STRUCTURAL CHARACTERIZATION OF AN RNA POLYMERASE RIBOZYME AND FUNCTIONAL IMPROVEMENT BY IN VITRO ENCAPSULATED SELECTION

by

Leslie Ka Lok Cheng
B.Sc., Simon Fraser University, 2006
Molecular Biology and Biochemistry

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

MASTER OF SCIENCE

In the Department of
Molecular Biology and Biochemistry

© Leslie Ka Lok Cheng 2009

SIMON FRASER UNIVERSITY

Summer 2009

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.
Approval

Name: Leslie Ka Lok Cheng  
Degree: Master of Science  
Title of Thesis: Structural Characterization of an RNA Polymerase Ribozyme and Functional Improvement by In Vitro Encapsulated Selection

Examining Committee:

Chair: Dr. Sharon M. Gorski  
Associate Professor of Molecular Biology and Biochemistry  
Senior Scientist of the B.C. Genome Sciences Centre

Dr. Peter J. Unrau  
Associate Professor of Molecular Biology and Biochemistry  
Senior Supervisor

Dr. Dipankar Sen  
Professor of Molecular Biology and Biochemistry  
Committee Member

Dr. Melanie A. O’Neill  
Assistant Professor of Chemistry  
Committee Member

Dr. Edgar C. Young  
Assistant Professor of Molecular Biology and Biochemistry  
Internal Examiner

Date Defended/Approved: July 7, 2009
Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website <http://ir.lib.sfu.ca>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, BC, Canada
Abstract

Structural Characterization of an RNA Polymerase Ribozyme and Functional Improvement by In Vitro Encapsulated Selection

At the heart of the “RNA World” hypothesis is the requirement for an RNA polymerase ribozyme capable of replicating itself and other RNA molecules. The most recent variant, isolated from multiple generations of in vitro selection, is the B6.61 RNA polymerase ribozyme that is capable of extending an RNA primer in a template-dependent fashion by 20 nucleotides. It contains a 5' ligase core domain that confers phosphodiester bond formation ability and a poorly understood 3' accessory domain that is proposed to attribute its polymerization ability. In my research presented in this thesis, we study the structure of the accessory domain of B6.61 by mutagenesis, chemical probing and crosslinking experiments and provide molecular constraints that are useful for modeling the three-dimensional structure of the ribozyme. Secondly, I have used in vitro encapsulated selection in an attempt to improve the overall processivity of B6.61.

For my loved ones.
Acknowledgements

My sincere thanks go to my senior supervisor, Dr. Peter Unrau, for his continued support and encouragement in the past years. He has given me many opportunities and has just taught me so much. My thanks also go to Hani Zaher, a previous graduate student in the Unrau laboratory, for working with me when I began as an undergraduate, and offering his help whenever I needed it. I would like to thank members of my supervisory committee, Dr. Dipankar Sen and Dr. Melanie O’Neill, for their helpful advice and support. My appreciation goes to Dr. Edgar Young and Dr. Sharon Gorski for taking time out of their busy schedule to attend my oral defense.

I would like to thank Sunny Wang, a previous graduate student in the Unrau laboratory, for her collaboration in the characterization of the B6.61 ribozyme (Chapter 3). My thanks go to members of the Unrau laboratory for their friendship and support and also to members of the Sen and Davidson laboratories for lending me equipment and offering valuable technical advice.

This research was funded by Simon Fraser University, the Natural Sciences and Engineering Research Council of Canada, and the government of British Columbia.

Finally, I would like to express my deepest gratitude to my family and loved ones for their continued support and prayers. May all glory go to God in heaven for His unconditional love and abundant blessings. Thank You, Father, for everything.
# Table of Contents

Approval ...................................................................................................................... ii
Abstract ..................................................................................................................... iii
Dedication ................................................................................................................... iv
Acknowledgements .................................................................................................. v
Table of Contents ..................................................................................................... vi
List of Figures ........................................................................................................... ix
List of Tables ............................................................................................................ xii
List of Abbreviations .............................................................................................. xiii

## CHAPTER 1: The Biochemistry of RNA ..................................................................... 1

1.1 Ribonucleic Acid and the “RNA World” Hypothesis ............................................. 1
1.1.1 The Structure of RNA ....................................................................................... 1
1.1.2 Ribozymes and the “RNA World” Hypothesis ................................................. 2
1.2 In Vitro Evolution of RNA .................................................................................... 5
1.2.1 Systematic Evolution of Ligands by Exponential Enrichment (SELEX) .......... 5
1.2.2 In Vitro Selection of RNA Catalysts ................................................................. 6

## CHAPTER 2: Closing the Circle – Replicating RNA with RNA .................................... 12

2.1 Introduction ......................................................................................................... 12
2.1.1 From an Abiotic World to a Biotic RNA World ................................................ 13
2.1.1.1 Abiotic Chemistries ..................................................................................... 13
2.1.1.2 Replication Strategies in an RNA World .................................................... 14
2.1.2 RdRP Initiation Mechanisms ........................................................................... 17
2.1.3 RdRP Elongation Mechanisms ...................................................................... 18
2.2 Background: Small Molecule Chemistry in an RNA World ............................ 21
2.2.1 The Chemistry of Nucleotide Synthesis: Glycosidic Bond Formation .......... 21
2.2.2 The Chemistry of Ligation: Phosphodiester Bond Formation .................... 23
2.2.3 RNA-Catalyzed Template-Directed RNA Polymerization ......................... 24
2.3 Recent Results .................................................................................................... 25
2.3.1 Selection of the B6.61 RNA Polymerase Ribozyme ....................................... 25
2.3.2 The Evolutionary Power of Constructing Modular RNA Catalysts ............ 27
2.4 Challenges: Steering a Path Between the Scylla and Charybdis ....................... 31
2.5 Research Directions ......................................................................................... 32
2.5.1 Addressing Processivity ................................................................................. 32
2.5.2 Solving the Strand Displacement Problem ..................................................... 33
CHAPTER 3: Structural and Functional Studies of an RNA Polymerase Ribozyme

3.1 Introduction ........................................................................................................... 36
3.1.1 RNA Polymerase Ribozymes for the “RNA World” Hypothesis ..................... 36
3.1.2 Further Characterization of B6.61 ................................................................... 38
3.1.3 Collaborative Efforts ....................................................................................... 38
3.2 Materials and Methods ....................................................................................... 39
3.2.1 Oligonucleotides ............................................................................................. 39
3.2.2 Polymerization Assessments ............................................................................ 39
3.2.2.1 Polymerization Assay ................................................................................ 39
3.2.2.2 Kinetic Analysis .......................................................................................... 40
3.2.3 Tethering the Primer-Template to Tagged Ribozyme Constructs ................. 41
3.2.3.1 Tag Sequence ............................................................................................. 41
3.2.3.2 Chemical Synthesis of DNA-PEG-Crosslinker-RNA Construct ............... 41
3.2.3.3 Construction of Tagged Ribozyme Variants .............................................. 42
3.2.3.4 Verification of Hybridization by Native Gel Shift .................................... 43
3.2.4 Synthesis of Other Ribozyme Constructs ..................................................... 44
3.2.5 Diethylpyrocarbonate (DEPC) Probing ......................................................... 44
3.2.6 4-thio-Uridine-Mediated UV Crosslinks ....................................................... 45
3.2.7 Primer-Ribozyme Photo-Crosslinking Studies .............................................. 45
3.3 Results .................................................................................................................. 46
3.3.1 Increase in Catalytic Rate Promoted by Increased PT Local Concentration ........................................................................................................... 46
3.3.2 Trans Cooperation of the Two Modular Polymerase Domains ...................... 56
3.3.3 Refining the Secondary Structure of the Accessory Domain ......................... 62
3.3.4 The Minimal Core Motif of the Accessory Domain ....................................... 63
3.3.5 Restoration of Unimolecularity with the tD.1 Accessory Domain .................. 67
3.3.6 Diethylpyrocarbonate Probing of the tA.1 and tD.1 Accessory Domains ........ 69
3.3.7 Elucidation of Essential Residues for Polymerase Activity ............................ 71
3.3.8 4S-U-Mediated UV Crosslinking Between the Ligase Core and Accessory Domain .................................................................................................................. 73
3.3.8.1 4S-U-Containing Polymerase Ribozymes are Active ............................... 74
3.3.8.2 Generation of Crosslinks and Assessment of Their Activities ................. 76
3.3.8.3 Mapping of Crosslinks by Partial Alkaline Hydrolysis .............................. 77
3.3.8.4 Mutation Analysis of Crosslinked Residues ............................................. 81
3.3.9 Initial Attempts to Probe the Primer-Template Binding Site in B6.61 ............. 83
3.3.10 Three-Dimensional Modelling of the Accessory Domain ............................ 87
3.4 Discussion ............................................................................................................. 90
3.4.1 Modular Architecture of B6.61 ...................................................................... 90
3.4.2 A Structural Model of B6.61 ......................................................................... 91
3.5 Contributions ....................................................................................................... 94
CHAPTER 4: Improvement of an RNA Polymerase Ribozyme by In Vitro Encapsulated Selection ................................................................. 95
  4.1 Introduction ........................................................................................................ 95
  4.2 Materials and Methods ...................................................................................... 96
      4.2.1 Oligonucleotides .................................................................................... 96
      4.2.2 Construction of Selection Pool ............................................................. 97
      4.2.3 Encapsulation Procedure ...................................................................... 98
      4.2.4 Recovery of Aqueous Phase and Selection .......................................... 99
      4.2.5 Polymerization Assay .......................................................................... 101
      4.2.6 Optimization of Hybridization-Based Capture ................................... 102
      4.2.7 Assessment of Pool Diversity ............................................................... 102
      4.2.8 Kinetic Analysis ..................................................................................... 102
      4.2.9 Cloning .................................................................................................... 103
      4.2.10 Generating Recombinations of Helper Domains and Intact B6.61 ........ 103
  4.3 Results............................................................................................................. 103
      4.3.1 Construction of Selection Pool ............................................................. 103
      4.3.2 Selection Scheme and Proof of Principle ............................................. 106
      4.3.3 In Vitro Encapsulated Selection ........................................................... 113
          4.3.3.1 Initial Selection Led to an Artifact ............................................... 113
          4.3.3.2 Assessment of Activity and Diversity of RNA Pools from Selection Rounds ........................................................................... 116
          4.3.3.3 Initial Analysis of Isolates from Round 7 and Round 8 ............... 121
          4.3.3.4 Recombining Helper Domains with Intact B6.61 Sequence .......... 129
  4.4 Discussion....................................................................................................... 130
  4.5 Contributions .................................................................................................. 132

CHAPTER 5: Conclusions ...................................................................................... 133
  5.1 Structural and Functional Studies of an RNA Polymerase Ribozyme .......... 133
  5.2 Improvement of an RNA Polymerase Ribozyme by In Vitro Encapsulated Selection ........................................................................... 134

Appendix: First-Nucleotide Extension Rates for Ribozymes Used in Chapter 3 .... 135
  A.1 Parent, Hybridized and Truncated Ribozyme Constructs ............................... 135
  A.2 tL.1:(Mutation Constructs of tA.1) (P16:T21) ............................................. 136
  A.3 tL.1:(Mutation Constructs of tA.1) (P9:T21) ............................................. 136
  A.4 tL.1:(Deletion Constructs of tA.1) (P16:T21) ............................................. 137
  A.5 tL.1:(AP3 and APX Mutation Constructs of tD.1) (P16:T21) ....................... 138
  A.6 tL.1:(A-rich Loop Mutation Constructs of tD.1) (P16:T21) ......................... 139
  A.7 tL.1:(A-rich Loop Mutation Constructs of tD.1) (P9:T21) .......................... 140
  A.8 tL.1:(A151 and A161 Mutation Constructs of tD.1) (P16:T21) .................... 140
  A.9 tL.1:(Crosslink Sites Mutation Constructs of tD.1) (P9:T21) ....................... 141
  A.10 Unimolecular Ribozyme Constructs (P9:T21) ............................................ 142

List of References ............................................................................................... 143
List of Figures

Figure 1-1: Structure of RNA and its Components .......................................................... 1
Figure 1-2: Timeline of Events in the Biological History of Earth .................................. 3
Figure 1-3: SELEX Scheme .......................................................................................... 5
Figure 1-4: Generalized Conventional In Vitro Selection Scheme .................................. 7
Figure 1-5: Generalized In Vitro Encapsulated Selection Scheme ................................. 9
Figure 1-6: Application of Compartmentalization in High Throughput Sequencing .......... 10
Figure 2-1: Comparison of RNA-Catalyzed Polymerization and Cross-Replication by Ligation .............................................................................................................. 15
Figure 2-2: Comparison of Initiation Mechanisms ......................................................... 17
Figure 2-3: In Vitro Encapsulated Selection Scheme for B6.61 .................................... 26
Figure 2-4: Modularity of the B6.61 RNA Polymerase Ribozyme ................................. 29
Figure 2-5: In Vitro Selection for Strand-Displacing RNA Polymerase Ribozymes ....... 34
Figure 3-1: Obtaining the First Order Rate from an Extension Curve ............................ 40
Figure 3-2: Proposed Secondary Structure of the B6.61 RNA Polymerase Ribozyme .......................................................................................................................... 47
Figure 3-3: Verification of Hybridization by Native Gel Shift ...................................... 48
Figure 3-4: Sequencing Gel Resolution of Extension of Untethered and Tethered Primers – Part 1 ............................................................................................................. 49
Figure 3-5: Sequencing Gel Resolution of Extension of Untethered and Tethered Primers – Part 2 ............................................................................................................. 50
Figure 3-6: The Effect of PEG Linker Length in Tethered Primers on RNA Polymerization .................................................................................................................. 54
Figure 3-7: Extension of RNA-Only Tethered Primers .................................................. 55
Figure 3-8: Trans Cooperation of the Ligase Core and Accessory Domain .................... 56
Figure 3-9: Deletion Analysis of the Ligase Core .............................................................. 58
Figure 3-10: Ribozyme Constructs Used in this Study .................................................... 59
Figure 3-11: Hybridization of Trans Bimolecular Constructs Restores Cis Activity ....... 60
Figure 3-12: New APX Stem Demonstrated by Covariational Mutation Analysis ........ 63
Figure 3-13: Deletion Analysis of the Accessory Domain and Elucidation of its Minimal Core Motif.................................................................64
Figure 3-14: Activity of tL1:tD.1 with Different Primers and Templates .......................66
Figure 3-15: Mutational Confirmation of AP3 and APX in tD.1 ......................................67
Figure 3-16: Polymerase Activity of Unimolecular Ribozymes Containing a D.1 Accessory Domain........................................................................68
Figure 3-17: DEPC Probing of tD.1 and tA.1.................................................................70
Figure 3-18: Essential Residues for Polymerase Function ...........................................72
Figure 3-19: Activity of 45U-Containing Polymerase Ribozyme Constructs..................75
Figure 3-20: Time-dependent Formation of Crosslinks and Their Activities................77
Figure 3-21: Mapping of Crosslinks Between tL.1 and 45U-tD.1 by Partial Alkaline Hydrolysis..............................................................................78
Figure 3-22: Mapping of Crosslinks Between tL.1 and 45U-tA.1 by Partial Alkaline Hydrolysis..............................................................................80
Figure 3-23: Activity Assay of Mutants of Residues Involved in Crosslinking ...............82
Figure 3-24: Assessment of Crosslinking Ability of Sulfo-NHS-Diazirine Crosslinker ......83
Figure 3-25: Generation and Mapping of Crosslinks Between 89P9° and tD.1.................85
Figure 3-26: 3D Models of D.2 Accessory Domain.........................................................87
Figure 3-27: Insertions in the Ligase Core in Context with the 3D Model of the Class I Ligase Ribozyme.................................................................92
Figure 4-1: Design and Construction of Selection Pool..................................................105
Figure 4-2: Construction of Selection Pool as Monitored by Gel Electrophoresis...........106
Figure 4-3: Optimization of Hybridization-Based Selective Step ..................................108
Figure 4-4: Pre-Selection Polymerization Assessment ....................................................110
Figure 4-5: Assessment of Streptavidin Beads Capture Selective Step .........................111
Figure 4-6: In Vitro Encapsulated Selection Scheme ......................................................112
Figure 4-7: Characterization of Selection Artifact – Part 1 ............................................114
Figure 4-8: Characterization of Selection Artifact – Part 2 ..........................................115
Figure 4-9: Polymerization Assessment of RNA Pools from Selection Rounds – Part 1 ...............................................................................................117
Figure 4-10: Polymerization Assessment of RNA Pools from Selection Rounds – Part 2 .............................................................................................118
Figure 4-11: T1 RNase Digestion of Selection Pool RNA...............................................120
Figure 4-12: Polymerization Assessment of Isolates from Round 7 Pool .......................122
Figure 4-13: Sequence Alignment of Round 7 Isolates – Part 1.................................123
Figure 4-14: Sequence Alignment of Round 7 Isolates – Part 2.................................124
Figure 4-15: Polymerization Assessment of Isolates from Round 8 Pool .................126
Figure 4-16: Sequence Alignment of Round 8 Isolates ...........................................127
Figure 4-17: Activity Assessment of Round 8 Isolates .............................................129
List of Tables

Table 3-1: Primer and Template Sequences ................................................................. 51
Table 3-2: Summary of Activity of 4SU-Containing Polymerase Ribozyme
  Constructs ...................................................................................................................... 75
Table 4-1: Primer, Template and Biotinylated Probe Sequences ............................. 113
List of Abbreviations

3D Three-dimensional
A/C/G/T/U Adenosine / cytidine / guanosine / thymidine / uridine
APM N-acryloylaminophenylmercuric acetate
ATP/CTP/GTP/UTP Adenosine / cytidine / guanosine / uridine 5'-triphosphate
bp Base-pair
cDNA Complementary DNA
DEPC Diethylpyrocarbonate
DMF N,N-dimethylformamide
DNA Deoxyribonucleic acid
dsDNA Double-stranded DNA
DTT Dithiothreitol
EDTA Ethylenediaminetetraacetic acid
e.g. exempli gratia (Latin expression for “for example”)
et al. et alii (Latin expression for “and others”)
HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
i.e. id est (Latin expression for “that is (to say)"
IVC In vitro compartmentalization
mRNA Messenger RNA
nmol Nanomole
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleotide 5'-triphosphate</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PEG</td>
<td>Poly(ethylene glycol)</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>poly(A)</td>
<td>Poly(adenosine)</td>
</tr>
<tr>
<td>pRpp</td>
<td>Phosphoribosyl pyrophosphate</td>
</tr>
<tr>
<td>ppi</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>PT</td>
<td>Primer-template</td>
</tr>
<tr>
<td>RCTag</td>
<td>Reverse-complement tag</td>
</tr>
<tr>
<td>RdRP</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>Rz</td>
<td>Ribozyme</td>
</tr>
<tr>
<td>SELEX</td>
<td>Systematic evolution of ligands by exponential enrichment</td>
</tr>
<tr>
<td>Span 80</td>
<td>Sorbitan monooleate</td>
</tr>
<tr>
<td>SSC</td>
<td>Saline-sodium citrate</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>Sulfo-NHS-diazirine</td>
<td>Sulfosuccinimidyl 4,4'-azipentanoate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Sulfo-GMBS</td>
<td>N-(γ-maleimidobutyryloxy) sulfosuccinimide ester</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em></td>
</tr>
<tr>
<td>TCEP-HCl</td>
<td>Tris(2-carboxyethyl)phosphine hydrochloride</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>Tween 80</td>
<td>Polysorbate 80</td>
</tr>
<tr>
<td>^45U</td>
<td>4-thiouridine</td>
</tr>
<tr>
<td>^45UTP</td>
<td>4-thiouridine 5'-triphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
CHAPTER 1: The Biochemistry of RNA

1.1 Ribonucleic Acid and the “RNA World” Hypothesis

1.1.1 The Structure of RNA

Ribonucleic acid (RNA) is one of the two nucleic acid polymers that exist in natural biology. The subunits of a nucleic acid polymer are called nucleotides. In RNA, they are called ribonucleotides and consist of three basic components: a ribose sugar, a

![Figure 1-1: Structure of RNA and its Components](image)

(A) Molecular structure of the four standard nucleosides in RNA; “R” represents the ribose sugar to which the nitrogenous bases are attached (Adapted from Stryer et al., 2002). (B) Comparison of the structures of a DNA strand and an RNA strand. The nucleotide subunit is highlighted in colour: the base is coloured in yellow, the sugar in red and the phosphate in blue (Adapted from Stryer et al., 2002).
phosphate and a nitrogenous base. In contrast, deoxyribonucleic acid (DNA), the other naturally occurring nucleic acid polymer, contains deoxyribonucleotide subunits that contain deoxyribose instead of ribose. When RNA is transcribed from DNA, it contains four standard nucleotides – adenosine, cytidine, guanosine and uridine – with each containing their respective bases adenine, cytosine, guanine and uracil (Figure 1-1A). However, other nucleotides such as pseudouridine, ribothymidine and inosine also exist in post-transcriptionally processed RNAs such as tRNAs (Auffinger & Westof, 1998). As shown in Figure 1-1B, bases are attached to the 1' carbon of ribose. The phosphate that is attached to the 3' carbon is also attached to the 5' carbon of the next nucleotide; in this way, adjacent nucleotides are connected by phosphodiester linkages.

Because RNA is most commonly single-stranded, it can fold into a wide range of base-pairing-mediated secondary structures involving both stems and loops. These secondary structural elements define the shape of the RNA molecule, which ultimately dictate its function.

1.1.2 Ribozymes and the “RNA World” Hypothesis

The central dogma of molecular biology, first stated by Francis Crick in 1958, provides a framework for the flow of genetic information: DNA is transcribed into a messenger RNA, and the messenger RNA is then translated in protein (Crick, 1958). While the most commonly known function of RNA, based on the central dogma, is to serve as an intermediate molecule in the relay of genetic information, it was first found in the 1980s that RNA could make use of its elaborate secondary structures to elicit
catalytic properties (Kruger et al., 1982; Guerrier-Takada et al., 1983). In 1982, Kruger et al. found that the intron of a ribosomal RNA (rRNA) gene in *Tetrahymena* thermophila catalyzed its own self-splicing to produce the mature RNA; one year later, Guerrier-Takada et al. found that the RNA component in the ribonuclease P enzyme complex is the catalytic subunit. These findings have not only demonstrated the catalytic potential of RNA, but have also paved the road for further discovery of other naturally occurring RNA enzymes or ribozymes: the hammerhead ribozyme (Forster & Symons, 1987), the hairpin ribozyme (Feldstein et al., 1989), the Neurospora VS ribozyme (Saville & Collins, 1990), the HDV ribozyme (Wu et al., 1989) and the ribosome (Nissen et al., 2000). At the same time, *in vitro* selections have been used extensively for the past two decades to isolate RNA aptamers capable of binding a particular substrate with high affinity (Stoltenburg et al., 2007) and artificial ribozymes capable of catalyzing small molecule chemistries other than those catalyzed by natural ribozymes (Ellington et al., 2009).

Because of the dual ability of RNA to store genetic information and catalyze chemical reactions, it has been postulated that early life may have began with an “RNA

---

**Figure 1-2: Timeline of Events in the Biological History of Earth**

Proposed timeline of events on Earth before current protein-based biology. Times are indicated in billions of years before the present. (Adapted from Joyce, 2002).
world”, in which biology was largely based on RNA, that then evolved to become the present day protein-based world (Woese, 1967; Crick, 1968; Orgel, 1968; Gilbert, 1986); this was initially supported by the discovery of natural ribozymes and the isolation of artificial ribozymes capable of promoting a wide range of chemistries. To further substantiate this seemingly straightforward RNA-first hypothesis, researchers then considered whether it was possible for RNA to be formed prebiotically. While there was considerable support for activated ribonucleotides to polymerize into short RNA oligomers abiotically (Ferris et al., 1996), there was little evidence for the formation of ribonucleotides in prebiotic conditions (Joyce, 2002). This had then prompted researchers to propose the existence of alternative pre-RNA worlds that may have existed prior to an RNA world (Joyce, 2002). These pre-RNA worlds were based on other nucleic acid candidates such as threose nucleic acid (TNA), peptide nucleic acid (PNA) and glycol nucleic acid (GNA), in which their nucleotide backbone were thought to be synthesized much more easily in prebiotic conditions (Joyce, 2002). Despite this, the RNA-first hypothesis has recently enjoyed a renaissance by results from Powner et al. who demonstrated the intrinsic ability of prebiotic compounds to react and produce activated pyrimidine ribonucleotides in prebiotically plausible conditions (Powner et al., 2009; Szostak, 2009). This finding demonstrated a possible mechanism by which the first RNA molecules would have formed in an early Earth primordial soup.
1.2 In Vitro Evolution of RNA

1.2.1 Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

In vitro evolution of RNA began in the early 1990s with the isolation of RNA aptamers capable of binding a small molecule of interest (Ellington & Szostak, 1990). The methodology used was termed SELEX (systematic evolution of ligands by exponential enrichment) and involved iterative rounds of “selection” of RNA molecules, from a diverse pool, that elicited the desired binding capabilities. In SELEX, a diverse pool of single-stranded DNA molecules containing a 5’ T7 promoter and two constant regions flanking a random region is synthesized by standard cyanoethyl phosphoramidite chemistry and then PCR-amplified into a double-stranded DNA product. Upon T7 RNA polymerase-dependent transcription, the resulting RNA pool is passed through a column typically containing sepharose or agarose beads bound

![Diagram](image)

**Figure 1-3: SELEX Scheme**

A single-stranded DNA pool is synthesized on a DNA synthesizer and is PCR amplified to result in a double-stranded DNA pool. In the selection procedure, the pool is first transcribed into RNA and then allowed to bind the ligand of interest. Enriched RNAs are reverse transcribed into cDNA and then re-amplified by PCR. This procedure is repeated several times. The resulting DNA pools are then cloned to reveal individual isolates and then transcribed to result in aptamers (Adapted from Wilson & Szostak, 1999).
covalently to the ligand of interest. The flowthrough is discarded, and the bound RNAs are washed stringently and then eluted. Making use of the constant primer-binding sequences in the pool, the enriched RNAs are then reverse transcribed and re-amplified by PCR. The process (shown in Figure 1-3) is repeated multiple times and the resulting DNA pools are cloned to reveal individual isolates. Transcription of these isolates resulted in RNA molecules called aptamers that are capable of binding a given ligand.

SELEX has been used extensively for the past two decades to isolate both DNA and RNA aptamers able to recognize a wide range of targets including small molecules, nucleotides, amino acids, antibiotics and even large protein complexes like Taq DNA polymerase and single-celled protozoa such as trypanosomes (Stoltenburg et al., 2007).

1.2.2 In Vitro Selection of RNA Catalysts

Recognizing the success in isolating a large range of RNA aptamers, it was logical to then ask whether RNA could do more than simply bind a ligand. In particular, was it possible to isolate, using a similar iterative selection and reamplification procedure as in SELEX, RNAs capable of catalyzing a desired chemistry? Bartel and Szostak performed the first so-called “in vitro selection” experiment in 1993 that isolated for artificial RNA ligase ribozymes capable of catalyzing a phosphodiester bond between a short RNA oligonucleotide substrate and themselves (Bartel & Szostak, 1993). The selection was based on the isolation of modified ribozymes as a result of their own catalysis (Joyce, 2004); in this particular case, active ribozymes would have been lengthened due to the acquirement of the ligated oligonucleotide. Figure 1-4 depicts a generalized scheme for
this selection methodology. Starting with a single-stranded RNA pool containing constant regions at both termini for PCR amplification purposes, the first substrate (S1) is tethered covalently to either the 5' or 3' end. The tethered pool-S1 is incubated with the second tagged substrate (S2): active RNAs will catalyze a reaction between S2 and tethered S1, thus modifying its own sequence. A selective step (or multiple selective steps) can then be applied to enrich for active RNAs based on the newly acquired tag originating from S2. This tag could be derived from a biotinylated S2 substrate in which enrichment is based on the capture of active RNAs on streptavidin-coated beads as in the case for the selection of ligase ribozymes (Bartel & Szostak, 1993) and Diels-Alderase ribozymes (Tarasow et al., 1997). Another notable strategy is the use of single atom

![Diagram of RNA selection process]

**Figure 1-4: Generalized Conventional In Vitro Selection Scheme**

A diverse RNA pool is given the opportunity to catalyze a chemical reaction on a substrate (S1) tethered onto itself. Active RNAs that contain the modified substrate are enriched by a selective step based on the chemical modification that occurred, and re-amplified to be fed into another round of selection.
substitution on the S2 substrate – most commonly this is an O (oxygen) to S (sulfur) substitution. RNAs that have successfully incorporated the thio-tag on S2 could be enriched by their retarded mobility in a mercury-containing gel (Igloi, 1988). Pyrimidine nucleotide synthase (Unrau & Bartel, 1998), purine nucleotide synthase (Lau et al., 2004) and RNA polymerase (Johnston et al., 2001) ribozymes were isolated using this methodology. Others, such as kinase (Lorsch & Szostak, 1994) and alcohol dehydrogenase (Tsukiji et al., 2004) ribozymes were selected via thio-targeted modifications that create a better “handle” for their purification. Upon purification of active RNAs, they are reverse transcribed, PCR amplified, transcribed back into RNA, re-attached the S1 substrate and fed into another round of selection. One main disadvantage of ‘conventional’ in vitro selections described above is that isolated ribozymes are not true catalysts since they perform a self-modifying reaction that prevents multiple turnover (however, they can be engineered afterwards to work in trans). Also, the covalent linkage to the substrate requires that the RNA folds back to reach the active site. This limits the amount of random sequence available to be sampled for catalytic motifs. Finally, active RNAs tend to have complex secondary structures that may be problematic for reverse transcriptase during the RT-PCR reamplification step.

One way to bypass the problems associated with conventional in vitro selections is through the use of in vitro encapsulated selections, or in vitro compartmentalization (IVC). Borrowing ideas from natural selection in which DNA genotypes and RNA/protein phenotypes are correlated based on their compartmentalization within a common cell,
*in vitro* encapsulated selections separate the genotype and phenotype into different molecules and use water-in-oil vesicles to provide a similar correlation (Tawfik & Griffiths, 1998; Miller et al., 2006). The general scheme of *in vitro* encapsulated selections is described diagrammatically in Figure 1-5. A DNA pool containing a T7 promoter sequence is attached the first substrate (S1) through a covalent tether. The tethered DNA-S1 is then encapsulated with T7 RNA polymerase, the second substrate (S2) and nucleotide triphosphates (NTPs) in water-in-oil-vesicles, in which transcription of the DNA pool results in an RNA phenotype that is then challenged to catalyze a reaction between S2 and S1 tethered onto its own genome. Due to their coexistence

![Diagram of *in vitro* encapsulated selection scheme]

**Figure 1-5: Generalized In Vitro Encapsulated Selection Scheme**

The DNA pool (genotype) is attached the first substrate (S1) by a tether and encapsulated into water-in-oil vesicles along with T7 RNA polymerase, the second substrate (S2) and NTPs. Encapsulated transcription results in an RNA phenotype: active RNAs then catalyze a chemical reaction between S2 and S1, with S1 being tethered to its own genome. Modified DNA genes that encode for active catalysts are isolated and reamplified.
within a common vesicle, DNA genes that encode for active RNA catalysts will thus be modified; upon breakage of the emulsion and isolation of DNA, modified DNA genes are enriched based on the acquirement of a “handle” intrinsic to the reacted tethered product (P). These tagged sequences are purified, reamplified by PCR, and attached S1 before the next round of selection. Ribozymes isolated using this methodology (Miller et al., 2006) include Diels-Alderase (Agresti et al., 2005), RNA ligases (Levy et al., 2005) and polymerases (Zaher & Unrau, 2007). It has also been used successfully in protein selections (Griffiths & Tawfik, 2003; Cohen et al., 2004; Bernath et al., 2005; Levy & Ellington, 2008).

This emulsion technology has been applied very successfully in high throughput DNA sequencing. In vitro encapsulated selections reduce cross contamination of

Figure 1-6: Application of Compartmentalization in High Throughput Sequencing

(A) Double-stranded DNA is fragmented (first arrow) and ligated to adaptors strand and denatured; (B) Each single-stranded DNA fragment is then attached to one bead, and encapsulated into water-in-oil vesicles with other components required for amplification. Encapsulated PCR amplification takes place within these vesicles. (C) Each bead now carries tens of millions of copies of the starting DNA fragment. The immobilized PCR products are denatured and deposited into picolitre-sized wells on a 6.4-cm² slide that fit only one bead. (D) The wells are then filled with smaller beads (in brown) containing immobilized enzymes required for subsequent pyrophosphate sequencing. (Adapted from Margulies et al., 2005)
different DNA genotypes and RNA phenotypes by ensuring the RNA phenotype for a particular DNA genotype are contained in a common vesicle; similarly in a high throughput sequencing procedure developed by Margulies et al., each single DNA template is amplified with no contamination from other DNA templates by virtue of its encapsulation in a unique water-in-oil vesicle (Margulies et al., 2005). As shown in Figure 1-6, encapsulated PCR within these vesicles amplifies a single DNA template to tens of millions of copies before the vesicle is broken open and subjected to sequencing within picolitre-sized reactor wells (Margulies et al., 2005).
CHAPTER 2: Closing the Circle – Replicating RNA with RNA

2.1 Introduction

One of the most elemental processes in biology is the replication of a single cell into two. This intrinsically exponential process makes Darwinian evolution possible by ensuring a plentiful supply of offspring, the majority of which succumb to the rigours of natural selection. Copying a cell requires at its core the templated replication of an organism’s heredity material. In extant organisms, DNA serves as the repository of genetic information and its duplication is made difficult by the daunting size and complex structural organization of modern genomes. For this reason, a host of enzymes are required to ensure the fidelity of the genomic replication in all higher life forms. As a general principle however, replication fidelity is balanced by the requirement for spontaneous mutation in daughter cells implying an inverse correlation between error rate and genome size (Gago et al., 2009). Thus as genomes decrease in size, there is general trend towards higher error rates corresponding in turn to simpler replicative machineries (Kunkel & Bebenek, 2000; Kunkel, 2004). As genomes in the ancient past would have been considerably smaller than those found in existing organisms, the templated replication of their genomes would be correspondingly simpler.

In modern metabolism, the templated replication of RNA fits outside the central dogma of molecular biology (Crick, 1958) and plays an important role in the replication

of RNA viruses and RNA silencing in eukaryotes for example (reviewed in Ahlquist, 2002). Most importantly the catalysts for such replication, RNA-dependent RNA polymerases (RdRPs), are thought to have played a critical role in the early Earth when RNA could have potentially been the sole biological polymer (Gilbert, 1986). In an attempt to understand how RNA could have been replicated early in the evolution of life, we first discuss the relevant issues surrounding the abiotic world to the biotic “RNA world” transition, and then address and discuss several proposed methods of ribozyme-catalyzed RNA replication in the hope of expanding the possible repertoire of replication-related chemistries that may have existed in the “RNA world”. Finally, we provide a brief overview of present day RdRP initiation and elongation mechanisms that may be essentially the same now as those used in a potential “RNA world”.

2.1.1 From an Abiotic World to a Biotic RNA World

2.1.1.1 Abiotic Chemistries

The formation of the first RNA molecule is requisite to its replication. While there has long been evidence that activated ribonucleotides can be polymerized into short RNA oligomers by abiotic processes (Ferris et al., 1996), very little could be said until recently about how these first nucleotides were synthesized under prebiotic conditions. Previously, it was shown that prebiotic molecules could be combined to make a nitrogenous base (Robertson & Miller, 1995; Zubay & Mui, 2001; Hill & Orgel, 2002) and the ribose sugar individually (Ricardo et al., 2004; Gesteland et al., 2006), albeit with low yield. The glycosidic bond formation between the two components was
then catalyzed by heat in the case for purine nucleotides (Fuller et al., 1972) and never shown to occur under prebiotic conditions in the case for pyrimidine nucleotides (Orgel, 2004). In 2009, Powner et al. have showed that under prebiotic conditions, pyrimidine ribonucleotides could be formed from the same precursor molecules used to make a pyrimidine and ribose, with the exception that these molecules first reacted to generate an intermediate, 2-aminooxazole, that supplied atoms for both the pyrimidine and ribose in the final nucleotide product. This approach bypasses the requirement for glycosidic bond formation in abiotic chemistry and gives greatly renewed hope that abiotic processes could have given rise to activated nucleotides in a relatively simple (i.e. statistically probable) fashion.

2.1.1.2 Replication Strategies in an RNA World

Assuming that purine nucleotides could also be initially synthesized by a similar mechanism, the first pool of RNA molecules that would have formed must have contained sequences capable of sustaining their own replication. This ability of self-replication would have closed the “circle of life” thus making self-sustained, Darwinian-based life possible for the first time. One method by which this might have been accomplished is by polymerization promoted by RNA itself. All present day RdRPs replicate RNA molecules by the sequential addition of nucleotides, in a template dependent fashion, to a growing strand. Figure 2-1A depicts a generalized scheme for RNA-catalyzed polymerization of a templated primer. The RNA replicase ribozyme that catalyzes sequential phosphodiester bond formation is generated by the complete
extension of the RNA primer. The chemistry has been studied in depth by *in vitro* selection and will be elaborated in section 2.2.3.

In addition to polymerization, other replication systems have been proposed and tested in the laboratory. One system, proposed by Doudna et al. in 1993 and exploiting the chemistry of splicing, is a ribozyme-catalyzed system in which an RNA primer is lengthened by three nucleotides at a time: nucleophilic attack on tetranucleotides releases the 5'-nucleoside as a leaving group and extends the RNA primer by the remaining trinucleotides (Doudna et al., 1993). The system is viable to a small extent but suffers from numerous competing side reactions and low fidelity. The authors

![Diagram](image-url)

**Figure 2-1: Comparison of RNA-Catalyzed Polymerization and Cross-Replication by Ligation**

(A) Cartoon schematic scheme of ribozyme-catalyzed RNA polymerization. The complete extension of an RNA primer (pink) according to the sequence of a template (blue) by an RNA replicate ribozyme regenerates the RNA replicate. (B) Cartoon schematic of cross-replication of RNA ligase ribozymes: a ligase ribozyme (coloured in two shades of green) catalyzes the ligation of two orange oligonucleotides (Rz'-1 and Rz'-2) to generate a ligase ribozyme that catalyzes the ligation of two green oligonucleotides (Rz-1 and Rz-2) to regenerate the first ligase ribozyme. (Adapted from Lincoln & Joyce, 2009).
proposed that further improvement by in vitro selection might enable the development of a simple robust trinucleotide-based extension system for RNA polymerization.

Another system, developed by Kim and Joyce (Kim & Joyce, 2004) and further optimized by Lincoln and Joyce in 2009 (Lincoln & Joyce, 2009), is another ribozyme-based system in which an RNA ligase ribozyme hybridizes to two oligonucleotide substrates and catalyzes their ligation to each other (Figure 2-1B). The ligated RNA is a second ligase ribozyme that in turn hybridizes two other oligonucleotide substrates and catalyzes their ligation to regenerate the first ribozyme (Lincoln & Joyce, 2009). This cross-catalytic system proved successful in the absence of any protein components and was found to promote continuous exponential doubling of RNA (Lincoln & Joyce, 2009). Finally, a similar system developed and optimized by the Lehman laboratory divided the Azoarcus Group I ribozyme into four parts and made use of its intrinsic ligation capability to autocatalytically reassemble the ribozyme (Hayden & Lehman, 2006; Draper et al., 2008; Hayden et al., 2008). While these systems have elegantly demonstrated the ability of RNA to replicate and self-sustain itself without the addition of any other components and provided insight into the possibility of such systems existing in an RNA world, their main limitation is the requirement for polymer substrate strands to be freely available. These shorter substrate strands may have initially arisen from a limited RNA polymerase ribozyme, so it is plausible that the RNA-dependent replication in an RNA world may have begun as a combination of polymerization and ligation-mediated recombination. While ligation-based systems might have been quite important early on in the RNA world, one can imagine that polymerization-based
replication would have been increasingly important as the “RNA world” became more established, thus ultimately dictating the fact that all current replicators use the more robust method of polymerization. The remainder of this chapter will focus on the polymerization-based replication of RNA.

2.1.2 RdRP Initiation Mechanisms

We will now provide a brief overview of modern day RdRP initiation and elongation mechanisms so as to provide insight into how these processes may have influenced polymerization in an RNA World. In general, initiation is the mechanism by which polymerization begins: the incorporation of the first nucleotide into the polymerase enzyme active site. There are two main types (Ng et al., 2008): primer-independent (de novo synthesis) and primer-dependent initiation. As shown in Figure

Figure 2-2: Comparison of Initiation Mechanisms

(A) Primer-independent (de novo) initiation. (Adapted from Ng et al., 2008) (B), (C) and (D) – Primer-dependent initiation strategies: (B) “borrowing” a hydroxyl from a nearby protein residue (Adapted from Paul et al., 1998); (C) utilization of a short oligonucleotide from abortive cycling in de novo initiation or from a cleaved mRNA; (D) Template folds back to form a stable hairpin that is then extended (adapted from van Dijk et al., 2004).
2-2A, de novo synthesis begins with the base pairing of an initiation purine nucleotide triphosphate (often this is GTP (Ng et al., 2008)) to the 3' end of the RNA template. Since this interaction is not sufficient to stabilize the short initiating “duplex”, often other interactions such as base stacking of aromatic protein residues with the initiation nucleotide are employed (Butcher et al., 2001). As a result of instable “duplexes” that are requisite to the formation of a stable elongation complex, de novo synthesis often leads to short abortive transcripts (Ng et al., 2008).

Primer-dependent mechanisms of initiation cleverly “borrow” a hydroxyl from a nearby source for nucleophilic attack on the triphosphate of the incoming nucleotide. Figures 2-2B, C and D depict the mechanisms employed by three viral RdRPs in which a free hydroxyl was derived, respectively, from a nearby protein residue (Paul et al., 1998), a short oligonucleotide originating from abortive cycling (McClure, 1985) or from a cleaved mRNA (Hagen et al., 1995), and from the folding back of the RNA template to produce a hairpin and a free 3'-hydroxyl (Laurila et al., 2002; Laurila et al., 2005).

2.1.3 RdRP Elongation Mechanisms

A stable elongation complex is formed once the double-stranded RNA that results from initiation and several cycles of single nucleotide addition is sufficiently thermodynamically stable (Ng et al., 2008). Taking the elongation mode of the more well-studied DNA-dependent RNA polymerase protein enzymes as examples, we will describe in this section the common features of the elongation phase of RNA polymerase enzymes.
Once a stable elongation complex is formed by the generation of an 8 to 12 nucleotide long RNA product by both viral (Huang & Sousa, 2000) and bacterial (Borukhov & Nudler, 2008) RNA polymerases, synthesis of the remainder of the transcript is fully processive (i.e. translocates without stalling) (Yin & Steitz, 2004). This processivity can be attributed to a progressive “peeling off” of the RNA transcript from the template (Yin & Steitz, 2004): the concurrent formation of a base pair (with the template) in the growing RNA with the release a base pair at its 5' end. One special property about this elongation complex is that it is extremely stable at any “register” of the transcription bubble; for E. coli RNA polymerase, elongation complexes can be halted at a particular position and the overall structure will hold in place for extended periods of time (von Hippel, 1998). This stability arises from the highly negative free energy of the elongation complex $\Delta G^*_{\text{complex}}$ at any elongation register relative to the unbound state, where $\Delta G^*_{\text{complex}} = \Delta G^*_{\text{DNA-DNA}} + \Delta G^*_{\text{RNA-DNA}} + \Delta G^*_{\text{NA-polymerase}}$. The unfavourable melting of the DNA-DNA duplex ($\Delta G^*_{\text{DNA-DNA}}$) is compensated and further made favourable by the formation of the RNA transcript-DNA template duplex ($\Delta G^*_{\text{RNA-DNA}}$) and other interactions between the enzyme and nucleic acid, NA ($\Delta G^*_{\text{NA-polymerase}}$) (von Hippel, 1998). The same stabilization would similarly be observed for RNA-dependent RNA polymerases replicating a double-stranded RNA genome, in which the $\Delta G^*_{\text{NA-polymerase}}$ term would contribute more to the favourable thermodynamics of the elongation complex.

For double-stranded genomes, helicase activity is important to not only unwind the duplex at the growing end of the transcription bubble (or replication bubble, in the
case of RdRPs for double-stranded RNA genomes), but also to rewind the upstream replication fork. Yin and Steitz have shown, by the use of X-ray structures of pre- and post-translocation complexes of T7 RNA polymerase, that the rotation of a protein subdomain allows for the concurrent release of pyrophosphate and the unwinding of the downstream DNA duplex by one base pair to facilitate the translocation of the polymerase (Yin & Steitz, 2002; Yin & Steitz, 2004). Likewise, pyrophosphate dissociation provides the energy for a structural change upstream of the bubble to allow for the breakage of one base pair (the DNA-RNA transcript base pair) and the formation of another (the DNA-DNA base pair at the upstream replication fork) (Yin & Steitz, 2004). Thus, strand displacement is achieved by a helicase mechanism that supplies the energy required for a net change of one base pair directly from the NTP substrate, particularly from the bond energy between the α and β phosphates. Having established a mechanism for helicase activity, RNA polymerases must carry out a careful set of regulatory procedures at any template position attained by translocation. Particularly, there are three competing pathways for a transcription complex: [1] elongation (moving forward), [2] termination (dissociation of the RNA strand from the template), or [3] editing (moving backward). Using the third pathway as an example, this deals with the steps taken by an RNA polymerase enzyme when a misincorporation event occurs. Misincorporation leads to an altered conformation of the enzyme that disallows the next NTP to enter into the active site, thus leading to a stall in transcription (von Hippel, 1998). One solution utilized by E. coli polymerases is cleavage of a 3’ terminal oligonucleotide from the growing RNA transcript so that the resulting shorter RNA can
resume elongation from its 3' end (Surratt et al., 1991). This cleavage process is made possible using the *E. coli* transcription factors GreA and GreB (von Hippel, 1998).

### 2.2 Background: Small Molecule Chemistry in an RNA World

The ability of RNA to catalyze fundamental biological reactions in the ribosome (Nissen et al., 2000), RNase P (Pannucci et al., 1999; Marquez et al., 2006) and telomerase (Qiao & Cech, 2008) gives insight to its catalytic potential in an early “RNA world”. Given that protein enzymes catalyze the large majority of small molecule chemistries found in modern biology, one way to evaluate the ability of RNA to perform similar chemistries in an RNA world is to try to artificially evolve these functions in the laboratory. *In vitro* selections have been used extensively for almost two decades to isolate for nucleic acids capable of binding specifically to a given substrate or those capable of catalyzing specific chemical reactions (Ellington et al., 2009). In this section, we provide a short review of interesting ribozyme-catalyzed chemistries isolated by *in vitro* selection that shed light on the diverse catalytic ability of RNA in an “RNA world”.

#### 2.2.1 The Chemistry of Nucleotide Synthesis: Glycosidic Bond Formation

In order to support an RNA-catalyzed metabolism on an early Earth, RNA must have acquired the ability to build high-energy nucleotides from more ubiquitous sugars and bases (Joyce, 1989; Orgel, 1998). Despite the fact that it has been shown that ribonucleotides could be synthesized prebiotically (Powner et al., 2009), this chemistry must have been ribozyme-catalyzed in the biotic RNA world that followed. Based on the chemistry catalyzed by current nucleotide synthase protein enzymes in modern
metabolism, *in vitro* selections were used to isolate both pyrimidine nucleotide (Unrau & Bartel, 1998) and purine nucleotide synthase (Lau et al., 2004) ribozymes that were able to catalyze a glycosidic bond between a sulfur-containing base (4-thiouracil and 6-thioguanine, respectively) and an activated ribose, 5'-phosphoribosyl 1'-pyrophosphate (pRpp), tethered onto its 3' terminus.

Pyrimidine nucleotide synthase ribozymes (Unrau & Bartel, 1998) were selected based on two selective steps from a highly diverse RNA pool consisting of 228 random positions. The first was the gel purification of active ribozymes that have acquired a thio-nucleotidase on its 3' terminus as a result of its own catalysis. By introducing APM (N-acryloylaminophenylmercuric acetate), a mercury-containing thiophilic reagent, in the gel, the mobility of sulfur-containing active RNAs was retarded to facilitate their gel purification (Igloi, 1988). The second selective step involved the thio-targeted chemical modification of tethered 4-thiouridine by iodoacetyl-LC-biotin and the capture of modified ribozymes by streptavidin-coated magnetic beads. Finally, the remaining RNA was reverse transcribed, the resulting cDNA was PCR amplified and re-transcribed for the next selection round. The selection for purine nucleotide synthase ribozymes was performed similarly (Lau et al., 2004). These ribozymes have been minimized by nonhomologous recombination and reselection (Wang & Unrau, 2005) in order to elucidate their core motifs and furthered characterized biochemically (Chapple et al., 2003; Unrau & Bartel, 2003).
2.2.2 The Chemistry of Ligation: Phosphodiester Bond Formation

*In vitro* selection was used extensively to enrich for ribozymes having the ability to ligate 2 RNA molecules together by the formation of a phosphodiester bond. This chemistry is particularly interesting since it is the same as that in RNA polymerization: the nucleophilic attack of a 3'-hydroxyl of one substrate on the 5'-triphosphate of a second substrate. Initial efforts in the 1990s led to the isolation and secondary structural characterization of the Bartel Class I ligase ribozyme (Bartel & Szostak, 1993; Ekland & Bartel, 1995; Ekland et al., 1995). Different classes of ligase ribozymes were subsequently isolated by *in vitro* selection, including the *hc* ribozyme and its variants isolated from the *Tetrahymena* group 1 intron (Jaeger et al., 1999; Yoshioka et al., 2004) and smaller ligases such as the L1 ligase (Robertson & Ellington, 1999; Robertson & Scott, 2007).

Probably the most interesting set of RNA ligase ribozymes are those, evolved by the Joyce laboratory, capable of undergoing continuous *in vitro* evolution. Originally used to improve the catalytic efficiency of the Bartel ligase ribozyme, continuous *in vitro* evolution enriched for variants capable of ligating a functional T7 promoter sequence onto their 5' end and then, in the presence of reverse transcriptase and T7 RNA polymerase, simultaneously selected for and amplified these variants (Wright & Joyce, 1997). Similar protein-containing systems were developed using the DSL ligase ribozyme (Ikawa et al., 2004; Voytek & Joyce, 2007). The most recent additions are the cross-catalytic replicating ligase ribozymes mentioned earlier (Kim & Joyce, 2004; Lincoln & Joyce, 2009).
2.2.3 RNA-Catalyzed Template-Directed RNA Polymerization

Achieving success in enriching for robust ribozyme-catalyzed phosphodiester bond formation ability, researchers were then interested in further evolution of these ribozymes to isolate for an RNA polymerase ribozyme. Ekland and Bartel found that variants of the Bartel Class I ligase ribozyme could extend a primer by 6 nucleotides in 4 days according to the sequence of a template that was hybridized onto the ribozyme (Ekland & Bartel, 1996). Recognizing the catalytic potential of the Class I ligase ribozyme, Johnston et al. appended a 76-nucleotide random sequence to its 3' end and selected for RNA polymerization activity (Johnston et al., 2001). This led to the isolation of the Round 18 RNA polymerase ribozyme that contains the Class I ligase ribozyme sequence (now called the ligase core) responsible for phosphodiester bond formation ability and a new 3' accessory domain proposed to attribute its polymerization ability. The Round 18 ribozyme was capable of using nucleotide triphosphates in a template-dependent manner to extend 14 nucleotides of a \textit{trans} templated RNA primer in 24 hours (Johnston et al., 2001). Similar selections using the same selection pool to isolate for alternative functional variants were not successful (Lawrence & Bartel, 2005). Zaher et al. attempted to use a completely different selection strategy in which an RNA pool was tested for its ability to extend its 3' terminus with $^{45}$UTP using a poly(A) template (Zaher et al., 2006). Unfortunately, while it was found that the resulting RNAs were able to be modified with $^{45}$UTP, this modification did not stem from a polymerase activity; instead, the resulting ribozymes used $^{45}$UTP to cap their 5' terminus (Zaher & Unrau, 2006; Zaher et al., 2006).
2.3 Recent Results

2.3.1 Selection of the B6.61 RNA Polymerase Ribozyme

The most recent variant of artificial RNA polymerase ribozymes is the B6.61 RNA polymerase ribozyme selected from an in vitro encapsulated selection (Zaher & Unrau, 2007). In direct analogy to living cells, encapsulated selections select for DNA genotypes that encode for active RNA catalyst capable of acting on a substrate tethered onto its own DNA genome. This is in contrast to conventional selections that select for catalytic RNA able to act on a substrate tethered onto itself. The correlation between DNA genotypes and RNA phenotypes is made possible by their coexistence within a common water-in-oil droplet. Not only do encapsulated selections select for true catalysts that act on a trans substrate (as opposed to conventional selections that rely on active RNAs performing a cis self-modification), they also provide a bypass to a common problem observed in conventional selections in which active RNAs with complex secondary structures fail to be reverse transcribed into cDNA. The selection of B6.61 from the Round 18 ribozyme has demonstrated the power of encapsulated selections in isolating improved trans-acting ribozymes.

In the selection for B6.61, a diverse DNA pool containing approximately $9 \times 10^{14}$ different sequences of mutagenized Round 18 ribozyme was attached an RNA primer that contained deoxyribonucleotide residues to facilitate ligation onto the pool, and annealed an RNA template to the ligated primer. As shown in Figure 2-3, these sequences were encapsulated into water-in-oil vesicles in which T7 RNA polymerase-
Figure 2-3: In Vitro Encapsulated Selection Scheme for B6.61

A DNA pool was generated that contained a T7 promoter followed by mutagenized Round 18 ribozyme sequence. After attaching an RNA primer (orange) – template (green) to the pool, the DNA genomes are encapsulated and their genomes transcribed into RNA. Active RNA polymerase ribozyme extend the RNA primer found on their genome and in the process incorporate 4-thio-uridine residues in the growing strand. This allows selection of functional genomes using thiol sensitive mercury gels and hybridization-based capture using biotinylated oligonucleotides. Hybridized extended RNA primers are eluted from the beads and re-PCR amplified. The PCR product was then gel-purified, then re-PCR amplified, digested, ligated to the selection primer and hybridized to the selection template.

dependent transcription of the DNA pool occurred. The resulting transcript, the RNA phenotype, was then challenged to use nucleotide triphosphates supplemented with \(^{45}\)UTP to extend the primer-template substrate tethered onto its own genome. Two selective steps then enriched for extended genomes that encoded for active polymerase ribozymes. The first is the selection of genomes that have incorporated 4-thiouridine in their tethered primers as a result of polymerase ribozyme activity. Using an N-acryloylaminophenylmercuric acetate-containing gel that retards the mobility of sulfur-
containing nucleic acids (Igloi, 1988), DNA genomes that have incorporated 4-thio-uridine are conveniently shifted and gel-purified. The second selective step used biotinylated oligonucleotides whose sequence is complementary to extended primer sequence; streptavidin-coated magnetic beads were then used to enrich for hybridized extended primers. These selective steps were repeated for several rounds and the B6.61 RNA polymerase ribozyme was isolated from the round 6 pool. From the success of B6.61, it can be observed that selection of increasingly complex artificial ribozymes have placed more and more biologically relevant constraints on the selective process itself. By mimicking natural selection, the in vitro encapsulated selection of B6.61 allowed for a true trans correlation between the genotype and phenotype, and thus proved successful in enriching for improved polymerase variants.

2.3.2 The Evolutionary Power of Constructing Modular RNA Catalysts

Modularity is a common theme in many protein enzyme systems. Ribonuclease P (RNase P), that cleaves the 5' leader sequence of pre-tRNAs, is an excellent example. Consisting of an RNA component and at least one protein component, RNase P is an ancient enzyme whose RNA sequence is conserved from bacteria to modern eukaryotes (Kazantsev & Pace, 2006). It has been shown that the RNA module is the active catalyst in bacterial (Guerrier-Takada et al., 1983), archaeal (Pannucci et al., 1999) and eukaryotic (Kikovska et al., 2007) organisms since it is able to elicit its endonuclease function in the absence of its protein partner in vitro. The protein module(s) also play important roles. Since the RNA module is only active in vitro under high ionic strength, it is thought that one function of the protein modules is to neutralize the electrostatic
repulsion of the RNA so as to maintain its catalytically active structure, just as how increased levels of cations in vitro help to maintain the correct RNA fold (Kazantsev & Pace, 2006). They may also be responsible for substrate binding and/or catalysis. Thus, modularity is achieved when each component of a complex system folds and works independently; however, it is the collaborative effort of all components that instigates overall enzyme activity.

Recent results have shown that the ligase core and accessory domain of B6.61 are also modular domains that fold independently, yet act cooperatively, to extend a trans primer-template substrate (Cheng et al., 2009). In order to test this possibility of modularity, B6.61 was cut into two molecules at nucleotide position 106 to result in two molecules, L.1 representing the ligase core and A.1 representing the accessory domain, and then reassembled in trans. As shown in Figure 2-4A, polymerization ability was still retained (last 5 lanes), albeit roughly 100 fold slower when compared to that of the unimolecular ribozyme (first 5 lanes). This result supports the notion that the two domains need not be attached to one another; as long as each folds independently and then are placed together, the limited polymerization observed may represent the small number of relative functional states between the two trans domains.

The two domains were then assessed for their ability to reassemble in different orientations compared to the one present in the unimolecular ribozyme. A hybridization-based reassembly was used in which a 12-nucleotide “tag” sequence was inserted into different regions of the L.1 module (resulting in constructs x’tL.1, where x represents the nucleotide position after which the sequence was inserted) and a 12-
Figure 2.4: Modularity of the B6.61 RNA Polymerase Ribozyme

(A) Comparison of polymerization activity of trans bimolecular constructs (L.1+A.1) with that of B6.61. (B) Polymerization activity assay of different assemblies of the two hybridized trans bimolecular constructs as shown in (C). (C) Cartoon schematic of the four assemblies of hybridized trans bimolecular constructs.

nucleotide “RCtag” sequence (reverse complement of the “tag” sequence) was inserted at the 5' end of the A.1 module (resulting in the ribozyme construct termed tA.1). When 106tL.1 was hybridized with tA.1, the original assembly of the ligase core and accessory domain was regenerated, with the exception of an extra 12-nucleotide helix that now
bridges the two domains. The activity of the hybridized construct, denoted as $^{106}\text{tL.1:tA.1}$ (see Figure 2-4C for a schematic of its secondary structure), is shown in Figure 2-4B and was shown to be near identical to that of unimolecular B6.61. When three other variants of the tagged ligased core ($^{0}\text{tL.1}$, $^{44}\text{tL.1}$ and $^{79}\text{tL.1}$) were allowed to hybridize to tA.1 and then assessed for polymerase activity, it was found that all of these assemblies were more active than the non-hybridized L.1+A.1 trans system. In particular, the $^{0}\text{tL.1:tA.1}$ hybridized complex showed substantial recovery to cis activity (Figure 2-4B). As mentioned earlier, it is likely that there are many possible orientations between the two modules with each having varying levels of polymerization ability. In the case of B6.61, the best relative orientations out of the four orientations tested were when the accessory domain was at the 5' end ($^{0}\text{tL.1:tA.1}$) or 3' end ($^{106}\text{tL.1:tA.1}$) of the ligase core. This finding agrees with the three-dimensional model of the Bartel ligase ribozyme in which the supposed 5' and 3' termini of the ligase core of B6.61 on the Class I ligase ribozyme appear on the same face of the molecular structure thus providing a similar location for the placement of the accessory domain (Bergman et al., 2004).

The finding that an enzymatic system is modular can quickly bring about interesting and powerful insights into enzyme inter-domain relationships. For example, in the recent study of B6.61 (Cheng et al., 2009) described in Chapter 3, 4-thiouridine-dependent crosslinking studies were performed exploiting the modular nature of the ligase core and accessory domain. Since the two domains could be separated and re-hybridized back together with minimal impact on polymerization ability, one of the two domains could be transcribed in the presence of $^{45}\text{UTP}$ in replacement of UTP and then...
reassembled with the other domain that that is transcribed normally with UTP. UV irradiation of the hybridized system will generate covalent crosslinks (Favre et al., 1998) between the two domains, which can then be mapped to elucidate regions or specific nucleotides in each domain that are in close contact. The development of such technologies will greatly assist in the biochemical characterization of other RNA-based modular systems.

2.4 Challenges: Steering a Path Between the Scylla and Charybdis

The goal to achieve an RNA replicase ribozyme by in vitro evolution is not an easy one. Class I ligase-derived RNA polymerase enzymes suffer from both weak primer-template binding (Lawrence & Bartel, 2003) and unpredictable vigour with different primer-templates. If such embryonic replicative systems are to be evolved into robust RdRps, researchers in the field must address a number of important challenges. Like ancient mariners, a delicate path between the Scylla and Charybdis must be carefully steered to achieve this goal.

First, of overwhelming importance is the evolution of an RNA polymerase ribozyme that is able to extend a primer-template by ideally several thousand nucleotides, but essentially the length of the polymerase itself. This requirement is the minimal condition required for the construction of a self-replicating system. The 193-nucleotide B6.61 polymerase, which can extend an RNA primer by 20 nucleotides, is therefore only 10% of the way to meeting this requirement. Second, and currently unaddressed, a mechanism for strand displacement must be found. All replicating
systems have to deal with this important problem. Organisms with single-stranded RNA genomes must provide a way to separate the template and product strands after replication, while organisms with double-stranded RNA genomes have to open the RNA duplex so as to make the template strand accessible for replication and transcription of the genome.

Nature has solved both of these problems by evolving polymerases with intricate mechanisms to convert between the initiation and elongation phases of RNA synthesis as described earlier in section 2.2.2. While similar mechanisms might have existed in ribozyme chemistry in an early RNA world, they have not yet been shown to exist in in vitro-evolved RNA polymerases. In order to reach the goal of an RNA replicase ribozyme, a careful path must be steered between processivity and strand displacement. These processes just as for protein enzymes are highly likely to involve modular RNA functional domains.

2.5 Research Directions

2.5.1 Addressing Processivity

In order to further improve the B6.61 RNA polymerase ribozyme, it is necessary to address its main problem of low processivity stemming from a combination of weak primer-template recognition, primer-template binding and NTPs utilization. As a result of these limitations, B6.61 has very strong preferences for particular primer-template sequences and generally, it is difficult to predict its catalytic vigour with a given primer-template. Because of the modular nature of the ligase core and accessory domain as
summarized in section 2.3.2, it appears likely that the selection of a new domain to
B6.61 might improve its overall processivity. The function of this new “helper” domain
would possibly help to resolve the aforementioned limitations.

Recognizing the success of in vitro compartmentalization for the selection of
B6.61, a new selection could be designed that would enrich for active “helper” domain
sequences appended to its already-existing bimodular sequence. These “helper”
sequences would supply extra catalytic power to render the polymerase more
processive; for example, these could include improvement in primer-template and NTP
binding and utilization. Starting with a DNA pool containing a random “helper” domain
appended onto the rest of the B6.61 ribozyme sequence, an encapsulated selection
could be used to enrich for active helper domain sequences capable of improving the
extension ability of the B6.61 system. Recent efforts by myself on this important topic
are summarized in Chapter 4.

2.5.2 Solving the Strand Displacement Problem

To date, none of the in vitro-evolved RNA polymerase ribozymes have been
selected to enrich for a strand displacement function. However, as mentioned earlier,
this function is critical even for a completely processive RNA replicase ribozyme capable
of extending a primer-template by its own length. We have proposed, in Figure 2-5, a
possible in vitro selection strategy to select for RNA polymerase ribozymes capable of
undergoing strand displacement during primer-template extension. A pool of DNA or
Figure 2-5: In Vitro Selection for Strand-Displacing RNA Polymerase Ribozymes

A DNA or RNA pool (green) is tethered covalently to a selective strand (red) by a short oligonucleotide (yellow) and a polyethylene glycol linker (lavender). This strand is then hybridized to a template strand immediately downstream of a primer that also hybridizes to the template sequence. Displacement of the selective strand allows active variants (transcription products of the DNA pool if an encapsulated selection is performed, or RNA sequences from the RNA pool in a conventional selection) to be specifically retained based on the sequence of the selective strand.

RNA molecules is attached an RNA strand called the “selective strand” via a covalent tether and selected for sequences capable of extended a primer hybridized to a template that is also hybridized to the selective strand downstream of the primer hybridization site. Strand-displacing variants would expose the sequence of the selective strand; employing selective steps that recognize this sequence, one would then be able to enrich for active variants able to displace by polymerization or other potential helicase-related activities.
I hope that I have adequately reviewed recent advancements in RNA-catalyzed RNA replication and provided interesting prospects for further *in vitro* evolution of RNA polymerase ribozymes that would shed light on how similar ribozymes could have functioned in an early RNA world: closing the “circle” thus making life possible for the first time. The remainder of this thesis will focus on two independent projects performed in the Unrau laboratory. The first, summarized in Chapter 3, is my collaborative effort with Sunny Wang, another graduate student in the laboratory, in the detailed characterization of the structure and function of the B6.61 ribozyme. The second, found in Chapter 4, is, as mentioned earlier, my attempt to use *in vitro* encapsulated selection to improve the overall processivity of B6.61.
CHAPTER 3: Structural and Functional Studies of an RNA Polymerase Ribozyme

3.1 Introduction

3.1.1 RNA Polymerase Ribozymes for the “RNA World” Hypothesis

The “RNA World” hypothesis suggests that there was a period of evolutionary time before current DNA-protein biology in which RNA acted both as an information carrier and catalyst (Gilbert, 1986). Supported by the discovery of ribozymes isolated from in vitro selection that are capable of performing a large range of chemistries, it appears likely that a ribozyme-mediated metabolism in the “RNA World” was feasible (Chen et al., 2007). At the heart of this hypothesis is the requirement for an RNA replicase ribozyme that was capable of copying itself and other RNA molecules early in evolution; however, this chemistry has not yet been demonstrated in the laboratory. There has been nearly a decade of efforts to try and evolve an RNA polymerase ribozyme using in vitro selection. RNA ligase ribozymes that catalyze the formation of a 3',5'-phosphodiester linkage between two RNA strands were used as starting points to evolve RNA polymerase ribozymes, for the simple reason that they promote similar phosphodiester bond formation chemistries: the nucleophilic attack of a 3'-hydroxyl of one substrate on the 5'-triphosphate of the other substrate (Bartel & Szostak, 1993; McGinness & Joyce, 2003). Two classes of RNA ligase ribozymes were subjected to in

vitro evolution in order to evolve ribozymes with polymerization abilities (Johnston et al., 2001; McGinness et al., 2002; McGinness & Joyce, 2003; Lawrence & Bartel, 2005). The first class, isolated from a random pool derived from the scaffold of the natural *Tetrahymena* Group I ribozyme (Jaeger et al., 1999; Yoshioka et al., 2004), was the hc ligase ribozyme and its daughter variants (Voytek & Joyce, 2007). The second class, the Class I ligase ribozyme (Bartel & Szostak, 1993; Ekland & Bartel, 1995; Ekland et al., 1995), was isolated from a completely random pool.

The first RNA polymerase ribozyme capable of extending a trans primer-template by more than a few nucleotides was the “Round-18” RNA polymerase ribozyme evolved from appending a random “accessory domain” to the 3’ end of the Class I ligase ribozyme. It can extend an RNA primer in a template-directed manner by up to 14 nucleotides in 24 hours. Further evolution using *in vitro* encapsulated selection led to the emergence of B6.61, the best RNA polymerase ribozyme so far (Zaher & Unrau, 2007). Differing from the “Round-18” ribozyme only in its 5’ end and in a few residues in the accessory domain, B6.61 can extend an RNA primer by up to 20 nucleotides (Zaher & Unrau, 2007). B6.61 is uniformly 2 to 3 fold faster than the Round 18 ribozyme under a range of conditions that varied metal ion concentration, pH, primer and template sequence, primer and template concentration, and ribozyme concentration (Zaher & Unrau, 2007). Nevertheless, the ability of the 193-nucleotide B6.61 to extend 20 nucleotides is still 10-fold less than required for replication of an RNA molecule its own length.
3.1.2 Further Characterization of B6.61

With the goal to further evolve the B6.61 polymerase towards an RNA replicase ribozyme, we decided to characterize the secondary and tertiary structure of the RNA enzyme in detail in order to facilitate the building of a 3D model and the understanding of the mechanism of polymerization. B6.61 has two sequence domains: the 5' ligase core domain comprised of a variant of the Class I ligase ribozyme, and the 3' accessory domain appended to the ligase core during the selection for the “Round 18” ribozyme (Johnston et al., 2001). Currently, the three-dimensional architecture and catalytic mechanism of the ligase core have been well established (Bergman et al., 2004; unpublished X-ray data from the Bartel laboratory). However, the accessory domain remains largely unexplored. In this chapter, chemical probing, mutagenesis (point mutations, deletions and insertions), and crosslinking methods were used to reveal the minimal core motif of the accessory domain and to provide insight into the tertiary interactions between the accessory domain and the ligase core. In a second and complementary approach, we designed a “tag” hybridization system that increased the local concentration of the primer-template substrate by several orders of magnitude, in an attempt to explore its consequent effects on polymerization.

3.1.3 Collaborative Efforts

Before I talk about the core of this research, I would like to mention at the start that this project is a collaborative effort mainly between myself and Sunny Wang, a graduate student that just graduated from the Unrau laboratory. The specific details in
terms of which experiments were performed by Sunny and which were performed by myself are listed in section 3.5. The two of us, equally, generated all of the text and figures in this chapter. Finally, Dr. Ali Mokdad (Université de Montréal), performed the 3D modeling of the D.2 accessory domain.

3.2 Materials and Methods

3.2.1 Oligonucleotides

DNA oligonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer by standard cyanoethyl phosphoramidite chemistry. Short RNAs (less than 20 nucleotides) were purchased from Dharmacon and deprotected according to the company’s protocol; deprotected RNAs were dried by speed vacuum. Longer RNAs (between 20 and 50 nucleotides) were synthesized by run-off transcription of a construct in which a DNA strand containing the T7 promoter and the reverse complement of the RNA is hybridized to a shorter strand that makes the T7 promoter region double-stranded.

3.2.2 Polymerization Assessments

3.2.2.1 Polymerization Assay

In order to assess the polymerization activity of ribozyme constructs used in this chapter, 5’-[32P]-radiolabeled RNA primer at 0.1 μM final concentration was first annealed to a tenfold excess of RNA template in 100 mM Tris-HCl at pH 8.5. Gel-purified ribozymes were then added to the mixture at a final concentration of 2 μM, unless otherwise indicated. Extension of the primer-template was initiated by the
addition of 200 mM MgCl₂ and 4 mM of each NTP and incubation at 22°C. The reaction was stopped by adding a four-fold excess of a stop mix containing 80% formamide, 40 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol and a four-fold excess of a competitor RNA oligonucleotide whose sequence hybridizes to that of the template. Samples were heated at 95°C to denature the primer-template before loading onto a 20% polyacrylamide gel. After exposure of the gel to a phosphor screen (Fuji), the gel was visualized by a Storm™ 820 phosphorimager (Amersham Biosciences).

3.2.2.2 Kinetic Analysis

![Figure 3-1: Obtaining the First Order Rate from an Extension Curve](image)

(A) The fraction of extended primer was determined by dividing the radioactivity in the background-subtracted blue box by that of the background-subtracted yellow box. Background radioactivity is obtained by quantifying a blank area of gel (left side of gel). (B) After obtaining a plot of fraction extended as a function of time, the data were fit to \( y = a(1-e^{-kt}) \) to determine the first-order rate, \( k \).

In order to determine the first order rate of a ribozyme construct from its extension profile of a primer-template substrate, kinetic analysis was performed, using ImageQuant v5.1, by quantifying the radioactivity above the unextended primer in each lane of a sequencing gel used to resolve extension and dividing by the total radioactivity in the lane. The fraction extended was plotted against time and fit to \( y = a(1 - e^{-kt}) \) using
GraphPad Prism v5.00, where ‘a’ is the initial fraction extended and ‘k’ is the first-order kinetic rate. Referring to Figure 3-1, the quotient of the background-subtracted blue box and the background-subtracted yellow box is the fraction of extended primer.

### 3.2.3 Tethering the Primer-Template to Tagged Ribozyme Constructs

#### 3.2.3.1 Tag Sequence

The 12-nucleotide RNA tag sequence was $^{5'}$ UAC GCA CUG GUU. When aligned, using the MegAlign 4.00 alignment software, to B6.61, the P9 primer and the T21 template (DNASTAR Inc.), both the tag sequence and its reverse complement had limited hybridization to any of the aforementioned (no more than 5 bp hybridization). The RCTag DNA oligonucleotide, designed to hybridize to the RNA tag sequence, was synthesized using an ABI 392 DNA/RNA synthesizer. The sulfhydryl moiety at the 3' end of the RCTag DNA oligonucleotide was derived from a 3'-Thiol-Modifier C3 S-S CPG column (Glen Research) and the following polyethylene glycol linker region was derived from the coupling the Spacer Phosphoramidite C18 (Glen Research) once, twice, or three times during DNA synthesis. Following DNA synthesis, the construct was deprotected overnight at 65°C in saturated ammonium hydroxide containing 100 mM dithiothreitol (DTT). Upon precipitation, the pellet was resuspended in 20 mM HEPES at pH 7.0 containing 0.5 mM TCEP-HCl (tris(2-carboxyethyl)phosphine hydrochloride).

#### 3.2.3.2 Chemical Synthesis of DNA-PEG-Crosslinker-RNA Construct

In order to synthesis the DNA-PEG-Crosslinker-RNA constructs (section 3.3.1), 1 nmol of 5'-modified 2'-amine substituted RNA primer (Dharmacon) dissolved in water
and 20 nmol of \(N\)-(\(\gamma\)-maleimidobutyryloxy) sulfosuccinimide ester (sulfo-GMBS) (Pierce) dissolved in DMF (N,N-dimethylformamide) were first mixed and lyophilized to dryness. The resulting pellet was resuspended in 2 \(\mu\)L PBS (phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na\(_2\)HPO\(_4\), 2 mM KH\(_2\)PO\(_4\)) at pH 7.2 and incubated at room temperature for 30 minutes. To purify nucleic acid away from unreacted sulfo-GMBS, a C18 SPICE column (Analtech) was first washed with 6 mL of anhydrous acetonitrile and then equilibrated with 6 mL of 50 mM ammonium acetate at pH 7.0. The crosslinking reaction was diluted with 500 \(\mu\)L of 50 mM ammonium acetate at pH 7.0 and loaded onto the C18 column. The column was washed vigorously with 20 mL of 50 mM ammonium acetate at pH 7.0 and then with 6 mL of 0.05% acetonitrile in 50 mM ammonium acetate at pH 7.0. RNA was then eluted with 2 mL of 20% acetonitrile in 50 mM ammonium acetate at pH 7.0, and then lyophilized to dryness. 2 nmol of 3'-PEG-sulphydryl-containing RCtag DNA oligonucleotide in 25 \(\mu\)L 20 mM HEPES buffer at pH 7.0 containing 0.5 mM TCEP-HCl was added to the dry pellet and allowed to react at room temperature overnight in the dark. The reaction was stopped by adding a stop mix containing 80% formamide, 0.025% bromophenol blue and 0.025% xylene cyanol and loaded onto a 20% polyacrylamide gel. Crosslinked DNA-PEG-RNA constructs were gel purified, eluted and ethanol precipitated.

### 3.2.3.3 Construction of Tagged Ribozyme Variants

This section details the steps used to construct tagged ribozyme variants \(^5\)tB6.61 (section 3.3.1). The terminal tag insertion constructs \(^0\)tB6.61 and \(^{193}\)tB6.61 were
constructed directly by PCR using a DNA primer containing an overhanging tag sequence. For example, the DNA primer for ⁰tB6.61 had the sequence ⁵'ttt taa tac gac tca cta tag gAA TAC GCA CTG GTT aat aca aaa gac aaa t (italicized is the T7 promoter; uppercase underlined is the 12-nt tag sequence; in uppercase “A”s are extra adenosine residues; and in lowercase are the first 16 nucleotides of the sequence of B6.61). Internally tagged ribozyme constructs (⁴⁴tB6.61 and ⁷⁹tB6.61 in the ligase core, ⁹⁷tB6.61 and ¹⁰⁶tB6.61 in the linker region, and ¹¹⁷tB6.61 and ¹⁵⁵tB6.61 in the accessory domain) were constructed by three successive PCR amplifications. Firstly, the double stranded DNA product for B6.61 was PCR amplified into two DNA fragments: the 1ˢᵗ piece consisted of B6.61 sequence from the 5' end up to the insertion site “x” followed by tag sequence; and the 2ⁿᵈ piece consisted of the RCtag sequence at its 5' end followed by the rest of B6.61 sequence starting at the (x+1)ᵗʰ residue. After purification by QIAquick columns (QIAGEN), the two fragments were mixed at 10 nM each in the absence of primers, and thermocycled for 30 cycles of PCR with an annealing step of 45°C for 1 minute and 50 seconds. The resulting product was then diluted and re-amplified using primers corresponding to their 5' and 3' termini to generate the final PCR product.

3.2.3.4 Verification of Hybridization by Native Gel Shift

In order to verify the hybridization between ⁷ᵗB6.61, the tethered primers and the template (section 3.3.1), 0.1 μM 5'-end radiolabeled RCtag-P9 and 1 μM gel-purified ribozyme were first mixed in ddH₂O, heated to 80°C for 2 minutes and cooled to 4°C in 15 minutes. 100 mM Tris-HCl at pH 8.5 and 200 mM MgCl₂ was added to the mixture,
incubated at 22°C for 10 minutes, and then added native loading dye and resolved on an 8% 19:1 acrylamide: bis(acrylamide) native polyacrylamide gel.

3.2.4 Synthesis of Other Ribozyme Constructs

All other ribozyme constructs mentioned in this chapter (mutated, truncated, etc...) were produced by PCR amplification using similar strategies described in section 3.2.3.3. All ribozyme constructs used in this chapter have been confirmed by their DNA and RNA lengths on agarose and polyacrylamide gels, respectively.

3.2.5 Diethylpyrocarbonate (DEPC) Probing

5'-end-[\textsuperscript{32}P]-radiolabeled ribozyme construct (1 µM) was incubated at 22°C in a 20 µL solution containing 50 mM sodium cacodylate at pH 7.5 in the presence of 1 mM EDTA. MgCl\textsubscript{2} was then added at the indicated concentrations. 1 µL diethylpyrocarbonate (DEPC) was added to the solution and resulting mix was incubated at 22°C for 1 hour in the dark. 4 volumes of stop buffer containing 200 mM Tris-acetate at pH 7.5, 375 mM sodium acetate, and 0.12 mM EDTA were added and the resulting mixture was ethanol precipitated. After precipitation, the pellet was dried by speed vacuum and was resuspended in 50 µL of 1 M aniline in 15% acetic acid. The resuspension was incubated at 60°C for 20 minutes in the dark and then dried to a pellet by speed vacuum. After three iterations of drying and dissolving, the sample was dissolved in 30 µL of loading dye containing 2 mM EDTA and electrophoresed on a sequencing gel with an alkaline hydrolysis and T1 RNase ladder of the unmodified RNA.
3.2.6 4-thio-Uridine-Mediated UV Crosslinks

In order to transcribe 4-thiouridine-containing ribozymes (section 3.3.8), 0.1 μM DNA template of tD.1 or tA.1 was added 40 mM Tris-HCl at pH 7.9, 2.5 mM spermidine, 26 mM MgCl₂, 0.01% Triton X-100, 8 mM GTP, 5 mM ATP, 5 mM CTP, 2 mM 45UTP and 3 U/μL T7 RNA polymerase and incubated at 37°C for 1.5 hours. The resulting RNA molecules, 45U-tD.1 or 45U-tA.1, contain 4-thiouridine in replacement of regular uridine.

To generate crosslinks between the ligase core and accessory domain of B6.61, tL.1 (at 1 μM final concentration) was assembled with one of 45U-tD.1 or 45U-tA.1 (also at 1 μM final concentration) in reaction buffer (100 mM Tris-HCl, pH 8.5, 50 mM MgCl₂) and heated at 80°C for 2 minutes and cooled to 4°C to encourage hybridization between the tag (in tL.1) and RTag (in 45U-tD.1 and 45U-tA.1) insertions. The 20 μL solution was then added into a single well of a polystyrene Costar® 96-well plate (Corning Incorporated) and placed on ice in a 4°C room. A UV transilluminator (UVP) was placed on the 96-well plate and the sample was irradiated at 356 nm for 10 to 20 minutes. Three volumes of stop solution (97% formamide, 20 mM EDTA, 0.025% bromophenol blue, and 0.025% xylene cyanol) were added to the reaction and the reaction was electrophoresed through a 10% or 12% polyacrylamide gel containing 6.67 M urea.

3.2.7 Primer-Ribozyme Photo-Crosslinking Studies

The following section details the steps required to synthesize a photoactive primer used for crosslinking studies (section 3.3.9). 5 nmol of N9P9 RNA primer that contains a 2'-NH₂ at its C8 residue was dissolved in 6 μL 1x PBS buffer at pH 7.2. The
resuspension was added to 750 nmol of sulfosuccinimidyl 4,4'-azipentanoate (sulfo-
NHS-diazirine) powder and incubated at 22°C for 1 hour in the dark. Chemical
crosslinking was stopped by addition of an 8-fold excess of a stop mix containing 80% 
formamide, 40 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue and 100 
mM Tris-HCl. Crosslinked \( nP9 \) (\( nP9^+ \)) was phenol-chloroform extracted and ethanol 
precipitated prior to use. Crosslink reactions prepared as 10 μL solutions in reaction 
buffer (100 mM Tris-HCl, pH 8.5, 50 mM MgCl\(_2\)) were added into single wells of a 
polystyrene Costar® 96-well plate (Corning Incorporated) and placed on ice. A UV 
transilluminator (UVP) was placed on top of the 96-well plate and the sample was 
irradiated at 356 nm for the length of time indicated. Three volumes of stop solution 
(80% formamide, 40 mM EDTA, 0.025% bromophenol blue, and 0.025% xylene cyanol) 
were added to the well and the solution was electrophoresed through a polyacrylamide 
gel containing 6.67 M urea.

3.3 Results

3.3.1 Increase in Catalytic Rate Promoted by Increased PT Local Concentration

In order to assess the effects of increasing the local concentration of the primer-
template substrate to B6.61, a hybridization system was used that linked the enzyme 
and substrate via a short DNA-RNA helix. A 12-nucleotide “tag” sequence that is 
inserted by PCR into various regions of the B6.61 ribozyme (secondary structure shown 
in Figure 3-2A) is complementary to a 12-nucleotide “RCTag” (reverse-complementary-
Figure 3-2: Proposed Secondary Structure of the B6.61 RNA Polymerase Ribozyme

(A) B6.61 consists of two domains: the 5' ligase core and the 3' accessory domain. Residue numbers and names of key secondary structural elements are labelled. The colouring system is consistent with alignments shown in this chapter (Adapted from Zaher and Unrau, 2007). (B) Demonstration of the tB6.61 tagged ribozyme construct. Hybridization of the inserted tag sequence hybridizes to the RTag DNA oligonucleotide that is linked to the primer-template via a flexible linker. Orange arrows indicate the location of the tag insertion in other constructs.
tag) DNA oligonucleotide that is linked flexibly via a polyethylene glycol (PEG) linker to the 5' end of an RNA primer substrate that in turn hybridizes to the template. This overall assembly tethers the primer-template to specific regions of B6.61 and thus increases its local concentration relative to the ribozyme. These tag-inserted ribozyme constructs “x” B6.61” are named based on the xth nucleotide of B6.61 after where the “tag” sequence was inserted. Figure 3-2B shows the assembly of the 0tB6.61 construct.

Figure 3-3 is native gel shift assessment of the hybridization event between radiolabelled RCTag-P9 primer (that contains the P9 RNA primer tethered to the RCTag DNA oligonucleotide), the T21 template and 4 tagged “tB6.61 constructs.

![Image]

**Figure 3-3: Verification of Hybridization by Native Gel Shift**

8% 19:1 acrylamide:bis(acrylamide) native polyacrylamide gel that resolved the hybridization event between the radiolabelled RCTag-P9 RNA primer with the T21 template (that hybridizes with the P9 portion of the modified primer) and with 4 tagged “tB6.61 constructs (x = 0, 193, 44, 79). The three-part hybridization event between the RCTag-P9 primer, the T21 template and the tagged ribozyme variants was also assessed.
Figure 3-4: Sequencing Gel Resolution of Extension of Untethered and Tethered Primers – Part 1

(A) Cartoon schematic of tethered primer-templates. (B) 20% sequencing gel resolution of concurrent ribozyme-mediated extension of untethered primer P9 and tethered primer RCTag-P9 according to the sequence of the T21 template by B6.61, 6tB6.61, 4tB6.61 and 79tB6.61. For all modified constructs, an increase in catalytic rate was observed for the tethered primer relative to the untethered primer.
Figure 3-5: Sequencing Gel Resolution of Extension of Untethered and Tethered Primers – Part 2

(A) Cartoon schematic of tethered primer-template.  (B) 20% sequencing gel resolution of concurrent ribozyme-mediated extension of untethered primer P9 and tethered primer RCTag-P9 according to the sequence of the T21 template by \(^{97}\)tB6.61, \(^{106}\)tB6.61, \(^{117}\)tB6.61, \(^{155}\)tB6.61 and \(^{193}\)tB6.61.
Table 3-1: Primer and Template Sequences

“d(Nₙ)” and “D” denote deoxyribonucleotide residues, “(18S)” and [n] denotes a “Spacer 18” polyethylene glycol linker (where n represents the number of linkers, “P” denotes primer and “T” denotes template.

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P9</td>
<td>5' CUG CCA ACC 3'</td>
</tr>
<tr>
<td>RCtag-[1]-P9</td>
<td>5' d(AAC CAG TGC GTA) (18S) CUG CCA ACC 3'</td>
</tr>
<tr>
<td>RCtag-[3]-P9</td>
<td>5' d(AAC CAG TGC GTA) (18S) CUG CCA ACC 3'</td>
</tr>
<tr>
<td>RCtag-[A]ₙ-P9</td>
<td>5' GG AAC CAG UGC GUA AAA AAA CUG CCA ACC 3'</td>
</tr>
<tr>
<td>RCtag-[U]ₙ-P9</td>
<td>5' GG AAC CAG UGC GUU UUU UUU CUG CCA ACC 3'</td>
</tr>
<tr>
<td>P12</td>
<td>5' CUG CCA ACC GUG 3'</td>
</tr>
<tr>
<td>D4P12 (P16)</td>
<td>5' d(CACG) CUG CCA ACC GUG 3'</td>
</tr>
<tr>
<td>⁴tB6.61</td>
<td>3' ... d(UUG GUC ACG CAU) ... 5'</td>
</tr>
<tr>
<td>T21</td>
<td>3' GAC GGU UGG CAC GCU UCG CAG 5'</td>
</tr>
<tr>
<td>T32</td>
<td>3' GAC GGU UGG CAC GCG AAC UGA CCA UGC... 5'</td>
</tr>
</tbody>
</table>

After verification of the desired hybridization events, polymerase activity assays were performed using the modified primer-templates. For a fair comparison, activity assays were performed in the presence of both unmodified primer (also termed “untethered primer”) and modified primer (also termed “tethered primer”) in the same reaction tube. The validity of these experiments is confirmed by the fact that the concurrent polymerization of the two primers by B6.61 was identical to those performed with only either one of the two primers present (results not shown). As shown in the extension timecourse for B6.61, a lower activity (approximately two-fold) was observed for the RCtag-P9 primer compared to the P9 primer (first 7 lanes of Figure 3-4). We speculate that the extra DNA sequence and PEG linker at the 5' end of the tethered primer may hinder its normal binding to the tag-less B6.61 ribozyme and thus inhibit extension to some extent. While the ⁰tB6.61, ⁴⁴tB6.61 and ⁷⁹tB6.61 variants showed slightly slower rates than B6.61 with the untethered primer (bottom half of Figure 3-4), they all showed a 5 to 10 fold increase in rate (as determined by procedures...
described in section 3.2.2.2) and longer extension products with the tethered primer, albeit with different extension patterns (last 21 lanes; top half of Figure 3-4).

As seen in Figure 3-5, \(^{106}\)tB6.61, \(^{117}\)tB6.61 and \(^{193}\)tB6.61 showed polymerization rates similar to that of B6.61, with no obvious improvement when using the RCTag-P9 primer. On the other hand, \(^{98}\)tB6.61 and \(^{155}\)tB6.61 constructs had very little activity with either the P9 or RCTag-P9 primer, indicating that a tag insertion at these sites interrupted some functionally important structural elements. Note that \(^{44}\)tB6.61 and \(^{79}\)tB6.61 have their tag sequences inserted in the loop regions that were found previously to be able to tolerate long sequence strands (Ekland et al., 1995). \(^{117}\)tB6.61 and \(^{155}\)tB6.61 constructs also contain their tag sequences in the proposed loop regions of the accessory domain (Johnston et al., 2001), although the tolerance of these loops for length is unknown. The absence of polymerase activity for the \(^{155}\)tB6.61 construct indicated that the AL3 loop (Figure 3-1A) is intolerant to insertions and thus may play an important structural or functional role for attributing polymerization ability.

Similar comparison patterns between the extension of P9 and RCTag-P9 were observed with varied pH, Mg\(^{2+}\) concentrations, ribozyme concentrations and extension temperatures (results not shown). Changing the length of the PEG linker from an 18- to 56-carbon PEG linker had no effect to “tagged” ribozymes but decreased the activity of unmodified B6.61 (Figure 3-6). This is consistent with our previous hypothesis that extra 5’ sequence in the primer substrate may inhibit its binding to B6.61. Finally, when an oligonucleotide entirely made of RNA was used as the tethered primer (RCTag-[A]\(_6\)-P9
and RCTag-[U]$_6$-P9 constructs in Table 3-1), in which the DNA tag oligonucleotide was
replaced by corresponding RNA sequence and the PEG linker was replaced by a string of
6 adenosines or 6 uridines, a similar enhancement of rate was observed with the
$^0$tB6.61, $^{44}$tB6.61 and $^{79}$tB6.61 constructs (Figure 3-7). Interestingly, B6.61 extended
these RNA primers faster than when asked to extend RCTag-P9 (compare first 5 lanes
with the next 10 lanes of Figure 3-7A). The reasons for this difference are still unknown.

The increased polymerization rate observed when using tethered primers
demonstrated that this tag-based hybridization system is able to provide several
suitable binding sites for the primer-template complex to effectively raise its local
concentration. This, in turn, has provided a remedy, to some extent, for their low
affinity for one another. However, considering that this technique can achieve an
approximately $10^6$ fold increase in the local concentration of the primer-template based
on its localization in a small spherical volume dictated by the length of the tethered
primer, a mere 10-fold increase in catalytic rate (from 0.6 min$^{-1}$ (Zaher & Unrau, 2007) to
$\sim$6 min$^{-1}$), which is still 60 fold slower than an RNA ligase ribozyme (360 min$^{-1}$, Bergman
et al., 2000), is still very low. In an independent study in the Bartel laboratory where the
primer-template was co-localized with the “Round 18” RNA polymerase ribozyme in
micelles by hydrophobic anchors, there was also only a 3 to 20 fold increase in catalytic
rate (Müller & Bartel, 2008). Since the location of the tag insertion site seems to have
the largest impact on polymerization, we suggest that the orientation of the PT complex
presented to the ribozyme, rather than its effective local concentration, plays a more
fundamental role on its catalytic rate.
Figure 3-6: The Effect of PEG Linker Length in Tethered Primers on RNA Polymerization

(A) Cartoon schematic of tethered primer-template. (B) 20% sequencing gel resolution of the concurrent ribozyme-mediated extension of untethered primer P9 and tethered primer RCTag-[1]-P9 or RCTag-[3]-P9 (Table 3-1) according to the sequence of the T21 template by B6.61, tB6.61, and t5B6.61. For all tagged ribozyme constructs, an increase in catalytic rate was observed for the tethered primer relative to the untethered primer.
Figure 3-7: Extension of RNA-Only Tethered Primers

20% sequencing gels to resolve the concurrent ribozyme-mediated extension of untethered primer P9 and RNA-only tethered primers RTag-[A]_P9 and RTag-[U]_P9 according to the sequence of the T21 template by B6.61, tB6.61, dTB6.61 and dtB6.61. For all modified constructs, an increase in catalytic rate was observed for the tethered primer relative to the untethered.
3.3.2 Trans Cooperation of the Two Modular Polymerase Domains

Figure 3-8: Trans Cooperation of the Ligase Core and Accessory Domain

(A) 20% sequencing gel resolution of the extension of the P16:T21 primer-template by a fixed concentration of the L.2 ligase core domain and increasing concentration of the A.2 accessory domain. (B) 20% sequencing gel resolution of the extension of the P16:T21 primer-template by a fixed concentration of the A.2 accessory domain and increasing concentration of the L.2 ligase core. (C) Sequence alignment of L.2 and A.2 relative to full-length B6.61 RNA polymerase ribozyme. Colour-coded structural elements in B6.61 are consistent with those in Figure 3-1A.
The ligase core of B6.61 was derived from the Class I ligase ribozyme (Bartel & Szostak, 1993; Ekland & Bartel, 1995; Ekland et al., 1995), which is known to have a compact structure (Bergman et al., 2004). However, very little is known about the very important accessory domain that confers polymerization ability. Without it, the ligase core is unable to extend even one nucleotide of a trans primer-template using NTPs (lanes 5 to 8 of Figure 3-8A). To test whether the two domains can function as separate modules in trans, the full-length B6.61 polymerase was separated into two components using standard PCR procedures.

B6.61 was first divided at nucleotide position 110, resulting in two molecules: L.2 is the 5' RNA molecule up to nucleotide 110 representing the ligase core, and A.2 is the 3' RNA molecule representing the accessory domain (Figure 3-8C). With fixed 0.5 μM concentration of L.2, adding A.2 in trans brings about polymerization of a trans primer-template; increasing the concentration of A.2 proportionally increases polymerization activity (Figure 3-8A). The same effect was found with the converse experiment where L.2 was added and increased against a fixed concentration of A.2 (Figure 3-8B). Since no polymerization was observed when each molecule was present alone, these results imply that the two domains can be viewed as separate modules and each module is required for polymerization activity. Furthermore, their mutual cooperation is requisite for activity.

In addition to nucleotide position 110, different cut sites were also tested. One of the best trans systems is when B6.61 was cut at position 106 to result in two molecules: L.1 (ligase core up to 106) and A.1 (accessory domain from 107 to the end).
Figure 3-9: Deletion Analysis of the Ligase Core

(A) 20% sequencing gel resolution of the extension of the P16:T21 primer-template by a fixed concentration of the A.1 accessory domain in trans with varying deletion modules of the ligase core domain. (B) Sequence alignment of A.1 accessory domain and L.1, L.2, L.3, L.4, L.5, L.6, L.7 and L.8 ligase core domains with B6.61.

Several ligase core modules (Figure 3-9B) with variable 3' ends (at nucleotide positions 96, 100, 101, 102, 103, 106, 110 and 126) were tested for their ability to act in trans with A.1 to extend a primer-template. As shown in Figure 3-9A, L.1 (that contains a 3' end at nucleotide position 106) was the shortest ligase core construct that, when in trans with A.1, retained polymerization activity (roughly 100 fold slower than that of unimolecular B6.61 at the same concentration and under the same conditions, as shown in Figure 3-11A). L.6 (that contains a 3' end at nucleotide position 101) was the shortest construct capable of instigating detectable polymerization.
Interestingly, while the A97 to C103 single-stranded region (shown in Figure 3-2A and Figure 3-9B) was once a strand of a helix called the P2 stem required for phosphodiester bond formation in ligation (Bergman et al., 2004), our studies showed that partial deletion of this region did not completely abort activity, suggesting that this region in B6.61 no longer requires a stem. This is consistent with the finding that hybridizing an external 7-nucleotide RNA oligonucleotide to this region (to restore the stem) inhibited polymerization and interfered with fidelity (Zaher & Unrau, 2007). Despite this, the single-stranded region between C96 and C100 is definitely important for catalysis, as demonstrated by a loss in activity due to its deletion (Figure 3-9A).

Figure 3-10: Ribozyme Constructs Used in this Study

(A) Alignment of the 4 types of ribozyme constructs used in this study: full-length, non-hybridized trans, hybridized trans and truncated hybridized trans. (B) Cartoon schematic of the assembly of these ribozyme constructs.
Exploiting the use of the tag:Rctag helix system used in primer-template localization (section 3.3.1), we then attempted to reassemble the two separated domains of B6.61 through hybridization. The 12-nucleotide tag sequence was inserted at the 3' end of L.1 resulting in the construct tL.1. The Rctag sequence was inserted at the 5' end of A.1, resulting in the construct tA.1. The sequence alignment of tL.1 and tA.1, along with a cartoon schematic of their hybridization, is shown in Figure 3-10. This figure also shows other ribozyme constructs studied in this chapter, and will be referenced as these constructs are mentioned.

![Image of figures A, B, and C]

**Figure 3-11: Hybridization of Trans Bimolecular Constructs Restores Cis Activity**

(A) Comparison of polymerization activity of trans bimolecular constructs (L.1+A.1) and hybridized trans bimolecular constructs (tL.1:tA.1) with that of B6.61. (B) Polymerization activity assay of different assemblies of the two hybridized trans bimolecular constructs as shown in (C). (C) Cartoon schematic of the four assemblies of hybridized trans bimolecular constructs.

60
Hybridization between tL.1 and tA.1 (hereafter termed tL.1:tA.1) was found to recover polymerase activity lost from separating the two domains. Figure 3-11A shows the direct comparison of polymerization activity of the full-length B6.61 ribozyme, L.1 and A.1 mixed in trans (hereafter denoted as L.1+A.1) and tL.1 and tA.1 hybridized in trans (tL.1:tA.1). While the non-hybridized trans constructs are roughly 100 fold slower than B6.61, hybridization between the two modules restores activity to that of the unimolecular B6.61.

Finally, different orientations of the two domains were assessed by using ligase core constructs with tag sequences inserted in locations other than after the 106\textsuperscript{th} nucleotide (the 3' end). Three other constructs (\textsuperscript{0}tL.1, \textsuperscript{44}tL.1 and \textsuperscript{79}tL.1) that contain the tag sequence at its 5' end, after the 44\textsuperscript{th} nucleotide position, and after the 79\textsuperscript{th} nucleotide position, respectively, were allowed to hybridize to tA.1. Interestingly, all hybridized constructs were more active than the non-hybridized L.1+A.1 trans system, but only the \textsuperscript{0}tL.1:tA.1 and tL.1:tA.1 constructs show substantial recovery to cis activity (Figure 3-11B, C). These results show that the accessory domain need not be placed solely at the 3' end of the ligase core, as in the orientation present in B6.61. The functional assembly between the two modules can tolerate some orientation flexibility. However, as certain orientations between the two modules did show better polymerization than others, a favourable orientation between the two modules will definitely improve polymerization, just as what was observed earlier with the orientation of the tethered primer-template complex in section 3.3.1.
3.3.3 Refining the Secondary Structure of the Accessory Domain

The proposed secondary structure of the accessory domain (Figure 3-2A) was predicted from a very limited alignment of active sequence variants (Johnston et al., 2001; Zaher & Unrau, 2007). Our data suggest that B6.61 may consist of two self-sustaining sub-modules both in cis and in trans; therefore, subsequent structural studies of the accessory domain were performed using the hybridized trans bimolecular construct tL.1:tA.1. For convenience of comparison, the numbering of residues in the accessory domain component tA.1 remains the same as for B6.61 (Figure 3-2A). Free Fe-EDTA radical cleavage analysis of tA.1, performed by Sunny Wang, revealed that A135, C136, G167, G168 and U169 were highly protected, whereas C167 and A166 were only slightly protected (data not shown). Moreover, A135 and A166 were protected from DEPC modification in the presence of Mg$^{2+}$ (Figure 3-17B), and G167 and G168 were resistant to T1 digestion (data not shown). Therefore, we predicted that there might exist an unknown helix, named APX that consists of four nucleotides A135, C136, C137 and U138 hybridizing to A166, G167, G168 and U169. To establish the existence of this helix, two point mutants were generated: C136G and G168C. The two mutants each had significantly lower (undetectable and ~20 fold lower, respectively) activity compared to tA.1 (tA.2 and tA.3, respectively, in Figure 3-12A), while the covariational mutant tA.4 that contains both the C136G and G168C point mutations, which restored the integrity of the helix, rescued polymerase activity (Figure 3-12A). From this point on, the numerical first-order extension rates of ribozyme constructs used are summarized in tables in Appendix 1. Based on the fact that the loop structures in the
proposed secondary structure of the accessory domain were consistent with our free Fe-EDTA probing results (not shown), the refined secondary structure of the accessory domain in Figure 3-12B is thus proposed.

Figure 3-12: New APX Stem Demonstrated by Covariational Mutation Analysis

(A) Polymerase activity assay of B6.61 compared with two mutants tA.2 and tA.3 that disrupt the proposed APX stem and covariational mutant tA.4 that rescues the disruption. (B) Newly-proposed secondary structure of the accessory domain with the new APX stem highlighted in purple.

3.3.4 The Minimal Core Motif of the Accessory Domain

In order to study the functional contribution of each structural element in the tA.1 construct, we progressively deleted its sequence elements using standard PCR procedures. Generally, we decided on the criterion that if deletion of a structural element slowed down both the catalytic rate and the extension processivity by more than three fold, it is essential for polymerase activity. The alignment of the deletion constructs tested and their polymerase activities are shown in Figure 3-13.
Figure 3-13: Deletion Analysis of the Accessory Domain and Elucidation of its Minimal Core Motif

(A) 20% polyacrylamide sequence gel resolution of polymerase activity assays of hybridized trans bimolecular constructs with various deletion mutants on the accessory domain, using the P16:T21 primer-template. Progressive deletion led to the design of the minimal motif of the accessory domain and resulted in construct td.2. (B) Sequence alignment of B6.61 with all deletion mutants of the accessory domain tested. Residue numbers are indicated at the top. The two stems are indicated with a colour scheme consistent with the secondary structure shown in (C). The C134G mutation that expanded the APX helix from 4 to 5 basepairs is indicated as an green residue in the sequence of td.1 and td.2. (C) Secondary structure of the hybridized trans bimolecular construct tl.1:td.1. The two stems, AP3 and APX, in td.1 have been confirmed by covariational mutation analysis.
Both the tA.5 mutant that removed the AP1 helix and the tA.6 mutant that removed three nucleotides in a small proposed loop had slightly slower activity compared to the tA.1 construct (Figure 3-13A, B). Mutant tA.8 had all nucleotides after the 180\textsuperscript{th} nucleotide removed and had activity comparable to tA.1, whereas mutant tA.7 had all nucleotides after the 173\textsuperscript{th} nucleotide removed and had activity 10 to 20 fold lower relative to tA.1. Combining the deletion of the AP1 helix with the tA.8 and tA.7 mutants resulted in constructs tA.9 and tA.10, respectively, as shown in Figure 3-12B. The activity of tA.9 was approximately 10 to 20 fold lower than that of tA.1, whereas the activity of tA.10 was significantly lower. It seemed that tA.10 had too much sequence deleted, thus inhibiting full polymerase activity. Based on these results, we decided to eliminate the structural elements AP1, AL1, AP2, AP4 and AL4 from the construct and ambitiously designed a construct named tD.1 (Figure 3-13B, C). Neglecting the RCtag sequence and the additional adenosine linker at its 5' end, the sequence of tD.1 begins with the APX helix (which was extended by one further base pair with a C134G mutation (Figure 3-13C). The activity of the \textit{trans} bimolecular hybridization construct tL.1:tD.1 was almost identical to that of tL.1:tA.1 when extending the P16:T21 primer-template. This conclusion remained the same when different templates were tested (Figure 3-14A). Only the length of primer was shown to have a limited effect, as the polymerization rate of tL.1:tD.1 was approximately 2 fold slower than that of the tL.1:tA.1 when a shorter primer P9 was used (compare first 15 lanes of Figure 3-13A with Figure 3-14B). tD.2 was a mutant tested later in which the single-stranded overhang at the 3' end of tD.1 was deleted and was shown to be as active as tD.1 (Figure
Figure 3-14: Activity of tL1:tD.1 with Different Primers and Templates

(A) 20% polyacrylamide sequencing gel resolution of comparison of polymerase activity of B6.61, tL.1:tA.1 and tL.1:tD.1 with the same primer but different templates. (B) 20% polyacrylamide sequencing gel resolution of comparison of polymerase activity of B6.61, tL.1:tA.1 and tL.1:tD.1 with a shorter P9 primer.

3-13A, B). Therefore, we identify tD.2, which contains only 60% of the original sequence length, as the minimal core motif of the accessory domain module. Since tD.1 was generated much earlier than tD.2, tD.1 was used in subsequent analysis to explore the structure and function of the accessory domain (sequence alignment in Figure 3-10B).

To further confirm the existence of the APX stem in tD.1, two nucleotides on either side were mutated to interrupt the five base-pair helix: G134C; C136G and G168C; C170G. Both mutants showed significantly slower activity compared to that of tD.1 (more than 500 and 200 fold slower, respectively; Figure 3-15A, C). Again, their compensatory complementary mutant G134C; C136G; G168C; C170G rescued activity as expected (12 fold slower than tD.1, Figure 3-15A, C). In a similar way, the AP3 stem on tD.1 was also confirmed by the mutagenesis analysis using mutants: G153C, C159G and G153C; C159G (Figure 3-15B, C).
Figure 3-15: Mutational Confirmation of AP3 and APX in tD.1

(A) 20% polyacrylamide sequencing gel resolution of polymerase activity assay of tL1:tD.1 compared with two mutants tL1:tD.13 and tL1:tD.14 that disrupt the APX stem and covariational mutant tL1:tD.15 that rescued the disruption. P16:T21 was used as the primer-template. (B) 20% polyacrylamide sequencing gel resolution of polymerase activity assay of tL1:tD.1 compared with two mutants tL1:tD.16 and tL1:tD.17 that disrupt the AP3 stem and covariational mutant tL1:tD.18 that rescues the disruption. P16:T21 was used as the primer-template. (C) Proposed secondary structure of the tD.1 accessory domain.

3.3.5 Restoration of Unimolecularity with the tD.1 Accessory Domain

To test whether a unimolecular RNA polymerase ribozyme that contains the tD.1 accessory domain could be created, we synthesized a molecule containing the L.1 and D.1 (tD.1 without the RCtag sequence) sequences joined end-to-end. The resulting RNA was completely inactive (results not shown). Since the hybridized tL.1:tD.1 construct had full activity, we hypothesized that extra single-stranded sequence, which could play the role of maintaining the distance between two domains as provided by the tag:RCtag helix, might be required to restore polymerase activity.
Figure 3-16: Polymerase Activity of Unimolecular Ribozymes Containing a D.1 Accessory Domain

(A) 20% polyacrylamide sequencing gel resolution of polymerase activity assays B6.61, tl.1:tD.1 and 8 unimolecular ribozyme constructs containing the D.1 accessory domain. Unimolecular ribozymes contain a variable number of adenosines (n) as a linker between the ligase core and D.1 accessory domain where n is defined diagrammatically in (B). The primer-template used was P9:T21. (B) Proposed secondary structure of unimolecular ribozymes containing a 5' ligase core and 3' D.1 accessory domain separated by a variable length adenosine linker. Note that since there already exist 3 and 2 adenosines upstream and downstream, respectively, of the variable linker insertion site, an n=x construct actually contains (x+5) adenosines in this linker region.

68
Based on the cartoon schematic shown in Figure 3-16B, a series of unimolecular constructs were generated in which \( n \) adenosine residues were inserted as variable length space between the 5' ligase core domain and 3' D.1 accessory domain. According to this scheme, a unimolecular construct with \( n = 0 \) would exactly be the inactive end-to-end construct mentioned earlier. Increasing \( n \) from 2 to 16 quickly brought about detectable polymerization at \( n = 3 \), and maximum activity was achieved when \( n = 5 \) (Figure 3-16A). For \( n > 6 \), activity slowly decreased but remained detectable even at \( n = 16 \) (Figure 3-16A). It is important to mention that, when using the P9:T21 primer-template, the maximum activity observed at \( n = 5 \) was still 2 to 3 times slower compared to that of the hybridized construct tL.1:tD.1 (lanes 4 to 6 of Figure 3-16A) and approximately 5 times slower than that of the unimolecular B6.61 (first three lanes of Figure 3-16A). This study demonstrated that the D.1 sequence retained its accessory domain function in a unimolecular ribozyme, but that the functional cooperation between the ligase core and the D.1 accessory domain could be markedly inhibited by their linker length.

3.3.6 Diethylpyrocarbonate Probing of the tA.1 and tD.1 Accessory Domains

Diethylpyrocarbonate (DEPC) is commonly used to modify the N-7 of adenosines in single-stranded and unpaired regions. In our study, either tA.1 or its minimal version tD.1 was treated with DEPC in the absence or presence of \( \text{Mg}^{2+} \) in an attempt to probe the differences in modification (and thus cleavage) patterns observed (Figure 3-17). DEPC probing studies were performed by Sunny Wang.
Figure 3-17: DEPC Probing of tD.1 and tA.1

(A) Summary of DEPC probing results from (B): arrows point to DEPC cleavage sites (dark bands in (B)), blue circles indicate protected adenosines in the presence of 50 mM Mg\(^{2+}\), and yellow circles indicate adenosines protected from cleavage at all times. (B) 15% (left) and 12% (right) polyacrylamide sequence gel resolution of DEPC-induced cleavage of 5'-end-[\(^{32}\)P]-radiolabelled tD.1* and tA.1*. Adenosines were mapped with a partial T1 RNase digestion (T1) and partial alkaline hydrolysis (H). "D" indicates denaturing conditions were used (RNA was incubated with DEPC at 95°C in the dark for 15 minutes), in which all adenosines would be damaged.
As shown in Figure 3-17, tD.1 and tA.1 showed nearly identical cleavage patterns for their shared residues. Among these residues, almost all the adenosines in or near a loop region, with the exception of A140 in the A-rich loop, were modified by DEPC in the absence of Mg$^{2+}$. However upon the addition of Mg$^{2+}$, all previously cleavable adenosines, with the exception of the supposedly bulged A151, became protected.

It also appears that while adenosines in the A-rich loop (A143 to A146), the AL3 loop (A156, A158) and the A165 bulge are not involved in any obvious base pairing, their N-7 (or more generally, their Hoogsteen faces) may be involved in some type of tertiary interaction upon addition of Mg$^{2+}$. Further investigations of the A151 bulge, A165 bulge and the A rich loop are worthwhile (see mutagenesis studies in section 3.3.7). It is also important to mention that all DEPC probing results presented are identical when either tD.1 or tA.1 was hybridized to the tL.1 ligase core (results not shown), thus further confirming the fact that the ligase core and accessory domain are modular structures that fold independently of the other.

3.3.7 Elucidation of Essential Residues for Polymerase Activity

As alluded earlier in section 3.3.6, we then proceeded to study the functional importance of interesting adenosines, based on DEPC probing results, found in the tD.1 minimal motif. Taking advantage of its reduced length, numerous point mutation, insertion and deletion constructs of tD.1 were used to elucidate its essential residues for polymerase activity. Important mutant constructs and their polymerase activities are shown in Figure 3-18.
Figure 3-18: Essential Residues for Polymerase Function

(A) Summary of DEPC probing (Figure 3-16) and mutagenesis analysis ((B) and (C)) results. (B) Polymerase assay of mutants of tD.1 and tA.1 pertaining to the A151 bulge. “ins.160U161” indicates that a U residue, designed to base-pair with A151 bulge, is inserted between 160A and 161C. The primer-template used was P16:T21. (C) Polymerase assay of mutants of tD.1 pertaining to the A165 bulge and the A rich loop. “del” indicates a deletion; “ins. 165CC166” indicates that two C residues are inserted between A165 and A166; “ins. 140AA141” indicates that two A residues are inserted between A140 and A141. The primer-template used was P16:T21.
As shown in Figure 3-18, the A165 bulge could be mutated to any other residue with limited effects on polymerase activity. A 100-fold and 20-fold decrease in activity was observed when two cytidines were inserted right after it or when it was deleted, respectively (Figure 3-18C). It appears that the backbone of the bulge, rather than the identity of the base, may play an important role in polymerase function.

The A-rich loop (G139 to A146) was found to be essential for polymerase activity. When any one of the residues were deleted (Figure 3-18C, e.g. tD.10), mutated (Figure 3-18C, e.g. tD.11) or when residues were inserted (Figure 3-18C, e.g. tD.12), polymerase activity dropped by at least 150 fold compared to that of the tD.1 parent. Since changing the length of the loop or mutating any residues in the loop proved detrimental to activity, this A-rich loop appears to be extremely important for function.

Mutation of the A151 bulge to any other base (U (tD.4 in Figure 3-18B), C (data not shown), or G (not shown)) decreased polymerase activity by 300 fold (Figure 3-18B). Polymerase activity dropped by at least 500 fold when it was forced to base pair with an inserted U (tD.3), and dropped by 20 fold when this residue was deleted (tA.11). Its potential roles will be reviewed later in section 3.3.10 and in the Discussion section.

3.3.8 4S-U-Mediated UV Crosslinking Between the Ligase Core and Accessory Domain

4-thio-uridine (4S-U)-mediated crosslinking is typically used to unearth tertiary contacts between an RNA molecule and another molecule. Similarly, it can be used to reveal tertiary contacts between two regions of the same RNA molecule. However, its applications are limited mainly due to the fact that 4S-U is conventionally incorporated
into RNA by chemical synthesis. Therefore, due to high cost and labour, both the length of the modified RNA and the number of sites containing $^{45}$U are constrained. It becomes extremely difficult when one want to globally explore all possible tertiary contacts, especially when little knowledge for potential interaction sites is available.

Exploiting the hybridization-based system developed in section 3.3.2, we developed an easy and low-cost assay to determine all tertiary contacts between two RNA molecules. One of the two RNA molecules is transcribed using $^{45}$UTP from a DNA template and the other RNA molecule is transcribed normally with UTP. Upon mixing these molecules at an appropriate ratio, UV irradiation at 356 nm will generate covalent crosslinks between $^{45}$U residues on one molecule and other spatially nearby residues on the other molecule (Favre et al., 1998). Following a crosslink mapping procedure later described, it is then possible to resolve tertiary contacts implied by the crosslinks between the two RNA molecules. The converse experiment, where the second RNA molecule is $^{45}$U-modified, would aid to confirm these crosslinks and to resolve other possible tertiary contacts.

3.3.8.1 $^{45}$U-Containing Polymerase Ribozymes are Active

Before using $^{45}$U-containing ribozymes for crosslinking, it is necessary to confirm that these modified RNAs are still active so that the results are relevant. $^{45}$U-transcribed B6.61 (hereupon termed $^{45}$U-B6.61) and $^{45}$U-transcribed domain modules ($^{45}$U-tL.1, $^{45}$U-tA.1 and $^{45}$U-tD.1) were tested for their polymerization ability relative to that of normally transcribed B6.61. As shown in Figure 3-19, $^{45}$U-tL.1:tA.1, tL.1:$^{45}$U-tA.1 and
tL.1:45U-tD.1 (Figure 3-19B) were all only approximately 10 fold slower than non-thiolated B6.61. When both domains contained 45U, the resulting constructs, 45U-tL.1:45U-tA.1 and 45U-tL.1:45U-tD.1, were approximately 50 fold slower than B6.61; this is consistent with the fact that 45U-transcribed B6.61 (45U-B6.61) was also approximately 50 fold slower than non-thiolated B6.61. These results are summarized in Table 3-2.

![Image](image_url)

**Figure 3-19: Activity of 45U-Containing Polymerase Ribozyme Constructs**

(A) Polymerase assay of 45U-modified ribozyme constructs in comparison to non-thiolated constructs. The primer-template complex used was P16:T21. (B) Polymerase assay of tL.1:45U-tD.1 in comparison to tL.1:tD.1. The primer-template complex used was P9:T21.

**Table 3-2: Summary of Activity of 45U-Containing Polymerase Ribozyme Constructs**

<table>
<thead>
<tr>
<th>System</th>
<th>Hybridized Trans Bimolecular Constructs</th>
<th>Ligase Core</th>
<th>Accessory Domain</th>
<th>Unimolecular Constructs</th>
<th>Rate of polymerization relative to B6.61</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>tL.1</td>
<td>-</td>
<td>45U-tA.1 45U-tD.1</td>
<td>-</td>
<td>1/10</td>
</tr>
<tr>
<td>II</td>
<td>45U-tL.1</td>
<td>tA.1  tD.1</td>
<td>-</td>
<td>-</td>
<td>1/10</td>
</tr>
<tr>
<td>III</td>
<td>45U-tL.1</td>
<td>45U-tA.1 45U-tD.1</td>
<td>-</td>
<td>-</td>
<td>1/50</td>
</tr>
<tr>
<td>IV</td>
<td>-</td>
<td>-</td>
<td>45U-B6.61</td>
<td>-</td>
<td>1/50</td>
</tr>
</tbody>
</table>

75
3.3.8.2 Generation of Crosslinks and Assessment of Their Activities

In our initial attempts at generating $^{45}$U-dependent crosslinks, both the $^{45}$U-tL.1:tA.1 and tL.1:$^{45}$U-tA.1 systems were used. However, results for $^{45}$U-tL.1:tA.1 showed that $^{45}$U-tL.1 readily formed intra-molecular crosslinks with residues in its own sequence (results not shown), and thus interfered with analysis of bimolecular inter-molecular crosslinks of which we are interested. Therefore, the remainder of our studies focused on generating and understanding crosslinks between the ligase core and $^{45}$U-containing accessory domain (tL.1:$^{45}$U-tA.1 or tL.1:$^{45}$U-tD.1).

Performed by Sunny Wang, UV irradiation of tL.1:*$^{45}$U-tD.1 at 356 nm (where $^{45}$U-tD.1 is 5'-end-[P]-radiolabelled) and electrophoresis on a polyacrylamide gel revealed two time-dependent bands (CL3 and CL4) with higher mobility than *$^{45}$U-tD.1 (Figure 3-20A). A negative control reaction was performed in parallel in which there was no UV irradiation and as expected, there were no bands with higher mobility on the polyacrylamide gel (results not shown). Two bands (CL1 and CL2) also appeared when *tL.1:$^{45}$U-tD.1 was UV-irradiated (right side gel of Figure 3-20A); Figure 3-21A shows that these two bands are likely chemically identical (with the exception of the location of the radiolabel) to the ones generated from UV irradiation of tL.1:*$^{45}$U-tD.1 since they have the same mobilities. Also, CL1 and CL3 appear with the same time dependence; so do CL2 and CL4 (Figure 3-20A). Because of their higher mobility and their lack of appearance in the negative control reaction, these bands likely correspond to two different covalent crosslinks between tL1 and tD.1. C3 and C4 were synthesized
Figure 3-20: Time-dependent Formation of Crosslinks and Their Activities

(A) Timecourse of UV irradiation of tl.1:*45U-tD.1 an *tl.1:*45U-tD.1 on an 8% and 6% polyacrylamide gel, respectively. (B) Polymerase assay of B6.61, crosslinked constructs CL3 and CL4 and negative control negCL3 and negCL4. Ribozymes were used at 200 nM. The primer-template complex used was P16:T21.

on a large scale and then gel-purified. At the same time, the negative control reactions were run on a large scale and the region of gel corresponding to the location that the covalent crosslinks would run was cut and nucleic acid (if any) was eluted (resulting in negCL3 and negCL4). The activities of these gel-purified constructs were then assessed (Figure 3-20B). While both negCL3 and negCL4 showed zero activity, both CL3 and CL4 retained low levels of polymerization. This suggests that the two crosslinks observed represent functional and active tertiary contacts between the tl.1 and tD.1.

3.3.8.3 Mapping of Crosslinks by Partial Alkaline Hydrolysis

CL1, CL2, CL3 and CL4 were synthesized and gel purified. Since CL1 and CL2 have a radioactive phosphate on the 5' end of tl.1, they were used to map their respective crosslink sites on the ligase core; conversely, CL3 and CL4, which have a radioactive phosphate on the 5' end of *45U-tD.1, were used to map their respective crosslink sites on the tD.1 accessory domain.
Figure 3-21: Mapping of Crosslinks Between tL.1 and $^{45}$U-tD.1 by Partial Alkaline Hydrolysis

(A) Lane 1: UV irradiation of *tL.1,$^{45}$U-tD.1 generated crosslinked molecules CL1 and CL2. Lane 2: UV irradiation of tL1,$^{45}$U-tD.1 generated crosslinked molecules CL3 and CL4. “*” represents a 5′-[32P]-radiolabel. (B) Sequencing polyacrylamide gels used to map location of crosslink. The T1 RNase ladder (T1) and partial alkaline hydrolysis adder (H) of the uncrosslinked radiolabelled RNA were used as a reference. Crosslinked residues are indicated in blue (for tL.1) and green (for $^{45}$U-tD.1) (C) Secondary structure of the tL.1:$^{45}$U-tD.1 construct with proposed tertiary contacts concluded from (B) denoted by lines between the two domains.
In order to map the location of the crosslink site, each crosslinked molecule (CL1, CL2, CL3 or CL4) was subjected to partial alkaline hydrolysis. Because of the covalent nature of the crosslink, crosslinked molecules will exhibit a normal hydrolysis ladder on a subsequent polyacrylamide gel from its 5'-radiolabelled end up from the bottom of gel to the crosslink site (Figure 3-21B). The hydrolysis ladder will then appear compressed up to the band corresponding to the uncut RNA. The crosslink site on the radiolabelled strand could therefore be mapped by making use of a partial T1 RNase digestion ladder and alkaline hydrolysis ladder of 5'-radiolabelled uncrosslinked strand. For example, the crosslinked residue in CL1 that contained a radiolabelled tL.1 strand mapped to the C25 residue of the tL.1 strand (Figure 3-21B, left gel). This same crosslink, which corresponds to CL3 when $^{45}$U-tD.1 was radiolabelled, was found to map to the $^{45}$U155 residue in $^{45}$U-tD.1 (Figure 3-21B, right gel). These results imply that there was a crosslinking event between C25 in tL.1 and $^{45}$U155 in tD.1 (Figure 3-21C). Using the same logic, analysis of crosslinked molecules CL2 and CL4 (Figure 3-21B) revealed that there was a crosslink between C25 with $^{45}$U156.

Similar experiments were performed using the tL.1:$^{45}$U-tA.1 system, in which tL.1:$^{45}$U-tA.1 and tL1:*$^{45}$U-tA.1 were used to generate crosslinks. Crosslinks were, again, mapped by partial alkaline hydrolysis. As seen in Figure 3-22A, two covalent crosslinks were also generated: CL5 and CL6 (that contains a radiolabel in the tL.1 strand) and CL7 and CL8 (that contains a radiolabel in the $^{45}$U-tA.1 strand). Again, CL5 and CL7 likely correspond to chemically identical crosslinks due to their identical mobility on a polyacrylamide gel (Figure 3-22A); the same applies for CL6 and CL8.
Figure 3-22: Mapping of Crosslinks Between tl.1 and 45U-tA.1 by Partial Alkaline Hydrolysis

(A) Lane 1: UV irradiation of *tl.1:45U-tA.1 generated crosslinked molecules CL5 and CL6. Lane 2: UV irradiation of tl1:*45U-tD.1 generated crosslinked molecules CL7 and CL8. "*" represents a 5'-[32P]-radiolabel. (B) Secondary structure of the tl.1:45U-tA.1 construct with proposed tertiary contacts concluded from (C) denoted by lines between the two domains. (C) Sequecing polyacrylamide gels used to map location of crosslink. The T1 RNase ladder (T1) and partial alkaline hydrolysis adder (H) of the uncrosslinked radiolabelled RNA were used as a reference. Crosslinked residues are indicated in blue (for tl.1) and green (for 45U-tA.1).
Analysis of crosslinked molecules CL5 and CL7 (Figure 3-22C) revealed a crosslink between C25 of the ligase core and $^{4S}$U154/$^{4S}$U155 of the accessory domain. Likewise, analysis of crosslinked molecules CL6 and CL8 (Figure 3-22C) revealed a crosslink between U25 of the ligase core and $^{4S}$U131 of the accessory domain. Satisfyingly, the C25 to $^{4S}$U154/$^{4S}$U155 interaction was identical to the one mapped earlier using tL.1:$^{4S}$U-tD.1. The second crosslink is unique to the $^{4S}$U-tA.1 construct and was located in a dispensable region that was shown to be unessential for polymerase activity.

3.3.8.4 Mutation Analysis of Crosslinked Residues

In order to assess whether residues involved in crosslinking between the two domains are important for polymerase activity, mutational analysis was performed. As shown in Figure 3-23A, mutation of U154 to cytidine in the context of the tD.1 accessory domain did not affect polymerase activity. On the other hand, a U155C mutation proved detrimental as activity dropped by more than 20 fold. When both residues were mutated (U154C; U155C), activity was undetectable. We then decided to test whether replacement of the AL3 loop, (see Figure 3-2A for secondary structure) in which U154 and U155 are located, with stable tetraloops such as GNRA (Heus & Pardi, 1991) and UNCG (Cheong et al., 1990) affects polymerase activity. Figure 3-23B shows that these mutant constructs had minimally detectable activity; thus despite the increased thermodynamic stability conferred by the GNRA and UNCG hairpin motifs, this suggests that the AL3 triloop that contains U155 is essential for polymerase activity. An insertion in the AL3 loop was also detrimental to activity (see activity of $^{15S}$tB6.61 in Figure 3-5).
Figure 3-23: Activity Assay of Mutants of Residues Involved in Crosslinking

(A) 20% polyacrylamide sequencing gel resolution of polymerase activity assessment of U154 and U155 mutants in the context of tD.1. The primer-template used was P9:T21. (B) 20% polyacrylamide sequencing gel resolution of polymerase activity assessment of tD.1 mutants that contain modified AL3 loops. The primer-template used was P9:T21. (C) 20% polyacrylamide sequencing gel resolution of polymerase activity assessment of C26 mutants in the context of tL.1. The primer-template used was P12:T21. (D) 20% polyacrylamide sequencing gel resolution of polymerase activity assessment of U25 mutants in the context of tL.1. The primer-template used was P9:T21.
In Figure 3-23C and D, mutational analysis was performed on the U25 and C26 residues in the ligase core. Mutant tL.1 constructs were hybridized to tD.1 to assess for altered polymerase activity. When either U25 or C26 was deleted, activity dropped more than 300 fold compared to tL.1:tD.1. However, mutation of these residues to any other base did not seem to have an effect on activity. This means that both U25 and C26 can tolerate a point mutation but cannot be deleted, which could either translate to a length constraint for the J3/4 single-stranded region in the ligase core (see Figure 3-2A for the secondary structure of B6.61) or a specific backbone element requirement (ribose sugar or phosphate) to confer polymerase activity.

3.3.9 Initial Attempts to Probe the Primer-Template Binding Site in B6.61

![Diagram](image_url)

**Figure 3-24: Assessment of Crosslinking Ability of Sulfo-NHS-Diazirine Crosslinker**

(A) Molecular structure of sulfo-NHS-diazirine. The NHS moiety reacts with primary amines and diazirine moiety is photoreactive. (B) Timecourse of reaction of nP9 primer with sulfo-NHS-diazirine monitored on a 23% polyacrylamide gel. "*nP9" is unreacted and "**nP9" is reacted RNA primer. (C) UV crosslinking reaction of 5'-[32P]-radiolabelled T21 with nP9 (negative control) and nP9" to produce no crosslinked molecule and CL9 crosslinked molecule, respectively. (D) Mapping of location of crosslink in CL9. "T1" represents T1 RNase digestion, "H" represents partial alkaline hydrolysis.
Our next goal was to attempt to use crosslinking to elucidate the precise active site in B6.61. To do this, we decided to synthesize a modified primer-template with a photoactive diazirine functional group on the C8 residue of the P9 RNA primer. To do this, we started with a modified P9 RNA primer whose C8 residues contains a 2'-NH$_2$ (P9); by coupling this primary amine with sulfo-NHS-diazirine (Figure 3-24A), we were able to synthesize (Figure 3-24B), with near 100% yield, a modified P9 RNA primer (called P9$^*$) that contains a photoactive diazirine near its 3'-OH terminus, the location of active site chemistry.

To test whether the photoactive RNA primer (P9$^*$) was UV-active and capable of crosslinking to RNA residues deliberately placed near its photoactive diazirine moiety, we hybridized it with 5'-end-[${}^{32}$P]-radiolabelled T21 template and then UV-irradiated the mixture and resolved the RNA on a polyacrylamide gel. As shown in Figure 3-24C, 10 minutes of UV irradiation was sufficient to produce a higher mobility band (called CL9) relative to the one corresponding to the radiolabelled T21. Upon gel purification of CL9, the crosslinked residue was mapped by partial alkaline hydrolysis (see section 3.3.8.3) and satisfyingly, it mapped to the G14 residue of T21, which is exactly the residue that base pairs with the photoactive C8 residue on P9 (Figure 3-24D).

Upon confirming that P9$^*$ was capable of producing crosslinks, we then attempted to produce crosslinks between P9$^*$ and either domains of B6.61. As shown in Figure 3-25A, radiolabelled P9$^*$ (*P9$^*$) crosslinked with all of tL.1, L.1, tD.1, tA.1 and A.1 to produce crosslinked molecules CL14, CL13, CL10, CL12 and CL11.
Figure 3-25: Generation and Mapping of Crosslinks Between *nP9* and tD.1

(A) UV-dependent crosslink reactions between *nP9* (radiolabelled) and a range of domain modules, in the presence of T21. CL9 corresponds to the primer-to-template crosslink. Other crosslinks are labeled.

(B) Partial alkaline hydrolysis mapping of strongest crosslink from (A), CL10. Upon gel purification of CL10 from (A), the tD.1 strand was re-kinased with γ-[32P]-ATP and run on a gel. Two bands are visible, CL10a and CL10b; they were gel-purified and subjected to partial alkaline hydrolysis. The two bands mapped to the 5′AGU region of the RCTag component of tD.1. “T1” represents T1 RNase digestion and “H” represents a partial alkaline hydrolysis. (C) More UV-dependent crosslinking reactions revealed that *nP9* only crosslinked to the RCTag component of tD.1, and not to D.1 (tD.1 without RCTag).

The most obvious crosslink was the one between *nP9* and T21, labeled as CL9 on the gel in Figure 3-25A. In addition, the two strongest crosslinks with polymerase ribozyme
modules were the ones with tD.1 and tA.1. Given the advantage of tD.1 having a much shorter sequence length, its crosslink (CL10) was gel-purified and recovered. In order to map the location of crosslink on the tD.1 strand, a radioactive 5'-[32P]-phosphate was added to the 5' end of CL10 with γ-[32P]-ATP and polynucleotide kinase; two species, CL10a and CL10b, that had gel mobilities higher than that of tD.1 were purified (gel purification gel not shown; see top of Figure 3-25B for differences in their gel mobility). Finally, a partial alkaline hydrolysis mapping procedure was employed and as shown from the gel in Figure 3-25B, there exists a gap in the hydrolysis ladder of CL10a and CL10b that corresponds to the underlined residues in the RCtag sequence in tD.1: 5' AAC CAG UGC GUA 3'. Since this sequence solely serves as a mode of connection between tD.1 and tL.1 and is thus not functional sequence, these results are not too informative.

To avoid crosslinking to the RCtag sequence, we then attempted to generate crosslinks between *N9P9 and D.1, tD.1 sequence that lacks the RCtag component. Unfortunately, the polyacrylamide gel in Figure 3-25C shows that crosslinks are only generated in reactions that contain tD.1 alone (CL10) or those that contain tL.1:tD.1 (CL10 and CL14). In reactions that contain D.1 alone or tL.1 in trans with D.1, no detectable higher-mobility crosslinks were observed. Despite the fact that crosslink CL11 was formed between *N9 and A.1 (Figure 3-25A), it is much too weak and there are simply too little counts in CL11 (and thus insufficient number of moles of nucleic acid) available for phosphorylating with γ-[32P]-ATP and partial alkaline hydrolysis to resolve the location of crosslink. Perhaps this work could be re-optimized and furthered by future students in the Unrau laboratory at Simon Fraser University.
3.3.10 Three-Dimensional Modelling of the Accessory Domain

Through a collaborative effort with Dr. Ali Mokdad from Dr. François Major’s laboratory (Université de Montréal), several preliminary three-dimensional models of the D.2 (minimal core tD.2 excluding the RCTag sequence) accessory domain have been successfully modeled based on kinetic data of mutant constructs and optimal thermodynamics. Shown in Figure 3-26 are three proposed models of its tertiary structure, along with their colour-coded secondary structure for comparison’s sake.

Figure 3-26: 3D Models of D.2 Accessory Domain

(A); (B); (C) 3D models 1, 2 and 3, respectively, for the D.2 minimal accessory domain. Helical elements APX and AP3 are labeled in purple and magenta, respectively. The AL3 loop is in yellow, the AJX/3 loop (A-rich loop) is in white and all single-nucleotide bulges are in cyan. (D) Secondary structure of the D.2 minimal accessory domain (tD.2 excluding RCTag sequence). The colour of structural elements are consistent with those labeled in (A), (B) and (C).
As seen in Figure 3-26, all three models are consistent with the fact that the ligase core (which would presumably extend off of the 5’ end of the D.1 accessory domain) would be on the face of the modeled accessory domain that exposes the “bottom” strand of AP3 (U152, G153, U154) according to the secondary structure shown in Figure 3-26D) and AL3 loop. This is reassuring since we have shown, by crosslinking studies, that residues in these structural elements (U154 in AP3; U155 in AL3) are involved with possible tertiary contacts with the ligase core.

In all three models, the A165 bulge is pointed towards the A-rich loop (rather than away as shown in Figure 3-26D). It also appears that the main purpose of this bulge is to provide a sharp turning point between the AP3 helix and APX helix. This model is consistent with the mutation analysis presented in Figure 3-18C, where mutation of A165 to any other residue does not affect polymerase activity, whereas deletion of the bulge or insertion of residues immediately after the bulge decreases activity. This may mean that a one-nucleotide bulge is strictly sufficient to confer an optimal tertiary spatial relationship between AP3 and APX. DEPC probing data of this residue (section 3.3.6, Figure 3-17) saw that it was cleaved by DEPC-induced damage, but was protected upon addition of 50 mM MgCl₂. Based on this result and on the fact that A165 looped inwards to the A-rich loop (as opposed than outwards), we hypothesize that A165 forms complicated tertiary interactions, induced in the presence of Mg²⁺, with the A-rich loop.

The A151 bulge points outwards from the AP3 helix and is thus highly exposed in all three models. This is consistent with DEPC probing data as this residue is readily cleaved by DEPC-induced damage and is never protected even with the addition of 50
mM MgCl$_2$. According to mutational analysis of this residue (section 3.3.7, Figure 3-18B), activity was near zero when A151 was forced to base-pair to a uridine in a 160U161 insertion mutant (insertion of U between A160 and C161) or in an A151U mutant. The activity of a mutant where A151 was deleted was also lower – approximately 20-fold slower compared to wildtype tD.1. This is consistent with a hypothesis proposing that the A151 bulge is important to establish the correct spatial orientation between the two “segments” of the AP3 helix, thus correctly positioning the AL3 loop (or more precisely, U154 and U155) to interact with the ligase core.

Analysis of residues in the A-rich loop suggested that its residues are fundamentally important since mutation, deletion or insertion of residues in the A-rich loop proved detrimental to activity (Figure 3-18C, tD.10 through tD.13). While these results may have suggested that the A-rich loop is directly involved with phosphodiester bond formation and thus associated, somehow, with the ligase core, all three models proposed by the Major laboratory had the A-rich loop pointing away from the face of the accessory domain (the face containing U154 and U155) that was confirmed, by crosslinking studies, to have a tertiary contact with the ligase core. If the models are correct, this could either mean that (1) the A-rich loop contains a strict sequence that forms a crucial and complicated tertiary structure (through a combination of non-canonical hydrogen bonding and stacking interactions) that properly positions the AL3 loop and the “bottom” strand of the AP3 helix to interact with the ligase core or (2) it is fundamentally involved in supplementing an elongation ability intrinsic in RNA polymerases to the phosphodiester bond formation ability catalyzed by the ligase core.
Finally, taking into account the existence of a crosslink between U25 in the ligase core and U131 in the full-length accessory domain (section 3.3.8.3, Figure 3-22), model 1 seems to be the most consistent with this crosslink data. This is because based on the facts that (1) the U131 residue is situated only a few residues from the 5' end of the AP3 helix, (2) its crosslink partner U25 is only one residue away from C26 (which crosslinks to U154 and U155), U131 must be spatially close to the AL3 loop in which U154 and U155 can be found. Thus, when one imagines an extension of a few residues from the 5' end of all the proposed models shown in Figure 3-26, the only model that can possibly generate a close contact between U131 and the AL3 loop is model 1. Perhaps further modeling of the full-length accessory domain can help us determine which of the three models in Figure 3-26 is most likely.

3.4 Discussion

3.4.1 Modular Architecture of B6.61

Our studies have clearly shown that B6.61 is a modular ribozyme that consists of two distinct domains, the ligase core and accessory domain, that work together to perform an enzymatic activity. The modular architecture of B6.61 shadows a common theme in biological evolution: complex functions are usually evolved via the combination of modular, simpler functions. For example, the study of protein polymerases has revealed their modular architectures (Arnold et al., 1995; Brautigam & Steitz, 1998; McGinness & Joyce, 2003). In the case of E. coli DNA polymerase III, its core can extend only tens of nucleotides. However, by interacting with the β-clamp
protein whose job is to surround the template to prevent its dissociation, the modular enzyme complex can then extend over 50,000 nucleotides (Benkovic et al., 2001). Modular systems that exist in protein polymerases are the product of Darwinian evolution, whereas the modular system appearing in B6.61 came from the systematic evolution of RNA by in vitro selection (Johnston et al., 2001). The step-wise evolution of the B6.61 RNA polymerase ribozyme may therefore serve as a model of how modular catalysis would have emerged in an early RNA world.

In B6.61, the addition of the accessory domain to the ligase core introduced a polymerization ability for trans primer-templates that was not present in the ligase core alone. Given the fact that these two domains are quite flexible in terms of their relative orientations in cis, trans or hybridized trans, it appears plausible that further evolution of the accessory domain may aid to improve the overall processivity of B6.61. Alternatively and more promisingly, appending a second random pool on the ribozyme and performing reselection by in vitro encapsulation will likely allow for the isolation of a third helper module to the ribozyme system (Chapter 4). This new three-domain RNA polymerase ribozyme would hopefully help to resolve the current limitations of B6.61 to bring closer the goal to isolate an RNA replicase ribozyme.

3.4.2 A Structural Model of B6.61

The three-dimensional structure of the Class I ligase ribozyme (1QXI, RCSB020188), modelled using Fe(II)-EDTA probing data and comparative sequence analyses (Bergman et al., 2004), is shown in Figure 3-27A. Although this structure
Figure 3-27: Insertions in the Ligase Core in Context with the 3D Model of the Class I Ligase Ribozyme

(A) Three-dimensional model of the Class I ligase core (1QXI (RCSB020188, (Bergman et al., 2004))) with single-stranded regions that contain tag insertions in $^0$tL.1, $^{44}$tL.1, $^{79}$tL.1 and $^{106}$tL.1 (tL.1) coloured. The active site (green/white) and the J3.4 loop that contains C26 is also coloured (pink). (B) Cartoon schematic of tag-inserted tL.1 hybridized to tA.1.

contains more residues and contains some different residues compared to the ligase core of B6.61, it is likely to resemble the 3D structure of the ligase core of B6.61.

However, caution and skepticism should always be kept when making any conclusions about the ligase core based on this model. At first glance, it appears that the relative positions of the tag insertions on the 3D model of the Class I ligase ribozyme are consistent with our data. When any of tL.1 ($^{106}$tL.1), $^0$tL.1, $^{44}$tL.1 and $^{79}$tL.1 were
hybridized to tA.1, all four resulting hybridization constructs retained polymerase activity (Figure 3-11). However, only the tL.1:tA.1 and $^0$tL.1.tA.1 systems had catalytic activity similar to that of unimolecular B6.61. What is interesting is that the single-stranded regions of the Class I ligase core (yellow and orange, respectively, on Figure 3-11A) that would contain a tag sequence in $^0$tL.1 and tL.1 appear on the same side of the phosphodiester bond formation active site ($^5$'phosphate shown in green, $^3$'-hydroxyl shown in white), whereas those that would contain a tag sequence in $^{44}$tL.1 and $^{79}$tL.1 appear on the other side of the active site.

This 3D model is also consistent with the location of the C26 residue in the J3/4 single-stranded region between the P3 and P4 helices of the ligase core (see Figure 3-2A for the secondary structure of B6.61). The C26 residue, that is proposed to form a tertiary contact with U154 and U155 in the accessory domain, was modelled to be pointing outwards from the face of the ligase core that contains the single-stranded region that would contain a tag sequence in tL.1. This is consistent with the fact that the accessory domain would be situated at that face of the ligase core; also, considering that the location of active site chemistry is in the same vicinity, the accessory domain could be validly hypothesized to stabilize the primer-template substrate near the active site by forming a close tertiary contract with the ligase core.

The combination of the 3D structure of the Class I ligase ribozyme together with that of the minimized D.2 accessory domain (section 3.3.10, Figure 3-26) would provide enough information to build an approximate structural model for B6.61. The ingredients used to build this more complete model would come from length
constraints observed in unimolecular constructs L.1-[A]₇-D.1 containing variable linker lengths (section 3.3.5) and confirmed tertiary contacts between the two domains (section 3.3.8). The final result may serve as a blueprint to further our understanding of structure-function relationship in RNA polymerase ribozymes, and most importantly, as a modular structure to build upon in an attempt to isolate or systematically engineer a truly processive RNA replicase ribozyme.

### 3.5 Contributions

Peter Unrau, Sunny Wang and I planned the experiments, analyzed results, wrote the manuscript and generated the figures. Sunny Wang and I performed the experiments. Specifically, I performed the chemical synthesis of the DNA-PEG-crosslinker-RNA tethered primers, verified their sequence by native gel shift assay, synthesized the activated P9₇ photoactive primer, verified the crosslinking ability of the photoactive primer, performed all polymerization assays in this study and quantified and analyzed all kinetic data. Sunny Wang synthesized all ribozyme constructs in this study by PCR, transcription and gel purification, performed DEPC probing, Fe-EDTA radical cleavage experiments and all UV crosslinking experiments and mappings. Our collaborator, Dr. Ali Mokdad from Dr. François Major’s laboratory (Université de Montréal), performed the 3D modeling of the D.2 accessory domain.
CHAPTER 4: Improvement of an RNA Polymerase Ribozyme by In Vitro Encapsulated Selection

4.1 Introduction

The “RNA World” hypothesis speculates that early life on Earth contained a period of time when RNA played not only the role of storing genomic information but also that of catalyzing biological reactions (Gilbert, 1986). This premise has not only been strengthened by the plethora of both naturally occurring ribozymes such as the ribosome (Cech, 2000), RNase P (Guerrier-Takada et al., 1983) and telomerase (Qiao & Cech, 2008) and artificially selected ribozymes able to perform various small molecule chemistries (Ellington et al., 2009), but also by recent findings that DNA-dependent enzymes, such as RNA polymerases, originally thought to solely recognize DNA substrates (in the case of RNA polymerases, as templates for transcription), have been found to able to utilize RNA as well (Wassarman & Saecker, 2006; Lehmann et al., 2007).

An RNA replicase ribozyme has been termed the holy grail (McGinness & Joyce, 2003) of the “RNA World” hypothesis for the simple reason that an RNA molecule capable of self-replication provides a solution towards the “chicken or egg” paradox that arises from the mutual requirement of DNA and protein in their syntheses. Recently, a self-replicating ribozyme system was produced in which two partially complementary ribozymes catalyze each other’s synthesis by ligating two oligonucleotides to form the other enzyme (Lincoln & Joyce, 2009). While this system provides the first example of RNA catalyzing its own synthesis, it is unlike modern protein-based nucleic acid
replicators that sequentially polymerize nucleotides in a template-dependent manner. Furthermore, for such a system to exist in an early “RNA World”, the oligonucleotide substrates must have been synthesized by some non-ligation-based mechanism such as template-directed polymerization. Thus, a robust RNA polymerase ribozyme is still required to substantiate the “RNA world” hypothesis.

As mentioned earlier in section 2.3.1 and 3.1.1, the most recent variant of RNA polymerase ribozymes to date is the B6.61 ribozyme isolated from in vitro encapsulated selection (Zaher & Unrau, 2007). It is able to extend a trans primer-template by up to 20 nucleotides in 24 hours. While B6.61 is currently the best RNA polymerase ribozyme, it suffers from weak primer-template binding (Lawrence & Bartel, 2003) and unpredictable vigour with different primer-templates. In this study, we have appended a third random “helper” domain to the 5' end of the class I ligase core of a modified B6.61 variant, \(^{44}\text{tB6.61}\), and performed further in vitro encapsulated selection with hopes that the extra sequence in the new domain would help address and resolve the aforementioned limitations.

4.2 Materials and Methods

4.2.1 Oligonucleotides

DNA oligonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer by standard cyanoethyl phosphoramidite chemistry. 3'-Biotinylated probes were synthesized using 3'-BiotinTEG CPG columns (Glen Research). RNA oligonucleotides
were purchased from Dharmacon and deprotected according to the company’s protocol; deprotected RNAs were dried by speed vacuum.

4.2.2 Construction of Selection Pool

Construction of the selection pool was performed using a procedure optimized by Zaher and Unrau (Zaher & Unrau, 2005). The double-stranded PCR product that is used to transcribe the $^{44}$tB6.61 construct (Cheng et al., 2009) was diluted and used as a template for PCR reamplification in order to add an Ear I digestion site. Two primers were used. The forward primer had the sequence $^{5'}$ GTA ATG CCT CTT CTA ACg gaa tac aaa aga caa atc tgc where bolded residues is the Ear I recognition site, italicized residues is the three nucleotide overhang upon Ear I digestion and lowercased residues are those that make up the 5' sequence of $^{44}$tB6.61; the reverse primer had the sequence $^{5'}$ GGA GCC GAA GCT CCG GG that is, exactly, the reverse complement of the 3' sequence of $^{44}$tB6.61. The resulting PCR product is termed “pre-Ear I-$^{44}$tB6.61”. A second PCR product was made that corresponded to the 5' random sequence that is to be appended to the 5' end of $^{44}$tB6.61. An oligonucleotide having the sequence $^{5'}$ CGT AGT GCT CTT CAG TTN$_{60}$ cct ata gtg agt cgt at (where N$_{60}$ represents a 60-nucleotide random sequence, bolded residues is the Ear I recognition site and lowercased residues hybridize to the forward primer described as follows) was used as a template for PCR amplification. The forward primer had the sequence $^{5'}$ CTG CTA AAC CCC GTG GCC ATC TCC TTC TTT Aat acg act cac tat agg, where bolded residues is the Btg I recognition site, underlined residues is the T7 promoter and lowercased residues hybridize to the
oligonucleotide template as described earlier; and the reverse primer had the sequence
5’ CGT AGT CCT CTT CAG TT, where bolded residues is the Ear I recognition site and
italicized residues is the three nucleotide overhang upon Ear I digestion. The resulting
PCR product is termed “pre-Ear I-Helper”. The two double-stranded constructs “pre-Ear
I-44b6.61” and “pre-Ear I-Helper” were each digested to completion with Ear I at 37°C
overnight and phenol-chloroform extracted, ethanol precipitated and loaded onto an
8% 29:1 acrylamide:bis-acrylamide native gel. Upon gel purification, the two double-
stranded constructs were mixed in a 1:1 ratio and ligated using T4 DNA ligase to
generate the full-length selection pool. The pool was digested to completion with Btg I
at 37°C for 2 hours and then treated with calf intestinal phosphatase to
dephosphorylate terminal phosphates. Finally, the Btg I-digested pool was ligated to
the 5’-phosphorylated selection DNA/RNA primer D6P11 (Table 4-1), and annealed to a	
twofold excess of the selection template T21 (Table 4-1) to generate the PT-pool.

4.2.3 Encapsulation Procedure

For early rounds, when the emulsion volume was large, encapsulated
transcriptions were performed in batches of 50-mL falcon tubes. In each falcon tube, a
2.5-mL transcription reaction was prepared containing 12.5 nM PT-pool in 40 mM Tris-
HCl at pH 7.9, 2.5 mM spermidine, 50 mM MgCl2, 0.01% Triton X-100, 10 mM DTT, 8 mM
GTP, 2 mM ATP, 2 mM CTP and 2 mM UTP. The resulting mix was incubated on ice at
4°C for 30 minutes before addition of 10 U/μL T7 RNA polymerase, after which 47.5 mL
of oil surfactant mixture, stored at 4°C, containing 0.5% Tween 80 (Sigma), 4.5% Span 80
(Sigma) in heavy mineral oil (Paraffin oil while, Anachemia) was immediately added.
DEPC-treated 6 mm glass beads (Fisher Scientific) were added to bring the volume of the resulting mixture to completely fill that of the falcon tube, then the falcon tube was taped using duct tape to a mini vortexer (VWR) and vortexed at full speed for 5 minutes. In later rounds when the emulsion volume was reduced to batches of 1-mL emulsion in 1.7-mL Eppendorf tubes, the emulsion was prepared by vortexing the tube at top speed for 1 minute after the addition of oil surfactant phase, and then homogenized using a dispersing tool (IKA WORKS) for 20 seconds. T7-dependent transcription was initiated by incubating emulsion prepared from both procedures at 37°C for 3 hours and then inactivated by heat-killing the enzyme at 65°C for 15 minutes. The emulsion was then allowed to incubate at 22°C overnight to allow for phenotype-dependent primer-template extension. After round 4, when an APM gel purification was added into the selection, encapsulated transcription was performed in the presence of an additional 1 mM 45UTP (TriLink Biotechnologies).

4.2.4 Recovery of Aqueous Phase and Selection

Overnight incubations were centrifuged at 13,200 rpm for 10 minutes to result in a white pellet and clear oil supernatant. Upon removing the oil supernatant, the white pellet was resuspended by addition of twofold excess of 30 mM EDTA and an equivalent volume of ethyl acetate to that of the emulsion, and then vortexed vigorously. The mixture was centrifuged briefly and the top organic phase was removed. In order to assess transcriptional yields, an aliquot of the aqueous phase was loaded onto a 2%
agarose gel along with a range of unrelated RNAs with varying concentrations. Finally, the recovered aqueous phase was phenol-chloroform extracted to remove T7 RNA polymerase enzyme and ethanol precipitated before resuspension in water.

In early rounds, an artificial fully extended selection primer (D4P35) ligated onto a marker double-stranded DNA construct of unrelated sequence but same length as the selection pool and fourfold molar excess (in relation to the template) of a competitor RNA oligonucleotide whose sequence is reverse complement to that of the selection template was added to the resulting aqueous phase, heated to 99°C for 5 minutes and loaded onto a denaturing 6% polyacrylamide gel. A region of the gel corresponding to the region between the ligated unextended primer and the ligated extended primer was cut and the nucleic acid was eluted overnight and ethanol precipitated. Post-round 4, when an APM gel purification was implemented into the selection, the standard gel purification step was replaced with an APM gel in which genomes that have been extended and incorporated a 4-thio-uridine would migrate slower in the mercury-containing gel. Upon gel purification of the shifted thio-containing nucleic acid, it was eluted in the presence of 1 mM DTT and ethanol precipitated.

Upon resuspension of the gel-purified nucleic acid in water, it was allowed to hybridize to a 10-fold excess of biotinylated DNA oligonucleotide capture probe that is complementary to the extended primer and applied onto streptavidin magnetic Dynabeads® (Invitrogen) and washed four times with 0.1x SSC (saline-sodium citrate) buffer. Captured, and thus hybridized, nucleic acid was eluted with 50 mM KOH,
followed by immediate neutralization with 1 M Tris-HCl and ethanol precipitation. The recovered nucleic acid was PCR re-amplified using two primers (forward primer: $^5$' CTG CTA AAC CCC GTG GC [bolded is Btg I site] and reverse primer: $^5$' GGA GCC GAA GCT CCG GG), digested with Btg I, ligated to the D6P11 selection primer and annealed to the T21 template before proceeding to the next round. Post-round 4, an extra denaturing gel purification step was introduced after PCR re-amplification to ensure removal of a parasitic artifact that is conveniently 20 to 30 base pairs longer than the selection pool.

4.2.5 Polymerization Assay

For extension under optimal conditions, gel-purified 5'-radiolabeled RNA primer at 0.1 μM final concentration was annealed to a tenfold excess of template and added 100 mM Tris HCl at pH 8.5. Gel-purified ribozymes were then added to the mixture at a final concentration of 2 μM (unless otherwise indicated). Extension was initiated with 200 mM MgCl$_2$ and 4 mM of each NTP and incubation at 22°C. When performing extensions under selection conditions, 40 mM Tris HCl at pH 7.9, 2.5 mM spermidine, 50 mM MgCl$_2$ and 0.01% Triton X-100 was added to the primer-template; extension was initiated with 8 mM GTP, 2 mM ATP, 2 mM CTP and 2 mM UTP. Extensions under either conditions were stopped by adding a fourfold excess of a mix containing 40 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol, 80% formamide, and a four-fold excess of an RNA oligonucleotide whose sequence hybridizes to the template. Samples were heated to 95°C and loaded onto a 20% polyacrylamide sequencing gel.
4.2.6 Optimization of Hybridization-Based Capture

The RNA sequence $^5$d(CAC G)(18S)(18S) (18S)CA CGC UGC CAA CCG UGC GAA GCG UCU ACU CGG UCC (D4P35) was subjected to partial alkaline hydrolysis in the presence of 37.5 mM NaHCO$_3$ at 90°C and then hybridized to several complementary biotinylated oligonucleotides with variable 3’ ends. The mixture was applied to streptavidin magnetic Dynabeads® (Invitrogen) and washed four times with 0.1x SSC (saline-sodium citrate) buffer. Upon elution of the hybridized nucleic acid with 50 mM KOH, it was run, together with the wash, in a 20% polyacrylamide sequencing gel.

4.2.7 Assessment of Pool Diversity

Calf intestinal alkaline phosphatase-treated RNA pools for chosen selection rounds were phosphorylated with γ-$[^{32}P]$-ATP using T4 polynucleotide kinase (Invitrogen) and gel purified on a 6% denaturing polyacrylamide gel. The ethanol-precipitated eluent was treated with T1 RNase (Fermentas) and the partial digestion patterns of the RNA pools were then resolved on a 12% polyacrylamide sequencing gel.

4.2.8 Kinetic Analysis

Using ImageQuant v5.1, kinetic analysis was performed by first quantifying the radioactivity above the unextended primer in each lane of a sequencing gel used to resolve extension and dividing by the total radioactivity in the lane. The fraction extended was plotted against time and fit to $y = a(1 - e^{-kt})$ using GraphPad Prism v5.00, where ‘a’ is the initial fraction extended and ‘k’ is the first-order kinetic rate. Figure 3-1 details the steps taken to obtain the first-order extension rate.
4.2.9 Cloning

Cloning of isolates from pools was performed using the TOPO-TA cloning kit (Invitrogen). Sequencing was performed by Macrogen Inc., Korea.

4.2.10 Generating Recombinations of Helper Domains and Intact B6.61

In order to generate ribozymes that contain helper domains from unique isolates fused with an intact B6.61 sequence, three successive PCR amplifications were performed. Firstly, two PCR amplifications were performed: the first used the primers $5^\prime$ CTG CTA AAC CCC GTG GC and $5^\prime$ GAC AAA TCT GCC CTC AG to amplify the helper domain out of individual isolates and the second used the primers $5^\prime$ CTG AGG GCA GAT TTG TC and $5^\prime$ GGA GCC GAA GCT CCG GG on an intact double-stranded PCR template of the B6.61 sequence to amplify B6.61. After purification of each PCR product by QIAquick columns (Invitrogen), the two fragments were mixed at 10 nM each in the absence of primers, and thermocycled for 30 cycles of PCR. The resulting product was diluted and then re-amplified using primers corresponding to their termini ($5^\prime$ CTG CTA AAC CCC GTG GC and $5^\prime$ GGA GCC GAA GCT CCG GG) to generate the final PCR product. Ribozymes were then transcribed using the PCR product as a template.

4.3 Results

4.3.1 Construction of Selection Pool

In order to improve the processivity of the B6.61 RNA polymerase ribozyme, we hoped that the addition of a third “helper” domain to the 5' end of its already modular
structure consisting of a ligase core and accessory domain would provide extra sequence sufficient to improve one or more of NTPs binding, NTPs utilization, primer-template binding, general primer-template utilization, and fidelity. We imagined that the extra sequence would fold back to interact with pre-existing ribozyme sequence. From current results by Sunny Wang and I (Chapter 3), we have shown that the primer-template substrate can localize at particular regions of the B6.61 ribozyme via a short 12-nt double-stranded RNA duplex and increase catalytic rate (Cheng et al., 2009). In particular, the $^{44}\text{tB6.61}$ construct that locks the primer-template into the L5 loop of the ligase core displayed the greatest increase in catalytic rate – roughly ten fold.

Considering the fact that this 12-nt insertion did not damage polymerase activity and that it represented a region in the tertiary structure that is involved with primer-template binding, we are hopeful that retaining this longer loop in our selection pool would provide an extra scaffold for a proposed helper domain to interact with in order to improve polymerase activity.

Selection of truly trans catalysts has been made simpler with the use of in vitro compartmentalization. Starting with two double-stranded PCR products – one for the random helper domain that contains 60 random nucleotide positions and one for the $^{44}\text{tB6.61}$ construct, each was engineered by PCR to contain an Ear I restriction site at their 3' and 5' ends, respectively. The random helper PCR product also contained a Btg I restriction site upstream of a T7 promoter. After digestion with Ear I, the two digested elements were gel-purified from the short double-stranded DNA fragments that resulted from digestion and were then joined using T4 DNA ligase with only three
nucleotides of fixed sequence. Approximately 0.3 nmol of the resulting full-length construct was digested with Btg I and ligated to the D6P11 DNA/RNA primer with T4 DNA ligase. Finally, a two-fold excess of the selection template T21 was annealed to the ligated pool to produce the “Round 0” PT-pool, that had an approximate diversity of $10^{14}$ different sequences. Figure 4-1 shows the synthetic scheme of the final PT-pool and Figure 4-2 is a 2% agarose gel detailing all the steps taken to synthesis the pool.

**Figure 4-1: Design and Construction of Selection Pool**

Using PCR, a Btg I restriction site was added to the 5' end of the random helper domain (green) and Ear I restriction sites were added to the 3' end and 5' end of the random helper domain and $^{44}$B6.61 (ligase core is in pink and accessory domain is in blue), respectively. Upon digestion with Ear I followed by gel purification, the two elements were joined by T4 DNA ligase. Further digestion with Btg I, ligation of the selection DNA/RNA primer D6P11 (DNA in orange, PEG linker in lavender and RNA primer sequence in magenta) and hybridization of the selection template T21 (dark green) resulted in the final PT-pool as the input nucleic acid for *in vitro* encapsulated selection.
Figure 4-2: Construction of Selection Pool as Monitored by Gel Electrophoresis

As seen on the 2% agarose gel, the lanes corresponding to pre-Ear I digestion (Ear I-) of the random helper domain and of 44tB6.61 depict the double-stranded DNA product after PCR reamplification. Ear I+ lanes correspond to the completely digested products. Upon gel purification to eradicate the short dsDNA fragments that would interfere with ligation of the helper to 44tB6.61, the two elements were mixed in a 1:1 ratio (Ligase-) and supplied with T4 DNA ligase to facilitate ligation (Ligase+). Btg I digestion then removed another short dsDNA fragment thus leaving an open space for ligation of the D6P11 primer and hybridization of the T21 template (not shown).

4.3.2 Selection Scheme and Proof of Principle

Initially, two selective steps were implemented in the selection process. The first was a standard gel purification step that separated lengthened DNA genomes as a result of ribozyme-dependent primer-template extension from unextended genomes. Upon adding four-fold excess of a competitor RNA molecule that is complementary to the selection template (this excess amount of competitor was shown to be sufficient to denature the primer from the template; results not shown) and dividing the recovered aqueous phase into two parts, one was added radiolabeled unextended selection primer D6P11 ligated to an unrelated double-stranded DNA of the same length as the selection
pool and the other part was added radiolabeled extended selection primer mimic D4P35 ligated onto the same double-stranded DNA marker. Upon electrophoresis, gel purification of the region between the top of the bottom radioactive band (ligated D6P11) and the top of the top radioactive band (ligated D4P35) would, in turn, purify genomes that have been extended. After round 4, the gel purification step was modified to utilize a stacking APM-containing denaturing gel. Radiolabeled $^{45}$S-UTP-extended D6P11 (from B6.61-dependent extension) ligated onto the aforementioned double-stranded DNA marker was used as the gel purification reference: the region between the radioactive band and the APM interface in the stacking gel was excised.

The second selective step was one that relied on the successful hybridization event between extended RNA primers and a biotinylated DNA oligonucleotide. In order to optimize the capture of DNA genomes attached to an RNA primers extended past a certain number of nucleotides, we sought to determine the minimal hybridization requirements between the biotinylated probe and the RNA. Particularly, we were interested in determining the minimal number of hybridized nucleotides upstream (5') of the $n^{th}$ extension in order to capture RNA primers extended by $n$ nucleotides. A pool of RNA molecules was generated by the partial alkaline hydrolysis of a synthetic “mimic” RNA molecule (D4P35) whose sequence is the fully extended primer (according to a longer template, T35), resulting in varying lengths of RNA corresponding to every possible extension. The ladder was tested its ability to be captured by different biotinylated probes with varying 3' ends onto streptavidin magnetic beads, and samples before and after the capture step were run on a denaturing sequencing gel.
Figure 4-3: Optimization of Hybridization-Based Selective Step

(A) Experimental scheme. (1) A pool of 35 different RNA molecules generated by the partial alkaline hydrolysis of a synthetic “mimic” RNA molecule (the T35 template-based fully extended selection primer) was (2) hybridized onto five different biotinylated DNA oligonucleotides whose names reflect their 3’ ends (e.g. the “-4” probe starts hybridizing 4 nucleotides upstream of the extension start site; the “0” probe starts hybridizing at the extension start site; the “+4” probe starts hybridizing after 4 nucleotides of extension). In order to assess which of the 35 RNA molecules each biotinylated probe hybridizes, the hybridization reactions were captured on streptavidin magnetic beads, washed stringently and eluted. (3) The input hydrolyzed mimic, wash and eluent are resolved on a denaturing sequencing gel. (B) 20% polyacrylamide denaturing sequencing gel resolving the input hydrolyzed mimic [-], wash [W] and eluent [E] for each capture experiment for each biotinylated probe.
From Figure 4-3, it appears that at least ten nucleotides of hybridization between the extended RNA primer and any biotinylated probe are required for successful capture. For example, the “-4” probe starts hybridizing 4 nucleotides upstream of the extension start site, but starts capturing RNA molecules that have been “extended” by at least 6 nucleotides. Keeping this criterion in mind, we first decided on the minimum number of nucleotides of primer extension we wanted out our selection to enrich and based on this, decided on the most suitable biotinylated probe.

We decided to use the “tethered” primer-template system developed earlier (Chapter 3) to perform a polymerization assessment of the capabilities of the “Round 0 pool”. Upon T7-dependent transcription of the initial selection pool, the resulting RNA pool (hereafter termed the “Round 0 pool”) was gel purified and resuspended in water. Using the selection primer and template as is (“Round 0” lanes in Figure 4-9), the Round 0 pool can extend seven nucleotides under optimal conditions (high pH and high Mg\(^{2+}\) concentration), but can barely extend one nucleotide under selection conditions (lower pH and lower Mg\(^{2+}\) concentration). Exploiting the fact that localizing the primer-template in the L5 loop of the \(^{44}\text{tB6.61}\) construct increased catalytic activity by approximately 10-fold (Cheng et al., 2009), we decided to use this exact system to assess how efficient the \(^{44}\text{tB6.61}\) construct and the Round 0 pool (which also contains the same expanded L5 loop) are in extending the selection primer-template under selection conditions. The sequencing gel in Figure 4-4 shows that under optimal conditions, both \(^{44}\text{tB6.61}\) and the Round 0 pool can extend 14 nucleotides to the end of the template sequence. Under selection conditions, both constructs can now extend 10
nucleotides in 24 hours under selection conditions. We hope that our newly appended helper domain will be able provide that extra catalytic “push” currently established by localizing the primer-template to the L5 loop. With this goal in mind, we will need to use the “0” biotinylated probe in the selection process in order to enrich for at least 10 nucleotides of extension. Afraid that there are only a small number of sequences in the starting pool with this catalytic capability, we ultimately decided to be less stringent and use the “-4” biotinylated probe that can enrich for at least 6 nucleotides of extension.

Figure 4-4: Pre-Selection Polymerization Assessment

(A) “Tethered” primer-template construct used for polymerization assessment. The primer-template P11:T21 is tethered to either 4tB6.61 or the Round 0 pool by a 12-nucleotide DNA-RNA helix. (B) 20% sequencing gel resolving polymerase activity of B6.61, 4tB6.61 and the Round 0 pool using the “tethered” primer-template under optimal conditions (100 mM Tris-HCl @ pH 8.5, 200 mM MgCl₂) and selection conditions (40 mM Tris-HCl @ pH 7.9, 50 mM MgCl₂). The primer-template used was RTag-P11:T21.
Finally, the capture ability of the streptavidin magnetic beads (Invitrogen) was tested with the “-4” probe together with either the selection primer (D6P11) or the extended primer “mimic” (D4P35). In addition, D6P11 and D4P35 were ligated onto the Btg I-digested round 0 pool in order to examine the effects on hybridization by the additional double-stranded DNA sequence. Radiolabeled primer (D6P11), mimic (D4P35), ligated primer (Pool-D6P11) and ligated mimic (Pool-D4P35) were allowed to hybridize to the “-4” biotinylated probe and captured on streptavidin magnetic beads. This was performed identically to what would happen during selection. Scintillation counting was then performed on the wash, eluent and remaining streptavidin beads of each capture experiment (Figure 4-5). Clearly, the “-4” probe fails to capture D6P11 nor Pool-D6P11 but captures both D4P35 and Pool-D4P35 at an approximate 80% efficiency.

![Figure 4-5: Assessment of Streptavidin Beads Capture Selective Step](image)

Scintillation counting assessment of the capture of radiolabeled primer (D6P11), mimic (D4P35), ligated primer to pool (Pool-D6P11) and ligated mimic to pool (Pool-D4P35) revealed that the “-4” probe fails to capture any variant of the unextended primer but captures the mimic and ligated mimic at an approximately 80% efficiency.
Figure 4-6: In Vitro Encapsulated Selection Scheme

The bottom cycle (black arrows) corresponds to the selection procedure used for rounds 1 to 4 and the top cycle (grey arrows) corresponds to that for rounds 5 to 8. For rounds 1 to 4, T7-dependent encapsulated transcription of DNA genomes within water-in-oil vesicles resulted in an RNA phenotype. The genomes of active RNA catalysts that can extend a primer-template substrate tethered to the DNA genome are enriched firstly by size in a gel purification step and secondly by hybridization to a biotinylated probe complementary to the extended primer followed by capture on streptavidin magnetic beads. The hybridized extended RNA primers are eluted from the beads and re-PCR amplified. The PCR product was then Btg I-digested, ligated to the selection primer D6P11 and annealed to the selection template T21. For rounds 5 to 8, encapsulated transcription was performed in the presence of 35UTP and active RNA catalysts were asked to use 35UTP for templated-primer extension. Genomes of active catalysts were purified firstly in an APM denaturing gel based on their newly acquired 4-thio-uridine residues in the extended primer as a result of ribozyme-based extension; and secondly in the same hybridization-based capture as for rounds 1 to 4. Hybridized extended RNA primers are eluted from the beads and re-PCR amplified. The PCR product was then gel-purified, then re-PCR amplified, Btg I-digested, ligated to the selection primer and hybridized to the selection template.
Table 4-1: Primer, Template and Biotinylated Probe Sequences

“d(Nₙ)” and “D” denote deoxyribonucleotide residues, “(18S)ₙ” denotes three “Spacer 18” polyethylene glycol linkers, “B” denotes a 3’ biotin-triethyleneglycolyl-glyceryl, “P” denotes primer and “T” denotes template.

<table>
<thead>
<tr>
<th>Oligo</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>D6P11</td>
<td>5’ d(CAC GAG) (18S)₃ CAC GCC GCC AA 3’</td>
</tr>
<tr>
<td>D4P35</td>
<td>5’ d(CAC G) (18S)₃ CAC GCC GCC AAC CGU GCC AAG CGU CUA CUC GGU CC 3’</td>
</tr>
<tr>
<td>P11</td>
<td>5’ CAC GCC GGC AA 3’</td>
</tr>
<tr>
<td>P12</td>
<td>5’ CU GCC AAC CGU G 3’</td>
</tr>
<tr>
<td>T21</td>
<td>5’ GA CGG UUG GCA CGC UUC GCA G 3’</td>
</tr>
<tr>
<td>T35</td>
<td>5’ GUG CGA CGG UUG GCA CGC UUC GCA GAU GAG CCA GG 3’</td>
</tr>
<tr>
<td>“-4” probe</td>
<td>3’ B-GG TTG GCA CGC TTC GC 5’</td>
</tr>
<tr>
<td>“-2” probe</td>
<td>3’ B-TTG GCA CGC TTC GC 5’</td>
</tr>
<tr>
<td>“+0” probe</td>
<td>3’ B-G GCA CGC TTC GC 5’</td>
</tr>
<tr>
<td>“-4” probe</td>
<td>3’ B-CGC TTC GCA GAT GAG CCA GG 5’</td>
</tr>
<tr>
<td>“+8” probe</td>
<td>3’ B-TC GCA GAT GAG CCA GG 5’</td>
</tr>
</tbody>
</table>

After optimizing both selective steps, the selective scheme shown in the bottom cycle of Figure 4-6 (black arrows) is proposed. Six cycles of in vitro selection was initially performed with this scheme using D6P11 as the selection primer and T21 as the selection template (Table 4-1).

4.3.3 In Vitro Encapsulated Selection

4.3.3.1 Initial Selection Led to an Artifact

The first six rounds of in vitro encapsulated selection began smoothly. In order to maintain the high diversity of the pool, 20 mL of encapsulated transcription (which translates to 400 mL of emulsion) was performed in the first round. Encapsulated transcription was monitored by running an aliquot of the recovered aqueous phase onto an agarose gel. After confirming to have decent transcription levels (at least ~2 μM RNA transcripts), the recovered post-transcription aqueous phase was placed through the gel purification step and hybridization-based capture step for enrichment of active
sequences. The second (5 mL transcription, 100 mL emulsion), third (200 μL transcription, 4 mL emulsion), fourth, fifth and sixth rounds (all 100 μL transcription, 2 mL emulsion) proceeded afterwards. However, after checking polymerase activity of transcribed RNA from each round, activity was decreasing after round 3. The size of the pool was also increasing from round 4 and reached a definite, longer length by round 6.

The catalytically inactive, longer artifact was characterized by PCR analysis and hybridization-based studies. Single primer PCR re-amplification of round 5 DNA (using only one of the standard forward primer for PCR re-amplification of the DNA pool, the standard reverse primer for pool re-amplification, and a non-biotinylated variant of the capture; see section 4.2.4 for sequences) revealed that the higher mobility artifact was amplified with only the reverse primer present (Figure 4-7A). Standard double primer PCR with all three permutations of the three primers revealed that when the “normal”

![Image](image.png)

Figure 4-7: Characterization of Selection Artifact – Part 1

(A) 2% agarose gel resolving a series of PCR re-amplifications. Re-amplification of round 5 DNA using different combinations of primers revealed the reverse primer alone is required to reamplify the higher mobility artifact (white arrows). Only when both the forward and reverse primers are supplied, the correct length pool is re-amplified (yellow arrow). (B) 12% denaturing gel showing PCR products from re-amplification of Round 3, 4, 5 and 6 using forward primer and radiolabeled reverse primer. Rounds 3 and 4 PCR products show DNA of the correct length, while round 5 and 6 products are clearly longer.
forward and reverse primers were used for amplification, two products appeared – the
correct lower mobility product corresponding to the pool (Figure 4-7A yellow arrow) and
the incorrect higher mobility artifact (Figure 4-7A white arrow). When the “normal”
reverse primer was used together with a non-biotinylated variant of the “-4” capture
probe, a higher mobility artifact was also observed; this most likely corresponds to the
same artifact as when only the reverse primer was used. At this point, we concluded
that the longer artifact observed has the reverse primer sequence at both of its ends.

In order to understand more about the artifact, the dominant bands in the “R3*”
and “R6A*” lanes in the gel in Figure 4-7B were carefully gel-purified. The DNA was then
allowed hybridize to the “-4” biotinylated probe and applied onto streptavidin magnetic
beads. The wash, eluent and beads were then assessed by scintillation counting.

**Figure 4-8: Characterization of Selection Artifact – Part 2**

Hybridization-based capture experiments for gel-purified round 3 (R3*) and round 6 (R6A*) DNA using the
“-4” biotinylated capture probe and streptavidin magnetic beads revealed that round 3 DNA was not
captured while round 6 DNA was captured at an approximate 40% efficiency.
Figure 4-8 shows that the round 6 DNA pool had acquired sequence that allowed it to be captured by the “-4” biotinylated probe. Reflecting back on the selection scheme, the first gel purification step would have enriched for these artifacts based on length and the second hybridization-based capture step would have also enriched for them based on sequence. In order to circumvent this undesired enrichment, the dominant band the “R4*” lane in the gel in Figure 4-7B was gel-purified and placed through four more selective rounds (round 5B, 6B, 7 and 8) using a modified selection scheme (top cycle with grey arrows in Figure 4-6). The first major modification was the addition of 1 mM $^{45}\text{UTP}$ in the encapsulated transcription mix. Active RNA catalysts were consequently asked to extend the selection primer tethered to its DNA genome using $^{45}\text{UTP}$. Those that were able to add a 4-thiouridine residue to its genome were enriched in an APM-containing denaturing stacking polyacrylamide gel that retards the mobility of sulfur-containing moieties to facilitate their gel purification. The eluted RNA is subjected to the same hybridization-based selective step and PCR re-amplification. To ensure the higher mobility artifact does not reappear anymore, the PCR product is then gel purified in a denaturing gel and then re-PCR amplified. Finally, the new PCR product is $Btg$ I-digested, ligated to D6P11 primer and hybridized to T21 template.

### 4.3.3.2 Assessment of Activity and Diversity of RNA Pools from Selection Rounds

Starting with the gel-purified round 4 DNA from the gel in Figure 4-7B, four additional selection rounds were performed using the scheme as described above (shown in the top cycle in Figure 4-6). The polymerization activity of the pool was
assessed for catalytic improvement relative to B6.61 using P11:T21 and P12:T21 primer-templates under optimal and selection conditions. P11 is the RNA segment of the selection primer D6P11 and P12:T21 is a PT substrate that B6.61 can extend quite well.

Figure 4-9: Polymerization Assessment of RNA Pools from Selection Rounds – Part 1

20% polyacrylamide sequencing gel resolution of extension reactions catalyzed by B6.61 and RNA pools from selection rounds under optimal and selection conditions. The primer and template used were P11 (the RNA portion of the selection primer D6P11) and T21, respectively.
Figure 4-10: Polymerization Assessment of RNA Pools from Selection Rounds – Part 2

20% polyacrylamide sequencing gel resolution of extension reactions catalyzed by B6.61 and RNA pools from selection rounds under optimal and selection conditions. The primer-template substrate used was P12:T21 and is one that B6.61 can extend quite well.

As seen in Figure 4-9, extension of the selection primer-template by the selection pools from rounds 3 through 8 under optimal conditions is slower than that by pools from rounds 0 through 2. Also, all the pools tested are slower than B6.61. Despite the apparent decrease in catalytic activity as the selection progressed, extension assays under selection conditions produced a somewhat different conclusion. While B6.61 can
barely extend one nucleotide, selection pools for rounds 2, 3, 4, 5B and 7 can clearly extend two nucleotides. Although there was nevertheless minimal consumption of the radiolabeled primer, this increase in polymerization may represent the existence of interesting helper domains that may help to improve the processivity of the ribozyme.

When considering the extension assay using the P12:T21 substrate under both optimal and selection conditions (Figure 4-10), the same pattern is seen compared to that in Figure 4-9A: selection pools from rounds 3 through 8 are slower than those from rounds 0 through 2, which are all slower than B6.61. It appears that the selection process may have specifically improved polymerization ability relative to the starting pool solely using the selection primer-template under non-optimal selection conditions.

In order to understand more the diversity levels of the pools from each selection round, transcribed RNA from the selection pools that saw an increase in polymerization using the selection primer-template under selection conditions were partially digested with T1 RNase that cuts specifically after guanosine residues (Pace et al., 1991). For a completely random sequence, such as the round 0 pool, the digestion pattern on a denaturing gel should appear as a smear as a result of its high diversity. However, pools from later selection rounds, where the diversity had supposedly decreased as a result of the selection process, should possess a discrete band pattern above a smeary background. As seen in Figure 4-11, the T1 RNase digestion pattern for 5’-[32P]-radiolabeled round 0 RNA is a smear for the random helper domain, as expected. The same smeary pattern is also observed for RNA from rounds 2, 3 and 5.
Figure 4-11: T1 RNase Digestion of Selection Pool RNA

12% polyacrylamide sequencing gel resolution of partial T1 RNase digestion of rounds 0, 2, 3, 5B, 7 and 8 RNA. Since the RNA was 5'-[32P]-radiolabeled, the gel is read 5' to 3' from bottom to top. Hence, the top segment corresponds to the constant ligase core and accessory domain from \(tB6.61\) and the bottom segment corresponds to the random 5'-helper domain. Three timepoints were taken for each digestion reaction: 2, 5 and 10 minutes.
However, discrete band patterns emerged in round 7 RNA and have become quite prominent in round 8 RNA. Since the round 7 pool can extend the selection primer-template by two nucleotides under selection conditions while round 8 pool cannot extend at all, the discrete bands may correspond to an undesired catalytically inactive artifact. Based on these results, we have decided to clone isolates from round 7 to screen for improved polymerase activity in unique sequences and also from round 8 to understand the primary sequence and secondary folds of the possible artifact(s).

4.3.3.3 Initial Analysis of Isolates from Round 7 and Round 8

32 unique round 7 sequences and 19 unique round 8 sequences were cloned into DH5α cells and amplified by colony PCR. All isolates were sequenced and screened for polymerase activity. In order to perform a quick and crude screening for activity, clones were transcribed at 37°C by T7 RNA polymerase, heated to 65°C for 5 minutes to heat-inactivate the protein enzyme, and immediately added to a primer-template extension reaction as described earlier in section 4.2.5.

As seen in Figures 4-13 and 4-14, the round 7 pool contained a diverse set of isolates. Out of 32 clones, 9 sequences (7.04, 7.06, 7.09, 7.12, 7.13, 7.15, 7.17, 7.27 and 7.28) were very similar in their 5' helper domain regions and displayed large differences in their ligase cores and accessory domains when compared to B6.61. In fact, careful analysis of their sequences revealed that they contain insertions of 4 to 7 nucleotides of the sequence of the “-4” biotinylated probe used during selection. And as expected, since these sequences hybridize to the selection primer and to the P12 primer, the
Figure 4-12: Polymerization Assessment of Isolates from Round 7 Pool

20% polyacrylamide sequencing gel resolution of extension reactions catalyzed by B6.61, \(^{44}\text{tB6.61}\), the round 7 pool and individual isolates from the round 7 pool under optimal extension conditions. The primer and template used were P12 and T21, respectively. (A) Polymerization assessment of isolates 7.01 through 7.11; (B) polymerization assessment of isolates 7.12 through 7.22; (C) Polymerization assessment of isolates 7.23 through 7.32. Since transcription reactions of isolates were heat-killed and immediately added to extension reactions, “HK T7” is a negative control in which heat-killed T7 RNA polymerase was added to the extension.
Figure 4-13: Sequence Alignment of Round 7 Isolates – Part 1

Alignment of S’ sequence of round 7 isolates RNA (between nucleotides 1 and approximately 120), together with the sequences of B6.61 and 86B6.61. The AAC motif between the helper domain and ligase core is at position 94 to 96. The ligase core begins at position 97.
Figure 4-14: Sequence Alignment of Round 7 Isolates – Part 2

Alignment of 3’ sequence of round 7 isolates RNA (from nucleotides ~120 to the 3’ end), together with the sequences of B6.61 and 44tB6.61. The accessory domain “begins” at approximately nucleotide 120.
sequencing gels (as shown in Figure 4-12) that resolved the extension of the P12:T21 primer-template show a smear of the radiolabelled primer for extension reactions catalyzed by these constructs. The sequences might have been generated by a strange and uncharacterized recombination mechanism during the PCR re-amplification step after each round that was seen previously during the selection of B6.61 by Hani Zaher.

In addition to the aforementioned parasitic sequences, all other sequences are either inactive or are less active than B6.61 and 44tB6.61. With the exception of point mutations (that are known to happen during PCR amplification) in the supposedly constant ligase core and accessory domain, there are, interestingly, whole regions that were deleted in certain constructs. For example, the expanded L5 loop that was purposely engineered into the selection pool was deleted in 8 constructs (7.03, 7.05, 7.07, 7.10, 7.11, 7.16, 7.20 and 7.22). Whether this is a result of contamination of the initial 44tB6.61 PCR product used to generate the selection pool is unknown. Also, 5 constructs (7.16, 7.19, 7.23, 7.29 and 7.32) strangely had a deletion of one strand of the P6 helix of the ligase core (“bottom” strand in Figure 3-2A) and interestingly, the disruption of this fundamental secondary structure element in 7.19, 7.29 and 7.32 did not abort activity completely (as seen in Figure 4-12). As expected, 7.16 and 7.23 are completely inactive. The best isolate of all round 7 sequences, albeit less active than B6.61, is 7.08. Apart from having a distinct helper domain compared to other isolates, it contains several point mutations in its accessory domain and most interestingly contains a unique deletion again within the ligase core: the L6 loop and parts of the P6 helix flanking the loop are deleted.
Round 8 isolates were less diverse than those from round 7; this is consistent with the T1 RNase digestion assay earlier (Figure 4-11). A quick scan of the sequences in Figure 4-16 will quickly lead one to conclude that there are two main types of sequences occurring 8 and 7 times, respectively, out of 19 isolates. The first set (8.01, 8.03, 8.04, 8.10, 8.16, 8.17, 8.19 and 8.20) contains the same helper domain with the exception of a few point residue differences. Most interestingly, while they contain the deliberately engineered expanded L5 loop, these sequences contain a 10-nucleotide deletion corresponding to the “bottom strand” of the P6 helix (see Figure 3-2A for the secondary structure of the B6.61 ribozyme). The remainder of their sequence corresponding to the accessory domain remains largely unchanged relative to that 44tB6.61, with the exception of a few point mutations.

![Image of gel electrophoresis](image)

**Figure 4-15: Polymerization Assessment of Isolates from Round 8 Pool**

20% polyacrylamide sequencing gel resolution of extension reactions catalyzed by B6.61 and 9 unique isolates from the round 8 pool under optimal extension conditions. The primer and template used were P12 and T21, respectively.
Figure 4-16: Sequence Alignment of Round 8 Isolates

Alignment of sequence of round 8 isolates RNA together with the sequences of B6.61 and 44tB6.61
Of this first set, there were only 4 unique sequences, 8.03, 8.10, 8.19 and 8.20 and according to the sequencing gel shown in Figure 4-15, all four were barely active, with an activity more than 300 fold slower than that of B6.61.

The second set of 7 sequences (8.06, 8.07, 8.08 8.09, 8.11, 8.12 and 8.14) contains the exact same sequence. To be specific, they contain a helper domain completely different from the one observed in the first set. In addition, they all lack the $^{44}t$ insertion in the L5 loop but otherwise, contain an intact ligase core. The sequence of their accessory domain is quite different compared to the wildtype sequence; based on the large number of modifications, it seems unlikely that these variants are active. Sequence 8.06 was tested for activity and as expected, it is completed inactive as shown in Figure 4-15.

Four other sequences (8.02, 8.05, 8.13 and 8.15) show minimal conservation to each other and to the two aforementioned sets of sequences in their helper domains. 8.02 and 8.15 have a similar ligase core and accessory domain to the second set of 7 sequences, while 8.05 and 8.13 are completely different with a great deal of mutations throughout. Interestingly, even though the second set of 7 sequences was completely inactive, 8.02, which had the same ligase core and accessory domain, was still somewhat active (Figure 4-15). Section 4.3.3.4 will detail our attempts in determining whether it is its helper domain improving the processivity of its inactive ligase core and accessory domain. On the other hand, 8.05, 8.13 and 8.15 are all inactive.
4.3.3.4 Recombining Helper Domains with Intact B6.61 Sequence

Since a large number of the isolates screened contained heavily mutated ligase cores and accessory domains that would render most of these sequences inactive, it was interesting to recombine their helper domains with an intact B6.61 sequence in order to more accurately examine the effects of having the new helper domain. This is especially true for constructs that are active even though they have a large number of mutations in their ligase core and accessory domain sequences. The helper domains in isolates 7.08, 7.32, 8.02 and 8.06 were recombined with intact B6.61 (resulting in constructs H7.08-B6.61, H7.32-B6.61, H8.02-B6.61 and H8.06-B6.61) and assessed for their ability to extend the P12:T21 primer-template (Figure 4-17).

Figure 4-17: Activity Assessment of Round 8 Isolates

20% polyacrylamide sequencing gel resolution of extension reactions catalyzed by B6.61, B6.61, gel-purified ribozyme isolates 7.08, 7.32, 8.02, 8.06 and recombined ribozyme constructs containing the helper domain from the aforementioned isolates fused to intact B6.61 sequence (H7.08-B6.61, H7.32-B6.61, H8.02-B6.61 and H8.06-B6.61). The primer and template used were P12 and T21, respectively.
As shown in Figure 4-17, gel-purified transcribed RNA from isolates 7.08, 7.32, 8.02 and 8.06 have the same activity as those when the transcription reaction was added immediately to a primer-template extension reaction after heat inactivation of T7 RNA polymerase (section 4.3.3.3, Figure 4-12, Figure 4-15). Recombined ribozyme constructs all have dramatically improved activity relative to their parent constructs prior to recombination. In particular, H7.08-B6.61, H7.32-B6.61 and H8.02-B6.61 all have activities comparable to that of B6.61, while H8.06-B6.61 is roughly 20 fold slower. For the first three constructs mentioned, it appears that the presence of a 5′-helper domain had minimal effect to the polymerization ability of B6.61; however, it is worth noting that these three constructs show improved polymerization relative to the initial input pool for selection (Figure 4-10A, R0 lanes).

4.4 Discussion

The results of this \textit{in vitro} encapsulated selection were not as fruitful as one would have hoped. The first type of artifacts that dominated the selection were those that had insertions of the sequence of the “-4” biotinylated probe in the selection pool. These artifacts were inactive due to their ability to sequester the primer-template by hybridization. These may have arisen by several mechanisms, one being contamination of the PCR amplification step with unremoved biotinylated oligonucleotide and then incorporation into the selection pool by the action of protein enzymes (such as T4 DNA ligase) used during the \textit{in vitro} selection. Nevertheless, it is clear that these artifacts were selected for by the second hybridization-based selective step.
The second type of artifact, found in both the round 7 and 8 pools, had a large number of mutations and deletions in supposedly constant regions (the ligase core and accessory domain) of the selection pool. One possible explanation for this occurrence is the accumulation of amplification errors by Taq DNA polymerase in the multiple rounds of selection. The error rate of Taq DNA polymerase, that lacks a 3' to 5' exonuclease proofreading ability, was measured to be approximately 1 in 9000 nucleotides (Tindall & Kunkel, 1988). Considering that eight rounds of selection were performed, with each round of selection requiring 20 to 22 cycles of PCR to reamplify the DNA pool for the next round, the pool has “experienced” at least 160 cycles of PCR. In combination with a possibly contaminated initial \( ^44 \)tB6.61 double-stranded DNA PCR product with unmodified B6.61 sequence, this large number of cycles may have introduced the mutations observed in the ligase core and accessory domain of round 7 and round 8 isolates. Other possible mechanisms by which these artifacts may have arisen are through experimental blunders. Some possibilities include contaminated buffer used for PCR that may have affected the fidelity of Taq DNA polymerase (a mutation rate of 0.66% per position per PCR was proved to occur with altered concentrations of MgCl\(_2\), MnCl\(_2\) and dNTPs in a PCR reaction; Cadwell & Joyce, 1992) or extended UV light exposure to the selection pool, when using visualizing nucleic acid on a gel by UV shadow, that could possibility introduce mutations as well.

Because there were multiple isolates with the same mutations, it is likely that these mutations were specifically selected for during the selection. Interestingly, a great deal of these mutations corresponds to structurally important regions of the ligase core.
that were shown previously to be critical for phosphodiester bond formation activity (Ekland & Bartel, 1995). Confirmed by the minimal activity of such constructs as shown in Figure 4-12 and 4-15, there is no easy explanation of how they managed to survive the selection. Since these artifacts emerged despite the added APM gel purification step that should only select for thio-incorporated sequences as a result of ribozyme-catalyzed primer-template extension, a possible explanation for the survival of these mutagenized, inactive sequences is the acquirement of a thio-containing moiety from $^{45}$UTP by some other mechanism. One possibility is that some RNA sequences were able to use $^{45}$UTP to cap the ends of their genomes (similar to those described by Zaher and Unrau in 2006) (Zaher & Unrau, 2006). Another possibility is that upon breakage of the emulsion and running of the APM denaturing gel, some genomes were still hybridized to 4-thiouridine-containing RNA that was transcribed using encapsulated transcription thus resulting in their purification. These hypotheses could easily be tested in the future; with further optimization of the selection scheme to avoid the enrichment of the aforementioned artifacts, it may be possible to use in vitro encapsulated selections to isolate for a truly processive RNA replicase ribozyme in the near future.

4.5 Contributions

Peter Unrau and I planned the experiments and analyzed results. I wrote this chapter, generated the figures and performed all the experiments.
CHAPTER 5: Conclusions

5.1 Structural and Functional Studies of an RNA Polymerase Ribozyme

In the research presented in Chapter 3, we have characterized the structure and function of the B6.61 RNA polymerase ribozyme. From primer-template localization studies in which the primer-template substrate was localized at a high concentration to different specific regions of B6.61, we have indirectly confirmed that regions in the ligase core are close to the phosphodiester formation active site since localization of the primer-template in these regions brings about an increase in catalytic rate. This finding is consistent with the relative location of these regions and the active site in its published 3D model (Bergman et al., 2004). Next, we proved that the ligase core and accessory domain are modular domains that fold independently and are able to act in trans to elicit a polymerase function. We also found that because the two domains are modular, as long as domains in trans are reassembled via hybridization to restore an active relative orientation (of which there are several), activity is very similar to that of the wildtype unimolecular B6.61. Exploiting this convenience, mutagenesis studies were performed to refine the secondary structure and determine the minimal core motif of the poorly understood accessory domain, chemical probing and further mutagenesis were used to elucidate critical residues in the accessory domain, and UV crosslinking studies were employed to reveal close contacts between the ligase core and accessory domain. This data allowed us to collaborate with the Major laboratory at l’Université de Montréal to generate a preliminary three-dimensional model of the accessory domain.
5.2 Improvement of an RNA Polymerase Ribozyme by *In Vitro* Encapsulated Selection

In my research detailed in Chapter 4, *in vitro* encapsulated selection was used in an attempt to isolate for improved variants of the B6.61 RNA polymerase ribozyme. A highly diverse random “helper” pool was appended onto the 5′ end of constant 44tB6.61 ribozyme sequence and subjected to *in vitro* encapsulated selection in which the random DNA pool was appended a primer-template substrate, encapsulated into water-in-oil vesicles along with T7 RNA polymerase and NTPs and transcribed into an RNA phenotype correlating to the DNA gene with which it is compartmentalized. Active RNA phenotypes that can extend the primer-template tethered onto its own DNA genome are enriched by selection of the DNA gene itself by two selective steps. The first was a gel shift-based selective step and the second was a hybridization-based selective step for extended genomes as a result of ribozyme-dependent polymerization. Genes encoding for active RNAs were retained and were re-amplified and re-attached a primer-template substrate before entering the next selection round.

Despite that fact that the overall pool diversity seemed to be decreasing as inferred by partial T1 RNase digestion of the resulting selection pools, the analysis of isolates from the round 7 and round 8 pools revealed that there was no improvement of polymerization ability relative to B6.61. This may unfortunately indicate an evolutionary dead-end in the *in vitro* selection for RNA polymerase ribozymes. Either new selection strategies or new pools should be tested in the future to assess the possibility of new and viable evolutionary paths towards improved RNA polymerases or RNA replicases.
Appendix: First-Nucleotide Extension Rates for Ribozymes Used in Chapter 3

The first-nucleotide kinetic rates presented in this appendix are those for ribozyme constructs used in Chapter 3. The experimental conditions used to determine the first-nucleotide extension rate are described in section 3.2.2.

A.1 Parent, Hybridized and Truncated Ribozyme Constructs

<table>
<thead>
<tr>
<th>Ribozyme Constructs</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
<th>First-Order Rate (min⁻¹) (for P9:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6.61</td>
<td>0.23</td>
<td>0.003</td>
</tr>
<tr>
<td>tL.1:tA.1</td>
<td>0.21</td>
<td>0.003</td>
</tr>
<tr>
<td>tL.1:tD.1</td>
<td>0.19</td>
<td>0.002</td>
</tr>
</tbody>
</table>
### A.2 tL.1:(Mutation Constructs of tA.1) (P16:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tA.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tA.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C136G (tA.2)</td>
<td>Mutational disruption of one side of APX helix</td>
<td>0.22</td>
</tr>
<tr>
<td>G168C (tA.3)</td>
<td>Mutational disruption of other side of APX helix</td>
<td>0.07</td>
</tr>
<tr>
<td>C136G, G168C (tA.4)</td>
<td>Covariational mutation to restore integrity of APX helix</td>
<td>0.19</td>
</tr>
<tr>
<td>U131A</td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td>A151D</td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td>U155A</td>
<td></td>
<td>0.14</td>
</tr>
<tr>
<td>A156C</td>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td>G174A</td>
<td></td>
<td>0.23</td>
</tr>
</tbody>
</table>

### A.3 tL.1:(Mutation Constructs of tA.1) (P9:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tA.1</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tA.1</td>
<td></td>
</tr>
<tr>
<td>C136D</td>
<td>0.003</td>
</tr>
<tr>
<td>G139D</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>C141D</td>
<td>0.002</td>
</tr>
<tr>
<td>A143D</td>
<td>0</td>
</tr>
<tr>
<td>A165D</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>A165U</td>
<td>0.003</td>
</tr>
<tr>
<td>A165G</td>
<td>0.003</td>
</tr>
<tr>
<td>A165C</td>
<td>0.002</td>
</tr>
</tbody>
</table>
A.4  **tL.1:(Deletion Constructs of tA.1) (P16:T21)**

<table>
<thead>
<tr>
<th>Deletion Constructs of tA.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tA.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tA.5</td>
<td>Deletion of AP1 stem</td>
<td>0.11</td>
</tr>
<tr>
<td>tA.6</td>
<td>Deletion of three nucleotides in the AJ3/2 loop</td>
<td>0.16</td>
</tr>
<tr>
<td>tA.7</td>
<td>Full-length accessory domain with truncated 3’ end (ends at nucleotide A173)</td>
<td>0.12</td>
</tr>
<tr>
<td>tA.8</td>
<td>Full-length accessory domain with truncated 3’ end (ends at nucleotide G180)</td>
<td>0.13</td>
</tr>
<tr>
<td>tA.9</td>
<td>Full-length accessory domain up to G180, AP1 stem deleted</td>
<td>0.07</td>
</tr>
<tr>
<td>tA.10</td>
<td>Full-length accessory domain up to A173, AP1 stem deleted</td>
<td>0.01</td>
</tr>
<tr>
<td>tA.11</td>
<td>Deletion of AP2, AP4 and AL4</td>
<td>0.23</td>
</tr>
<tr>
<td>tD.1</td>
<td>Accessory domain with the following deleted: AP1, AL1, AP2, AP4, AL4</td>
<td>0.19</td>
</tr>
<tr>
<td>tD.2</td>
<td>Deletion of A171, U172, A173, C174, C175</td>
<td>0.11</td>
</tr>
<tr>
<td>tD.3</td>
<td>A171~ C175 mutated to U171, U172. APX has 7 bp</td>
<td>0.16</td>
</tr>
</tbody>
</table>
A.5  tL.1:(AP3 and APX Mutation Constructs of tD.1) (P16:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tD.1</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>tD.14</td>
<td>G134C, C136G double mutant on one side of new APX stem</td>
<td>0</td>
</tr>
<tr>
<td>tD.16</td>
<td>G168C, C170G double mutant on other side of new APX stem</td>
<td>0.01</td>
</tr>
<tr>
<td>tD.15</td>
<td>G134C, C136G, C168C, C170G) to restore new APX helix</td>
<td>0.12</td>
</tr>
<tr>
<td>tD.18</td>
<td>G153C mutation at AP3</td>
<td>0.01</td>
</tr>
<tr>
<td>tD.19</td>
<td>C159G mutation at AP3</td>
<td>0.12</td>
</tr>
<tr>
<td>tD.20</td>
<td>G153C, C159C mutations at AP3</td>
<td>0.15</td>
</tr>
<tr>
<td>DB</td>
<td>C163A at AP3</td>
<td>0.04</td>
</tr>
<tr>
<td>DV</td>
<td>AL3 loop replaced with AUAC</td>
<td>0</td>
</tr>
<tr>
<td>DH</td>
<td>Delete A171 through C175 and AL3 loop replaced with GUGA</td>
<td>0.01</td>
</tr>
<tr>
<td>DI</td>
<td>Delete A171 through C175 and AL3 loop replaced with UACG</td>
<td>0.01</td>
</tr>
<tr>
<td>DK</td>
<td>Add a string of 6 adenosines between RCTag and DT accessory domain</td>
<td>0.21</td>
</tr>
<tr>
<td>DD8</td>
<td>A171U of DK, increase APX stem length to 6 bp</td>
<td>0.15</td>
</tr>
</tbody>
</table>
A.6  tL.1:(A-rich Loop Mutation Constructs of tD.1) (P16:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tD.1</td>
<td>Two adenosines inserted between U138 and G139</td>
<td>0.19</td>
</tr>
<tr>
<td>DD14</td>
<td>Two adenosines inserted between G139 and A140</td>
<td>0.16</td>
</tr>
<tr>
<td>DD13</td>
<td>Two adenosines inserted between G139 and A140</td>
<td>0.10</td>
</tr>
<tr>
<td>DD9</td>
<td>G139C</td>
<td>0.01</td>
</tr>
<tr>
<td>DD10</td>
<td>G139 deleted</td>
<td>0.01</td>
</tr>
<tr>
<td>DD11</td>
<td>C141 deleted</td>
<td>0.02</td>
</tr>
<tr>
<td>DD1</td>
<td>C141 and G142 deleted</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>DQ</td>
<td>G142 and A143 deleted</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>DD12</td>
<td>A143 deleted</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>DD2</td>
<td>Four adenosines inserted between G142 and A143</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>DP</td>
<td>A143 and A144 deleted</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>DZ</td>
<td>Deletion of A140, C141, G142, A143, A144</td>
<td>0</td>
</tr>
<tr>
<td>tD.1</td>
<td></td>
<td>0.19</td>
</tr>
</tbody>
</table>
### A.7 tL.1: (A-rich Loop Mutation Constructs of tD.1) (P9:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tD.1</td>
<td></td>
<td>0.0021</td>
</tr>
<tr>
<td>DD18</td>
<td>Deletion of A140</td>
<td>0</td>
</tr>
<tr>
<td>DD19</td>
<td>Deletion of G142</td>
<td>0</td>
</tr>
<tr>
<td>DD20</td>
<td>C141U</td>
<td>&lt;0.0010</td>
</tr>
<tr>
<td>DD21</td>
<td>A143U</td>
<td>0</td>
</tr>
<tr>
<td>DD22</td>
<td>A151U</td>
<td>0</td>
</tr>
</tbody>
</table>

### A.8 tL.1: (A151 and A161 Mutation Constructs of tD.1) (P16:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tD.1</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>DJ</td>
<td>Insert U between A160 and C161 to force base-pair to A151</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
### A.9  tL.1:(Crosslink Sites Mutation Constructs of tD.1) (P9:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min⁻¹) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>**tD.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD23</td>
<td>U154C</td>
<td>0.0021</td>
</tr>
<tr>
<td>DD24</td>
<td>U155C</td>
<td>0.0016</td>
</tr>
<tr>
<td>DD25</td>
<td>U154C, U155C</td>
<td>&lt;0.0010</td>
</tr>
<tr>
<td>DD26</td>
<td>U154C, A165U</td>
<td>0.0016</td>
</tr>
<tr>
<td>DD27</td>
<td>U155C, A165U</td>
<td>&lt;0.0010</td>
</tr>
<tr>
<td>DD28</td>
<td>U154C, U155C, A165U</td>
<td>0</td>
</tr>
</tbody>
</table>
### A.10 Unimolecular Ribozyme Constructs (P9:T21)

<table>
<thead>
<tr>
<th>Mutation Constructs of tD.1</th>
<th>Description</th>
<th>First-Order Rate (min(^{-1})) (for P16:T21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6.61</td>
<td></td>
<td>0.0056</td>
</tr>
<tr>
<td>tL.1:tD.1</td>
<td></td>
<td>0.0021</td>
</tr>
<tr>
<td>L.1-(A)(_2)-D.1</td>
<td>3 adenosines between domains</td>
<td>0</td>
</tr>
<tr>
<td>L.1-(A)(_4)-D.1</td>
<td>4 adenosines between domains</td>
<td>&lt;0.0010</td>
</tr>
<tr>
<td>L.1-(A)(_5)-D.1</td>
<td>5 adenosines between domains</td>
<td>0.0018</td>
</tr>
<tr>
<td>L.1-(A)(_6)-D.1</td>
<td>6 adenosines between domains</td>
<td>0.0019</td>
</tr>
<tr>
<td>L.1-(A)(_7)-D.1</td>
<td>7 adenosines between domains</td>
<td>0.0018</td>
</tr>
<tr>
<td>L.1-(A)(_9)-D.1</td>
<td>9 adenosines between domains</td>
<td>0.0017</td>
</tr>
<tr>
<td>L.1-(A)(_{11})-D.1</td>
<td>11 adenosines between domains</td>
<td>&lt;0.0010</td>
</tr>
<tr>
<td>L.1-(A)(_{17})-D.1</td>
<td>17 adenosines between domains</td>
<td>&lt;0.0010</td>
</tr>
</tbody>
</table>
List of References


B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R.,
Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson,
"Genome sequencing in microfabricated high-density picolitre reactors." Nature
437 (7057): 376-80.


McGinness, K. E., Wright, M. C. and Joyce, G. F. (2002). "Continuous in vitro evolution of
a ribozyme that catalyzes three successive nucleotidyl addition reactions." Chem

Miller, O. J., Bernath, K., Agresti, J. J., Amitai, G., Kelly, B. T., Mastrobattista, E., Taly, V.,


among RNA-dependent RNA polymerases." Curr Top Microbiol Immunol 320:
137-56.


