possible on hand to test out certain conditions or even drastically change the polymerase of the selection. Many of the selection techniques involve the differentiation between polymerases according to their activities. Two key differences between these two polymerases that we hope to exploit are the difference in size of their gene and the difference in incorporation rate. These differences will become critical in a control experiment that shows that CSR can selectively enrich a gene pool in gene encoding polymerases with a desired activity.

The wild-type versions of both Taq polymerase and Stoffel Fragment were expressed and isolated according to literature protocol.(Desai & Pfaffle, 1995) Taq appeared as a 94 kDa protein and Stoffel Fragment appeared as a 61 kDa protein (Figure 5.2). As mentioned in the literature protocol, full length Taq can degrade to a ~61 kDa protein that was named the Stoffel Fragment. Faint bands in the Lane 3 having roughly the same electrophoretic mobility as Stoffel Fragment are assigned as degradation products that retain their thermostability and are not heat denatured.

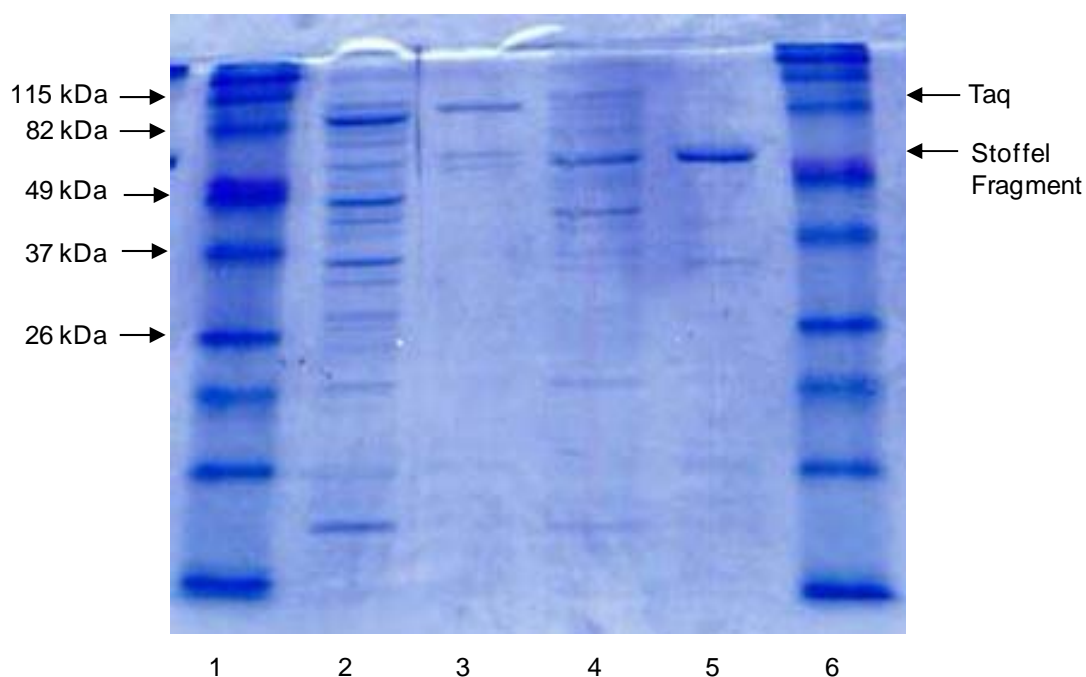


Figure 5.2 SDS polyacrylamide gel depicting Taq and Stoffel expression. Lane 1 + 6: Invitrogen Prestained Protein Markers. Lane 2: Cell lysate of *E. coli* expressing Taq polymerase. Lane 3: Heat-treated cell lysate containing full length Taq polymerase. Lane 4: Cell lysate of *E. coli* expressing Stoffel Fragment. Lane 5: Heat-treated lysate containing Stoffel Fragment. Approximate masses of the protein ladder are provided on the right.

5.4.2 Site-directed mutagenesis

One of the modified nucleoside triphosphates used in this work for the directed evolution of Taq polymerase is 2'-F-araTTP. Selecting for a mutant polymerase that can incorporate a sugar-modified nucleotide could be the simplest polymerase selection to perform. The key interaction that allows the DNA polymerase to distinguish its native substrate, dNTP, for a structurally similar NTP occurs between the incoming 2' substituent of the incoming nucleoside triphosphate or deoxynucleoside triphosphate and amino acid E615. Polymerases evolved to incorporate the ribonucleotides frequently have the E615G mutation.(Ong *et al.*, 2006; Xia *et al.*, 2002) While there's no guarantee that this mutation will improve the incorporation 2'-F-araTTP, it should be tested. Even if

this single point mutant should fail, it was suggested that nearby compensatory mutations can restore activity.

The E615G mutation was introduced using a variation on Stratagene's Quikchange method (Figure 5.3). Complementary primers containing the necessary base mismatches to introduce the mutation were used in this site-directed mutagenesis. Plasmid pTaq was used as template for the linear amplification. The mutagenic primers **ODN 5.5** and **ODN 5.6** are annealed to the plasmid and extension is carried out over the entire circular plasmid stopping only when the 5' end of the primer is reached. As both strands are being amplified and the product strands hybridize, the resulting product is a double-nicked circular plasmid. The template pTaq having originated from bacterial cells is N-6 methylated and susceptible to restriction enzyme digestion by Dpn I. (Lacks, Mannarelli, Springhorn, Greenberg, & De La Campa, 1987) With the original template being selectively digested, what remains is the new plasmid containing the base mutation that introduces the E615G mutation. This remaining double-nicked circular plasmid can be transformed into the cell.

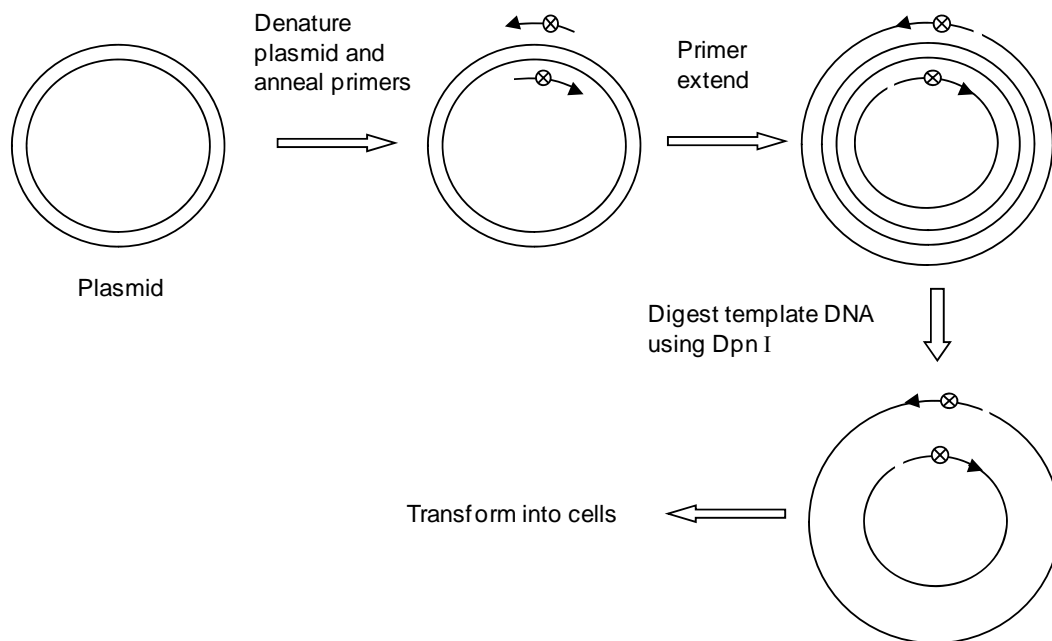


Figure 5.3 General scheme depicting site-directed mutagenesis based on Stratagene's Quikchange Kit.

Twenty cycles were used to linearly amplify the doubly-nicked circular plasmid containing the E615G nucleotide mutation (Figure 5.4). Dpn I was used to degrade the original plasmid. Transformation of the new plasmid afforded colonies. These colonies were used to inoculate media, and select plasmids were sent for sequencing. A plasmid was obtained that contains the appropriate mutation to encode for an E615G mutant. Methods used in the expression of Taq and Stoffel Fragment, however, failed to yield detectable amounts of polymerase. At the time of the writing of this thesis, expression of the E615G mutant polymerase had not been successfully detected.

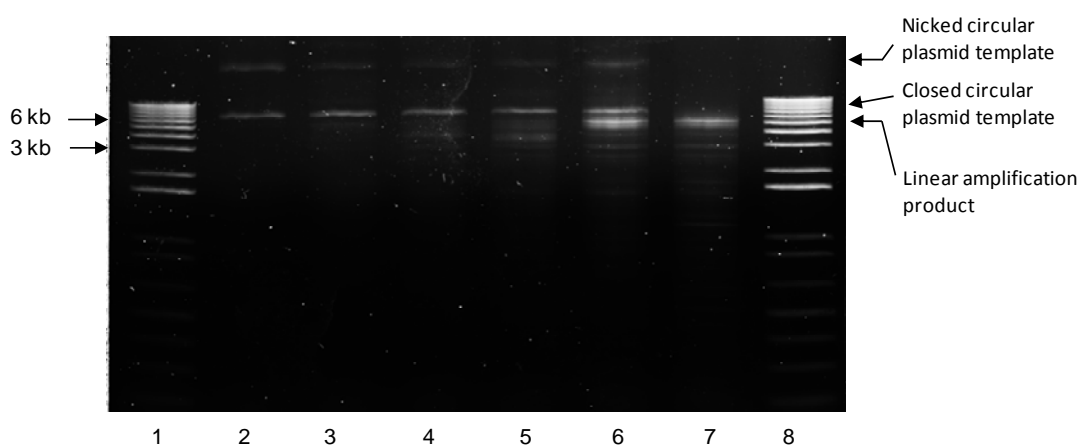


Figure 5.4 Cycle-dependent production site-directed mutant plasmid. Lane 1: Invitrogen's 1 kb Plus Molecular Weight Ladder. Lanes 2-6 show cycles 4, 8, 12, 16, 20. Lane 7 shows Dpn I-treated amplicon.

5.4.3 CSR controls

5.4.3.1 Reverse micelle size in relation to *E. coli*

The size of the reverse micelle is critical to this directed evolution for the following reasons. If the reverse micelle is too large, this could be problematic as there will be fewer total reverse micelles than in a CSR containing smaller reverse micelles the concentration of the molecules within the reverse micelle will decrease due to the large volume. The easiest method of assuring reverse micelles were appropriately formed is to visualize them using a light microscope as demonstrated by Holliger and coworkers.(Ghadessy *et al.*, 2001) The size of the reverse micelle is of interest as well as

the size of the reverse micelle in relation to the encapsulated *E. coli* cells. While, independently, the reverse micelles and the *E. coli* can be visualized easily, it is very difficult to see the *E. coli* within the reverse micelle. To address this, enhanced green fluorescent protein (EGFP) was used to improve visualization of the cell within the reverse micelle.

Reverse micelles are formed by stirring an aqueous solution with an oil phase containing surfactants (Figure 5.5). Reverse micelle size was calibrated accordingly in relation to the size of the *E. coli* expressing EGFP. The fluorescence of the *E. coli* cells can be visualized using UV light. Using UV light and just enough white light so that cell fluorescence is not washed out, the *E. coli* cells can be visualized within the reverse micelle (Figure 5.6). Each of the micelles contains one or no cells. Emulsification of individual *E. coli* cells expressing EGFP allowed visualization of both the whole cell and the interface between the aqueous phase and oil phase. The size of the reverse micelles in relation to the *E. coli* cell was acceptable. The average reverse micelles diameter appears to be $\sim 8\text{ }\mu\text{m}$, which gives an average volume of $\sim 0.3\text{ pl}$. The variability in diameter amongst the reverse micelles appears to be between $5\text{-}10\text{ }\mu\text{m}$, which is also acceptable for the isolation of individual cells. A more precise measurement of the variability between cells was not performed, but could be performed in the future using dynamic light scattering. Also, it appeared that the cells were individually isolated and a few empty reverse micelles could also be seen. As the reverse micelles have passed a visual inspection, chemical inspection must also be performed.

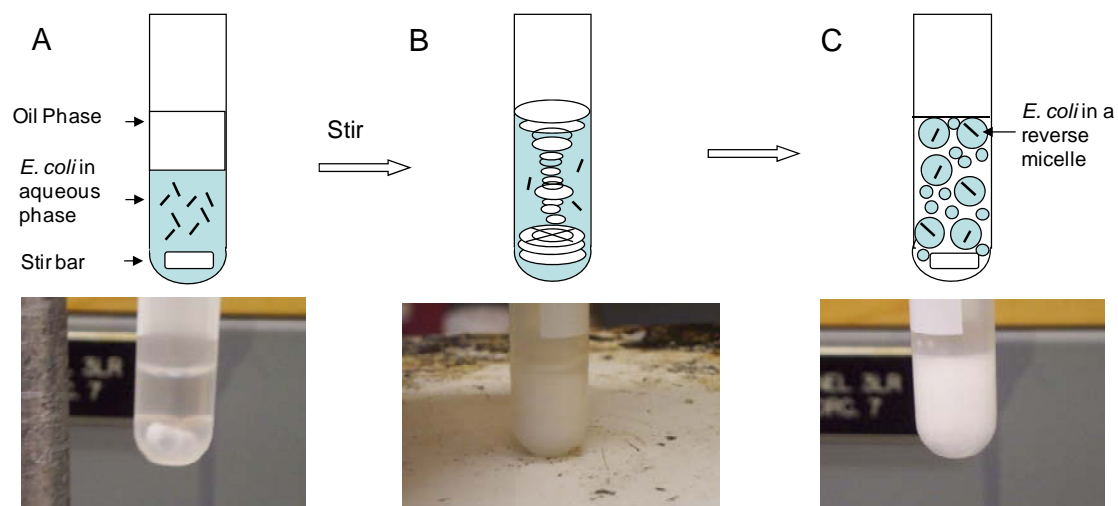


Figure 5.5 Formation of reverse micelles. **A** A surfactant-containing oil phase sits on top of a suspension of cells. Included in the two phases is a stir bar. **B** The two phases are stirred. **C** Because of the added surfactants, the two phases form an emulsion consisting of water-in-oil reverse micelles.

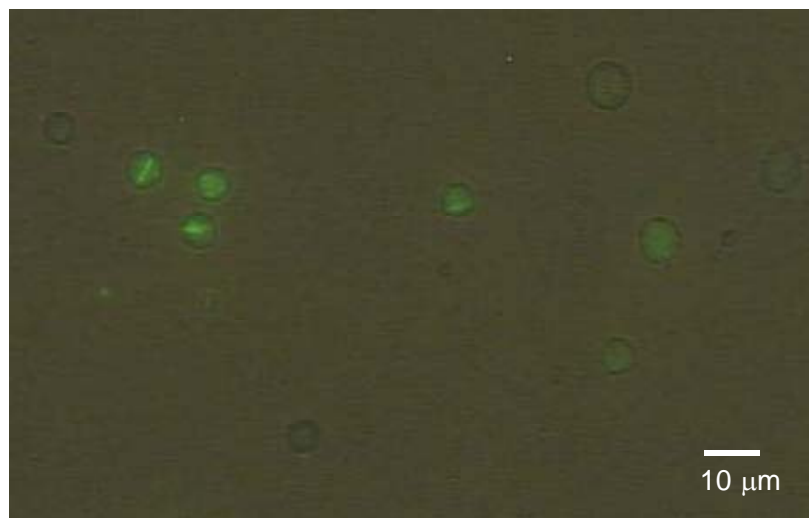


Figure 5.6 Sixty times magnified view of *E. coli* cells expressing EGFP. Cells were illuminated using white light to visualize the borders of the aqueous compartment. Cells were also irradiated with UV light to induce fluorescence and to visualize the *E. coli* cells. Aqueous compartments contain one or no *E. coli* cells.

5.4.3.2 Determination of cross-reaction between droplets

The previous section confirmed the physical size and stability of the reverse micelles before and after thermocycling using visual means. This is not sufficient proof to guarantee the isolation of CSR component as the chemicals, primers, polymerases cannot be inspected visually. If chemicals and biomolecules can pass from one reverse micelle to the other, the spCSR will fail. Therefore, a different, more sensitive, test must subsequently be done. Ghadessey *et al.* showed that amplification of a specific target DNA can be suppressed if the PCR components were isolated in reverse micelles prior to thermocycling. (Ghadessey *et al.*, 2001) A PCR reaction was performed to simulate failed isolation of chemicals and biomolecules by the reverse micelles. To do this, the PCR reaction, along with its controls, was made to simulate an absolute collapse of the reverse micelles.

Two PCR amplification reactions named Reaction A and Reaction B were used to determine the amount of transport of reaction components between droplets. Reaction A has a short target but no polymerase added and Reaction B has a long target and contains commercial polymerase (Figure 5.7). As expected, Reaction A that was lacking polymerase did not produce product (Figure 5.8). Reaction B that did contain polymerase did produce the expected 0.8 kb amplicon. The 1:1 mixture of Reaction A and Reaction B produced exclusively the shorter 160 bp product.

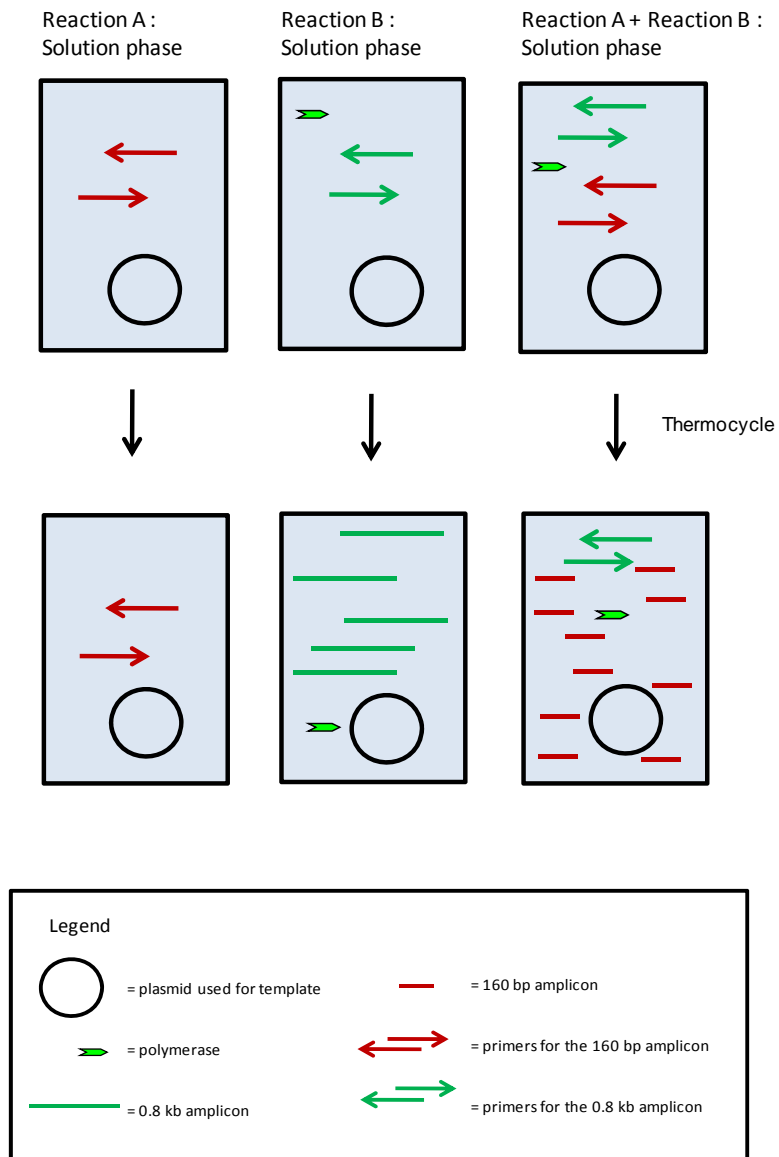


Figure 5.7 Solution phase PCRs of Reaction A, Reaction B and a homogeneous mixture of the two reactions. In the mixture, the shorter product is exclusively produced.

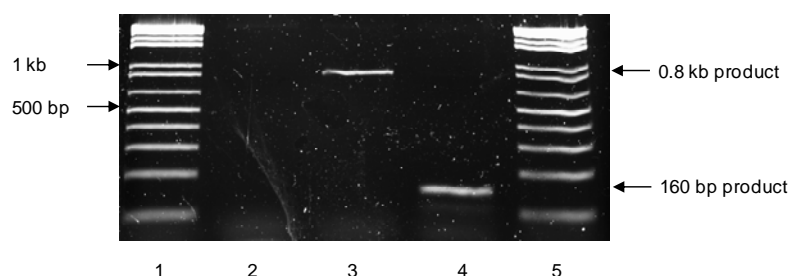


Figure 5.8 Compartmentalized self-replication of targets of differing size. Two PCRs were prepared separately with each amplifying a target of different lengths. Prior to thermocycling the reactions were mixed; then emulsified. Lane 1 + 5: Invitrogen's 1kb Plus Molecular Weight Marker. Lane 2: Reaction A targeting a 160 bp fragment with no polymerase added. Lane 3: Reaction B targeting a 0.8 kb fragment with polymerase added. Lane 4: 1:1 mixture of Reaction A and Reaction B.

The solution phase mixture of the two reactions was meant to demonstrate the outcome resulting from the failure by the reverse micelles to prevent leakage of chemicals and biomolecules. The primers used to produce the 0.8 kb amplicon flank the primer binding regions for the primers that produce the 160 bp amplicon. Thus, both the 0.8 kb amplicon and the 160 bp amplicon can serve as a template for the amplification of the 160 bp amplicon in subsequent cycles. As the primers used to amplify the 160 bp target can hybridize to the 0.8 kb amplicon, this will hinder the production of additional 0.8 kb amplicon in that cycle. The size and the nature of the 160 bp target make it far more easily amplified over the 0.8 kb amplicon. The product of the solution phase mixture is exclusively the shorter 160 bp amplicon. Therefore, the presence of the 160 bp amplicon will serve as an indicator that leakage of chemicals or biomolecules has occurred. At worst, the reverse micelles themselves may have collapsed. To test the integrity of the reverse micelles, Reaction A and Reaction B were individually emulsified using the same method shown in Figure 5.5. Emulsified Reaction A and emulsified Reaction B were used as controls to ensure that PCR occurs as anticipated within the emulsion (Figure 5.9). Equal volumes of the two emulsions were then thoroughly mixed and thermocycled along with the two emulsion controls. Emulsion was collapsed by centrifuging the emulsions, decanting excess oil phase, and adding water-saturated diethyl ether to breakdown the emulsion. The diethyl ether was removed and the aqueous phase was extracted once more using water-saturated diethyl ether. Remaining ethyl was removed by briefly incubating the aqueous reaction in a 65 °C heat block until bubbling

subsides. The amplicon products of the aqueous layers were purified using Qiagen's Nucleotide Clean Up Kit. Products were visualized on a 2 % agarose gel.

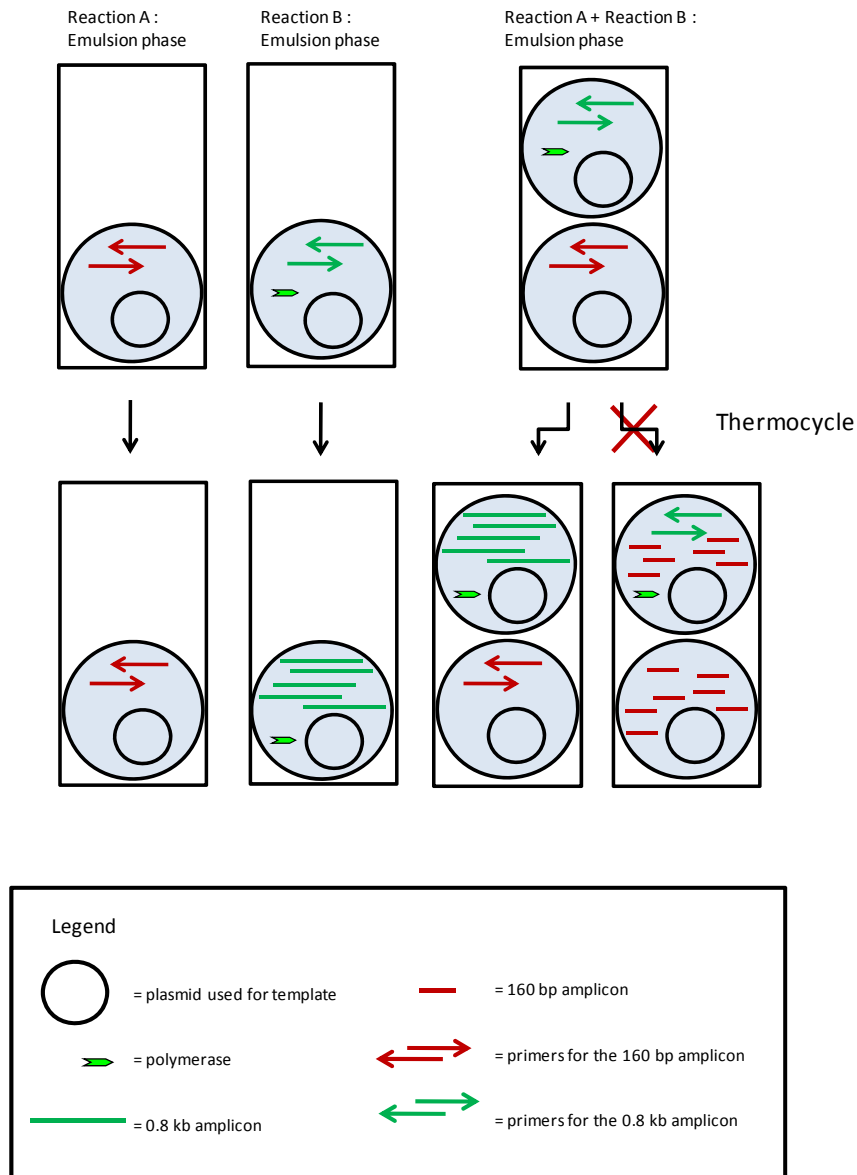


Figure 5.9 Scheme showing the possible products formed from stable or unstable reverse micelles. Reaction B produced a 0.8 kb amplicon represented by the green lines. Reaction A contains no polymerase and fails to yield any amplicon. If the mixture of the two emulsions stays intact and no components can cross the oil phase barrier, only the larger 0.8 kb amplicon will be produced. If there is exchange of components, either primers or polymerase, the shorter 160 bp amplicon will be produced.

The controls showed the same products as the solution phase (Figure 5.10). Therefore, it can be concluded that emulsification did not inhibit amplification as the control products were nearly identical to the products identified in the solution phase PCR. The emulsified Reaction A and Reaction B were then mixed in a one to one ratio and thermocycled showed that the 0.8 kb target was exclusively produced. An insignificant amount of the 160 bp amplicon can be detected. This indicates that the reverse micelles did not collapse during thermocycling and small molecules such as primers, amplicon or polymerases did not diffuse between intact reverse micelles to any significant amount. The isolation of chemicals and biomolecules using reverse micelles was found to be acceptable.

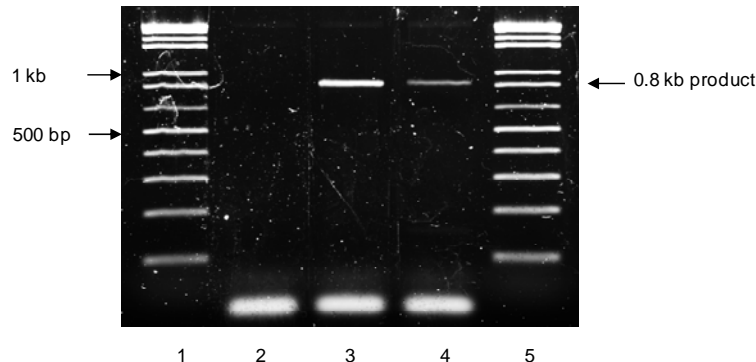


Figure 5.10 Compartmentalized self-replication of targets of differing size. Two PCRs were prepared separately with each amplifying a target of different lengths. Prior to thermocycling the reactions were emulsified and then mixed. Lane 1 + 5: Invitrogen's 1kb Plus Molecular Weight Marker. Lane 2: PCR targeting a 160 bp fragment with no polymerase added. Lane 3: PCR targeting a 0.8 kb fragment with polymerase added. Lane 4: 1:1 mixture of the two PCRs.

5.4.3.3 Isolation of individual cells through emulsification

In section 5.4.3.1, the formation of reverse micelles of appropriate size to isolate individual *E. coli* cells was confirmed. In the preceding section, the integrity of those reverse micelles and their ability to isolate their components even through thermocycling was also confirmed. In this section, the findings of both of those experiments must simultaneously hold true. The process of emulsification itself was used to isolate and

encapsulate individual *E. coli* cells. After encapsulation, the components within the reverse micelle must remain within the reverse micelles throughout the thermocycling. Proper controls were produced to facilitate detection of the failure of these processes. One final requirement is the isolation of individual cells using the emulsification process itself. If the cells cannot be isolated using the emulsification during the polymerase selection, individual polymerase mutants will not be isolated and genes encoding for both active and inactive mutants will be amplified by the active mutants. If multiple cells are present in one reverse micelle, the result would be that the gene pool will not be as completely enriched for the genes encoding for active mutants. However over several generations of selection, even genetic contamination from other species in the micelle should ultimately be weeded out in favour of the selected activity.

Ghadessy *et al.* used a system of whole cell PCRs that used cells expressing either Taq or Stoffel Fragment to verify the isolation of individual cells by emulsification.(Ghadessy *et al.*, 2001) In this section, we verified that the emulsion that we produce effectively isolates individual cells. To stay consistent with Ghadessy's work, we also used *E. coli* cells expressing Taq and *E. coli* cells expressing Stoffel Fragment were used. This experiment examines whether or not, in a selection under CSR conditions, one of two very similar polymerases can amplify its own gene and enrich the gene pool in its own gene. Taq, of course, is meant to represent an active mutant polymerase where as Stoffel Fragment is representing an inactive or low activity mutant polymerase. In actuality, both polymerases are active, but under a certain selection pressure used for this control experiment, Taq managed to differentiate itself due to its high nucleotide incorporation efficiency.

The potential for cross contamination due to neighbouring cells being within the same reverse micelle was examined by using two different polymerases to amplify their own gene under CSR conditions. Polymerases and their respective genes used in these tests were Taq and Stoffel Fragment. The first test is to differentiate the two polymerases by their catalytic ability without an emulsion i.e. solution phase. Taq is the larger of the two polymerases and is encoded by the larger 2.5 kb gene. Stoffel Fragment has a lower rate of nucleotide incorporation and is encoded by the shorter 1.6 kb gene. Amplification using short extension times in the PCR cycling allows for the selective amplification of

the longer Taq gene target by Taq polymerase. Stoffel Fragment, while its Stoffel Fragment gene target is much shorter, is unable to exponentially amplify its own gene using 45 second extension times as it is slower than Taq polymerase at elongation. To show that both polymerases are being expressed, a pair of primers was used to produce a short 0.8 kb amplicon. This target size is short enough that the Stoffel Fragment can amplify the target as well as Taq polymerase (Figure 5.11). However, when a pair of primers is used such that the polymerase is required to amplify its own gene, Taq succeeds and Stoffel Fragment fails. To be as consistent as possible between the two systems, the same pair of primers is used for both Taq and Stoffel Fragment, and this pair of primers hybridizes on regions of pUC18 that flank either the Taq or Stoffel Fragment. Mixing followed by thermocycling of the two individual solution phase reactions in a 1:1 ratio results in both Taq and Stoffel Fragment genes being amplified, presumably by Taq, with a higher amount of the shorter Stoffel Fragment gene being produced.

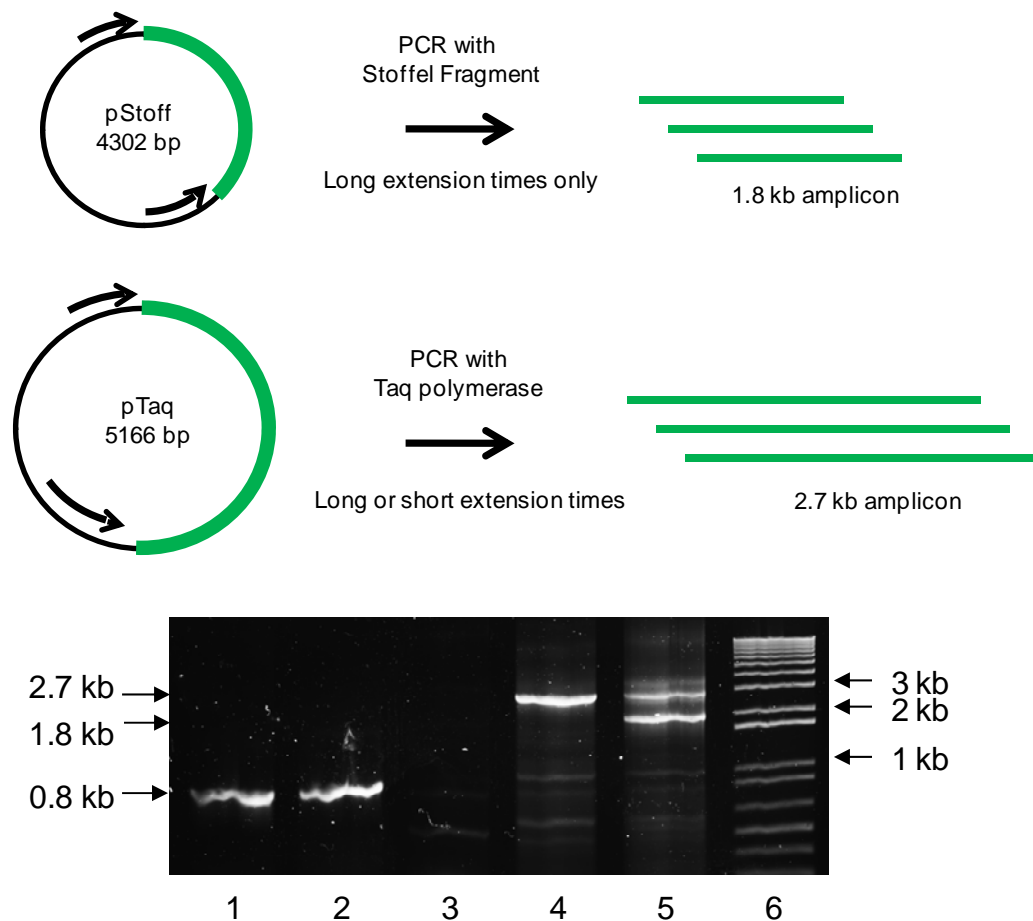


Figure 5.11 Solution phase PCR. Top. Representation of the two control plasmids pStoff and pTaq and the condition that support amplification of their polymerase gens. Bottom. Whole cell PCRs using *E. coli* cells containing either a plasmid bearing a Stoffel Fragment gene or a Taq gene. Lane 1: A PCR mixture containing Stoffel Fragment-expressing cells with a 0.8 kb target. Lane 2: A PCR mixture containing only Taq-expressing cells with a 0.8 kb amplicon target. Lane 3: A PCR mixture containing only Stoffel Fragment-expressing cells. Lane 4: A PCR mixture containing only Taq-expressing cells. Lane 5: A 1:1 mixture of the Taq-expressing cells PCR mixture and the Stoffel Fragment-expressing cells PCR mixture. Lane 6: Invitrogen's 1 kb molecular weight ladder.

The separation of individual cells using emulsification was examined using a 1:1 mixture of cells expressing Taq and cells expressing Stoffel Fragment. The efficiency of the isolation of individual cells through emulsification can be qualitatively determined by examining the ratio of PCR products. Stoffel Fragment has a lower rate of nucleotide incorporation than Taq polymerase and was shown to be unable to amplify its gene using

short extension times (Figure 5.11, Lane 3). As the emulsification for the whole cell PCRs followed the visual control described in section 5.3.7.1, we assume that the emulsification of the cells effectively isolated the *E. coli* cells prior to thermocycling. A 1:1 mixture of cells expressing Taq and cells expressing Stoffel Fragment were first emulsified prior to thermocycling to isolate individual cells. Analysis of the whole cell PCR products shows that the emulsified Stoffel Fragment is not capable of amplifying its gene to any significant extent (Figure 5.12, Lane 2). Emulsified Taq, on the other hand, amplifies its own gene exclusively due to the physical barrier separating Taq polymerase from the Stoffel Fragment gene.

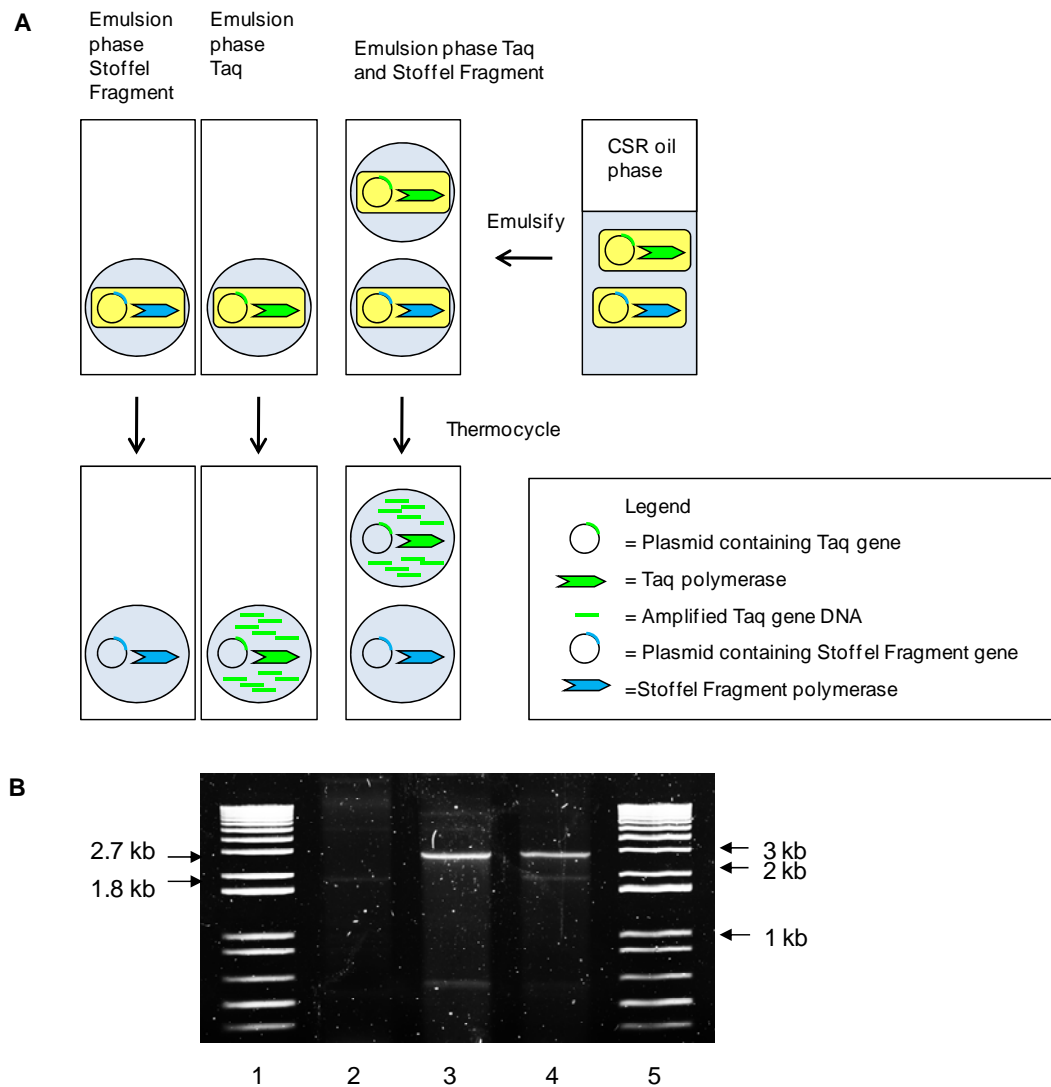


Figure 5.12 Compartmentalized self-replication of two different genes. **A** Scheme showing how the emulsification isolates cells prior to thermocycling. Successful isolation of the individual cells followed by thermocycling as represented would result in only the Taq gene being amplified and the reverse micelles containing cells expressing Stoffel Fragment would have no amplicon. **B** *E. coli* cells expressing either Stoffel Fragment or a Taq were emulsified and thermocycled. Lanes 1 + 5: Invitrogen's 1 kb molecular weight ladder. Lane 2: Stoffel Fragment-expressing cells were emulsified and thermocycled. Lane 3: Taq expressing cells emulsified and thermocycled. Lane 4: 1:1 ratio of Taq expressing cells and Stoffel expressing cells mixed prior to emulsion and thermocycling.

5.4.5 Shuffled Taq library

Even with the cassette libraries in hand, we considered that demanding the polymerase incorporate multiple modified nucleotides may be a very high selection pressure. Reducing the selection pressure to a single event such as elongation of a primer bearing a single modified nucleotide would facilitate amplification of long (>1 kb) DNA targets. The dA^{ime}-modified primer described in section 2.4.1 was intended for use in the amplification of a long target – the entire Taq gene. Holliger and coworkers have been very successful in evolving polymerases using CSR that involves designing the primers used in the CSR to introduce the selection pressure.(Ghadessy *et al.*, 2004; Loakes *et al.*, 2009; Loakes & Holliger, 2009) The selection pressure that was introduced was elongation of primers designed with base pair mismatches, modified bases, and hydrophobic base pairs. Because the desired activity is physically localized to the primer: template region, which is independent of the length of the CSR target, the polymerase can replicate the entire gene and not just a small region within it, as in spCSR, without too much difficulty.

One of the benefits of performing a CSR that targets the entire gene is that mutations can be introduced throughout the entire gene and the CSR process would be able to amplify those mutations. With the cassette libraries, we target areas where direct contact between the substrates and enzymes are made. The necessary structural or chemical change introduced by mutations that will provide the desired activity may be more subtle than direct alteration of the catalytic amino acids and/or the amino acids proximal to the active site. Making distal mutations may provide the subtle changes to the active site that lead to the desired activity. Two methods of introducing mutations are error-prone PCR and shuffling.(Stemmer, 1994) Shuffling involves the fragmentation of two or more closely related genes. The double-stranded fragments, when used in PCR conditions will dissociate and anneal to another piece of DNA that may have originated from a different gene. Multiple cycles of elongation, melting and annealing result in full length genes bearing DNA sequences from more than one of the original genes. Shuffling was chosen over error-prone PCR due to the high amount of control over the frequency of introducing mutations. Zhao and Arnold report a very low frequency of mutation of 0.05 % using specific polymerases.(Zhao & Arnold, 1997)

Normally, shuffling is performed using different genes having high level of homology. For example, d'Abbadie *et al.* shuffled three polymerase genes from the genus *Thermus*.(d'Abbadie *et al.*, 2007) The gene for Taq was amongst these genes. This so called molecular breeding allows for entire regions of gene to be swapped and still produce a functional protein with new characteristics. In our hands, shuffling is first being used to introduce random mutations at a very low frequency. Molecular breeding at a later time is possible if we acquire the appropriate genes or in between rounds to reintroduce diversity during the selection. The first step in shuffling is the fragmentation of the DNA. Once the fragments of appropriate size (50-100 bp) are isolated from larger fragments, they can be recombined using amplification where the fragments serve as both primer and template. To obtain fragments, pTaq was digested with DNase I. The expression vector (2.7 kb) and the Taq gene (2.5 kb) are roughly the same size and would be difficult to purify away from each other so both components were digested. Careful monitoring of the digestion ensured that the fragment sizes are within the desired 50-100 bp range. Fragments migrating faster than the 100 bp marker on 2 % agarose gel were excised and purified using Qiagen's QIAQuick Gel Extraction Kit.

The Taq gene fragments were recombined using conditions that promote a very low rate of mutation. The gene encoding for wild-type was carefully digested with DNase I to produce DNA fragments roughly 50-100 nt in length (Figure 5.13). These fragments were gel purified by excising the gel corresponding to that particular size range. This was done to minimize the amount of large fragments. The purified fragments were reassembled using DNA polymerases. The products of the reassembly appear as a large smear (Figure 5.14). Presumably, within this smear of DNA lie multiple copies of the Taq gene each with slight mutations.

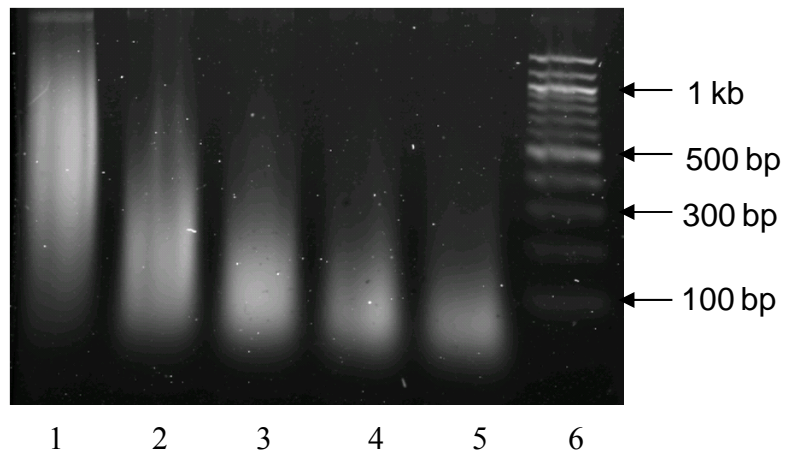


Figure 5.13 Time-dependent fragmentation of DNA by DNase I. Lanes 1-5: 1, 2, 3, 4, and 5 minutes of digestion, respectively.

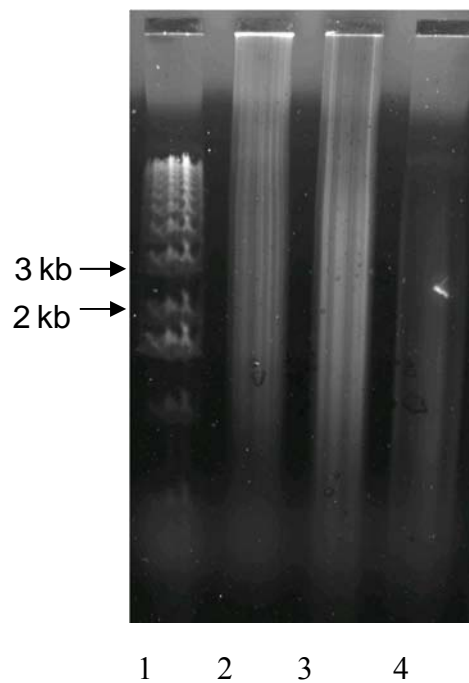


Figure 5.14 Polymerase-mediated reassembly of gene fragments. Lane 1 Invitrogen 1kb Plus Ladder. Lane 2: Taq-mediated shuffling with exogenous DNA fragments. Lane 3: Taq-mediated shuffling 4: Pfu-mediated shuffling.

Using the appropriate primers one can amplify the Taq gene from the reassembly mixture. The library generated by Pfx-mediated reassembly and Taq-reamplified appeared as a single band (Figure 5.15). The larger amount of Taq-mediated reassembly and Taq-reamplified is hypothesized to be due to more reassembled product being produced by the faster Taq polymerase and therefore more initial template for the reamplification is present.

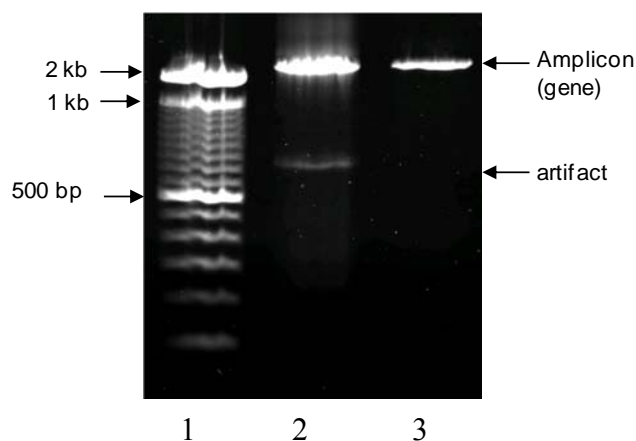


Figure 5.15 Amplification of the shuffled Taq gene. Lane 1: Invitrogen's 100 bp Ladder. Lane 2: Taq-reassembled Taq reamplified shuffled library. Lane 3: Pfx-reassembled Taq reamplified shuffled library.

While the reassembly appears to work, the level of mutation would be difficult to assess using sequencing. As briefly mentioned in Chapter 4, a thorough study of the level of mutation could be done using the mismatch binding protein, MutS, to do a proper analysis. The library has been cloned into the expression vector pUC18 and is ready for use in a selection as shown by a small amount of activity (Figure 5.16). RNA was left intact as a visual reference of how much cellular material was needed to be added to produce an amplicon.

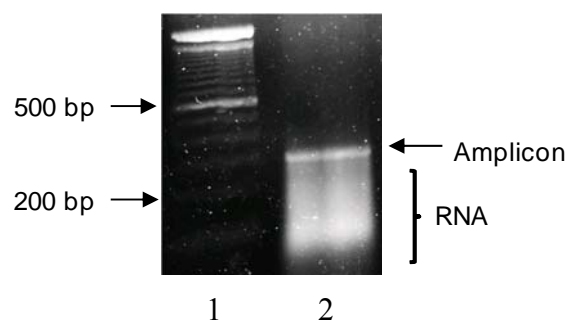


Figure 5.16 Activity of the shuffled Taq library. Lane 1: Invitrogen's 100 bp Ladder. Lane 2: Whole cell PCR using the shuffled library.

5.4.6 CSR preselection of the active site using natural nucleoside triphosphates

With all of the necessary controls done, the actual selection can begin. As with all selections, this one starts with a library. The choice of libraries depends on the selection pressure that was applied. We chose to begin with the active site library constructed using cassette mutagenesis. The reason for choosing this library over the shuffled Taq library is due to the availability of substrates that introduce a desired selection pressure. Recall that the production of the shuffled library introduced mutations throughout the entire gene. Therefore the target for self-replication would be the entire gene. In the case of full length Taq, this target would be about 2.5 kb. Using modified nucleoside triphosphates in the replication of this target would require the mutant polymerase to incorporate about four hundred modified nucleotides to be successful. That is far too high a selection pressure. The alternative way of introducing a selection pressure is by using modified primers and demanding the polymerase extend past modified bases. This is an excellent approach, however, we have a very limited number of modified primers and the phosphoramidites necessary to make a particular primer are not always available. Libraries constructed by either cassette mutagenesis or by shuffling have led to evolved polymerases in the literature. Both techniques are valid and while the following experiments focus on the cassette library, the shuffled library is ready for use in selection if the selection using the cassette library is not successful.

The cassette libraries are the better choice for selection in our hands as we have a large repertoire of modified dNTPs that may make unique and different contacts with various polymerase domains. Since the mutations are localized and concentrated to a small region, this library is perfect for short patch compartmentalized self-replication (spCSR). The selection pressure of incorporating modified nucleotides is attenuated by demanding that the self-replication only occur at the defined mutated region. In the case of the active site, the mutated region is only a stretch of thirty-nine bases. The selection pressure can further be attenuated by selecting modified nucleotides that are known to be incorporated. With the library selected, one way of improving the chances of finding a mutant polymerase is to use a library with enriched activity. To obtain this library with enriched activity, we first perform a round of “preselection” using natural dNTPs to help remove clones that are completely inactive. This “preselection” is actually a true selection for activity, but for the majority of this chapter “selection” refers to the rounds that use modified nucleotide substrates, be them modified nucleoside triphosphates or modified primers.

Preselection products were isolated, purified and amplified according to the scheme in Figure 5.17. Note the selection primers and the amplicon contain sequences, represented by the purple lines in Boxes A-D, which are not found in the plasmids. Since selection products do not have a defined electrophoretic mobility due to the varying number of modified bases, it is not possible to isolate the selection products from the plasmid DNA. Introduction of foreign sequences provides a means to selectively amplify the selection or preselection products exclusively in the presence of contaminating plasmid. Any selection primers left over from the selection must be removed. This method of removing unwanted chemicals and biomolecules prior to amplification of the selection products, sometimes referred to in this thesis as first amplification, is critical to ensure that genes encoding for inactive mutant polymerases are not amplified. Otherwise, the gene pool may not become enriched in genes encoding for desired mutant polymerases. This purification process is critical to this preselection round and all subsequent selection rounds.

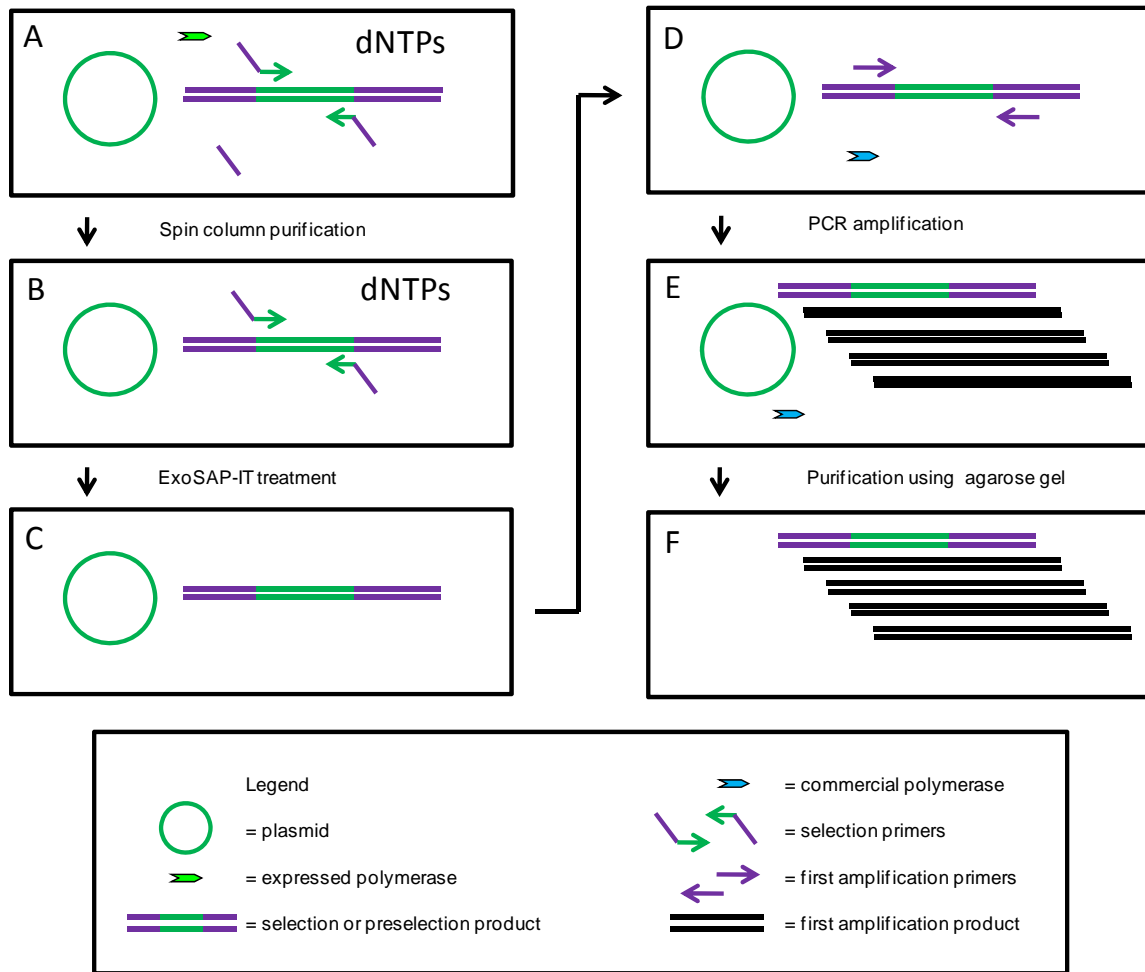


Figure 5.17 **A.** Important biomolecules present in the spCSR selection reaction. Plasmid is represented by the green circle. Polymerase is represented by the green chevron. Green and purple arrows represent primers. Green and purple lines represent spCSR selection product. Note that the selection products are not one defined size. **B.** The biomolecules from A purified using Qiagen's PCR Purification Kit. Polymerase is removed, but dNTPs, primers and plasmids are present in trace amounts. **C.** The components of B are treated with ExoSAP-It to remove trace dNTPs from the selection and primers. **D.** New primers represented by purple arrows and Vent (exo-) represented by the blue chevron are added to the remaining plasmid and amplicon in a PCR reaction mixture. **E.** The reaction is thermocycled, and selection products are amplified exclusively. **F.** The new amplicon is agarose gel purified to remove plasmid and primer.

First, polymerase, chemical reagents, the majority of dNTPs, and primers are removed using Qiagen's Nucleotide Clean Up Kit. Trace amounts of dNTPs, plasmids and primers from the preselection CSR could have gotten through this first clean-up process. The trace primers and plasmid, in particular, could cause amplification of genes that encode for inactive polymerases. In this preselection round, the dNTPs are natural and will not cause any problems in the first amplification, but the use of modified nucleotide in the selection rounds could inhibit the commercial polymerases used in the first amplification. Trace amounts of selection primers and dNTPs are degraded by treating the mixture with ExoSAP-IT from the USB Corporation. ExoSAP-IT, containing Exonuclease I and Shrimp Alkaline Phosphatase, was used to degrade the single-stranded selection primers and deoxynucleoside triphosphates that carry through from the spCSR. Double-stranded amplicon products do not get degraded. The products generated from the degradation of the selection primers and nucleoside triphosphates are dinucleotides, nucleosides, inorganic phosphate and pyrophosphate. A 15 minute heat treatment at 80 °C inactivated these two enzymes and the remaining DNA was ready to be used in the first amplification. The primers of the first amplification were specifically designed to hybridize to the selection products on the regions corresponding to the foreign sequences introduced by the primers used in spCSR (purple lines). Since the plasmids bearing either the desired or undesired gene do not contain these foreign sequences, they cannot be used as templates for the first amplification. Amplified products were cloned into pTaq to create a new library.

Clones from this new library were examined. Colonies were grown on an agar plate to isolate individual clones. Colonies were picked at random, used to inoculate media and tested for activity. Eight of the ten showed activity and their respective plasmids were purified and sent for sequencing (Figure 5.18). Six of the eight active clones contained mutations in the partially degenerate motif A region (Figure 5.19).

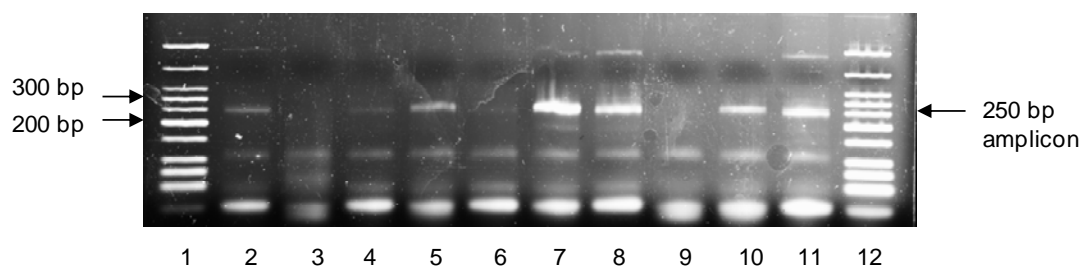


Figure 5.18 Activities of preselected clones of the active site library. Lanes 1 and 12: NEBs Low Molecular Weight Markers. Lanes 2-11: PCR products indicating the activities of active site preselection clones 1-10, respectively.

Cl one i denti ty	Sequence
Wild-type Taq	GGGTGG CTA TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTGCTGGCC L G V A S D Y S R I E L R
1 *	GGGTGG CTA TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTGCTGGCC
3	GGGTGG CTA TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTG AGG GTGCTGGCC L616L
4	GGGTGG CTA TTG GTG <u>ICI</u> CTG GAC TAT AGC CAG ATA GAA <u>CAC</u> AGG GTGCTGGCC A608S E615E L616H
5	GGGTGG CTA TTG GTG <u>ACC</u> CTG GAC TAT <u>ACC</u> CAG ATG GAG CTC <u>AAG</u> GTGCTGGCC A608T S612T I614M R617L
6	GGGTGG <u>CAA</u> TTG GTG <u>GCG</u> CTG GAC TAT <u>ATC</u> CAG ATA GAG CTC AGG GTGCTGGCC L605Q A608A S612I
7	GGGTGG CTA <u>TTC</u> GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC <u>AGA</u> GTGCTGGCC G607R R617R
9	GGGTGG CTA TTG GTG <u>ICI</u> CTG GAC TAT AGC CAG ATA GAA <u>CAC</u> AGG GTGCTGGCC A608S E615E L616H
10 *	GGGTGG CTA TTG GTG GCC CTG GAC TAT AGC CAG ATA GAG CTC AGG GTGCTGGCC

Figure 5.19 Active site sequences from active clones isolated from CSR using natural nucleoside triphosphates. The base mutations are underlined, and the amino acid mutations are indicated below the nucleotide sequence. Clones marked with an asterisk have the same sequence as wild-type Taq gene. Bases outside of the mutated region are shown in blue.

5.4.7 Combined library construction

Since the goal of the work in this thesis is to increase the value and utility of base-modified nucleotides in selection, the design of the polymerase libraries should also reflect that goal. The Watson-Crick base pairing is primarily accomplished by a complex mechanism mediated by the finger subdomain of the polymerase. Specific interactions are made between the O-helix of the finger subdomain, the DNA and nucleoside triphosphate substrates. We hypothesize that modifications in the O-helix would have the most drastic effect on nucleobase-modified nucleotide incorporation. Therefore, a partially mutagenized cassette of DNA encoding for the O-helix was cloned into the preselected active site plasmid library in place of the DNA encoding for the wild-type O-helix. The O-helix library cassette was cloned into the plasmids containing the preselected active site library (Figure 5.20). Preselected libraries were combined by first treating the plasmid library containing preselected active site with Sac II and Nhe I. The amplicon containing the preselected O-helix library was also treated with Sac II and Nhe I.

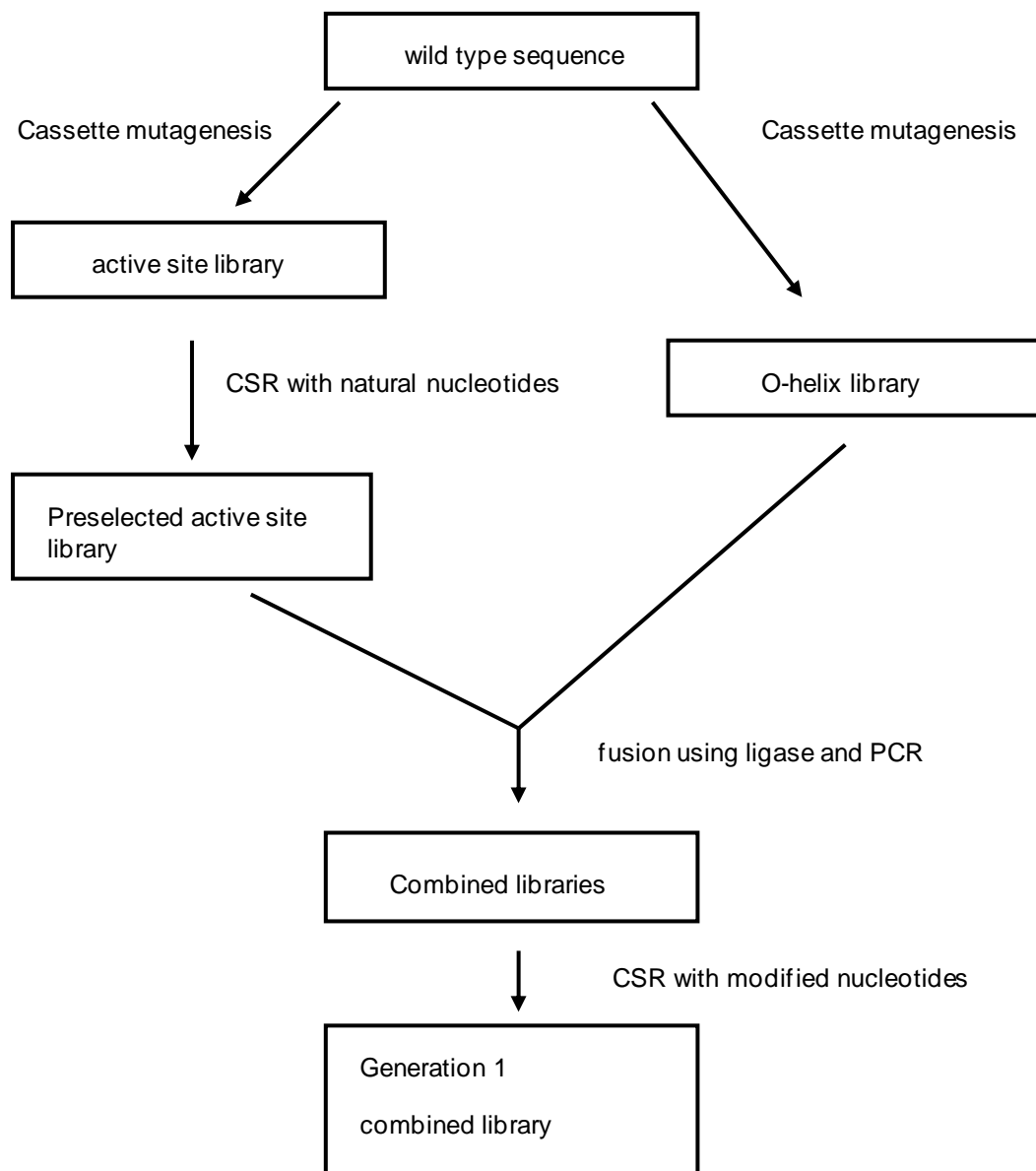


Figure 5.20 Flowchart describing the building of the Generation 1 libraries. The active site residue 605-617 were partially mutagenized to create an active site library. After preselection of the active site library, it was fused to the O-helix library. The library resulting from the combination was used for CSR with modified nucleotides.

5.4.8 CSR using unnatural nucleosides

With both cassette and shuffled libraries in hand, we first started with the spCSR using the cassette library, made by combining the preselected active site library and the O-helix library, and modified nucleoside triphosphates. A pair of primers different from the pair used in the active site preselection was used. This new pair of primers flanked the region of DNA that contains both the active site and the O-helix. Therefore, the “short patch” target DNA would contain the genetic information for both active site mutations and O-helix mutations. Incorporating modified nucleotides into the selection scheme involves some modification of the protocol to accommodate the products read-through of the selection products (Figure 5.21). We made one deviation from the Ong *et al.* version of the spCSR that follows the scheme shown in Figure 1.18. The deviation is illustrated in step 5 of Figure 5.21 where we replaced the unnatural substrate, ribonucleoside triphosphate, from the original spCSR with one of our modified nucleoside triphosphates. We’ve introduced dA^{imm}TP, dC^{aa}TP, or 2'-F-araTTP *in lieu* of its natural counterpart in the amplification mixture. Step 8 was added in to convert modified DNA sequences to unmodified DNA sequences and follows the clean-up and reamplification scheme shown in Figure 5.17.

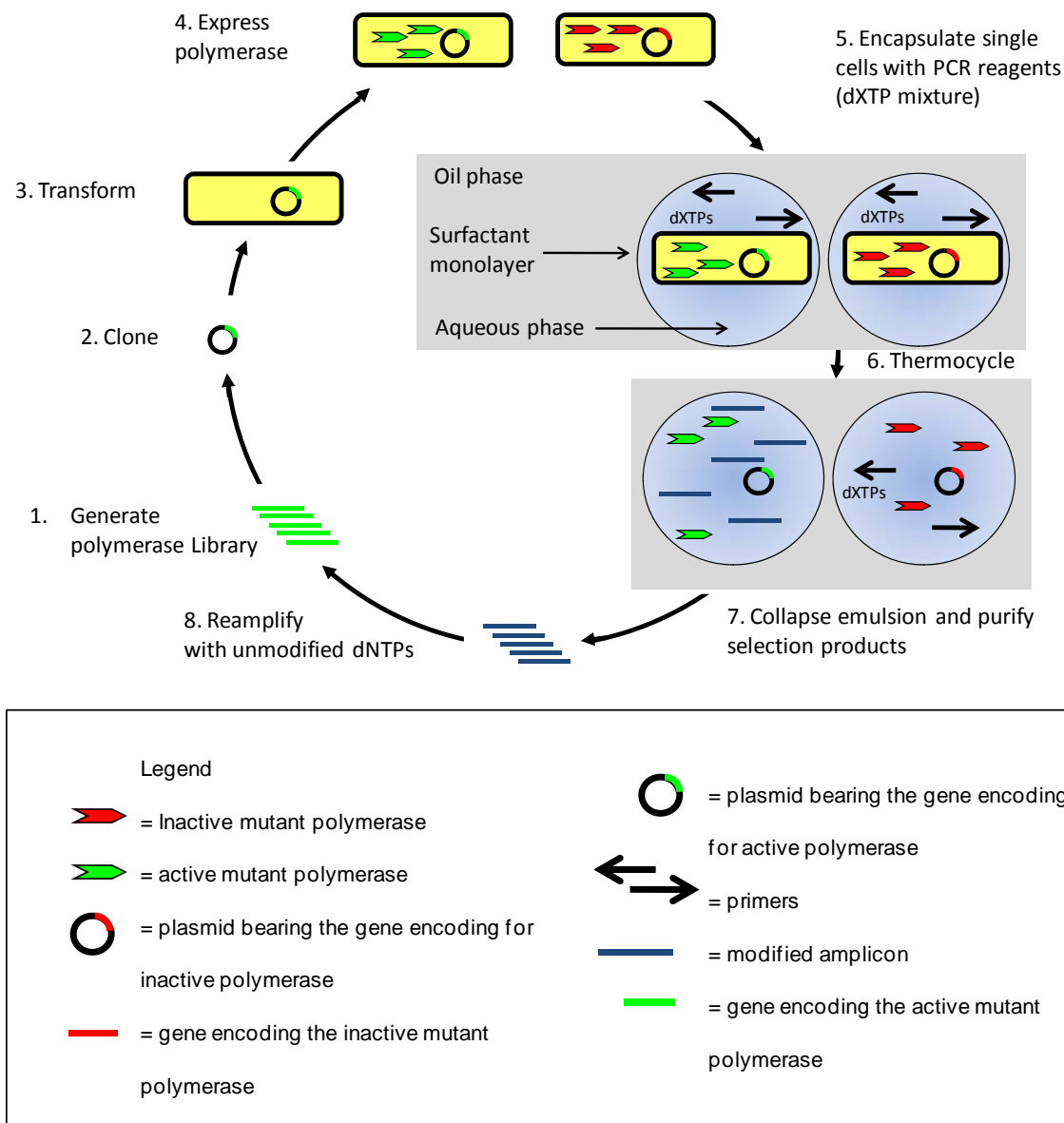


Figure 5.21 Revised CSR scheme to accommodate unnatural nucleic acids. Step 5 shows the introduction of modified nucleoside triphosphates into spCSR.

5.4.9 CSR using unnatural nucleoside triphosphates and first amplification product

Three unnatural nucleoside triphosphates were used in the CSR selections. The first is the 2'-fluoro-2'-deoxyarabinothymidine triphosphate (2'F-araTTP). This modification is on the sugar and is the easiest to predict the critical interactions and possibly the mutations necessary. The key contacts are made primarily at Motif A which contains the active site residues. The 5-aminoallyl-2'-deoxycytidine triphosphate

(dC^{aa}TP) was used. This substrate is very weakly accepted by Taq, but is capable for use in PCR. Lastly, 8-(4-imidazolyl)aminomethyl-2'-deoxyadenosine (dA^{aa}TP) was used. The base-modified substrates interact mainly with the O-helix. However, specificity can also be relaxed with mutations located in the Motif A region.

Successful amplification using modified nucleoside triphosphates cannot be visualized directly. In the event of high efficiency incorporation, the polymerase may extend the primer far beyond the target length of the library. For the positive control using natural nucleoside triphosphates, the 5 minute extension time is enough time for the polymerase to extend the entire length of the plasmid. It is necessary to purify all dsDNA from the CSR reaction without size discrimination. Unfortunately, this means that the original plasmid will copurify with the amplicons. This is a major contamination of genes encoding for inactive polymerases. To avoid the contamination of the gene pool with unamplified undesirable genes, an extra sequence of DNA was added to the 5' end of each of the CSR primers. This extra sequence is foreign to the expression vector. For first amplification, a new set of primers composed of these foreign sequences will be used so that there will not be mispriming or hybridization to contaminating plasmid. Upon amplification of the purified CSR product, only the amplicons successfully produced in the preceding PCR and necessarily composed of the nucleotides added to the CSR mix will be amplified in the first amplification. Each of the CSR first amplifications produced amplicons to varying degrees (Figure 5.22). Positive control produced the expected product and amplification product was absent in the negative control.

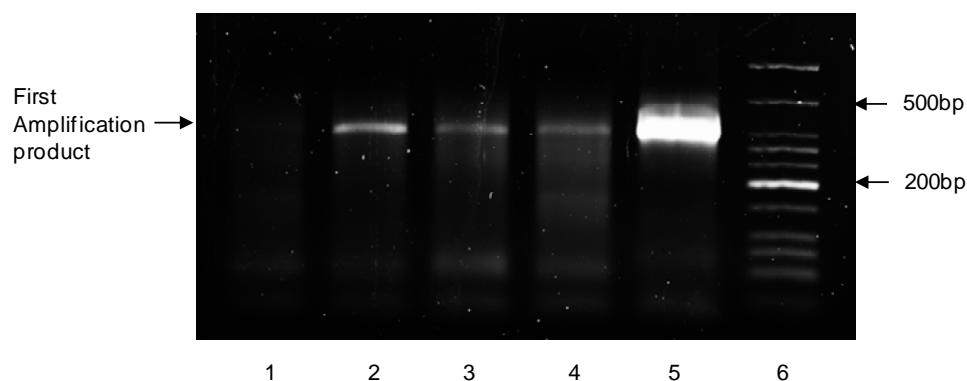


Figure 5.22 First amplification of the isolated selection product. Lane 1: negative control (no dNTPs added), Lane 2: amplification of 2'-F-araT amplicon, Lane 3: amplification of dC^{aa} amplicon, Lane 4: amplification of dA^{imm} amplicon, Lane 5: positive control (natural dNTPs) and Lane 6: NEB's low molecular weight ladder.

5.4.10 Activity of generation 1

The activity of generation one was assessed by allowing the libraries to perform PCR using dNTPs. While the incorporation of all four dNTPs is an activity that was not selected for, it is an indication of expression and activity. The library used in the selection for a FANA monomer-incorporating polymerase showed no ability to perform amplification using dNTPs (Figure 5.23). The library used in the selection for a dC^{aa}TP incorporating polymerase showed the ability to incorporate dNTPs and produced amplicon. The library used in the selection for a dA^{imm}TP incorporating polymerase showed weak, but detectable ability to incorporate dNTPs as seen by a faint band in the agarose gel. The positive control library showed the most amount of activity as expected, but in this reaction and all others that produced amplicon there was some artifacts being produced. The source of these artifacts is unknown.

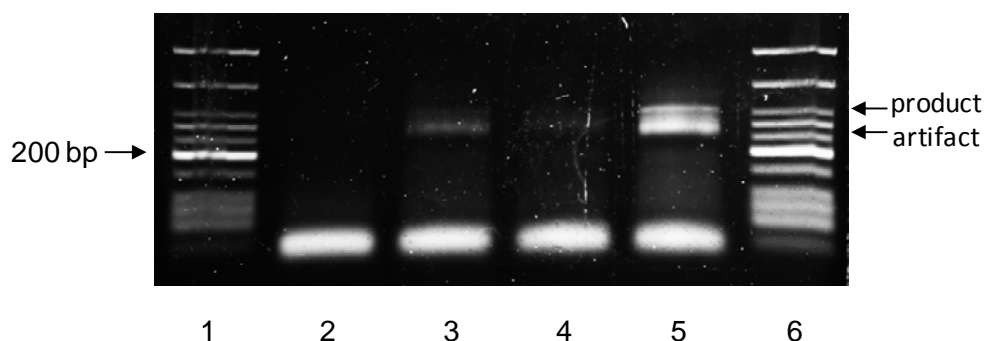


Figure 5.23 Library activities under standard PCR conditions. Lane 1 and 6: NEB's low molecular weight ladder, Lane 2: 2'F-araT library, Lane 3 dC^{aa} library, Lane 4 dA^{imm} library, Lane 5 dNTP library.

5.4.11 First amplification product of round 2

The 2'F-araT, dC^{aa}, dA^{aa}, and the positive control libraries were taken into the second round of selection whether activity was detected or not. Each selection produced first amplification products (Figure 5.24). Amplicon products were the expected 350 bp in size. Compared to the first round, it seems that the product ratios of the amplicons produced by the modified nucleoside triphosphate selection libraries relative to the dNTP positive control has increased indicating that there may be more desired activity compared to the first round.

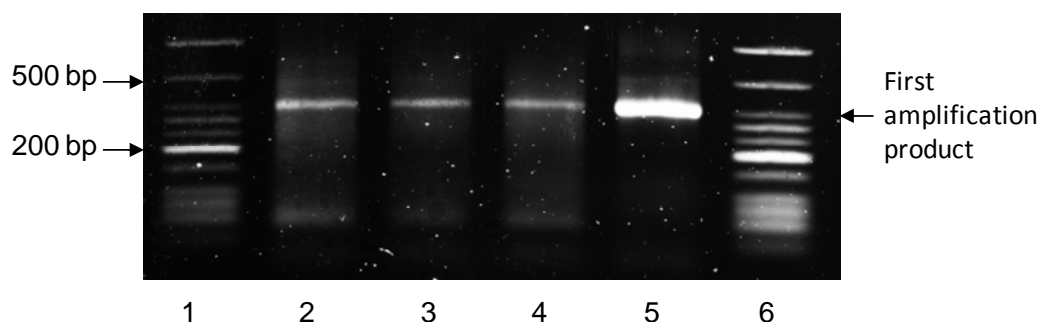


Figure 5.24 First amplification of the isolated selection product from Round 2. Lane 1 + 6: NEB's Low Molecular Weight Ladder. Lane 2: amplification of 2'F-araT amplicon. Lane 3: amplification of dC^{aa} amplicon. Lane 4: amplification of dA^{imm} amplicon. Lane 5: positive control (natural dNTPs)

5.5 Discussion

5.5.1 General discussion

Compared to the Dz20-49 selection described in Chapter 3, the directed evolution of protein DNA polymerases attempted in this chapter is, by far, much more difficult. First, the DNAzyme selection of Chapter 3 involves a single intramolecular transphosphorylation; given enough time in the selection buffer, this intramolecular transphosphorylation would occur spontaneously. Second, each isolated selection system, the minimum components to carry out a selection reaction or event, needs only one molecule, the DNAzyme, in cleavage buffer. Third, there is little concern about cross-reaction as the selected DNAzymes are product inhibited. In comparison, our spCSR involves the incorporation of modified nucleotides multiple times on a modified template strand. In the construction and amplification of the selection product, an unnatural dsDNA biopolymer, spCSR requires each individual selection system to contain multiple components including a single type of expressed mutant polymerase, modified nucleoside triphosphates, original plasmid template and amplicons produced be perfectly physically isolated from any neighboring systems in a reverse micelle. These factors make the need for a well conceived selection scheme as well as careful controls absolutely critical.

Using unnatural nucleoside triphosphates in spCSR requires the polymerase to perform not one, but two tasks. Both of these tasks must be performed multiple times (up to 30 times) per selection round. The first task is the incorporation of multiple modified nucleotides to produce a given gene target sequence. Many of these modified nucleotide incorporations are sequential. In a single round of selection, the polymerase may be required to do hundreds of incorporations of the unnatural nucleotide. The second task occurs in the second and all subsequent cycles of spCSR. This task is the successful read-through of a modified template strand indicated by the successful synthesis of a modified strand across from the modified template strand effectively producing doubly-modified dsDNA. Failure to perform this second task will result in linear amplification, which may or may not produce sufficient enrichment.

The intramolecular nature of the bound ribophosphodiester bond-cleaving DNAzymes and fact that the successfully released DNAzyme is product-inhibited eliminates the possibility of a cross-reaction. The protocol for polymerase evolution uses a multicomponent system and the cross reaction of even a single molecule could have dramatic negative effects on the selection. The role of the reverse micelle is the isolation of a single phenotype and its corresponding genotype. If polymerase, gene, or amplicon pass from one reverse micelle to a neighboring micelle, then genetic material encoding an inactive mutant will be amplified. The worst case scenario would be that the reverse micelles themselves are unstable under the thermocycling conditions. If the reverse micelles are unstable, they would fuse and the emulsion would collapse during thermocycling. However, section 5.4.3 describes how, with careful optimization, the integrity of the reverse micelles through thermocycling and the successful isolation of individual cells can be achieved.

5.5.2 Taq libraries

Libraries were constructed using cassette mutagenesis and shuffling. Cassette mutagenesis restricts the mutations to predefined areas of the polymerase. Prior knowledge of the structure of the enzyme must be known for this type of mutagenesis to be used in selection. Shuffling can be used to make mutations throughout the gene and can be customized to give a very low level of mutagenesis (~0.05 %).(Zhao & Arnold, 1997) Shuffling will also be a powerful tool for diversifying between rounds or if one wants to combine mutations between known isolated mutant polymerases. Both the shuffled and cassette libraries would be valid for our selection as we do know where the critical contacts between the polymerase and the modifications on the substrates are, but we don't know if the structural changes needed are large or small. As seen with previous selections of Taq, both types of libraries have had success.(d'Abbadie *et al.*, 2007; Ghadessy & Holliger, 2004; Ghadessy *et al.*, 2001; Loakes *et al.*, 2009; Loakes & Holliger, 2009; Ong *et al.*, 2006) However, depending on the library chosen, the selection scheme would be very different. If we would like to use modified dNTPs in the selection we would have to lower the stringency as much as possible. spCSR would be

ideal for the incorporation of modified nucleotides, and the extent of mutagenesis can be a higher in the focused region of the gene. If only one event is required of the polymerase such as read-through of a 3'-modified primer, then the mutations can be made throughout the gene making shuffling a viable option to be used in CSR that involves the amplification of the whole gene. Since we have a sizable repertoire of modified dNTPs and not the appropriate modified primers or modified phosphoramidite to produce the modified primers, we chose spCSR using cassette mutagenesis. The shuffled library is ready for use if we seek an alternate method of selecting mutant polymerases.

5.2.3 Site-directed mutagenesis

While the sequencing did show that the mutation was produced, significant amounts of protein were not detected. A potential cause of this is unexpected mutations possibly produced in the site-directed mutagenesis. The whole gene would need to be sequenced to determine if there are other mutations. Alternatively, the E615G mutation can be introduced using cassette mutagenesis to avoid distal mutations in the gene. The next possible cause is the nature of this expression system. Taq expression using pUC18 is considered leaky and protein with low to no toxicity can be slowly expressed. The low toxicity and acceptable stability of the E615G mutant is thought to be similar to that of the native protein, but if the mutant is toxic or unstable, it may be susceptible to degradation within the cell. In order to address this, the mutant gene should be cloned into an expression vector that uses tight regulation of expression such as a pET vector.

5.5.4 Visualization of the compartmentalized cell

The dimensions of the droplet in relation to the size of the cell are very important to the selection. First, ensuring a minimal droplet size results in a higher number of droplets in a given reaction volume. This is critical as we want to have a very high droplet to cell ratio as this statistically places a one cell per droplet. In our 200 μ l CSR reaction ideally we would produce 10^{10} droplets containing 2×10^8 cells. Therefore, there is a 2 % chance that any one droplet would have a single cell. Isolation of

individual cells is critical to optimize enrichment. Second, the size of the drop affects the local concentration of a molecule within the droplet. One molecule in a 5 μm diameter droplet is at a concentration of ~ 18 nM.

Visualization of the droplet is important, but more important is the controlling of the size of the droplet to encapsulate whole individual cells. Ideally, both the droplet and the encapsulated cell can be simultaneously visualized. Unfortunately, the visualization of the *E. coli* cell within the reverse micelle is difficult using visible light. Therefore, *E. coli* cells expressing Enhanced Green Fluorescent Protein (EGFP) were used to enhance visualization of the cell. Using UV light, the cells could be clearly seen, but the periphery of the reverse micelle cannot be visualized. Therefore, a combination of UV light and white light was used to visualize both simultaneously. First, UV light was turned on and then white light was slowly increased just until the periphery of the reverse micelle can be seen without washing out the fluorescence given off by the *E. coli* cell.

At a 2:1 oil phase to aqueous phase ratio, it is difficult to produce a monolayer of reverse micelles. To circumvent this, the reverse micelles were diluted into mineral oil to disperse them. Well dispersed reverse micelles containing *E. coli* cells expressing EGFP can be seen clearly under the microscope (Figure 5.5). Both reverse micelles containing *E. coli* cells and no *E. coli* cells can be seen. The diameter of the reverse micelle is slightly larger than the *E. coli* cell. Ideally, the diameter should be equivalent to the length of a single *E. coli* cell to make the reverse micelle volume as small as possible. Minimizing reverse micelle size will increase the total number of micelles in the vial. Increasing the total number of reverse micelles increases the statistical probability of finding one or zero *E. coli* cells per reverse micelle. The average reverse micelle size appeared to be ideal for our purposes.

5.5.5 CSR controls to measure cross reaction between droplets

Contaminations from neighboring reverse micelles would have a negative effect on CSR. If any of the active mutant polymerase or genes encoding inactive mutant polymerases leak through to neighboring reverse micelles, genes encoding for inactive mutants will be enriched in the gene pool. In the worst case scenario, the enrichment of

gene encoding for inactive mutants may dominate the gene pool and the selection will never lead to the isolation of active mutant polymerases.

To ensure that components are staying within their respective reverse micelle during the thermocycling, an experiment using two PCR reactions referred to as Reaction A and Reaction B were used. The product from Reaction A would be a 160 bp amplicon and Reaction B would be a 0.8 kb amplicon. Polymerase was not added to the first Reaction A, but was included in Reaction B. The different methods of mixing these two reactions, with and without emulsification, is meant to simulate the CSR conditions and to help us identify when the isolation of components have failed. For the first test we would look at how much of each product would be produced and verify this is an adequate test for visualizing cross reactions. When mixed together in solution phase, the two reactions produced exclusively the shorter 160 bp product even though the PCR conditions satisfied the amplification of the longer target. The ease of replication of the shorter product dominated the PCR. A solution phase mixture of the two reactions represents an extreme case of the reverse micelle cross reaction or complete collapse.

The assessment of reverse micelle stability was performed by first emulsifying both mixtures effectively sealing the reactions within billions of microscopic droplets. Individually emulsified Reaction A and Reaction B were mixed and the mixture was thermocycled. Emulsions containing only one reaction each were used as controls to ensure that the amplifications were not inhibited due to the emulsification. After breaking down the emulsions and cleaning up the products, the products show the 0.8 kb target was almost exclusively produced. There is the slightest trace of a 160 bp product, but can be considered insignificant for this test. This test shows that the reverse micelles remain intact and there is virtually no cross reaction.

5.5.6 Physical separation of cells using an emulsion

When wanting to isolate a particular cell that expresses a desired clone or mutant, investigators have a few techniques at their disposal. Such techniques include Fluorescence-Activated Cell Sorting (FACS) sorting or plating and spot picking are methods of separating and isolating cells. Depending on one's needs, separating and

isolating cells by emulsion is a faster method and can handle a very large number of cells. Emulsification is a powerful method of physically isolating the individual cell from its neighbor when handling of the individual cell (i.e. inoculation, protein scale up, sequencing) is not needed at this point, but isolated systems utilizing the activity of the protein is.

The elimination of cross reaction discussed in section 5.5.5 is a wasted effort if emulsification does not isolate the individual cell. To establish that compartmentalization was effective, the *E. coli* cells were isolated using an emulsion to compartmentalize the individual cells each in their own reverse micelle. If the cells expressing mutants cannot be separated using the emulsion, then the genetic material of inactive mutants will also be enriched and the selection may fail. Much like the cross reaction tests discussed in section 5.5.5, the same basic tests can be performed to determine the successful separation and isolation of individual *E. coli* cells.

To perform this test that mimics the contamination of an active polymerase and an inactive polymerase, two versions of the same polymerase were used. Taq and Stoffel Fragment were used to represent an active and an inactive mutant, respectively. Actual CSR will be performed in this experiment as the two polymerases will be amplifying their own gene. The two differences between the systems are that Stoffel Fragment is a slower polymerase than full length Taq and the Stoffel Fragment gene is shorter than the Taq gene. If the performance of the two polymerases is equivalent, an equal amount of amplicon would be produced as the two polymerases self-replicate their own gene. Similar to the experiment described in section 5.5.5, in a homogeneous mixture of reactions, the smaller Stoffel Fragment gene target would preferentially be amplified in a mixture of targets. However, in isolated systems, the Taq polymerase is the faster polymerase between the two and given the right conditions such as short extension times, Taq will out produce Stoffel Fragment. Due to Stoffel Fragment's slower extension rate, extension times can be altered so that the Stoffel Fragment will not be able to amplify its own 1.8 kb gene target. Taq, on the other hand, manages to amplify its own gene with the given extension time. Using a forty-five second extension time effectively nullified exponential amplification from Stoffel Fragment while allowing Taq to exponentially amplify its own gene. For an expression control, a different set of primers were used that

target a 0.8 kb sequence on the plasmid. Both expressed Stoffel Fragment and expressed Taq are able to amplify this short target with roughly the same efficiency given the same extension time.

Using a one to one mixture of Taq expressing cells to Stoffel Fragment expressing cells, the efficiency of single cell isolation through emulsification was examined. As with the cross reaction experiment, the first step is deliberately establishing what a failed isolation or collapsed emulsion would produce. To show the preferential amplification of the shorter Stoffel Fragment gene, cells expressing either Stoffel Fragment polymerase or Taq polymerase were used in a solution phase PCR. Mixing the two reactions in equal proportions resulted in an amplification that produced Stoffel Fragment gene as its major product. In this case where both products could be amplified, the higher amount of 1.8 kb amplicon indicates that a larger amount of the Stoffel Fragment gene (remembering that the Stoffel gene is only two-thirds the length of the Taq gene) is preferentially produced. For the mixtures of the emulsifications, we can see that the major product is the Taq gene, which tells us that isolation of single cells was efficient and the Stoffel Fragment gene is not being amplified by the Taq polymerase. The lower amount of amplicon compared to the Taq only control is partially due to the fact that only half the reverse micelles contained Taq polymerase. A trace amount of Stoffel Fragment gene was produced in the Stoffel Fragment only control, which means that the trace amounts in the mixture, was not due to any significant cross reaction or poor isolation of single cells.

Unlike the chemical cross-reaction control described in section 5.4.3.2, emulsification of the mixture was not done prior to mixing of the cells expressing either Taq polymerase or Stoffel Fragment polymerase. Instead, the emulsification itself was used to isolate the individual cells thereby isolating individual reactions. Since failure to isolate reactions results in the preferential amplification of the Stoffel Fragment gene, the ratio of the products produced will be indicative of the efficiency of single cell isolation. The major PCR product from the emulsified mixture of cells expressing either Taq or Stoffel Fragment is the amplicon corresponding in size to the Taq gene due to Stoffel Fragment's low elongation rate (Figure 5.11). Trace amounts of the amplicon corresponding to the Stoffel Fragment gene are also present. The interpretation of this

result is that the isolation of the individual cells through the use of emulsification appears successful. The amount of Stoffel Fragment gene amplicon appears slightly higher than the amount in the control. This may be due to leakage of components, but is more likely due to more than one cell being present in a single droplet. The contamination is minimal and these controls are meant to help minimize background and side reactions rather than take extreme measures to eliminate them altogether. There will always be a very small statistical probability that a reverse micelle will contain two *E. coli* cells.

5.5.7 Library construction, expression and CSR

Combining two libraries may cause too great of a mutational load for proteins that are expressed from the library to retain activity for CSR. With this in mind, the elimination of known detrimental mutations was undertaken. Any deleterious mutations of active site residue D610 must be avoided in the selection for unnatural nucleoside triphosphate incorporation. Residue D610 is intolerant to mutation and previous studies by Loeb have shown that any mutation to this residue renders the polymerase inactive. The first priority was to do a modest selection for those polymerases capable of natural nucleotide incorporation. This is a variation on the complementation selection done by Loeb to select for active mutants, but it also will help verify the efficiency of the gene pool enrichment.

A library of mutant polymerases was created and enriched in active polymerases using spCSR. In the event that creating libraries that strayed too far from wild-type rendered the library too inactive for enrichment, a stepwise development of the library was taken. First was the creation of an active site library according to Loeb. This library was then used in a CSR, which demanded that the polymerase be able to amplify its own gene under normal PCR conditions. Performing CSR without selection pressure will remove any mutants that are inactive such as mutants containing mutations to the D610 residue and will enrich the gene pool in active polymerases.

The amplicon products produced by the active polymerases will not be of defined length. A truly efficient mutant may be able to extend the entire plasmid with natural or unnatural dNTPs. No sizing restriction should be placed upon the products at this point.

Under standard PCR conditions, the 5 minute extension time used is long enough for wild-type Taq to extend the entire plasmid. As seen in section 5.5.6, forty-five seconds is sufficient to amplify a 2.7 kb target that contains the entire Taq gene. Assuming that the only products will be the expected target length may exclude products produced by highly efficient polymerases. Successful polymerases can have an activity that varies from being capable of just extending the library region to being highly efficient and able to replicate the entire plasmid. It then would be safe to assume that the desired products could be any length between 350 and 5100 nt. The reamplification of the trace amounts of variably sized products is designed to produce amplicons of predetermined length. At this point, purification according to size can be done. With these concepts in mind, a single round of preselection was carried out to enrich the gene pool in polymerases capable of wild-type activity and perhaps other activities. The resulting gene pool contained a high percentage of active clones from randomly selected colonies.

5.5.8 Analysis of the isolated and active site preselection CSR clones

After one round of preselection, 10 colonies were picked at random. The sequences of the eight polymerases found to be active are shown in Figure 5.19. Of these, two were inactive and eight showed varying amounts of activity. The plasmids of the eight active clones were isolated and sent for sequencing. To summarize the activities of the 10 clones, clones 1 and 10 are wild-type, clones 2 and 8 were inactive. Of the weakly active mutants, clone 3 has a single silent mutation and clone 5 has four mutations. Of the mutants showing significant activity, clones 4 and 9 both had the same three mutations; clone 7 has a single amino acid mutation and one silent mutation. Clone 6 showed high activity and has two mutations and one silent mutation. Except for one of the mutations, all mutations occurring in the active site of active mutants were previously reported in the literature. (Patel & Loeb, 2000a)

The one mutation not previously reported is L616H that appears twice in clone 4 and 9. Since this mutation was found from two independent colonies both displaying polymerase activity, the probability of this being a sequencing error is low. From Loeb's findings, the two most intolerant residues are D610 and E615.(Patel & Loeb, 2000a)

Both of these residues point into the active site. Most other residues located in Motif A point away from the active site. Another interesting finding from Loeb's work is that there are a high number of low activity mutants with increasing number of mutations. The population with the highest number of low activity mutants is those with two mutations.

5.5.9 The O-helix library

The plasmids containing the preselected active site library was used to clone the O-helix library as well. The O-helix is responsible for proper base pairing and is highly mutable. This polypeptide is an excellent candidate for mutagenesis and clones containing mutations in this area may bestow upon the polymerase the ability to make non-canonical base pairs using a modified base and a natural base. Even if mutations in the O-helix alone do not bestow desired qualities, they may compensate for mutations done to the active site that do bestow favorable activities but are detrimental to the catalytic rate or vice versa. This library combination of the active site and O-helix libraries was used in the ongoing selection for an evolved Taq polymerase capable of incorporating modified nucleotides using a modified DNA template.

5.5.10 CSR using modified nucleoside triphosphate

Similar to the DNAzyme selection of Chapter 3, the products of the selection should not be directly visualizable for several reasons. First, the product may be produced in trace amounts. Second, the doubly modified dsDNA may have unknown electrophoretic mobility different from unmodified DNA and different modified sequences would also migrate differently. Lastly, high efficiency transcription will produce modified DNA larger than the expected target size (350 bp) as the template is a plasmid and extension could produce a 5.1 kb product. Instead, detection of the selection product should be done using the product from the first amplification of the selection products. As described in section 5.4.6, design of the primers used in the selection included a DNA sequence that was foreign to the plasmids that facilitated exclusive

amplification of selection products in the presence of contaminating plasmid DNA. These primers contain sequences at their 5' end that are not found on the plasmid. Therefore, only primers and successfully synthesized modified DNA will contain these sequences. To ensure that the primers from the selection are not carried through to the first amplification, ExoSAP-It was used to degrade the single-stranded primers that were not consumed in the spCSR. Incomplete degradation of the selection primers can be detected by the appearance of amplicon in the first amplification of the negative control. First amplification primers were designed to anneal to the regions of the selection products that contain a sequence foreign to the expression vector. Therefore, the only template that is amplified will be the products of the selection as any plasmids will not have a primer hybridization region. The first amplification will produce a product of defined size (350 bp).

A possible outcome for the first amplifications is that no amplicon is formed. To determine whether such an observation was a result of the selection pressure being too high or if some other component, chemical or biological, of the spCSR was not operating properly, a positive control was used. For a positive control, an spCSR using only natural dNTPs was performed simultaneously. Since direct detection of product is difficult, this positive control will help show if some condition unrelated to the modified nucleoside triphosphate could have caused the selection to fail. These conditions could be failed polymerase expression, some failure in thermocycling, and potential mispriming. Fortunately, all modified nucleoside triphosphate spCSR produced amplicons in the first amplification. Therefore, all components of the spCSR, from the nucleoside triphosphates to the thermocycling program, are sufficient for amplification during the selection.

For a negative control, selection conditions were done without the use of exogenous dNTPs in the CSR reaction mixture. This control will determine if the endogenous nucleoside triphosphates are in sufficient quantities to produce contaminating amounts of unmodified product that will go into the first amplification. Another possibility this negative control will test is if the unextended primers from the selection are sufficiently degraded by ExoSAP-It. If primers from the selection carry through to the first amplification reaction, the plasmid that is also present in the first

amplification reaction can be amplified. No significant amount of product was detected in the first amplification of the negative control. This suggests that the concentration of endogenous dNTPs from the *E. coli* cell is insufficient to support spCSR. The absence of product in the negative control of the first amplification provides sufficient evidence that neither the endogenous nucleoside triphosphates nor the primers from the selection will cause contamination from genes encoding undesired mutants.

Small amounts of products can be seen from the first amplifications of the selections. At this point the first amplification products, being composed entirely of unmodified DNA, can be gel purified according to size. A second PCR was performed to produce sufficient amounts of DNA for cloning. Using natural nucleotides, the wild-type activity of these generation 1 libraries were assayed. This assay is merely an estimate of activity as it involves PCR amplification using the four natural dNTPs. The activities of the libraries using natural dNTPs are not entirely indicative of enrichment as there is the highly unlikely possibility of the mutant polymerases that compose the library discriminate against the incorporation of the natural nucleotide counterpart of the modified nucleotide used in the spCSR. If the selected mutant polymerases did discriminate against the natural counterparts of the modified nucleotides used in the selections, the selected mutant polymerases will not be able to support amplification under standard conditions, which uses all four natural nucleotides. The 2'F-araTTP G1 library failed to show any activity for natural nucleoside triphosphates even though it lead to the most first amplification product and therefore the most selection product was produced by this library.

The second round of selection showed a slightly higher amount of first amplification products. There also seems to be an increase in the ratio first amplification product from the modified nucleoside triphosphate selections to the first amplification product from the positive control. This is what we would hope to see. As the positive control most likely contains such a high percentage of active mutants, we wouldn't expect drastic enrichment as the spCSR reactions are practically saturated with active mutants. For the modified nucleoside triphosphate selections, however, enrichment represented by a higher amount of first amplification product is expected. Thus far, the selection seems to be progressing as expected and should be continued.

Chapter 6: Summary, Conclusions and Future Directions

6.1 Summary

This thesis explores the role of modified nucleosides, including their potential and their limitations. Chapter 2 introduces a synthetic scheme for producing the dA^{ime}TP which is a well-established nucleotide for use in DNAzyme selections. In efforts to find a variable for improving the rate of modified DNAzyme self-cleavage, a selection that assessed our ability to fine-tune nucleotide modifications was described in Chapter 3. To do this we mimicked the selection for Dz10-66 replacing the well-established dA^{ime}TP with the then untested dA^{imm}TP. The use of a shorter linker was thought to be a strategy that would produce a DNAzyme with constrained catalytic residues. Minimizing the allowable conformations of a particular functional group reduces the amount of energy needed to arrange the residues in their active state. After 20 rounds of selection, the fastest DNAzyme, Dz20-49, was measured to have a rate constant of $3.5 \pm 0.4 \cdot 10^{-3} \text{ min}^{-1}$. This rate constant is far lower the rate constant of Dz20-49's immediate predecessor Dz10-66 ($0.63 \pm 0.04 \text{ min}^{-1}$, 37 °C). In addition to the slower rate, Dz20-49 does not seem to possess the same thermal stability shown by Dz10-66. When tested for self-cleavage at 37 °C, the rate ($2.0 \pm 0.5 \cdot 10^{-3} \text{ min}^{-1}$) was roughly half that at 24 °C. The modified nucleoside dA^{imm} proved to be an inferior substitute for dA^{ime} with regards to improving catalytic rate constants and thermostability.

Chapter 4 describes cloning of the final generations of various DNAzyme selections. The cloning of the final DNAzyme generations and analysis of the sequences raise the question of bias from post-selection manipulation such as transcription into unmodified DNA followed by PCR. Suspicions of biased amplifications are supported by the inefficient amplification of the Dz20-49 sequence augmented with all of the appropriate modifications. The same sequence without modifications amplifies cleanly and produces a higher amount of amplicon. Even the Dz20-49 analogs constructed with dA^{ime} or dA^{imp} produce far more amplicon, implying that the discrete changes in linker length that have an effect on catalytic competence also have an effect on amplifiability. These preliminary tests using dA^{imp}TP, however, show that the longer linker is worthy of further studies. Also described in this chapter is the PCR amplification using 5-modified dU analogs to produce doubly modified dsDNA decorated with the modification dU^{ga}, dU^{ph} or dU^{aa} on both strands and the modifications' ability to bestow site specific

restriction enzyme digestion resistance. Introduction of the modifications dU^{ga} or dU^{ph} into the cell via cloning and electroporation addressed whether the cell will tolerate the modification and replicate it with fidelity or if it will trigger a repair mechanism that destroys the plasmid. The sequencing of the harvested plasmids revealed that the modified nucleoside dU^{ga} and dU^{ph} were well tolerated or repaired and replicated with fidelity. This study also hints at the potential for these particular modifications to be used in a therapeutic modified DNAzyme.

Limitations using modified nucleotides prevent access to DNAzymes with higher catalytic efficiency, higher substrate specificity, or novel function. One of these limitations is the inability to incorporate modified nucleotides with the same efficiency as natural nucleotides. Unfortunately, such a limitation has shown that incorporation of these beneficial modifications increases the production of truncation products for heavily modified clone sequences. Chapter 4 showed that sequenced full-length product isolated from numerous selections shows few sequential modified adenosines. In general, there appears to be a pressure that reduces the number of these crucial modifications below what is statistically expected. It is speculated that the polymerases used in our selections are the cause of this bias. Therefore, the evolution of a polymerase that can incorporate modified nucleotide without a bias in efficiency would be desirable. Chapter 5 looks at the steps taken towards this directed evolution of Taq using spCSR. Reverse micelle compartments containing single EGFP-expressing *E. coli* cells were constructed and visualized. The minimization of cross reaction between reverse micelles and the ability to isolate individual cells through emulsification was verified. The generation of a Taq library and successful isolation of active mutants bearing wild-type activity as a control was achieved. The selection for an evolved polymerase capable of incorporating and reading through modified nucleotides is currently underway.

6.2 Conclusions

The modified synthetic scheme used to produce $\text{dA}^{\text{ime}}\text{TP}$ was successful, and Curtis Lam successfully modified the synthetic scheme further to gain access to a small library of 8-modified 2'-deoxyadenosine triphosphate analogs. The selection of Dz20-49 was a success in terms of isolating an active DNAzyme, but did not succeed in producing a faster DNAzyme than its predecessor Dz10-66. What was learned, however, was valuable information about the large effects of minute changes in linker length. If shortening the linker length by one methylene group had such a drastic detrimental effect, could then lengthening the linker by one methylene group have a beneficial effect? Although the main goal was not achieved, these results do add to the knowledge of the modified DNAzyme community, and hopefully will help lead researchers towards asking the important questions. One of the questions that arose from the selection of Chapter 3 is: what is the unknown origin of the bias against the crucial 8-modified dA? Chapter 4 clearly illustrates that the dA^{imm} in modified DNA causes difficulties in read-through and amplification when compared to both the established dA^{ime} and the yet selection-untested dA^{imp} . While this shows yet another failing of the dA^{imm} modification, it also shows that the dA^{imp} modification has untapped potential. Outside of the selection context and a little more in the DNAzyme application context, well tolerated dU^{ga} and dU^{ph} show favorable characteristics of being amenable to PCR amplification, bestowing resistance to restriction enzyme digestion and compatibility with a biological system. This experiment keeps the goal of potential therapeutics in focus while discrete chemical changes are simultaneously examined. The directed evolution of Taq for incorporation of our modified nucleotides has been met with many difficulties. While the ultimate goal of isolating an active mutant polymerase has yet to be achieved, many of the small goals that lead to this main goal have been achieved, and encouraged further study. In our hands, we have shown that thermostable reverse micelle compartments can be made and can be used to control reaction conditions. This technology can be used beyond the selection of polymerases and may one day even be integrated into our selection of DNAzymes.

6.3 Future directions

Since evaluation of the nucleotide bearing the short linker, the dA^{imm}TP, proved that this particular modified nucleoside triphosphate is an inferior modified nucleoside for DNAzyme selections, the next experiment would be to test the dA^{imp}TP. Performing a selection using an imidazole tethered to the nucleobase with a propyl linker will help complete the picture on the effects of linker length for the imidazole functional group tethered to the 8 position of 2'-deoxyadenosine. Since the short linker system did not increase catalytic rate as expected, one should not assume that the longer linker would be detrimental to activity.

The enrichment of the gene pool in active mutants shows that spCSR is a viable method of isolating genes encoding for active mutants capable of amplifying their own gene. With an enriched polymerase library in hand, one could attempt replacing natural nucleoside triphosphates with a nucleoside triphosphate that has been chemically modified. The selection undertaken has yet to be finished, but shows many encouraging results and should be continued. Also, modified primers can be made and the shuffled library can be used in a CSR selection for the same improved polymerase activity.

Directed evolution can be continued on other polymerases once the work on Taq is complete. We have in our possession genes for Family B polymerase Pfu and Family Y polymerase Dpo4. Since there is far less structural data on these two polymerases than what is available on Taq, a low level of mutation throughout the gene should be attempted versus mutating specific regions. We have started working with a shuffled library of Dpo4. This is a potential project for a future student.

References

- Agresti, J. J., Kelly, B. T., Jaschke, A., & Griffiths, A. D. (2005). Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *Proc. Natl. Acad. Sci. USA*, 102(45), 16170-16175.
- Aigner, A. (2007). Applications of RNA interference: current state and prospects for siRNA-based strategies in vivo. *Appl. Microbiol. Biotechnol.*, 76, 9-21.
- Arber, W., & Linn, S. (1969). DNA Modification and Restriction. *Annu. Rev. Biochem.*, 467-500.
- Barnard, E. A. (1969). Ribonucleases. *Annu. Rev. Biochem.*, 38, 677-732.
- Barnes, W. M. (1992). The Fidelity of Taq Polymerase Catalyzing PCR is Improved by an N-Terminal Deletion. *Gene*, 112(1), 29-35.
- Baum, D. A., & Silverman, S. K. (2008). Deoxyribozymes: useful DNA catalysts in vitro and in vivo. *Cell. Mol. Life Sci.*, 65, 2156-2174.
- Blouin, S., Mulhbach, J., Penedo, J. C., & Lafontaine, D. A. (2009). Riboswitches: Ancient and Promising Genetic Regulators. *ChemBioChem*, 10(3), 400-416.
- Boudsocq, F., Iwai, S., Hanaoka, F., & Woodgate, R. (2001). Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic pol eta. *Nucleic Acids Res.*, 29(22), 4607-4616.
- Braithwaite, D. K., & Ito, J. (1993). Compilation, Alignment, and Phylogenetic-Relationships of DNA-Polymerases. *Nucleic Acids Res.*, 21(4), 787-802.
- Brakmann, S. (2005). Directed evolution as a tool for understanding and optimizing nucleic acid polymerase function. *Cell. Mol. Life. Sci.*, 62, 2634-2646.
- Brautigam, C. A., & Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.*, 8(1), 54-63.
- Breaker, R. R., & Joyce, G. F. (1994). A DNA enzyme that cleaves RNA. *Chem. Biol.*, 1, 223-229.
- Breaker, R. R., & Joyce, G. F. (1995). A DNA enzyme with Mg²⁺-dependent RNA phosphoesterase activity. *Chem. Biol.*, 2, 655-660.
- Butcher, S. E. (2001). Structure and function of the small ribozymes. *Current Opinion in Structural Biology*, 11(3), 315-320.
- Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 2(1), 28-33.
- Cahova, H., Havran, L., Brazdilova, P., Pivonkova, H., Fojta, M., & Hocek, M. (2008). Aminophenyl- and nitrophenyl-labeled DNA. Synthesis by polymerase incorporation of nucleoside triphosphates and electrochemical properties. *Chemistry of Nucleic Acid Components*, 10, 178-181.
- Cahova, H., Pohl, R., Bednarova, L., Novakova, K., Cvacka, J., & Hocek, M. (2008). Synthesis of 8-bromo-, 8-methyl- and 8-phenyl-dATP and their polymerase incorporation into DNA. *Org. Biomol. Chem.*, 6(20), 3657-3660.

- Canny, M. D., Jucker, F. M., Kellogg, E., Khvorova, A., Jayasena, S. D., & Pardi, A. (2004). Fast cleavage kinetics of a natural hammerhead ribozyme. *J. Am. Chem. Soc.*, 126(35), 10848-10849.
- Capek, P., Cahova, H., Pohl, R., Hocek, M., Gloeckner, C., & Marx, A. (2007). An efficient method for the construction of functionalized DNA bearing amino acid groups through cross-coupling reactions of nucleoside triphosphates followed by primer extension or PCR. *Chem. Eur. J.*, 13, 6196-6203.
- Cech, T. R. (2000). Structural biology - The ribosome is a ribozyme. *Science*, 289(5481), 878-879.
- Chinnapen, D. J.-F., & Sen, D. (2004). A deoxyribozyme that harnesses light to repair thymine dimers in DNA. *Proc. Natl. Acad. Sci. USA*, 101, 65-69.
- Crick, F. H., Brenner, S., Watstobi.Rj, & Barnett, L. (1961). General Nature of Genetic Code for Proteins. *Nature*, 192(480), 1227-&.
- d'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., *et al.* (2007). Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.*, 25(8), 939-943.
- Delcardayre, S. B., & Raines, R. T. (1994). Structural Determinants of Enzymatic Processivity. *Biochemistry*, 33(20), 6031-6037.
- Desai, U. J., & Pfaffle, P. K. (1995). Single-step Purification of a Thermostable DNA Polymerase Expressed in *Escherichia coli* *Biotechniques*, 19(5), 780-&.
- Dewey, T. M., Zyzniewski, M. C., & Eaton, B. E. (1996). RNA world: Functional diversity in a nucleoside by carboxyamidation of uridine. *Nucleosides & Nucleotides*, 15(10), 1611-1617.
- Doherty, E. A., & Doudna, J. A. (2001). Ribozyme structures and mechanisms. *Annual Review of Biophysics and Biomolecular Structure*, 30, 457-475.
- Dowler, T., Bergeron, D., Tedeschi, A. L., Paquet, L., Ferrari, N., & Damha, M. J. (2006). Improvements in siRNA properties mediated by 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (FANA). *Nucleic Acids Res.*, 34(6), 1669-1675.
- Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346, 818-822.
- Eom, S. H., Wang, J. M., & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, 382(6588), 278-281.
- Fa, M., Radeghieri, A., Henry, A. A., & Romesberg, F. E. (2004). Expanding the Substrate Repertoire of a DNA Polymerase by Directed Evolution. *J. Am. Chem. Soc.*, 126, 1748-1754.
- Faulhammer, D., & Famulok, M. (1996). The Ca²⁺ Ion as a Cofactor for a Novel RNA-Cleaving Deoxyribozyme. *Angew. Chem. Int. Ed.*, 35, 2837-2841.
- Faulhammer, D., & Famulok, M. (1997). Characterization and divalent metal-ion dependence of in vitro selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides. *J. Mol. Biol.*, 269(2), 188-202.
- Fedor, M. J., & Uhlenbeck, O. C. (1992). Kinetics of Intermolecular Cleavage by Hammerhead Ribozymes. *Biochemistry*, 31(48), 12042-12054.
- Fedor, M. J., & Williamson, J. R. (2005). The catalytic diversity of RNAS. *Nat. Rev. Mol. Cell Biol.*, 6(5), 399-412.

- Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806-811.
- Franzen, S. (2010). Expanding the catalytic repertoire of ribozymes and deoxyribozymes beyond RNA substrates. *Curr. Opin. Mol. Ther.*, 12(2), 223-232.
- Geyer, C. R., & Sen, D. (1997). Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chem. Biol.*, 4(8), 579-593.
- Ghadessy, F. J., & Holliger, P. (2004). A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Eng. Des. Sel.*, 17(3), 201-204.
- Ghadessy, F. J., Ong, J. L., & Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA*, 98(8), 4552-4557.
- Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., *et al.* (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.*, 22(6), 755-759.
- Gilbert, W. (1986). Origin of Life - The RNA World. *Nature*, 319(6055), 618-618.
- Gourlain, T., Sidorov, A., Mignet, N., Thorpe, S. J., Lee, S. E., Grasby, J. A., *et al.* (2001). Enhancing the catalytic repertoire of nucleic acids. II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR. *Nucleic Acids Res.*, 29(9), 1898-1905.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell*, 35, 849-857.
- Hanes, J., & Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA*, 94(10), 4937-4942.
- Held, H. A., & Benner, S. A. (2002). Challenging artificial genetic systems: thymidine analogs with 5-position sulfur functionality. *Nucleic Acids Res.*, 30, 3857-3869.
- Held, H. A., Roychowdhury, A., & Benner, S. A. (2003). C-5 modified nucleosides: Direct insertion of alkynyl-thio functionality in pyrimidines. *Nucleosides Nucleotides*, 22, 391-404.
- Hobartner, C., & Silverman, S. K. (2007). Recent advances in DNA catalysis. *Biopolymers*, 87(5-6), 279-292.
- Hocek, M., & Fojta, M. (2008). Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. Construction and applications of base-functionalized nucleic acids. *Org. Biomol. Chem.*, 6, 2233-2241.
- Hollenstein, M., Hipolito, C., Lam, C., Dietrich, D., & Perrin, D. M. (2008). A highly selective DNazyme sensor for mercuric ions. *Angew. Chem. Int. Ed.*, 47(23), 4346-4350.
- Hollenstein, M., Hipolito, C., Lam, C., & Perrin, D. M. (2009a). A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M²⁺). *Nucleic Acids Res.*, 37, 1638-1649.
- Hollenstein, M., Hipolito, C. J., Lam, C. H., & Perrin, D. M. (2009b). A DNazyme with Three Protein-Like Functional Groups: Enhancing Catalytic Efficiency of M²⁺-Independent RNA Cleavage. *ChemBioChem*, 10(12), 1988-1992.

- Ikehara, M., & Kaneko, M. (1970). STUDIES OF NUCLEOSIDES AND NUCLEOTIDES-XLI PURINE CYCLONUCLEOSIDE-8 SELECTIVE SULFONYLATION OF 8-BROMOADENOSINE DERIVATIVES AND AN ALTERNATIVE SYNTHESIS OF 8,2'-S-CYCLONUCLEOSIDES AND 8,3'-S-CYCLONUCLEOSIDES. *Tetrahedron*, 26(18), 4251-4259.
- Jäger, S., & Famulok, M. (2004). Generation and Enzymatic Amplification of High-Density Functionalized DNA Double Strands. *Angew. Chem. Int. Ed.*, 43, 3337-3340.
- Jäger, S., Rasched, G., Kornreich-Leshem, H., Engeser, M., Thum, O., & Famulok, M. (2005). A versatile toolbox for variable DNA functionalization at high density. *J. Am. Chem. Soc.*, 127(43), 15071-15082.
- Jestin, J. L., Volioti, G., & Winter, G. (2001). Improving the display of proteins on filamentous phage. *Res. Microbiol.*, 152(2), 187-191.
- Joyce, G. F. (2004). Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, 73, 791-836.
- Kalota, A., Karabon, L., Swider, C. R., Viazovkina, E., Elzagheid, M., Damha, M. J., *et al.* (2006). 2'-Deoxy-2'-fluoro-beta-D-arabinonucleic acid (2' F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing. *Nucleic Acids Res.*, 34(2), 451-461.
- Khan, A. U. (2006). Ribozyme: A clinical tool. *Clinica Chimica Acta*, 367(1-2), 20-27.
- Khvorova, A., Lescoute, A., Westhof, E., & Jayasena, S. D. (2003). Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Structural Biology*, 10(9), 708-712.
- Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S., & Hirao, I. (2009). An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.*, 37(2).
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell*, 31, 147-157.
- Kuwahara, M., Hanawa, K., Ohsawa, K., Kitagata, R., Ozaki, H., & Sawai, H. (2006). Direct PCR amplification of various modified DNAs having amino acids: Convenient preparation of DNA libraries with high-potential activities for in vitro selection. *Bioorg. Med. Chem.*, 14(8), 2518-2526.
- Kuwahara, M., Hososhima, S.-i., Takahata, Y., Kitagata, R., Shoji, A., Hanawa, K., *et al.* (2003). Simultaneous incorporation of three different modified nucleotides during PCR. *Nucleic Acids Res Suppl*(3), 37-38.
- Kuwahara, M., Nagashima, J., Hasegawa, M., Tamura, T., Kitagata, R., Hanawa, K., *et al.* (2006). Systematic characterization of 2'-deoxynucleoside-5'-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Res.*, 34(19), 5383-5394.
- Kuwahara, M., Takahata, Y., Shoji, A., Ozaki, A. N., Ozaki, H., & Sawai, H. (2003). Substrate properties of C5-substituted pyrimidine 2'-deoxynucleoside 5'-triphosphates for thermostable DNA polymerases during PCR. *Bioorg. Med. Chem. Lett.*, 13, 3735-3738.

- Kuwahara, M., Tamura, T., Kitagata, R., Sawai, H., & Ozaki, H. (2005). Comparison study on PCR amplification of modified DNA by using various kinds of polymerase and modified nucleoside triphosphates. *Nucleic Acids Symp Ser (Oxf)*(49), 275-276.
- Lacks, S. A., Mannarelli, B. M., Springhorn, S. S., Greenberg, B., & De La Campa, A. G. (1987). Genetics of the Complementary Restriction Systems DPN-I and DPN-II Revealed by Cloning and Recombination in *Streptococcus pneumoniae*. *Ferretti, J. J. and R. Curtiss, Iii (Ed.). Streptococcal Genetics; Second Asm (American Society for Microbiology) Conference, Miami, Florida, USA, May 21-24, 1986. Viii+300p. American Society for Microbiology: Washington, D.C., USA. Illus, 31-42.*
- Lam, C., Hipolito, C., & Perrin, D. M. (2008). Synthesis and Enzymatic Incorporation of Modified Deoxyadenosine Triphosphates. *Eur. J. Org. Chem.*, 4915-4923.
- Latham, J. A., Johnson, R., & Toole, J. J. (1994). The Application of a Modified Nucleotide in Aptamer Selection - Novel Thrombin Aptamers Containing 5-(1-Pentynyl)-2'-Deoxyuridine. *Nucleic Acids Res.*, 22(14), 2817-2822.
- Lau, M. W., Cadieux, K. E., & Unrau, P. J. (2004). Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.*, 126, 15686-15693.
- Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., *et al.* (1993). High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl*, 2(4), 275-287.
- Leconte, A. M., Chen, L. J., & Romesberg, F. E. (2005). Polymerase evolution: Efforts toward expansion of the genetic code. *J. Am. Chem. Soc.*, 127(36), 12470-12471.
- Lee, S. E., Sidorov, A., Gourlain, T., Mignet, N., Thorpe, S. J., Brazier, J. A., *et al.* (2001). Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.*, 29(7), 1565-1573.
- Lerner, L., Hobbs, J., & Perrin, D. M. (2002). Incorporation of 8-histaminyldoxyadenosine [8-(2-(4-imidazolyl)ethylamino)-2'-deoxyriboadenosine] into oligodeoxyribonucleotides by solid phase phosphoramidite coupling. *Nucleosides Nucleotides*, 21, 651-664.
- Lerner, L., Roupioz, Y., Ting, R., & Perrin, D. M. (2002). Toward an RNaseA mimic: a DNAzyme with imidazoles and cationic amines. *J. Am. Chem. Soc.*, 124, 9960-9961.
- Li, J., Zheng, W., Kwon, A. H., & Lu, Y. (2000). In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme. *Nucleic Acids Res.*, 28, 481-488.
- Li, Y. F., & Sen, D. (1996). A catalytic DNA for porphyrin metallation. *Nature Structural Biology*, 3(9), 743-747.
- 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. *Proc. Natl. Acad. Sci. USA*, 96, 2056-2061.
- Liu, J., & Lu, Y. (2003). A Colorimetric Lead Biosensor Using DNAzyme-Directed Assembly of Gold Nanoparticles. *J. Am. Chem. Soc.*, 125, 6642-6643.

- Liu, J., & Lu, Y. (2004). Accelerated Color Change of Gold Nanoparticles Assembled by DNAzymes for Simple and Fast Colorimetric Pb²⁺ Detection. *J. Am. Chem. Soc.*, *126*, 12298-12305.
- Liu, J., & Lu, Y. (2007a). Colorimetric Cu²⁺ detection with a ligation DNAzyme and nanopairticles. *Chem. Commun.*, 4872-4874.
- Liu, J., & Lu, Y. (2007b). A DNAzyme catalytic beacon sensor for paramagnetic Cu²⁺ ions in aqueous solution with high sensitivity and selectivity. *J. Am. Chem. Soc.*, *129*, 9838-9839.
- Liu, J., & Lu, Y. (2007c). Rational design of "Turn-On" allosteric DNAzyme catalytic beacons for aqueous mercury ions with ultrahigh sensitivity and selectivity. *Angew. Chem. Int. Ed.*, *46*, 7587-7590.
- Liu, Z., Mei, S. H. J., Brennan, J. D., & Li, Y. (2003). Assemblage of Signaling DNA Enzymes with Intriguing Metal-Ion Specificities and pH Dependences. *J. Am. Chem. Soc.*, *125*, 7539-7545.
- Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T., & Holliger, P. (2009). Evolving a Polymerase for Hydrophobic Base Analogues. *J. Am. Chem. Soc.*, *131*(41), 14827-14837.
- Loakes, D., & Holliger, P. (2009). Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.*(31), 4619-4631.
- Ludwig, J., & Eckstein, F. (1989). Rapid and Efficient Synthesis of 5'-O-(1-Thiotriphosphates), 5'-Triphosphates and 2',3'-Cyclophosphorothioates Using 2-Chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, *54*, 631-635.
- Macickova-Cahova, H., & Hock, M. (2009). Cleavage of adenine-modified functionalized DNA by type II restriction endonucleases. *Nucleic Acids Res.*, *37*(22), 7612-7622.
- Masud, M. M., Kuwahara, M., Ozaki, H., & Sawai, H. (2004). Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX. *Bioorg. Med. Chem.*, *12*(5), 1111-1120.
- Mills, D. R., Peterson, R. L., & Spiegelman, S. (1967). An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule. *Proc. Natl. Acad. Sci. USA*, *58*(1), 217-&.
- Murphy, E., Freudenrich, C. C., Levy, L. A., London, R. E., & Lieberman, M. (1989). Monitoring Cytosolic Free Magnesium in Cultured Chicken Heart-Cells by use of the Fluorescent Indicator Fura-2. *Proc. Natl. Acad. Sci. USA*, *86*(8), 2981-2984.
- Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M., & Scott, W. G. (1998). The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.*, *5*, 587-595.
- Obayashi, T., Masud, M. M., Ozaki, A. N., Ozaki, H., Kuwahara, M., & Sawai, H. (2002). Enzymatic synthesis of labeled DNA by PCR using new fluorescent thymidine nucleotide analogue and superthermophilic KOD Dash DNA polymerase. *Bioorg. Med. Chem. Lett.*, *12*(8), 1167-1170.
- Ohbayashi, T., Kuwahara, M., Hasegawa, M., Kasamatsu, T., Tamura, T., & Sawai, H. (2005). Expansion of repertoire of modified DNAs prepared by PCR using KOD dash DNA polymerase. *Org. Biomol. Chem.*, *3*(13), 2463-2468.

- Ohmichi, T., Kuwahara, M., Sasaki, N., Hasegawa, M., Nishikata, T., Sawai, H., *et al.* (2005). Nucleic Acid with Guanidinium Modification Exhibits Efficient Cellular Uptake. *Angew. Chem. Int. Ed.*, 44, 6682-6685.
- Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., *et al.* (2001). The Y-family of DNA polymerases. *Molecular Cell*, 8(1), 7-8.
- Ohsawa, K., Kasamatsu, T., Nagashima, J. I., Hanawa, K., Kuwahara, M., Ozaki, H., *et al.* (2008). Arginine-modified DNA aptamers that show enantioselective recognition of the dicarboxylic acid moiety of glutamic acid. *Anal. Sci.*, 24(1), 167-172.
- Ong, J. L., Loakes, D., Jaroslawski, K. T., & Holliger, P. (2006). Directed Evolution of DNA Polymerase, RNA polymerase and Reverse Transcriptase Activity in a single polypeptide. *J. Mol. Biol.*, 361, 537-550.
- Patel, P. H., Kawate, H., Adman, E., Ashbach, M., & Loeb, L. K. (2001). A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.*, 276(7), 5044-5051.
- Patel, P. H., & Loeb, L. A. (2000a). DNA polymerase active site is highly mutable: Evolutionary consequences. *Proc. Natl. Acad. Sci. USA*, 97(10), 5095-5100.
- Patel, P. H., & Loeb, L. A. (2000b). Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.*, 275(51), 40266-40272.
- Patel, P. H., & Loeb, L. A. (2001). Getting a grip on how DNA polymerases function. *Nature Structural Biology*, 8(8), 656-659.
- Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., & Loeb, L. A. (2001). Prokaryotic DNA polymerase I: Evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.*, 308(5), 823-837.
- Peng, C. G., & Damha, M. J. (2007). Polymerase-directed synthesis of 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acids. *J. Am. Chem. Soc.*, 129(17), 5310-+.
- Peracchi, A. (2004). Prospects for antiviral ribozymes and deoxyribozymes. *Rev. Med. Virol.*, 14(1), 47-64.
- Perrin, D. M., Garestier, T., & Hélène, C. (1999). Expanding the catalytic repertoire of nucleic acid catalysts: Simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides*, 18, 377-391.
- Perrin, D. M., Garestier, T., & Hélène, C. (2001). Bridging the gap between proteins and nucleic acids: A metal-independent RNaseA mimic with two protein-like functionalities. *J. Am. Chem. Soc.*, 123, 1556-1563.
- Pradeepkumar, P. I., Höbartner, C., Baum, D. A., & Silverman, S. K. (2008). DNA-catalyzed formation of nucleopeptide linkages. *Angew. Chem. Int. Ed.*, 47, 1753-1757.
- Prakash, T. P., Krishna Kumar, R., & Ganesh, K. N. (1993). Synthesis and Conformational Studies of d(TpA) and r(UpA) Conjugated with Histamine and Ethylenediamine. *Tetrahedron*, 49, 4035-4050.
- Prakash, T. P., Puschl, A., & Manoharan, M. (2007). N,N'-Bis-(2-(cyano)ethoxycarbonyl)-2-methyl-2-thiopseudourea A Guanylylating Reagent for Synthesis of 2'-O-[2-(Guanidinium)ethyl]-Modified Oligonucleotides. *Nucleosides Nucleotides*, 26, 149-159.

- Purtha, W. R., Coppins, R. L., Smalley, M. K., & Silverman, S. K. (2005). General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. *J. Am. Chem. Soc.*, *127*, 13124-13125.
- Raines, R. T. (1998). Ribonuclease A. *Chem. Rev.*, *98*, 1045-1065.
- Ramsay, N., Jemth, A. S., Brown, A., Crampton, N., Dear, P., & Holliger, P. (2010). CyDNA: Synthesis and Replication of Highly Cy-Dye Substituted DNA by an Evolved Polymerase. *J. Am. Chem. Soc.*, *132*(14), 5096-5104.
- Roberts, R. W., & Szostak, J. W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA*, *94*(23), 12297-12302.
- Robertson, D. L., & Joyce, G. F. (1990). Selection In Vitro of an RNA Enzyme that Specifically Cleaves Single-Stranded-DNA. *Nature*, *344*(6265), 467-468.
- Roig, V., & Asseline, U. (2003). Oligo-2'-deoxyribonucleotides containing uracil modified at the 5-position with linkers ending with guanidinium groups. *J. Am. Chem. Soc.*, *125*(15), 4416-4417.
- Roth, A., & Breaker, R. R. (1998). An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Natl. Acad. Sci. USA*, *95*(11), 6027-6031.
- Roychowdhury-Saha, M., & Burke, D. H. (2006). Extraordinary rates of transition metal ion-mediated ribozyme catalysis. *RNA*, *12*(10), 1846-1852.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., *et al.* (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA-Polymerase. *Science*, *239*(4839), 487-491.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., *et al.* (1985). Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia. *Science*, *230*(4732), 1350-1354.
- Sakthivel, K., & Barbas, C. F. (1998). Expanding the potential of DNA for binding and catalysis: Highly functionalized dUTP derivatives that are substrates for thermostable DNA polymerases. *Angew. Chem. Int. Ed.*, *37*, 2872-2875.
- SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, *95*(4), 1460-1465.
- Santoro, S. W., & Joyce, G. F. (1997). A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*, *94*, 4262-4266.
- Santoro, S. W., Joyce, G. F., Sakthivel, K., Gramatikova, S., & Barbas, C. F. (2000). RNA cleavage by a DNA enzyme with extended chemical functionality. *J. Am. Chem. Soc.*, *122*, 2433-2439.
- Sawai, H., Nagashima, J., Kuwahara, M., Kitagata, R., Tamura, T., & Matsui, I. (2007). Differences in substrate specificity of C(5)-substituted or C(5)unsubstituted pyrimidine nucleotides by DNA Polymerases from thermophilic bacteria, archaea, and phages. *Chemistry & Biodiversity*, *4*, 1979-1995.
- Sawai, H., Ozaki, A. N., Satoh, F., Ohbayashi, T., Masud, M. M., & Ozaki, H. (2001). Expansion of structural and functional diversities of DNA using new 5-substituted deoxyuridine derivatives by PCR with superthermophilic KOD Dash DNA polymerase. *Chem. Commun.*, *24*, 2604-2605.
- Schlosser, K., Lam, J. C. F., & Li, Y. F. (2009). A genotype-to-phenotype map of in vitro selected RNA-cleaving DNazymes: implications for accessing the target phenotype. *Nucleic Acids Res.*, *37*(11), 3545-3557.

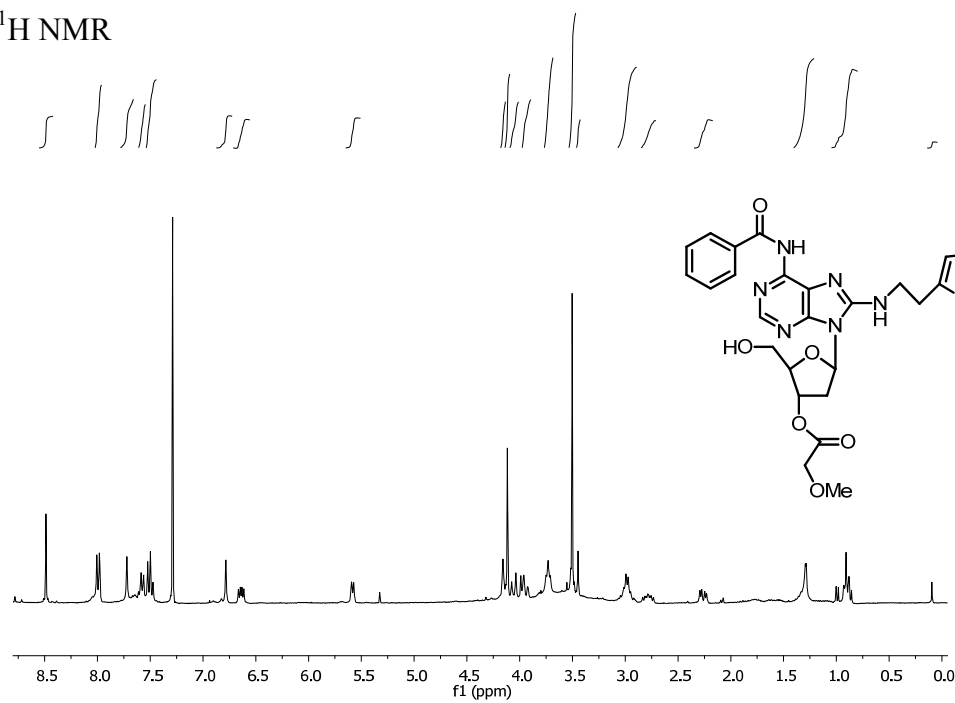
- Schlosser, K., & Li, Y. (2009). Biologically inspired synthetic enzymes made from DNA. *Chem. Biol.*, *16*, 311-322.
- Schmidt, N., Mishra, A., Lai, G. H., & Wong, G. C. L. (2010). Arginine-rich cell-penetrating peptides. *FEBS Letters*, *584*(9), 1806-1813.
- Schonbrunner, N. J., Fiss, E. H., Budker, O., Stoffel, S., Sigua, C. L., Gelfand, D. H., *et al.* (2006). Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3'-5' exonuclease activity. *Biochemistry*, *45*(42), 12786-12795.
- Seelig, B., & Jäschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.*, *6*, 167-176.
- Shoji, A., Kuwahara, M., Ozaki, H., & Sawai, H. (2007). Modified DNA aptamer that binds the (R)-Isomer of a thalidomide derivative with high enantioselectivity. *J. Am. Chem. Soc.*, *129*(5), 1456-1464.
- Sidorov, A. V., Grasby, J. A., & Williams, D. M. (2004). Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. *Nucleic Acids Res.*, *32*, 1591-1601.
- Silverman, S. K. (2004). Deoxyribozymes: DNA catalysts for bioorganic chemistry. *Org. Biomol. Chem.*, *2*, 2701-2707.
- Silverman, S. K. (2005). In vitro selection, characterization, and application of deoxyribozymes that cleave RNA. *Nucleic Acids Res.*, *33*, 6151-6163.
- Silverman, S. K. (2008). Catalytic DNA (deoxyribozymes) for synthetic applications - current abilities and future prospects. *Chem. Commun.*, *30*, 3467-3485.
- Singh, D., Kumar, V., & Ganesh, K. N. (1990). Oligonucleotides, Part 5: Synthesis and Fluorescence Studies of DNA Oligomers d(AT)₅ Containing Adenines Covalently Linked at C-8 with Dansyl Fluorophore. *Nucleic Acids Res.*, *18*(11), 3339-3345.
- Stanislawska-Sachadyn, A., & Sachadyn, P. (2005). MutS as a tool for mutation detection. *Acta Biochimica Polonica*, *52*(3), 575-583.
- Steitz, T. A. (1998). Structural biology - A mechanism for all polymerases. *Nature*, *391*(6664), 231-232.
- Stemmer, W. P. C. (1994). Rapid Evolution of a Protein In Vitro by DNA Shuffling. *Nature*, *370*(6488), 389-391.
- Suzuki, M., Avicola, A. K., Hood, L., & Loeb, L. A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.*, *272*(17), 11228-11235.
- Suzuki, M., Baskin, D., Hood, L., & Loeb, L. A. (1996). Random mutagenesis of *Thermus aquaticus* DNA polymerase I: Concordance of immutable sites in vivo with the crystal structure. *Proc. Natl. Acad. Sci. USA*, *93*(18), 9670-9675.
- Tabor, S., & Richardson, C. C. (1989). Selective Inactivation of the Exonuclease Activity of Bacteriophage-T7 DNA-Polymerase by In Vitro Mutagenesis. *J. Biol. Chem.*, *264*(11), 6447-6458.
- Tarasow, T. M., Tarasow, S. L., & Eaton, B. E. (1997). RNA-catalysed carbon-carbon bond formation. *Nature*, *389*(6646), 54-57.
- Tawfik, D. S., & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, *16*(7), 652-656.
- Thomas, J. M., Ting, R., & Perrin, D. M. (2004). High affinity DNAzyme-based ligands for transition metal cations - a prototype sensor for Hg²⁺. *Org. Biomol. Chem.*, *2*, 307-311.

- Thomas, J. M., Yoon, J. K., & Perrin, D. M. (2009). Investigation of the Catalytic Mechanism of a Synthetic DNAzyme with Protein-like Functionality: An RNaseA Mimic? *J. Am. Chem. Soc.*, *131*(15), 5648-5658.
- Thum, O., Jager, S., & Famulok, M. (2001). Functionalized DNA: A new replicable biopolymer. *Angew. Chem. Int. Ed.*, *40*(21), 3990-3993.
- Ting, R., Lermer, L., & Perrin, D. M. (2004). Triggering DNAzymes with light: A photoactive C8 thioether-linked adenosine. *J. Am. Chem. Soc.*, *126*(40), 12720-12721.
- Ting, R., Thomas, J. M., Lermer, L., & Perrin, D. M. (2004). Substrate specificity and kinetic framework of a DNAzyme with an expanded chemical repertoire: a putative RNaseA mimic that catalyzes RNA hydrolysis independent of a divalent metal cation. *Nucleic Acids Res.*, *32*, 6660-6672.
- Ting, R., Thomas, J. M., & Perrin, D. M. (2007). Kinetic characterization of a cis- and trans-acting M²⁺-independent DNAzyme that depends on synthetic RNaseA-like functionality - Burst-phase kinetics from the coalescence of two active DNAzyme folds. *Can. J. Chem.*, *85*(4), 313-329.
- Trautwein, K., Holliger, P., Stackhouse, J., & Benner, S. A. (1991). Site-Directed Mutagenesis of Bovine Pancreatic Ribonuclease - Lysine-41 and Aspartate-121. *Febs Letters*, *281*(1-2), 275-277.
- Travascio, P., Li, Y. F., & Sen, D. (1998). DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chemistry & Biology*, *5*(9), 505-517.
- Tsukiji, S., Pattnaik, S. B., & Suga, H. (2003). An alcohol dehydrogenase ribozyme. *Nat. Struct. Mol. Biol.*, *10*(9), 713-717.
- Tsukiji, S., Pattnaik, S. B., & Suga, H. (2004). Reduction of an aldehyde by a NADH/Zn²⁺-dependent redox active ribozyme. *J. Am. Chem. Soc.*, *126*(16), 5044-5045.
- Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage -T4 DNA-polymerase. *Science*, *249*, 505-510.
- Unrau, P. J., & Bartel, D. P. (1998). RNA-catalysed nucleotide synthesis. *Nature*, *395*, 260-263.
- Vaish, N. K., Fraley, A. W., Szostak, J. W., & McLaughlin, L. W. (2000). Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids. *Nucleic Acids Res.*, *28*, 3316-3322.
- Vaught, J. D., Bock, C., Carter, J., Fitzwater, T., Otis, M., Schneider, D., *et al.* (2010). Expanding the Chemistry of DNA for in Vitro Selection. *J. Am. Chem. Soc.*, *132*(12), 4141-4151.
- Vichier-Guerre, S., Ferris, S., Auburger, N., Mahiddine, K., & Jestin, J. L. (2006). A population of thermostable reverse transcriptases evolved from *Thermus aquaticus* DNA polymerase I by phage display. *Angew. Chem. Int. Ed.*, *45*(37), 6133-6137.
- Wang, Y., & Silverman, S. K. (2005a). Directing the outcome of deoxyribozyme selections to favor native 3'-5' RNA ligation. *Biochemistry*, *44*, 3017-3023.
- Wang, Y., & Silverman, S. K. (2005b). Efficient one-step synthesis of biologically related lariat RNAs by a deoxyribozyme. *Angew. Chem. Int. Ed.*, *44*, 5863-5866.

- Winkler, W. C., & Breaker, R. R. (2003). Genetic control by metabolite-binding riboswitches. *ChemBioChem*, 4(10), 1024-1032.
- Wu, Q. J., Huang, L., & Zhang, Y. (2009). The structure and function of catalytic RNAs. *Science in China Series C-Life Sciences*, 52(3), 232-244.
- Xia, G., Chen, L. J., Sera, T., Fa, M., Schultz, P. G., & Romesberg, F. E. (2002). Directed evolution of novel polymerase activities: Mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA*, 99(10), 6597-6602.
- Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., & Yanagawa, H. (2003). DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res.*, 31(19).
- Zaher, H. S., & Unrau, P. J. (2007). Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *Rna-a Publication of the Rna Society*, 13(7), 1017-1026.
- Zhao, H. M., & Arnold, F. H. (1997). Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.*, 25(6), 1307-1308.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13), 3406-3415.

Appendix A: NMR spectra for dA^{imm}TP

¹H NMR



¹³C NMR

