A COMPREHENSIVE STUDY OF RIBOZYME-MEDIATED NUCLEOTIDE AND SUGAR CHEMISTRIES

by

Matthew Lau
Bachelor's of Science (Honours), Simon Fraser University, 2003

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

DOCTOR OF PHILOSOPHY

In the Department of Molecular Biology and Biochemistry

© Matthew Lau, 2009

SIMON FRASER UNIVERSITY

Spring 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.
<table>
<thead>
<tr>
<th><strong>APPROVAL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name:</strong> Matthew Lau</td>
</tr>
<tr>
<td><strong>Degree:</strong> Doctor of Philosophy</td>
</tr>
<tr>
<td><strong>Title of Thesis:</strong> A Comprehensive Study of Ribozyme-Mediated Nucleotide and Sugar Chemistries</td>
</tr>
<tr>
<td><strong>Examining Committee:</strong></td>
</tr>
<tr>
<td><strong>Chair:</strong> Dr. Neil Brenda</td>
</tr>
<tr>
<td>Professor of Chemistry</td>
</tr>
<tr>
<td>Dr. Peter Unrau</td>
</tr>
<tr>
<td>Senior Supervisor</td>
</tr>
<tr>
<td>Associate Professor of Molecular Biology and Biochemistry</td>
</tr>
<tr>
<td>Dr. Dipankar Sen</td>
</tr>
<tr>
<td>Committee Member</td>
</tr>
<tr>
<td>Professor of Molecular Biology and Biochemistry</td>
</tr>
<tr>
<td>Dr. Christopher Beh</td>
</tr>
<tr>
<td>Committee Member</td>
</tr>
<tr>
<td>Associate Professor of Molecular Biology and Biochemistry</td>
</tr>
<tr>
<td>Dr. Andrew Bennet</td>
</tr>
<tr>
<td>Internal Examiner</td>
</tr>
<tr>
<td>Professor of Chemistry</td>
</tr>
<tr>
<td>Dr. Thorsten Dieckmann</td>
</tr>
<tr>
<td>External Examiner</td>
</tr>
<tr>
<td>Professor of Chemistry</td>
</tr>
<tr>
<td>University of Waterloo</td>
</tr>
<tr>
<td><strong>Date Defended/Approved:</strong> March 26, 2009</td>
</tr>
</tbody>
</table>
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 <www.lib.sfu.ca> at: <http://ir.lib.sfu.ca/handle/1892/112>) 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

The ability of RNA catalysts (ribozymes) to synthesize nucleotides as basic building blocks for the assembly of larger RNA polymers is an important aspect of the 'RNA World' hypothesis. To examine the ability of RNAs to catalyze this chemistry, we performed an *in vitro* selection in my first project and successfully isolated ribozymes capable of synthesizing a purine nucleotide (6-thioguanosine, $^6S\text{G}$) from tethered 5-phosphoribosyl 1-pyrophosphate (PRPP) and 6-thioguanine ($^6S\text{Gua}$). Compared to the previously selected pyrimidine synthase ribozymes, these ribozymes are 50-100 times more efficient. In a continuation of this work, we next deduced the secondary structure of two purine synthase ribozymes by performing a re-selection on two truncated pools for reactivity with $^6S\text{Gua}$. Interestingly we were able to isolate for functional sequences that were ~2.5 fold shorter than their full-length parental ribozymes, and with much simpler secondary structure relative to the pyrimidine nucleotide synthase ribozymes. The ability of RNAs to synthesize both purine and pyrimidine nucleotides strongly suggests that RNA could have preceded proteins in a hypothetical RNA based metabolism.

While the ribozymes isolated from our selection were successful in catalyzing purine nucleotide synthesis, they do require a pre-activated sugar substrate. This is not highly plausible in the RNA world since PRPP readily undergoes hydrolysis. To examine this phenomenon, in our third project we
implemented a selection using a RNA pool tethered to ribose-5-phosphate and selected them for reactivity with $^6$S Gua. After six selection rounds, a single RNA sequence dominated the selection. Interestingly, this ribozyme can produce three different products: two when tethered with PR and a third product when tethered with PRPP, which was determined to be $^6$S G. Our results suggest that this highly versatile ribozyme is capable of recognizing two similar substrates for the synthesis of different products, and thus demonstrates the promiscuous potential of ribozymes upon encountering of an alternative substrate.
To my dearest parents and Emily
ACKNOWLEDGEMENTS

I would like to express the deepest gratitude to my senior supervisor Dr. Peter Unrau for his patience, guidance, and for giving me the opportunity to work in his laboratory. He has always been supportive throughout my graduate career. I am grateful to my committee members, Dr. Dipankar Sen and Dr. Christopher Beh, for the time they have taken to offer me helpful advice in my work, as well as my internal examiner, Dr. Andrew Bennet, for providing me with useful suggestions related to various aspects of sugar chemistries.

I would also like to thank my external examiner, Dr. Thorsten Dieckmann, for taking the time to travel from Waterloo to attend my thesis defence.

I was very fortunate to have worked with three talented undergraduate students, Elaine Chiu, Danny Shum, and Nelly Said, and I would like to express my appreciation for their help in my projects. I would also like to thank Kelly Cadieux, Sunny Wang, as well as all the members of the Unrau and Sen laboratories, past and present, for their advices in my experimental work.

And lastly, to my parents, Loretta, Emily, Yuki and Sparky. Thank you for your continuous support and encouragement through the hard times in my graduate career, for I do not think I could have made it through without you.

This work has been funded from the Natural Sciences and Engineering Research Council of Canada and Simon Fraser University.
# TABLE OF CONTENTS

**Approval** ........................................................................................................................ ii

**Abstract** .......................................................................................................................... iii

**Dedication** ....................................................................................................................... v

**Acknowledgements** ....................................................................................................... vi

**Table of Contents** ......................................................................................................... vii

**List of Figures** .............................................................................................................. xi

**List of Tables** ................................................................................................................. xiii

**Abbreviations** ................................................................................................................ xiv

**Chapter 1: Overall Introduction** ..................................................................................... 1

1.1 The Prebiotic World and Early Life ................................................................................. 6

1.1.1 Template Mediated Replication ............................................................................. 8

1.1.2 Requirement for Compartmentalization ................................................................. 12

1.2 Nucleotide Synthesis .................................................................................................... 16

1.2.1 Abiotic Synthesis of Nucleobases ......................................................................... 17

1.2.1.1 Purine Bases .................................................................................................. 17

1.2.1.2 Pyrimidine Bases ......................................................................................... 18

1.2.2 Abiotic Synthesis of Ribose .................................................................................. 22

1.2.2.1 The Formose Reaction .................................................................................. 22

1.2.2.2 Alternatives to Ribose and Ribonucleotides ................................................... 26

1.3 Characteristics and Significance of Ribose Sugars ................................................. 30

1.3.1 The Instability of Ribose ....................................................................................... 30

1.3.2 Reactivity of Ribose and Other Sugars ................................................................. 33

1.3.3 Sugar Metabolism in Cells ..................................................................................... 35

1.3.3.1 Glycolysis Pathway ....................................................................................... 35

1.3.3.2 Pentose Phosphate Pathway .......................................................................... 37

1.3.4 Regulation of Cellular Carbohydrate Levels by RNAs ........................................ 38

1.3.4.1 Regulatory sRNAs .......................................................................................... 38

1.3.4.2 Regulation by a Ribozyme .............................................................................. 39

1.4 *In vitro* Ribozyme Selection and Evolution ............................................................... 40

1.4.1 Construction of Sequence Pools for Selection ....................................................... 43

1.4.2 The Selection Step ................................................................................................. 46

1.4.2.1 Aptamer Selection Strategies ........................................................................... 46

1.4.2.2 Ribozyme Selection Strategies ........................................................................ 47

1.4.3 RNA Evolution by *In Vitro* Selection ................................................................. 49

1.4.3.1 Evolution from a Biased Pool ......................................................................... 49
1.4.3.2 Evolution by Engineering of Functional RNAs ........................................ 51
1.4.3.3 Evolution by Random Mutagenesis ................................................. 51
1.4.3.4 Evolution of Natural Ribozymes ....................................................... 52

Chapter 2: In vitro Selection of Fast Purine Nucleotide Synthase Ribozymes ................................................................. 56

2.1 Introduction ............................................................................................ 56
2.2 Methods .................................................................................................. 58
  2.2.1 APPRPP and APP<sup>6S</sup>G Synthesis .................................................. 58
  2.2.2 Designing of Selection Pools ............................................................... 59
  2.2.3 RNA Ligation Protocol ...................................................................... 60
  2.2.4 Selection ........................................................................................... 61
  2.2.5 TLC Analysis .................................................................................... 62
  2.2.6 Kinetic Analysis ............................................................................... 63
  2.2.7 Analogues ......................................................................................... 63
2.3 Results .................................................................................................... 64
  2.3.1 Selection ........................................................................................... 64
  2.3.2 Sequencing Analysis ........................................................................ 68
  2.3.3 Magnesium Dependence .................................................................. 70
  2.3.4 Product Characterization ................................................................. 71
  2.3.5 Kinetics ............................................................................................ 73
  2.3.6 Substrate Specificity ......................................................................... 75
2.4 Discussion ................................................................................................ 77
2.5 Contributions .......................................................................................... 82

Chapter 3: Secondary Structure Characterization of Two Purine Nucleotide Synthase Ribozymes .................................................. 83

3.1 Introduction ............................................................................................ 83
3.2 Materials and Methods .......................................................................... 85
  3.2.1 Synthesis of Truncated Pools by Non-Homologous Recombination .......... 85
  3.2.2 Selection ........................................................................................... 86
  3.2.3 Substrate Analogues ......................................................................... 88
  3.2.4 UCUUU Randomization and Re-selection ......................................... 88
  3.2.5 Site Directed Mutagenesis ................................................................. 89
  3.2.6 Kinetics ............................................................................................ 89
3.3 Results .................................................................................................... 90
  3.3.1 Pool Construction by Non-Homologous Recombination ..................... 90
  3.3.2 Selection ........................................................................................... 93
  3.3.3 Sequencing Analysis ........................................................................ 95
    3.3.3.1 Alignment of Round 6 Recombined MA Pool Sequences ............. 95
    3.3.3.2 Alignment of Round 6 Recombined MF Pool Sequences ............. 96
  3.3.4 Site Directed Mutagenesis ................................................................. 99
    3.3.4.1 Mutagenesis of the Ribozyme MAa ........................................... 102
    3.3.4.2 Mutagenesis of the Ribozyme MFa and tMFa-50 ....................... 104
  3.3.5 Replacement of Loop Sequences in Predicted Stem-Loop Structures with the Nucleotide Sequence GNRA ................. 105
3.3.6 {G}CUUU Motif is Critical for Functionality of MFa, but not MAa...... 106
3.3.7 Kinetic Studies .................................................................................. 109
3.3.8 Magnesium Dependence ................................................................ 111
3.3.9 Substrate Specificity ......................................................................... 111
3.4 Discussion ............................................................................................... 113
3.4.1 Structural Comparison Between Purine and Pyrimidine Nucleotide Synthase Ribozymes ................................................ 113
3.4.2 Sequence Specificity of Truncated MA and MF Isolates ................... 114
3.4.3 Importance of the {G}CUUU Motif ..................................................... 116
3.4.4 Different Strategies for Ribozyme-Mediated Purine Nucleotide Synthesis .................................................................................... 117
3.4.5 Non-Homologous Random Recombination as a Tool for Secondary Structural Analysis and RNA Evolution ...................... 118
3.5 Contributions ........................................................................................... 120
Chapter 4: A Promiscuous Ribozyme Mediating a Sugar Aldehyde Reaction and an Unexpected Nucleotide Synthesis Chemistry .......... 121
4.1 Introduction .............................................................................................. 121
4.2 Materials and Methods ............................................................................ 124
4.2.1 Design of the RNA Pool .................................................................... 124
4.2.2 Substrate Ligation and Product Marker Synthesis ............................ 124
4.2.3 Selection ........................................................................................... 125
4.2.4 Precise Cleavage of the 3' End of pR1 ............................................. 126
4.2.5 Kinetic Analysis of pR1 ..................................................................... 127
4.2.6 Terminal Enzymatic Extension Assays ............................................. 127
4.2.7 TLC Analysis of PRPP Product ......................................................... 128
4.2.8 Purine Analogue.................................................................................. 129
4.2.9 Re-equilibration of the Two PR Products .......................................... 129
4.2.10 Reduction of Ribose-5-phosphate................................................... 130
4.2.11 Processing and Derivatization of pR1 with PdR ............................... 130
4.2.11.1 Synthesis of Shortened pR1 .................................................... 130
4.2.11.2 Synthesis of *pCAAGC-PdR ................................................... 130
4.2.11.3 Generation of Full Length pR1-PdR ......................................... 131
4.3 Results .................................................................................................... 132
4.3.1 Selection of the pR1 Sequence ......................................................... 132
4.3.2 The Synthesis of Two Thiol-containing Products Requires PR, but Not Triphosphates ......................................................... 135
4.3.3 Two Observed PR Products are Not Caused by Sequence Heterogeneity or Alternative Folding ....................................................... 135
4.3.4 pR1 Utilizes PRPP in Addition to PR and Specifically Utilizes $^{65}$Gua ................................................................. 138
4.3.5 Only the PRPP Product is a Conventional Nucleotide .......................... 139
4.3.6 Characterization of the PRPP Product .............................................. 140
4.3.7 Characterization of the PR Products .................................................. 142
4.3.8 Reduction of Ribose Prevents Product Formation ............................... 144
4.3.9 Deoxyribose is an Alternative Sugar Substrate for pR1 .................... 146
4.3.10 Kinetics ........................................................................................... 150
4.4 Discussion .................................................................................................................. 152
  4.4.1 Catalysis of Two Sequential Distinct Chemical Steps is Difficult for In Vitro Selected Ribozymes ................................................................. 152
  4.4.2 Ribozyme Promiscuity .................................................................................... 153
  4.4.3 Promiscuity of the pR1 Ribozyme ................................................................. 154
  4.4.4 Likely Mechanism of pR1-PR, pR1-PdR and pR1-PRPP Chemistries ................. 156
  4.4.5 Evolutionary Significance of Ribozyme Small Molecule Substrate Promiscuity ................................................................. 157
  4.5 Contributions ....................................................................................................... 161

Chapter 5: Concluding Remarks and Significance of Work ................................. 164
  5.1 Discovery of Ribozymes Mediating Efficient Purine Nucleotide Synthesis by In Vitro Selection ................................................................. 164
  5.2 Characterizing Purine Nucleotide Synthase Ribozymes by Non-Homologous Random Recombination ................................................................. 165
  5.3 Isolation of a Promiscuous Ribozyme Catalyzing Both Purine Nucleotide Synthesis and an Unusual Sugar Chemistry ................................. 167

Bibliography ............................................................................................................... 169
LIST OF FIGURES

Figure 1-1: Synthesis of organic molecules in hypothetical early world conditions. .................................................................................................................... 7
Figure 1-2: Predicted secondary structure of the B6.61 polymerase ribozyme. ................................................................................................................................. 11
Figure 1-3: C10 decanoic fatty acid derivatives used for vesicle construction. ................................................................................................................................. 15
Figure 1-4: Proposed prebiotic pathway for the synthesis of different purine bases. ............................................................................................................. 19
Figure 1-5: Proposed prebiotic pathway for the synthesis of pyrimidine bases ...................................................................................................................... 22
Figure 1-6: Synthesis of ribose by the autocatalytic formose reaction pathway. ......................................................................................................................... 24
Figure 1-7: Examples of possible sugar systems for synthesis of nucleic acid polymers. ................................................................................................. 27
Figure 1-8: Examples of potential backbone structures of early nucleic acid polymers. ................................................................................................. 29
Figure 1-9: Stabilization of ribose by phosphate esters or cyanide. ................................................................................................................................. 31
Figure 1-10: Stabilization of ribose by borate. ............................................................................................................................... 33
Figure 1-11: The Maillard reaction. ........................................................................ 34
Figure 1-12: The 10-step glycolysis pathway. ....................................................... 36
Figure 1-13: General scheme for typical in vitro selections. ................................ 42
Figure 1-14: Catalytic strategies employed by two different natural ribozymes. ................................................................................................. 53
Figure 2-1: Reactions catalyzed by pyrimidine and purine nucleotide synthases. ............................................................................................................. 57
Figure 2-2: Template design of the mutagenic pool. ............................................ 65
Figure 2-3: Overall in vitro selection scheme for purine nucleotide synthase ribozymes. ................................................................................................. 66
Figure 2-4: Purine ribozyme activity as a function of selection round. ............. 67
Figure 2-5: Alignments of round 10 isolates from biased and random sequence pools. .............................................................. 69
Figure 2-6: Overall in vitro selection scheme for purine nucleotide synthase ribozymes. .............................................................. 72
Figure 2-7: Catalytic rate ($k_{obs}$) for MA and RA as a function of $6S$Gua concentration. .............................................................. 74
Figure 2-8: Ribozyme incubation with 6-thioguanine or 6-thiopurine.................. 76
Figure 3-1: Synthesis of recombined pools by Non-Homologous Random Recombination............................................................ 91
Figure 3-2: In vitro selection scheme for truncated $6S$G synthase ribozymes. ............................................................................. 92
Figure 3-3: Reactivity of round 3 recombined MA and MF pool sequences with $6S$Gua. ................................................................. 94
Figure 3-4: Length distribution of recombined pools through 6 selection rounds. ......................................................................... 97
Figure 3-5: Sequencing analysis of round 6 recombined MA and MF pools....... 98
Figure 3-6: Secondary structure predictions of ribozymes MAa, M Fa and the 50-nt truncated version of MFa (tMFa-50). ...................... 101
Figure 3-7: Kinetic Analysis of MF and 4 other truncated isolates. ............... 110
Figure 3-8: Substrate specificity of MF, MAa and MFa................................. 112
Figure 4-1: Ribozyme-mediated nucleotide synthesis pathways studied. ....... 122
Figure 4-2: In vitro selection scheme for isolation of ribozymes that utilize PR and $6S$Gua................................................................. 133
Figure 4-3: The substrate dependent reactions of the pR1 ribozyme. ............ 136
Figure 4-4: The pR1-PRPP, but not the pR1-PR, reaction products are enzymatically recognized and co-migrated with a $6S$GMP standard in two dimensional TLC analysis. ............................................. 141
Figure 4-5: Interconversion of the two pR1-PR dependent products. ............ 143
Figure 4-6: Reduction of ribose prevents the synthesis of the PR products. .... 146
Figure 4-7: pR1 reaction with $6S$Gua using PR or PdR as tethered substrates. .................................................................................... 149
Figure 4-8: Magnesium dependence of pR1 ribozyme............................... 151
Figure 4-9: Reduction of the lower PR band product...................................... 157
Figure 4-10: Proposed reaction mechanisms for the synthesis of different pR1 products. .............................................................. 160
LIST OF TABLES

Table 3-1: Length of sequences and reaction time with $^6$S$_{Gua}$ for both recombined MA and MF pools in each selection round. ........................................ 96

Table 3-2: Mutants synthesized for structural analysis of MAa, MFa and 50-nt truncated version of MFa (tMFa-50). ................................................................. 100
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^6\text{S}\text{Gua}$</td>
<td>6-thioguanine</td>
</tr>
<tr>
<td>$^4\text{S}\text{Ura}$</td>
<td>4-thiouracil</td>
</tr>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>A, G, C, U</td>
<td>Adenosine, Guanosine, Cytidine, Uridine</td>
</tr>
<tr>
<td>APP$^{6\text{S}G}$</td>
<td>Adenylated 6-thioguanosine 5'-monophosphate</td>
</tr>
<tr>
<td>APPRPP</td>
<td>Adenylated 5-phosphoribosyl 1-pyrophosphate</td>
</tr>
<tr>
<td>APM</td>
<td>N-acryloylaminophenyl-mercuric acetate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dG, dC</td>
<td>Deoxyguanosine, Deoxycytidine</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide 5'-triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>{G}</td>
<td>Any nucleotide but G</td>
</tr>
<tr>
<td>GMP</td>
<td>Guanosine 5'-monophosphate</td>
</tr>
<tr>
<td>HGPRTase</td>
<td>Hypoxanthine guanine phosphoribosyltransferase</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>Imp A</td>
<td>Adenosine 5'-phosphorimidazolide</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>N</td>
<td>Any arbitrary nucleotide</td>
</tr>
</tbody>
</table>
nt  Nucleotide
P  Phosphate group
*P  $^{32}$P radiolabelled phosphate group
pCp  cytidine 5', 3' bisphosphate
PdR  Deoxyribose-5-phosphate
PAGE  Polyacrylamide gel electrophoresis
PCR  Polymerase chain reaction
PdR  Deoxyribose-5-phosphate
PR  Ribose-5-phosphate
PRPP  5-phosphoribosyl 1-pyrophosphate
R  Any purine nucleotide
RNA  Ribonucleic acid
RT  Reverse transcription
TLC  Thin layer chromatography
tRNA  Transfer RNA
u  Unified atomic mass unit
wt  Wild type
CHAPTER 1: OVERALL INTRODUCTION

Organisms, from simple prokaryotes to more complex eukaryotes and multi-cellular organisms, have changed and developed because of continuous evolution. Evolution includes adaptation to continuous environmental changes, nutrient availability and natural selection of the fittest, all of which were made possible by the inheritance of genetic information. All organisms could have evolved from the hypothetical progenitor organism known as life’s last universal common ancestor (LUCA), which is believed to have lived from ~3.5 to 3.8 billion years ago (Doolittle 2000). But how did “life” actually start in the first place? This question has been argued in continuous debates over the past century in the scientific field studying the origin of life, and is still highly controversial today. In fact, there is no simple definition as to the word “life”. As we look around ourselves today, all known living organisms seemingly share various common characteristics, such as the ability to grow, reproduce, evolve, maintain homeostasis, adapt to the environment and consume energy. However, do these characteristics apply to all life forms? Viruses and various parasitic bacteria, for example, cannot survive independently as they require a host to supply them with essential nutrients or proteins necessary for growth and reproduction (Wagenlehner, Naber et al. 2006). So are they considered as “living”? To further complicate matters, there are various entities that we would commonly consider “non-living”, that also demonstrate some of the characteristics described above. Fire is a classic example that feeds, grows, adapts to environment and
reproduces. Can fire also be classified as “living”? To define living from non-living matter, physicist Erwin Schrödinger in 1944 came up with the definition that all living matter has the ability to resist entropy, or maintain a state of order, whereas non-living matter cannot (Schrodinger 1944). Humans, for example, live and resist entropy by taking in nutrients and converting them to energy in a process known as metabolism. Once we die, metabolism stops, our body breaks down, and over time all of the atoms within our body are spread out and subjected to the entropy of the environment. Other definitions of life have also been reported, such as Gerald Joyce’s in defining life as a “self-sustaining system capable of Darwinian evolution” (Deamer, Fleischaker et al. 1994), and the more specific definition by astrobiologist Benton Clark, stating that “life reproduces, life uses energy and that these functions follow a set of instructions embedded within the organism” (Mullen 2002). The instructions described in Clark’s definition would be the genetic code, which today is encoded in living organisms by DNA.

DNA consists of four different nucleotides and it is through this nucleotide code that genetic information is stored and inherited. DNA has the ability to form numerous intramolecular and intermolecular non-covalent interactions, as well as copying the information encrypted in the DNA strands using various protein enzymes. However, does this mean that life initiated with DNA? Are there other examples of molecules that can carry genetic information? In modern cell-based biology, cells must be able to process the information encoded within its DNA to direct which proteins are needed to be synthesized in order to maintain
homeostasis. This process initiates from the transcription of DNA to generate a complementary mRNA strand, which is exported to the cytoplasm and translated into proteins by the ribosomal machinery. Since the transcribed RNAs contain all of the information from the DNA template in which it was copied from, RNAs in principle can also carry genetic information.

For decades, RNA was thought to act solely as a messenger between DNA and protein synthesis. It was not until 1965 that Holley and colleagues first reported the secondary structure of alanine tRNA from yeast, showing that RNA can also fold into complex secondary structures like DNA and proteins (Holley, Apgar et al. 1965). These results led scientists such as Francis Crick and Leslie Orgel to the hypothesis that life may have started with RNA molecules in which RNA can serve both as a genetic carrier and a catalyst, roles that are currently performed by DNA and proteins, respectively (Crick 1968). Supportive of this idea was the discovery of catalytic RNAs, or ribozymes, in independent work by Thomas Cech, who showed that the group I intron from *Tetrahymena thermophila* can catalyze self-splicing (Cech, Zaug et al. 1981), and Sidney Altman, who showed that the RNA component of RNAsse P can catalyze RNA cleavage in tRNA maturation (Guerrier-Takada, Gardiner et al. 1983). Further evidence in support of the RNA World hypothesis came later in 2000, when crystal structure studies showed that the active site of the ribosome lies deep within the RNA core, and no amino acid side chain comes within 18 Å of the active site (Nissen, Hansen et al. 2000). Further structural studies by Steitz and Moore in 2003 showed that peptide bond synthesis can be catalyzed using solely
the RNA components (Steitz and Moore 2003). Since then, the catalytic component of the ribosome has been recognized as being essentially a ribozyme that can catalyze protein synthesis. Given these results together, it seems feasible that RNAs would have predated proteins and that catalytic RNAs would play a fundamental role in the early world.

In the hypothesized RNA World, a riboorganism must be able to cope with many different challenges, such as the availability of resources and changing environmental conditions, in order to sustain life. At the minimum, it must evolve strategies for generating ribonucleotides, which are the basic building blocks for RNAs, to allow replication of the genetic information and synthesis of ribozymes to catalyze important early chemistries. Failure to do so will halt further propagation and will ultimately lead to the extinction of the organism. While recruitment of ribonucleotides from the prebiotic environment to meet the needs of a riboorganism may be a possible initial approach, over time, however, even an environment with a large reservoir of nucleotides would eventually become depleted. The ultimate solution, therefore, would be to evolve an RNA-dependent strategy to synthesize ribonucleotides such that there is a continuous supply of these important molecules.

The overall structure of a ribonucleotide comprises of three primary components, a base, a phosphate group and the ribose sugar. Currently, however, very little is known about how ribozymes can mediate the synthesis of a nucleoside from a base and a ribose sugar (Gesteland, Cech et al. 2006). The study of ribose chemistries is of particular interest to the origin of life because it is
the primary component contributing to the structure, catalytic properties and unique characteristics observed in RNAs. The importance of ribose, however, is not only limited to nucleotide synthesis. Various different ubiquitous co-enzymes, such as nicotinamide adenine dinucleotide, flavin adenine dinucleotide and coenzyme A, are believed to be remnants from the RNA world as suggested by the presence of ribose, ribose derivatives, or ribonucleotides in their structure (Benner, Ellington et al. 1989). The highly reactive ribose sugar is also involved in different chemistries that we observe in living systems today. In cells, ribose is a fundamental metabolite involved in different cellular pathways, most notably the pentose-phosphate pathway, which will be described in further detail in section 3, to drive nucleotide synthesis (Mathews 1996). Ribose, along with other sugars, can also be harvested for metabolic energy, thus serving as a potential food source for cells (Mathews 1996). It is therefore not surprising that ribose has been conserved from the hypothesized RNA world to modern cellular metabolism, and that nature has evolved different strategies to exploit this important compound.

In my thesis research, I have conducted a detailed study of ribozymes capable of mediating the small molecule chemistries involved in ribonucleotide synthesis. This included exploring how RNAs can manipulate and use different ribose sugar substrates to catalyze different chemistries with the bases. To gain a further understanding of the catalytic strategies employed by these ribozymes, including their substrate specificity and ability to achieve rate enhancement, we
have also delineated their secondary structures through various recombination and mutagenesis experiments.

1.1 The Prebiotic World and Early Life

Much effort has been invested over the last decade in understanding the conditions of the world when life originated. The earth is slightly more than 4.5 billion years old, but it was not until nearly one billion years had elapsed that life emerged (Joyce 2002). This timeline is supported by chemical evidence from carbon isotopes in Archean rocks, which suggests that life existed on Earth as early as 3.5 billion years ago (Altermann and Kazmierczak 2003). Researchers in the origin of life field have debated over the years as to what the environment was like during this time span, but most would agree that water and gases such as methane, ammonia, carbon monoxide, carbon dioxide and hydrogen sulphide were present (Pace 2001). These gases, under reducing conditions such as those found under the ocean in deep-sea vents, may have given rise to various prebiotic molecules such as formate, formaldehyde and methanol (Martin, Baross et al. 2008). Alternatively, as suggested by some researchers, prebiotic molecules may have been generated elsewhere and brought to earth by meteorites and comets (Bockelee-Morvan, Colom et al. 1991). This idea was strengthened by the discovery of organic compounds such as methanol, hydrogen cyanide, formaldehyde and ethanol in comets, along with some of the gases (i.e. methane and carbon dioxide) postulated to be present on the early Earth (Pizzarello 2004).

In 1952, Stanley Miller and Harold Urey performed an interesting
Figure 1-1: Synthesis of organic molecules in hypothetical early world conditions.

This system contains four main components, a water flask, a gas chamber connected to an electrode, a condenser and a trap, all of which are connected and sealed with glass tubes. The water in the water flask (bottom right) is first heated to water vapour to mimic evaporation from the ocean, which subsequently mixes with gases predicted to be present in the early atmosphere. Sparks were continuously fired in the chamber to simulate lightning, followed by cooling as the water mixture condenses and trickles back into the ocean where the cycle repeats. Between the condenser and the flask is a trap in which newly synthesized molecules can be collected and sampled (Miller and Urey 1959).

experiment in which they simulated early Earth conditions in by mixing water vapour (evaporation from the ocean) with methane, ammonia and hydrogen (gases believed to be present in the early atmosphere) inside a glass chamber (Figure 1-1) (Miller and Urey 1959). The chamber was connected to a pair of
electrodes and sparks were continuously fired (electric discharge) to simulate lightning. The bottom of the chamber was connected to a condenser, which simulated condensation in the atmosphere and subsequently flowed through a trap in which samples were taken for analysis of newly synthesized compounds. The water returned to a glass flask (the ocean), in which it was again heated to water vapour and remixed with prebiotic gases in a continuous cycle. After the experiment was conducted for one week, they found that a substantial percentage of the carbon (10-15%) was incorporated into simple organic molecules such as cyanide and formaldehyde. About 2% of the carbon in the system was converted into amino acids, predominantly glycine, alanine and \( \alpha \)-amino-\( n \)-butyric acid. Other more complex biological molecules, such as polyhydroxyl compounds, were also formed, but in much lower yield. From this work, Miller and Urey have shown that it is possible for various simple organic molecules to accumulate under reducing conditions in the early earth, which could have served as substrates for the synthesis of more complex biological compounds.

1.1.1 Template Mediated Replication

As suggested by the Miller and Urey experiment, the prebiotic atmosphere may have contained a good mixture of simple organic molecules, and such an environment would likely be an ideal scenario for the beginning of life. How these molecules came together to initiate life is still a mystery, but once life started, an organism would had to continuously evolve in order to sustain life. An important aspect of evolution requires the inheritance of genetic information from one
generation to another. A hypothetical riboorganism, for example, that contains an RNA sequence catalyzing a reaction critical for the survival of the organism, must evolve a strategy for generating more of this particular RNA. Therefore, one of the requirements for survival is being able to generate more of one’s self, or self-replication (Gesteland, Cech et al. 2006).

In the hypothesized RNA world, the replication system of a riboorganism would have likely been primitive, without the luxury of modern protein enzymes to assist it. To sustain efficient self-replication, the system would require, at the minimum, both a template and a source of ribonucleotide as RNA building blocks. To first explore the possibility of a catalyst-free self-replication system, Leslie Orgel’s group attempted various experiments using different activated derivatives of guanosine 5’ phosphates and examined their ability to oligomerize on a poly-cytidine (poly (C)) template (Tohidi, Zielinski et al. 1987). Most of the analogues, however, showed both very poor extension efficiencies and regioselectivity (i.e. mixtures of 2’ to 5’ and 3’ to 5’ linkages). To optimize the polymerization reaction, he substituted the 3’ OH with a more nucleophilic amino group and the 2’ OH with a methyl group on guanosine 5’-phosphorimidazolidate and consistent with his hypothesis, he observed that the new analogue oligomerized efficiently and regiospecifically in a 3’ to 5’ manner on both poly (C) and poly (dC) templates. While Orgel’s work illustrates a catalyst-free template copying strategy, the resulting polymerized products, however, are modified RNA polymers (i.e. phosphodiester linkages replaced with phospho-nitrogen bonds).
which require specific substrate modifications to achieve the 3' to 5' linkages found in nature today.

To address whether the addition of a RNA catalyst (ribozyme) may induce the synthesis of RNAs specifically in a 3' to 5' dependent manner, Ekland and Bartel in 1995 successfully isolated seven families of ribozymes that can join a ribonucleotide \textit{in trans} to their 5' terminus (Ekland, Szostak et al. 1995). While six of the ribozyme families yielded 2' to 5' linkages, there was one family, denoted as the class I ligases, which can catalyze the formation of the desired 3' to 5' phosphodiester bond. Next, they explored the potential of the class I ligase to act as a natural polymerase by adding a primer template complex and replacing the substrate with mononucleoside triphosphates (Ekland and Bartel 1996). Interestingly they observed that the primer sequence, which aligned to the template by complementary base pairing, was polymerized by up to six nucleotides. Although the class I ligase showed activity mimicking that of protein polymerases, it could only extend the primer by a limited number of nucleotides.

In an effort to find a ribozyme that can catalyze general RNA polymerization, Johnston et al. in 2001 performed another \textit{in vitro} selection using a RNA pool consisting of a mutagenized region relative to the catalytic core of the class I ligase, along with 76 random positions (Johnston, Unrau et al. 2001). After 18 rounds, they successfully isolated a ribozyme that could extend the primer up to 14 nucleotides. This ribozyme was not only active \textit{in trans}, but it also did not require a specific region on the ribozyme for hybridizing to either the primer or template. Further work by Zaher et al. in 2007 resulted in the isolation of the RNA
Figure 1-2: Predicted secondary structure of the B6.61 polymerase ribozyme.

The ribozyme can be separated into two domains, with the catalytic domain primarily responsible for mediating the polymerization reaction. The accessory domain, however, is required for efficient template dependent primer extension (Zaher and Unrau 2007).

polymerase ribozyme B6.61, which showed superior extension and fidelity relative to the Round 18 polymerase from which it evolved from (Figure 1-2) (Zaher and Unrau 2007). B6.61 polymerized all tested primer template complexes faster than its progenitor, and for one of the primer template complexes an extension of at least 20 nucleotides (nt) was observed. Although extension of ~20-nt is still far from the full length of ~180-nt of Round 18 or B6.61 polymerases, this study does show the potential of RNAs in catalyzing
continuous polymerization in a template dependent manner. Therefore, it is expected that in the near future a RNA-dependent RNA polymerase with high processivity that can extend a broad range of templates will be discovered, which will further solidify the hypothesis of riboorganisms existing as the first life form in the early world.

1.1.2 Requirement for Compartmentalization

Genetic inheritance is a critical requirement for sustaining life, but in order to efficiently replicate, as well as grow and maintain homeostasis, an early riboorganism would also need to accumulate essential building blocks. Such building blocks, if they were in the environment, were likely present in a low abundance such that it would be virtually impossible for an early organism to obtain sufficient amounts through random searching. Therefore, many researchers have suggested that there must have been a common strategy employed for concentrating early biological molecules. One hypothesis is that molecules may be synthesized, or attached onto mineral surfaces, and collected at one location. One example would be the work by Otroshchenko and Vasilyeva in 1976, in which they showed that different nucleoside monophosphates can be readily absorbed onto the surface of volcanic rocks (Otroshchenko and Vasilyeva 1977). Interestingly, they also noted that there was substantial desorption at elevated pH values (pH 11). The concentrated organic material (such as nucleoside monophosphates) released from the mineral surfaces, therefore, may possibly serve as building blocks for the synthesis of early RNA polymers. Mineral surfaces, however, are not only involved in concentration of molecules,
but can also provide sites for catalysis. This was demonstrated in work by Ferris et al., in which they showed that the clay montmorillonite with alkali metals, or \( \text{NH}_4^+ \), or \( \text{Ca}^{2+} \) as exchangeable ions, can actually catalyze the polymerization of long RNA polymers with activated nucleotides (Ferris and Ertem 1993; Ferris, Hill et al. 1996). This was an important result because prior to this work, it was not deemed possible to synthesize long oligomers by condensation in water, which would have been the most likely prebiotic solvent, due to competition from hydrolysis.

A second hypothesis has also been suggested, which involves the trapping and concentration of important molecules within encapsulated vesicle-like structures (Szostak, Bartel et al. 2001). All aerobic cellular organisms, dating back to the cyanobacteria, have evolved the use of such an encapsulation method to contain important molecules (Lazcano and Miller 1994). Within the cell membrane, important metabolites are contained and used as substrates for the synthesis of more complex products. Even viruses are enclosed in a primitive protein coat synthesized by the hijacked host cell machinery. If early life did begin with the use of encapsulation, the membrane systems that developed could have been as simple as that of viruses. Such early protocell systems would have been either autotrophic or heterotrophic. As an autotroph, the cell would synthesize its own metabolites *in situ* that are required for self-replication. Heterotrophs, on the other hand, would acquire their nutrients and energy from its environment by importing them across their cell membrane and entrapping them within.
In an effort to study the first primitive membrane systems, the Szostak lab began examining various methods of generating vesicles from prebiotic sources such as simple fatty acids. In their published work in 2003, they reported their findings of the increased vesicle formation rate from fatty acid micelles upon the addition of small amounts of montmorillonite (Hanczyc, Fujikawa et al. 2003). Interestingly, these mineral particles can be readily taken up by vesicles, thus illustrating the possibility that membrane building blocks and primitive oligonucleotides could have accumulated in early protocells (Hanczyc, Fujikawa et al. 2003).

For protocells to have been a viable route for early life, they must also have been capable of cell division. The simple division of one cell to form two cells is vital to the survival of a population for both cell propagation and exponential growth. Various strategies are possible for vesicle division, but they all require either a significant energy input for either the export of solutes or generating membrane asymmetry (Hanczyc, Fujikawa et al. 2003). In another model, vesicles perhaps do not need to divide, but instead can synthesize additional smaller vesicles from within. This was supported by the observation that upon the synthesis of new amphiphiles catalyzed by an encapsulated catalyst, new vesicles can be formed within the parent vesicle (Chen, Roberts et al. 2004). This scenario may also have been possible for a protocell, in which amphiphilic molecules that have diffused into the interior can similarly be converted to smaller vesicles, catalyzed by encapsulated minerals. These new daughter vesicles perhaps can then be released by either lysis or by passing
through large transient pores of the parent vesicle. Together, the different strategies for vesicle growth and propagation described above would support the heterotroph protocell model, as in all cases an external source of amphiphiles is required.

Under the predicted prebiotic conditions, the most plausible amphiphiles for vesicle formation would be saturated fatty acids containing short acyl chains, such as C10 decanoic acid (Figure 1-3). Interestingly, as reported in a more recent work by the Szostak group, vesicles made up of decanoic acid, decanol and glycerol mono-decanoate, were highly permeable to ribose, as well as to activated nucleotides that have a low net charge and polarity (Mansy, Schrum et al. 2008). To examine whether a self-replication system can be initiated in such a protocell, they designed a DNA primer that could bind to a template containing both a primer-binding region and a string of 15 deoxycytidine nucleotides (dC). The primer and template were subsequently encapsulated in the protocells, with activated nucleotides 2′ amino-guanosine 5′ phosphorimidazolide added to the extracellular space (as in Orgel’s work described previously (Tohidi, Zielinski et al. 1987)). The activated nucleotides were indeed able to diffuse into the

Figure 1-3: C10 decanoic fatty acid derivatives used for vesicle construction.
protocells to engage in template copying, with the majority of the products formed within 6 hours being the full length extended primer. Their results further support the heterotrophic protocell model in which building blocks for replication are imported from their environment, as opposed to an autotroph, which would have problems preventing leakage of internally synthesized metabolites. From this work, Szostak and co-workers have shown that it is highly possible for life to begin within a simple protocell, which would have started with simple self-replication, and gradually evolved to contain nucleic acid catalysts for synthesis of new biological compounds. While this is a highly attractive model, it does require the prebiotic environment to be rich in activated nucleotides for heterotrophs to replicate their genome efficiently. Activated nucleotides, however, are high energy molecules and are only known to be synthesized by protein enzymes in modern metabolism. While the synthesis of primitive oligomers may have been made possible by the use of a scarce source of prebiotically synthesized nucleotides (as will be discussed in the next section), over time a more efficient mechanism of nucleotide synthesis must have evolved to supply riboorganisms with sufficient building blocks for replication (Gesteland, Cech et al. 2006).

1.2 Nucleotide Synthesis

Nucleotides are essential to the RNA world hypothesis because they are the basic units from which larger RNA polymers are constructed. In modern metabolism, there are various nucleotide synthesis pathways catalyzed by protein enzymes to ensure a constant intracellular supply of this important
molecule (Mathews 1996). The scenario for early riboorganisms is drastically different, however, as prebiotic supplies of nucleotides were likely scarce in the early atmosphere, with no protein enzymes to assist in replenishing their numbers after consumption (Bartel and Unrau 1999). This problem motivated us to our initial project, which involved exploring the ability of RNAs to catalyze the synthesis of nucleotides (Chapter 2). However, a dilemma arises, given that even if such ribozymes were to be found that were capable of nucleotide synthesis, these ribozymes would still require an initial source of nucleotides for their own assembly. In the following sections, we will discuss the possible abiotic pathways in which various components of nucleotides could have been synthesized, and how these may have ultimately led to the emergence of the first nucleotides in the early world.

1.2.1 Abiotic Synthesis of Nucleobases

1.2.1.1 Purine Bases

The synthesis of purine bases under prebiotic conditions was first demonstrated by John Oro in 1961, when he successfully synthesized a small amount of adenine and hypoxanthine by refluxing an aqueous solution containing hydrogen cyanide (HCN) and ammonia (NH₃) (Oro 1961). The HCN polymerization steps leading to adenine is not thoroughly understood, but a possible prebiotic mechanistic series has been proposed (Figure 1-4) (Zubay and Mui 2001). This involves an initial step in which high concentrations of HCN in small bodies of water first slowly reacted to form diaminomaleonitrile
(nucleophilic additions involving 4 molecules of HCN to give the HCN tetramer).

Upon exposure to solar radiation (ultraviolet), the tetramer can be subsequently converted to aminoimidazole carbonitrile (AICN). AICN can then be synthesized into adenine by either heating with formamide (which forms from the hydrolysis of HCN), by reacting with formamidine, or by reaction with HCN. AICN can also be partially hydrolyzed to form aminoimidazole carboxamide (AICA), which can react with HCN or cyanogen (C$_2$N$_2$) to give hypoxanthine and guanine, respectively (Zubay and Mui 2001). This series of purine base synthesis reactions, along with the hydrolytic nature of HCN, would require large quantities of prebiotic HCN to be present in the early atmosphere (Shapiro 1995). However, various reports have suggested alternative pathways in which HCN may be required in lesser amounts. One example was work by Hill and Orgel, in which they obtained a respectable yield of adenine (18%) by heating at high temperatures (110°C) the HCN tetramer with ammonium formate (Hill and Orgel 2002). This was an interesting result because firstly, it argues against the assumption that hydrolysis of HCN would always be a major obstacle to adenine synthesis, since ammonium formate is the ultimate product from HCN hydrolysis. Secondly, it showed that purine base synthesis is possible with a requirement of only very low concentrations of HCN, as long as such concentrations permit the initial synthesis of HCN tetramers.

1.2.1.2 Pyrimidine Bases

The possible presence and abundance of pyrimidine bases in the early world is still currently a topic of debate. Unlike purine bases, both cytosine and
Figure 1-4: Proposed prebiotic pathway for the synthesis of different purine bases.

An initial input of four molecules of hydrogen cyanide (HCN) is first polymerized to form the tetramer diaminomaleonitrile, which can be converted to aminimidazole carbonitrile (AICN) by photoisomerization. AICN can either react with a molecule of HCN to give adenine, or undergo partial hydrolysis to aminimidazole carboxamide, which can subsequently form hypoxanthine or guanine upon reaction with either HCN or cyanogen, respectively.
uracil were not observed as products in the electric spark discharge experiments described earlier (Miller and Urey 1959). This led to the alternative hypothesis that perhaps other bases may have been used instead as the first genetic material (i.e. isocytosine, diaminopyrimidine, urazole, guanazole, or just purine bases alone) (Wachtershauser 1988; Piccirilli, Benner et al. 1990; Kolb, Dworkin et al. 1994). The first plausible prebiotic route for cytosine synthesis was first described in work by Sanchez and Orgel in 1966 in which they reported that cytosine can be synthesized with a low yield of 5% from cyanoacetylene and potassium cyanate by heating at 100 °C (Sanchez, Ferris et al. 1966). Cyanoacetylene is described as a possible prebiotic molecule because it is readily generated from methane-nitrogen mixtures through spark discharge reactions (Miller and Urey 1959). This reaction, however, suffers from the fact that cyanoacetylene is readily converted to give cyanoacetaldehyde upon hydration, where as cyanate similarly reacts with water and hydrolyzes to carbon dioxide and ammonia (Robertson and Miller 1995; Nelson, Robertson et al. 2001). In addition, the reaction requires a high concentration of cyanate (at 1M) in aqueous solution and it is unlikely that such a reservoir would have been present under prebiotic conditions (Ferris, Sanchez et al. 1968; Shapiro 1999). In continuation of their work, Sanchez and Orgel examined other plausible prebiotic compounds, and in 1968, they reported their finding that cytosine can be synthesized with similar yield using cyanoacetylene and the much more water soluble urea (at 1 M concentration) (Ferris, Sanchez et al. 1968). The likelihood
of this reaction occurring in a prebiotic environment was strengthened by work from Ferris, Zamek, Altbuch and Freiman, who later described an alternative route in which cyanoacetylene is replaced with cyanoacetaldehyde (Ferris, Zamek et al. 1974). They postulated that the more stable cyanoacetaldehyde might have been an intermediate in the synthesis of cytosine. They were unable to detect cytosine production using 0.1 M urea, but in 1995, Robertson and Miller extended upon this work and showed that cytosine can indeed be synthesized efficiently (with yields as high as 53%) from near saturating concentrations of urea (Figure 1-5) (Robertson and Miller 1995). They also described the low product yield observed under low urea concentrations was due to the competing hydrolysis reaction (Figure 1-5), and suggested that the favourable conditions conducive to cytosine production may have been present in evaporating lagoons or in pools on drying beaches on the early Earth.

Although the synthesis of cytosine from urea and cyanoacetaldehyde is highly efficient, its prebiotic plausibility has been vigorously challenged. In a report published by Robert Shapiro in 1999, he outlined some of the problems with the proposed pyrimidine synthesis pathways (Shapiro 1999). One principal objection was that the molecules that were postulated to be involved, such as urea, cyanate and cyanoacetylene (or cyanoacetaldehyde), also undergo rapid side reactions with common nucleophiles, thus greatly reducing their availability for cytosine synthesis. In addition, cytosine decomposes to uracil and other products by deamination, with a half life of ~340 years at 25°C (Shapiro 1999).
Figure 1-5: Proposed prebiotic pathway for the synthesis of pyrimidine bases.
Cyanoacetaldehyde can be synthesized to cytosine or uracil under high concentrations of urea. In a competitive hydrolysis reaction, cyanoacetaldehyde is broken down into formic acid and acetonitrile (Robertson and Miller 1995).

None of the proposed prebiotic cytosine synthesis routes known today, even those occurring in specialized locations, are fast enough to compensate for the decomposition of the base. To address these problems, various scientists in ongoing research efforts are attempting to find alternative substrates that offer greater stability, while other researchers are actively exploring alternative pathways for pyrimidine base synthesis.

1.2.2 Abiotic Synthesis of Ribose

1.2.2.1 The Formose Reaction

The formose reaction is an autocatalytic reaction that satisfies the prebiotic requirements and is the most studied reaction for the synthesis of
ribose. It involves a series of reactions that generate aldoses starting from formaldehyde (giving the name formose reaction, formaldehyde and aldose), and the reaction is catalyzed by base and divalent metal ions, such as calcium hydroxide (Figure 1-6) (Breslow 1959; Zubay and Mui 2001; Ricardo, Carrigan et al. 2004). The initial step of the reaction involves the condensation of two molecules of formaldehyde, resulting in the formation of glycoaldehyde. Under basic conditions, the 2-carbon glycoaldehyde loses a proton to form the nucleophilic enediolate form, which can then react with another molecule of formaldehyde to give the 3-carbon glyceraldehyde. In the pentose synthesis pathway, glyceraldehyde can react with another nucleophilic enediolate of glycoaldehyde in an aldo reaction to give various 5-carbon pentoses such as ribose. Subsequent nucleophilic attack of one of the hydroxyl groups on the carbonyl carbon results in the cyclic form of pentose sugars. Alternatively, the nucleophilic enediolate form of glyceraldehyde can react with another molecule of formaldehyde to give the 4-carbon tetrulose, which can form tetrose through an enediol intermediate. Tetrose can then fragment to give two molecules of glycoaldehyde and re-enter the formose reaction to fuel the autocatalytic cycle.

The formose reaction is a very attractive solution for prebiotic ribose synthesis, as this autocatalytic cycle requires only the input of formaldehyde. However, while most of the steps in the formose reaction are well understood, it is still open to discussion as to how glycoaldehyde can be generated from two formaldehyde molecules in the very first step of the reaction (Gesteland, Cech et al. 2006). The difficulty with this chemistry is that formaldehyde is commonly
Figure 1-6: Synthesis of ribose by the autocatalytic formose reaction pathway.

The autocatalytic cycle begins with glycoaldehyde forming an enediolate under basic conditions, and subsequently reacting with a formaldehyde molecule to give glyceraldehyde (3-carbon molecule, 2+1→3). Glyceraldehyde can then react with a glycoaldehyde in the enediolate form to generate ribose (3+2→5). In an alternative pathway to regenerate glycoaldehydes, glyceraldehydes can form the nucleophilic enediolate, which can react with a molecule of formaldehyde to give tetrulose (3+1→4). Tetrulose can form tetrose through an enediol intermediate, which can subsequently fragment to give two molecules of glycoaldehydes (4→2+2) (Gesteland, Cech et al. 2006).
known as an electrophile, but to directly form a carbon-carbon bond between two formaldehyde molecules would require the carbon atom from one molecule to be a nucleophilic center, while the carbon atom from the other to be an electrophilic center (Gesteland, Cech et al. 2006). This would necessitate the addition of one pair of electrons to one of the formaldehyde molecules in order for it to act as a nucleophile, which is an unusual situation for this electrophilic molecule. One possible solution to alleviate this problem would be the removal of a proton from formaldehyde, possibly by a strong base, and leaving behind the pair of electrons formerly used for the C-H bond. This results in the formation of an acyl anion and the electron pair can then be used to form a new carbon-carbon bond with another molecule of formaldehyde to give glycoaldehyde. Alternatively, formaldehyde can form short polymers with O–CH₂–O repeating units, and upon deprotonation, an RO-CH'-OR anion is formed (R indicates repeating formaldehyde units). The anion can now act as a nucleophile and react with another formaldehyde monomer, and subsequent depolymerization would give glycoaldehyde. The anions formed in either case are not highly stable, but small amounts of them would possibly be sufficient to initiate the synthesis of glycoaldehyde, which can be regenerated through the autocatalytic formose reaction cycle (Gesteland, Cech et al. 2006). A more conceptual idea, which negates the synthesis debate, has also been proposed in which glycoaldehydes are delivered to the early Earth, as both formaldehydes and glycoaldehydes are known to be present in the cosmos (Hollis, Lovas et al. 2000).
1.2.2.2 Alternatives to Ribose and Ribonucleotides

While the formose reaction is seemingly the most feasible prebiotic pathway for the synthesis of ribose, its downfall is the very poor yield of ribose that results from the reaction, which itself is highly unstable (Shapiro 1988). Many research efforts have been invested in attempts to optimize the formose reaction, the most notable being Steven Benner’s work (as discussed in the next section), in the discovery of using borate to enhance ribose stability (Ricardo, Carrigan et al. 2004). However, many scientists have dismissed the likelihood of ribose synthetic pathways outright, and have instead turned their attention to examining possible ribose alternatives that may be more prebiotically plausible.

In a detailed study published in 1999, Albert Eschenmoser compared various sugar systems of different forms, all of which had the nucleobases in the equatorial position, in an effort to address the question as to why nature selected the 3’ to 5’ pento-furanosyl structure as the sugar backbone for both RNA and DNA (Figure 1-7) (Eschenmoser 1999). The first sugar system in his study was the 4’ to 6’ hexo-pyranosyl system, which contain the presence of six carbon atoms as well as a six-membered pyranose ring. While this sugar system, in the 2’, 3’ deoxy form, actually showed enhancement in Watson-Crick base pairing relative to RNA, the opposite was true when a hydroxyl group was inserted at the 2’ position, resulting in a steric clash between the hydroxyl group and the neighbouring base. Because the hexo-carbon system proved to be too sterically bulky, they shifted their attention to the 2’ to 4’ pento-pyranosyl sugar system, which again contains a six membered ring, but with only five carbon atoms. Interestingly, all members of this system formed much stronger Watson-Crick
base pairing than RNA, and their nucleotides can form base pairing between different members of the family, but not with ribonucleotides (i.e. \(\beta\)-ribo-pyranosyl can cross-pair with \(\beta\)-xylo-pyranosyl, but not with \(\beta\)-ribo-furanosyl). It was also noted that all members of this system are over tolerant towards base pair mismatches, resulting in self-pairing of oligonucleotides (i.e. sequence of A\(_5\) self pairs with another sequence of A\(_5\) to form a duplex). While the strong base pairing of the pento-pyranosyl system may seem to be an attractive trait, it does generate various problems in a self-replicating system. First, templates for replication would be strongly inhibited by their polymerization products, as newly synthesized strands cannot easily separate from their parental strand due to strong Watson-Crick base pairing. Secondly, replication fidelity can be a problem if local concentrations are biased toward a particular set of nucleotides (i.e.
higher purine than pyrimidine nucleotides), because of the high tolerance of the system towards base pair mismatches. These findings support the notion that it is the optimization, and not maximization, of base pairing strength that is the criteria for the selection of furanosyl over the pyranosyl system. In addition, the flexible ribo-furanose backbone is beneficial for RNA in different biological functions, such as catalysis, which may also contribute to its selection by nature (Chen, Li et al. 2007).

While Albert Eschenmoser has done an exhaustive study of five and six carbon sugars, other researchers have performed similar work in examining nucleotides with different backbone systems. Most notably are the four carbon sugar threose nucleic acid (TNA), which is chemically simpler and forms more readily than ribose, and the peptide nucleic acid (PNA), which contain a highly stable neutral backbone constructed by amide bonds (Figure 1-8) (Joyce 2002). Both of these analogues, interestingly, can form stable structures and Watson-Crick base pairs with DNA, RNA and themselves. In work by Szostak and colleagues, they have demonstrated that sequence dependent DNA polymerization is possible with a TNA template using a DNA polymerase (Chaput, Ichida et al. 2003), and vice versa for TNA synthesis on a DNA template (Kempeneers, Vastmans et al. 2003). Similarly, it was reported that complementary PNA oligomers can be polymerized using hetero-sequences of DNA as templates (Schmidt, Christensen et al. 1997). Other possible backbone systems, such as the glycerol nucleic acid, alanyl nucleic acid, and the pyranosyl RNA system (as described above), have also been studied, but there are
Figure 1-8: Examples of potential backbone structures of early nucleic acid polymers. TNA = threose nucleic acid, PNA = peptide nucleic acid (Joyce 2002).

currently little experimental evidence for their postulated roles as strong potential predecessors to RNA (Joyce 2002).

However, if any of the RNA analogues described above were indeed responsible for carrying genetic information in what would be called pre-RNA life, then how did this pre-RNA world convert into one based on RNA? To address this question, Orgel and colleagues in 1997 reported their work showing that information carried by PNA can be transferred to RNA in a template directed manner using activated ribonucleotides as substrates (Schmidt, Nielsen et al. 1997). While this finding is encouraging, currently there is still no experimental evidence showing that the intrinsic catalytic properties of RNAs can be carried over to either TNAs or PNAs, or vice versa (Gesteland, Cech et al. 2006). This is likely due to the involvement of the sugar moiety or their phosphate backbone (or both) for structural interactions and catalysis. The first catalytic RNAs, therefore, were seemingly derived from the assembly of RNA polymers, rather than from conversion of a genetic carrier in another form (Gesteland, Cech et al. 2006).
1.3 Characteristics and Significance of Ribose Sugars

1.3.1 The Instability of Ribose

Sugars, in particular ribose, are highly unstable under alkaline conditions (pH >9.5) and undergo various fragmentation and dehydration reactions to give tar-like products. Compared to other sugars, ribose is much more labile, as demonstrated by its high rate of enediol formation, and reacts about 4 fold and 16 fold faster than the average of other pentoses and hexoses, respectively (Isbell, Frush et al. 1969). Under mild conditions, ribose also readily undergoes isomerisation into other aldo- and ketopentoses, such as arabinose and ribulose (Isbell, Frush et al. 1969). The instability of ribose further hampers the possibility of prebiotic ribose synthesis by the formose reaction, as the reaction gives a very small amount of ribose (estimated to be less than 1% relative to other products synthesized), and any that are synthesized are easily broken down to other tar-like products (Springsteen and Joyce 2004). Ongoing research efforts, however, have shown that the formose reaction can be modified such that the synthesis of ribose is favoured. One of these modifications was described in a report published by Langenbeck in 1954, in which he demonstrated that lead can act as a catalyst, by stabilizing the diol of glyceraldehyde, and enhances the yield of aldopentoses from formaldehyde by 20 fold, giving an efficiency of about 30% (Langenbeck 1954). Further work by the Zubay group in 1998 showed that ribose was the first aldopentose generated upon addition of the divalent ion to the formose reaction, with other aldopentoses (hypothesized to be secondary products of ribose) emerging afterwards (Zubay 1998). Work by Eschenmoser
Figure 1-9: Stabilization of ribose by phosphate esters or cyanide.

a. Ribose 2,4-diphosphate, a stable ribose derivative, can be synthesized from the reaction between the enediolate form of glycoaldehyde phosphate with formaldehyde to give glyceraldehyde 2-phosphate (2+1 → 3), followed by subsequent reaction with another molecule of glycoaldehyde phosphate enediolate (2+1+2 → 5). Ribose 2,4-diphosphate can similarly be generated starting from a mixture of glyceraldehyde 2-phosphate and glycoaldehyde (Mueller, Pitsch et al. 1990). b. Ribose, in its open chain conformation, can react with a molecule of cyanamide to give a bicyclic product with cyanamide joined at the 2’ oxygen and the anomeric carbon (Sanchez and Orgel 1970).

and colleagues have shown that when glycoaldehyde is replaced with glycoaldehyde phosphate as the starting substrate and subsequently reacted with formaldehyde, ribose 2,4-diphosphate can eventually be formed (Figure 1-9a) (Mueller, Pitsch et al. 1990). The same product can be obtained in solutions of glycoaldehyde and glyceraldehyde 2-phosphate in the presence of hydroxide minerals. Ribose 2,4-diphosphate is much more stable and less reactive than ribose, as the phosphate groups raises the pKa of the α-hydrogen and lowers the sugar's potential nucleophilicity. In another work by Sanchez and Orgel, they have shown that upon reaction of ribose with cyanamide, a bicyclic product is
formed, involving 2' OH and the anomeric carbon (Figure 1-9b) (Sanchez and Orgel 1970). Springsteen and Joyce have further extended this work and have shown that the cyclic compound is much more stable than free ribose, thus giving rise to the possibility that ribose-cyanamide may have served as a reservoir for ribose in the early world (Springsteen and Joyce 2004).

While varying degrees of success have been achieved in optimizing the formose reaction, it was not until a recent landmark publication in 2006 by Ricardo and Benner that strengthened the claim of the plausibility that ribose could have accumulated in the early world. They discovered that the addition of borate dramatically increased both the stability of ribose and the efficiency of its synthesis by the formose reaction (Ricardo, Carrigan et al. 2004). Borate has been known to form cyclic complexes with molecules containing hydroxyl groups on two consecutive carbon atoms (1,2-diols), with the boron atom carrying a negative charge in the complex state (Figure 1-10). In the formose reaction, glyceraldehydes are an example of a compound containing a 1,2-diol. Upon binding with borate, the glyceraldehydes are disfavoured towards the formation of the nucleophilic enediolate state that ultimately leads to the regeneration of glycoaldehydes in the autocatalytic cycle. As a result, the reaction is driven towards the synthesis of pentoses, as glyceraldehydes can now act solely as electrophiles and react with enediolate of glycoaldehydes. The newly synthesized ribose sugars can also bind with borate as they similarly contain a diol (2' OH and 3' OH). This disfavours the formation of the nucleophilic enediolate, and as a result the ribose is further stabilized in the borate complex form.
Figure 1-10: Stabilization of ribose by borate.

Borate can be found from evaporated minerals, such as ulexite and colemanite, and can stabilize intermediates of the formose reaction by forming complexes with vicinal diols. Curved lines indicate complex formation with another unit of the same diol compound (i.e. two molecules of glyceraldehydes forming a complex with borate) (Ricardo, Carrigan et al. 2004).

1.3.2 Reactivity of Ribose and Other Sugars

Ribose sugars, as described above, are highly reactive in their open chain form, which is consistent with their rapid degradation rate (Larralde, Robertson et al. 1995). For example, ribose reacts ~3.5 fold faster than other sugars with urazole (Dworkin and Miller 2000). Their high reactivity is attributed primarily to the aldehyde group at the anomeric carbon of the ribose in its acyclic form.

Aldehydes are particularly reactive due to the polar nature of the carbonyl group, as oxygen atoms are much more electronegative than carbon and will draw electrons towards it. This results in a polarized positive charge on the carbon atom, making it an excellent electrophilic site for reaction with a wide range of nucleophiles. Examples of nucleophiles that are reactive with aldehydes include cyanide, water (acidic or basic conditions, resulting in hemiacetal formation), enolates, and amine containing compounds, particularly primary amines (Mathews 1996). In reactions with amines, the nitrogen atom can readily act
Sugars, such as glucose and ribose, can undergo ring opening to their aldehyde form, which can subsequently react with nucleophilic primary amines of different amino acids (glycine shown in this case) to form Schiff bases (forms carbon- nitrogen double bond). The Schiff bases then undergo an Amadori rearrangement to form the more stable Amadori product (Munanairi, O'Banion et al. 2007).

as a nucleophile and attack the carbonyl center using its lone pair of electrons, ultimately resulting in the formation of a Schiff base. This chemistry is commonly observed between amino acids and carbohydrates, resulting in the changes in colour and flavour of foods (commonly known as the Maillard Reaction) (Figure 1-11) (Munanairi, O'Banion et al. 2007). The reaction rate is directly related to the pKa of the amine group, as reaction rates have been shown to increase linearly with pKa values of amino acids containing only an $\alpha$-amino group (i.e. no additional amine groups in the side chains) (Sandwick, Johanson et al. 2005).

Following the subsequent formation of the imine, the Maillard reaction product undergoes an Amadori rearrangement and regenerates a carbonyl moiety (Munanairi, O'Banion et al. 2007). Subsequent reactions involving

**Figure 1-11: The Maillard reaction.**
Sugars, such as glucose and ribose, can undergo ring opening to their aldehyde form, which can subsequently react with nucleophilic primary amines of different amino acids (glycine shown in this case) to form Schiff bases (forms carbon- nitrogen double bond). The Schiff bases then undergo an Amadori rearrangement to form the more stable Amadori product (Munanairi, O'Banion et al. 2007). The reaction rate is directly related to the pKa of the amine group, as reaction rates have been shown to increase linearly with pKa values of amino acids containing only an $\alpha$-amino group (i.e. no additional amine groups in the side chains) (Sandwick, Johanson et al. 2005). Following the subsequent formation of the imine, the Maillard reaction product undergoes an Amadori rearrangement and regenerates a carbonyl moiety (Munanairi, O'Banion et al. 2007). Subsequent reactions involving
rearrangement, fragmentation, dehydration and condensation reactions give rise to various degradation products known as advanced Maillard reaction end products, which are currently not well characterized (Smith, Taneda et al. 1994). Interestingly, work by Nissl and colleagues have shown that guanosine can similarly interact with ribose using the amino group at the 2 position in a Maillard reaction, in the presence of a primary amine containing compound (Nissl, Ochs et al. 1996). This chemistry will be of significance in Chapter 4, as it is the potential reaction mediated by a highly versatile ribozyme, which uses ribose sugar derivatives and purine bases as substrates for catalysis.

1.3.3 Sugar Metabolism in Cells

1.3.3.1 Glycolysis Pathway

Carbohydrates, such as ribose and glucose, are critical to cellular organisms for both nucleotide synthesis and as an energy source. In modern metabolism, various six-carbon sugars can be broken down to pyruvate for the generation of energy in the 10-step glycolysis catabolic pathway (Figure 1-12) (Mathews 1996). The first step involves the phosphorylation of glucose by hexokinase to form glucose 6-phosphate, with the investment of one molecule of ATP. Glucose 6-phosphate is subsequently isomerised to fructose 6-phosphate by phosphoglucoisomerase, followed by another phosphorylation step, with the investment of a second molecule of ATP, to give fructose 1,6-bisphosphate. In the fourth step, the six-carbon sugar is broken down to two triose phosphates (dihydroxyacetone phosphate and glyceraldehyde 3-phosphate (G3P)) by
Figure 1-12: The 10-step glycolysis pathway.

Hexose sugars, such as glucose, can be metabolised in a 10-step pathway, with each step catalyzed by a different protein enzyme. The first five steps are known as the energy investment phase, in which two ATP molecules are invested to form two molecules of glyceraldehyde 3-phosphate. The final five steps (steps 6-10) are known as the energy generation phase, in which triose phosphates are converted to ATPs, and NADH molecules are generated. Glycolysis can be used by both anaerobic and aerobic cells for harnessing of metabolic energy (Mathews 1996).

aldolase, with dihydroxyacetone phosphate subsequently isomerised by triose phosphate isomerase to give another molecule of G3P. In the first five steps described thus far, two ATP molecules have been invested, and it is only upon further processing of the two high energy G3P molecules from reactions 6 to 10, is energy generated. The energy generation phase begins by the oxidation and subsequent phosphorylation of G3P, with the formation of a thiohemiacetal intermediate, to give 1,3-biphosphoglycerate. This reaction is catalyzed by glyceraldehyde 3-phosphate dehydrogenase, and the hydride released from the oxidation is transferred onto the coenzyme NAD$^+$ to give NADH. The high energy 1,3-biphosphoglycerate molecule can then transfer its acyl-phosphate group to ADP by phosphoglycerate kinase, giving 3-phosphoglycerate and the synthesis of a molecule ATP. To generate the next high energy compound, 3-phosphoglycerate is first isomerized by phosphoglycerate mutase to 2-phosphoglycerate, followed by a dehydration reaction to give
phosphoenolpyruvate, which is in a thermodynamically unstable enol configuration. This drives the final energy generation reaction, catalyzed by pyruvate kinase, in which phosphoenolpyruvate dephosphorylates to give pyruvate and drives the synthesis of a second ATP molecule. Because the first five steps of the glycolysis pathway generate two molecules of G3P, this gives a net gain of two ATP molecules and two molecules of NADH after the completion of the glycolysis cycle. The resulting pyruvate molecules can be further harnessed for energy upon insertion into the citric acid cycle, ultimately leading to their oxidation to CO$_2$.

1.3.3.2 Pentose Phosphate Pathway

Sugars, however, can also be used in the anabolic pentose phosphate pathway for the synthesis of ribose 5-phosphate, which is required for the synthesis of nucleotides (Mathews 1996). In the oxidative phase, glucose 6-phosphate (from the phosphorylation of glucose) is first oxidized to 6-phosphoglucono-δ-lactone using glucose 6-phosphate dehydrogenase, resulting in the conversion of the carbon at the 1-position to a carbonyl group and the subsequent reduction of NADP$^+$ to NADPH. 6-phosphoglucono-δ-lactone is then hydrolyzed by 6-phosphoglucoisomerase to 6-phosphogluconate in the next step, causing the opening of the sugar chain. A carbon atom is then subsequently removed as CO$_2$ in the third step by 6-phosphogluconate dehydrogenase in an oxidative decarboxylation reaction, resulting in the formation of ribulose 5-phosphate and another molecule of NADPH from NADP$^+$. Ribulose 5-phosphate can undergo further isomerisation via an enediol intermediate by
phosphopentose isomerase, resulting in the formation of ribose 5-phosphate (PR). PR is an important precursor in various nucleotide synthesis salvage pathways as it can be activated by ATP using pyrophosphotransferases to form the high energy compound 5-phosphoribosyl 1-pyrophosphate (PRPP) (Mathews 1996). PRPP, catalyzed by phosphoribosyltransferases, can react with various bases to synthesize different nucleotides, which can subsequently be assembled into DNA and RNA polymers (Unrau and Bartel 1998; Lau, Cadieux et al. 2004). Reactions involving both PR and PRPP will be further elaborated in the following chapters.

1.3.4 Regulation of Cellular Carbohydrate Levels by RNAs

1.3.4.1 Regulatory sRNAs

As carbohydrates are critical to cells as both an energy source and as the building blocks for other essential macromolecules, it is important that their uptake and metabolism are tightly regulated. Common regulators include different proteins such as transporters and receptors, and enzymes, such as protein kinases. Interestingly, however, regulation of many carbohydrate metabolic pathways also involve non-coding small RNAs (sRNAs) (Gorke and Vogel 2008). In E. coli, sRNAs can regulate cellular carbohydrate content by both activation and repression at the translational level. The sRNA Spot42, for example, is a 109-nt RNA that is upregulated upon increase in intracellular glucose and shows high complementarity to the start codon region of the mRNA encoding galactose kinase (GalK). Galactose kinase is typically involved in the phosphorylation of intracellular galactose to galactose 1-phosphate, which can
then be further processed in a series of pathways to glucose-6-phosphate. Upon expression of Spot42, GalK synthesis is inhibited due to antisense binding of Spot42 to the 5’ end of galK, thus preventing the docking of the 30S ribosome onto the transcript. This is an example of translational repression, which results in both the accumulation of galactose and the shunting of the glucose synthesis pathway upon increased uptake of glucose from the environment.

The opposite is true in the case of the *E. coli* glucosamine 6-phosphate synthetase (GlmS) regulatory pathway (Gorke and Vogel 2008). In this system, the sRNA glmZ serves an activating role and triggers the translation of the glmS mRNA. This results in the synthesis of glucosamine 6-phosphate (GlcN-6-P), which is an essential precursor for the synthesis of bacterial cell wall components. When cells have a sufficient supply of GlcN-6-P, the glmS mRNA typically folds into a stable hairpin structure in which the Shine Dalgarno (SD) sequence is sequestered within the structure and GlmS synthesis is turned off. Upon depletion of amino sugars in an environment lacking external sources, GlmS synthesis is switched on through upregulation of glmZ, which alternatively base pairs in an antisense manner with the glmS mRNA at the sequence region upstream of the SD sequence. This results in the disruption of the hairpin structure, and exposure of the SD sequence for initiation of glmS translation.

### 1.3.4.2 Regulation by a Ribozyme

The regulation of carbohydrate metabolism by RNAs is not limited to sRNAs, however, as catalytic RNAs have similarly been shown to play an important role. In gram-positive bacteria, the GlmS system has apparently
evolved a very different strategy of RNA mediated regulation (Klein and Ferre-D'Amare 2006; Gorke and Vogel 2008). In this system, the \textit{glmS} mRNA encodes, in the 5' untranslated region (UTR), a cis-acting self-cleaving ribozyme with a \text{GlcN-6-P} binding site. Binding of \text{GlcN-6-P} causes the GlmS ribozyme to change its secondary structure, resulting in the subsequent cleavage of the mRNA with \text{GlcN-6-P} acting as a co-factor in the acid-base catalysis reaction. In the absence of \text{GlcN-6-P}, the GlmS ribozyme folds into its inactive native structure, thus allowing translation and synthesis of GlmS. The GlmS ribozyme is currently the only known natural ribozyme that controls gene expression (Klein and Ferre-D'Amare 2006), and is a classic example of a negative feedback mechanism involving RNAs in regulating intracellular carbohydrate concentrations.

1.4 \textit{In vitro} Ribozyme Selection and Evolution

A vital component of the RNA world hypothesis is the emergence of new functional RNAs over time, which would play a fundamental role in the synthesis of important organic molecules necessary for life. Research in the characterization of such ribozymes, however, has been challenging since only a limited number of natural functional RNAs are currently known. Due to the dearth of ribozymes found in nature today, there was considerable interest in discovering new functional RNAs by \textit{in vitro} systems. The first examples of such work were reported in 1990 from the laboratories of Larry Gold and Jack Szostak (Ellington and Szostak 1990; Tuerk and Gold 1990). Their chief interest of their research lay in identifying RNAs that can specifically bind to ligands of interest, and they have independently developed a technique known as the Systematic
Evolution of Ligand by Exponential Enrichment (or SELEX) to isolate single stranded nucleic acid sequences with unmodified nucleotides that could perform this very function. Such sequences, termed aptamers, fold into complex structures characterized by the presence of stems, loops, hairpins, bulges, pseudoknots, triplexes or quadruplexes (Stoltenburg, Reinemann et al. 2007). The binding affinity of an aptamer to its target is primarily dependent upon its folding and the positioning of its nucleic acid components in the active site, resulting in interactions such as aromatic ring stacking, hydrogen bonding, and van der Waals interactions with its ligand.

*In vitro* selections, however, are not limited to the isolation of ligand binders, but can also be used in the discovery of nucleic acid sequences that can catalyze novel biochemical reactions. In our discussion we will focus mainly on the selection of RNAs, although DNA aptamers or catalytic DNAs can also be isolated using similar approaches (Figure 1-13) (Wilson and Szostak 1999). The first step of an *in vitro* selection experiment typically involves the construction of high diversity DNA pools containing chemically synthesized DNA sequences with a varying number of random or mutagenized nucleotides. These sequences are flanked with constant primer binding sequences on both ends, with an additional T7 promoter sequence at the 5’ terminus. The DNA pools are then amplified for several cycles such that there are multiple copies of each unique sequence, followed by transcription to RNA. The nucleic acid sequences are then tested for activity by incubating with either a particular ligand (aptamer selection) or with
Figure 1-13: General scheme for typical *in vitro* selections.

Starting with a nucleic acid pool of large diversity (typically greater than $10^{15}$ different sequences), the sequences are first incubated with specific ligands or substrates. Inactive sequences are then removed from the pool through stringent selection steps, resulting in the amplification and enrichment of active sequences. The selection process is repeated until the active sequences dominate the pool and can be readily detectable by biochemical assays.

substrates involved in a reaction of interest (ribozyme selection). Then, using specific stringent selection methodology, the active sequences are partitioned away from the inactive species. The sequences of interest are then reverse transcribed back into DNA, enriched by PCR amplification, and the rounds of selection are repeated until they become the dominant species in the pool. Both selections for aptamers and catalytic nucleic acids are very time consuming, and involve a degree of high risk (there is a chance that no active sequences are
present in the initial pool), but it is currently the most powerful and direct
technique in the isolation of novel functional nucleic acids. In the following
sections, we will be describing in detail the major steps involved in a typical in
vitro selection, as well as examples of catalytic RNAs that have been
successfully isolated using this approach.

### 1.4.1 Construction of Sequence Pools for Selection

The very first step in an in vitro selection is the generation of an input RNA
pool, which typically contains a large number of different sequences ranging from
$10^{15}$ to $10^{16}$. This is made possible with the innovation of the automated DNA
synthesizer, which can chemically synthesize DNA chains containing regions of
completely randomized sequences (Zaher and Unrau 2005). This high diversity is
necessary because the more different sequences there are, the more likely that
the sequences of interest will be present in the pool. Although the range of $10^{15}$
to $10^{16}$ is seemingly a very high number, it will only represent all the possible
sequences for a pool containing ~25-nt long sequences. This can be calculated
by reasoning that for any particular position in a sequence, there can only be 4
possible nucleotide options, A, G, C, T (or U for RNA pool). If a sequence
contains 25 possible random positions, then a maximum of $4^{25}$ number of
different sequences are possible. For random pools containing long sequences
(100 random positions, for example), the chances of finding a particular
sequence in a $10^{15}$ diversity pool would be incredibly small ($\sim10^{-46}$, from $10^{15} / 
4^{100}$). Therefore, the most important parameter in a long sequence pool is not the
amount of sequence space that can be sampled, but instead the probability of

43
finding functional sequences within the restricted $10^{15}$ to $10^{16}$ sequences (Wilson and Szostak 1999). This fine balance between obtaining high sequence diversity and probability of having interesting sequences initially in the pool must be carefully considered in deciding on an optimal length for an oligonucleotide library.

In an aptamer selection, another important factor to consider would be the properties of the ligand, such as its size, charge, hydrophobicity and atom composition (Stoltenburg, Reinemann et al. 2007). Similarly for a ribozyme selection, features of the substrate, such as the presence of hydrophobic or hydrophilic side groups is considered, but just as importantly is the complexity of the reaction to be catalyzed. If the designed pool is too short, it may not contain enough sequences to form functional motifs required for catalysis. The complex secondary structure of the Round 18 polymerase ribozyme, for example, would never have been discovered in a sequence pool of 20 to 80-nts long (Johnston, Unrau et al. 2001). Using very long random sequences is certainly possible, but often these are more costly, technically more difficult to synthesize, and are less manageable in selections relative to shorter sequence libraries. Typically, for an aptamer selection, a random pool containing ~20 to 80 random nucleotides are used (Stoltenburg, Reinemann et al. 2007). For ribozyme selections, on the other hand, researchers often prefer to start with a longer sequence pool because it gives the library greater structural complexity and thus enhances the chance of discovering functional species (Wilson and Szostak 1999).
In an effort to establish the balance between structural complexity and sequence length, a bioinformatics study from the Schlick laboratory showed that the optimal sequence length for obtaining a structure with $x$ stems is $20x$ (i.e. a 100-nt sequence will most likely have 5 stems in its structure) (Gevertz, Gan et al. 2005). Interestingly, they also reported their results indicating that random sequences do not have a uniform distribution of RNA secondary topologies, likely due to the thermodynamics of RNA folds. In random pools containing 100-nts or less, for example, the structures formed were heavily favoured towards simple topologies, such as linear stem-loop and low branching motifs. This may explain why complex structural motifs are rare in many in vitro selected aptamers or ribozymes that originated from short random pools. With increasing length, the population of branched motifs increases while that of unbranched structures decreases. Therefore, longer RNA sequence pools may be necessary for generating higher topological structures, which are often required for ribozymes catalyzing chemistries that are more complex. However, it is likely that not all of the sequence found to be present in a functional RNA is critical for its function. Therefore, after the successful isolation of catalytic RNAs, deletion re-selections are often performed to deduce the actual sequence regions required for catalysis (Wilson and Szostak 1999; Chapple, Bartel et al. 2003; Wang and Unrau 2005). An example of such a re-selection will be described in my work in Chapter 3, which involves the deduction of the minimal core motifs required for purine nucleotide synthase ribozymes.
1.4.2 The Selection Step

The most critical part of an in vitro evolution scheme is the separation of inactive from active sequences. The selection step must be designed such that it will guarantee that the few reactive sequences present in the pool will survive and be amplified, yet be stringent enough to allow the separation away from inactive species. Conditions in the selection step, such as buffer pH, and the time required for the selection process, must all be taken into account. Low stringency selections will require more selection rounds, resulting in more time and cost (Wilson and Szostak 1999; Stoltenburg, Reinemann et al. 2007). Another important point of consideration is whether the selection scheme lends itself to the amplification of sequences that fulfil the selection requirements, yet do not perform the desired function. These sequences will thus become “parasites” and over time will dominate the selected pool over the true reactive species.

1.4.2.1 Aptamer Selection Strategies

In aptamer selections, the most common selection technique is separation by affinity column chromatography. Ligands of interest are typically immobilized on particular column matrix material, such as sepharose or agarose, and subsequently incubated with high diversity nucleic acid sequences to allow binding with immobilized targets. The column is then washed several times with low salt buffers to remove any non-specific binding sequences. To recover the active sequences from the column, typically compounds with high affinity to the ligand are added such that they will outcompete the aptamers for binding. Alternatively, higher concentrations of free ligand can be added, which similarly
competes with the immobilized targets for binding with the aptamers. Examples of affinity column selections include the isolation of nucleic acid aptamers for zinc finger proteins, thyroxine hormone, tetracycline, and S-adenosyl-methionine (Burke and Gold 1997; Berens, Thain et al. 2001; Liu and Stormo 2005; Levesque, Beaudoin et al. 2007).

1.4.2.2 Ribozyme Selection Strategies

The selection methods for ribozyme selection are much more diverse, as it depends heavily on the biochemical reaction for which the nucleic acid pools are selected to catalyze. One common technique, which is similar to the affinity column selections described previously for aptamer selection, involves the selection of RNAs for their ability to react with a molecule containing a tag, which can then be immobilized on a column matrix. An example of such a selection strategy would be the isolation of polynucleotide kinase ribozymes, in which a random RNA sequence pool was incubated with the substrate $\gamma$-thio-ATP (Lorsch and Szostak 1994). The reactive RNAs, with the transferred thiophosphate from $\gamma$-thio-ATP, could then be selected away from inactive sequences by the formation of disulfide bonds with thiopyridine-activated thiopropyl Sepharose beads (Lorsch and Szostak 1994).

Another common technique used in ribozyme selection involves gel shift assays in which a tag is incorporated into the reactive RNAs that have successfully catalyzed the reaction of interest. Incorporation of this tag results in a supershift and allows the separation from inactive species (unshifted material) when subjected to gel electrophoresis, thus allowing stringent isolation of
reactive RNAs. A classic example of such gel shift assays that is widely used for selection of DNAzymes and ribozymes involves the use of streptavidin and biotin. Streptavidin is a large tetrameric protein with a very high binding affinity (\(K_d\) of \(~10^{-15}\) mol/L) to biotin, which is one of the strongest known non-covalent interactions (Michael Green 1990). By incorporating active RNAs with a biotin tag (usually by using biotin- labelled substrates) and upon subsequent incubation with streptavidin, functional sequences can be stringently isolated from inactive ones upon separation through a polyacrylamide gel. This is due to the large size of streptavidin, which forms a complex with biotin- labelled RNAs and retards their migration through the gel, resulting in a supershift. Ribozymes that have been successfully discovered using this technique include the self-cleaving and substrate ligating ribozymes, the Diels-Alderase ribozyme, and the ribozymes catalyzing the decarboxylative Claisen condensation (Jaeger, Wright et al. 1999; Seelig and Jäschke 1999; Ryu, Kim et al. 2006). Another example of the gel shift technique exploits the interaction between sulphur and mercury. Compounds containing thiol groups interact strongly with mercury, and any RNAs with a thiol-tag would similarly result in a supershift when separated through a mercury containing gel (Igloi 1988). This strategy can be employed to isolate single-turnover ribozymes that would have added a thiol-tagged substrate to themselves. Examples of ribozymes isolated using this approach includes 2' autophosphorylating 2PTmin3.2 ribozyme (which use the \(\gamma\)-ATP substrate) (Saran, Held et al. 2006), the pyrimidine nucleotide synthase ribozymes (Unrau
and Bartel 1998), and the purine nucleotide synthase ribozymes (Lau, Cadieux et al. 2004); the latter will be described in later chapters.

Functional RNAs that are successfully isolated from the selection step are subsequently reverse transcribed and amplified for sequence enrichment. Commonly, pool sequences are designed such that their sequences have constant primer regions built into at their 5’ and 3’ termini, which allow for easy amplification by PCR (Wilson and Szostak 1999). In the *in vitro* selection work by Bartel and Szostak, they have cleverly taken advantage of this amplification step by designing the selection such that the only sequences that can be amplified are the catalytic sequences that can ligate a primer sequence at their 5’ terminus (Bartel and Szostak 1993). A similar strategy was later used in isolating additional families of ligase ribozymes (Ekland, Szostak et al. 1995). As with all selection schemes, after enrichment, the aptamers or catalytic sequences will have completed one selection cycle, and are henceforth ready for reinsertion into the next selection round. The entire selection process is continuously repeated until catalytic RNAs performing the desired function dominate the pool and are readily detectable.

1.4.3 RNA Evolution by *In Vitro* Selection

1.4.3.1 Evolution from a Biased Pool

Different *in vitro* selection methods can be employed for evolving a ribozyme that catalyzes a particular reaction to one that mediates a similar or completely different chemistry. One of such methods would be the use of a biased pool, in which certain regions of the pool sequences, as designed by the
researcher, will be mutagenized at a low frequency relative to a previously isolated ribozyme. The sequences generated by this approach are biased to conserve sequence elements believed to be critical for either the structural stability or functionality of the ribozyme. An example would be the evolution of a ligase ribozyme, which originated from a biased pool based on a scaffold region of the group I intron from *Tetrahymena* (Jaeger, Wright et al. 1999). To examine the likelihood of ribozymes arising from pre-existing ribozymes that perform completely new functions, Curtis and Bartel performed an *in vitro* selection in which they attempted to evolve kinase ribozymes from an aminoacylase ribozyme (Curtis and Bartel 2005). In their selection, they generated a mutagenized pool containing a mutation rate of 11% at 65 of the 90 possible positions, and observed the evolution of 23 new distinct kinase ribozymes with new folds, even from within such short mutational distances. Upon further sequence analysis of the starting mutagenized pool, they noticed that the probability of finding kinase ribozymes increased significantly with an increasing number of mutations relative to the parent. This suggests that there is a need of escape from the parental structure, as evident from the new folds observed in the isolated kinase ribozymes, in order for the evolution of new function to occur. Mutagenized pools, however, do not always offer an advantage over randomized pools, as RNA catalysts selected for a particular function may not always share sequence or structural similarities relative to the sequences from which they originated from. Such was the case in our selection for purine nucleotide synthase ribozymes, in which we observed no beneficiary effect from a selection
starting with a mutagenized pool relative to that with a randomized pool (Lau, Cadieux et al. 2004).

1.4.3.2 Evolution by Engineering of Functional RNAs

A slight variation to the mutagenized pool approach is the incorporation of an aptamer binding motif as part of the pool design. This was demonstrated in work by Lorsch and Szostak, in which they successfully isolated ribozymes that can catalyze self-phosphorylation of their 5′ terminus using \( \gamma \)-thio-ATP (Lorsch and Szostak 1994). These ribozymes were evolved from a pool containing a mutagenized ATP binding aptamer motif flanked with random nucleotides. Since substrate binding is an important requirement for most catalytic reactions, the presence of a sequence region that is biased towards substrate recognition may therefore favour the selection. In a further extension of this approach, Kumar and Joyce in 2003 reported their work in which they attempted to select for ribozymes catalyzing successive reactions by conjoining a self-cleaving ribozyme (Group I ribozyme) and a ligase ribozyme (R3C ribozyme) together (Kumar and Joyce 2003). While simply attaching the ribozymes together did not result achieve the desired outcome, likely due to inefficient substrate transfer between the two motifs, they were successful in evolving bifunctional ribozymes by introducing mutations and performing further rounds of selection.

1.4.3.3 Evolution by Random Mutagenesis

Another approach for evolving RNAs is by error-prone PCR, which incorporates random point mutations throughout a DNA sequence. This
technique takes advantage of the fact that the TAQ polymerase, isolated from bacterium *Thermus aquaticus*, has a high rate of nucleotide mis-incorporation during primer extension under high magnesium concentrations (Cadwell and Joyce 1992). The frequency of mutagenesis, however, is very low, with a 0.66% ± 0.13% chance of mis-incorporation per position over the course of 30 PCR cycles. An example of a selection employing this strategy would be the evolution of the dual function ribozyme described above, in which the starting pool was generated from the incorporation of random mutations to the conjoined catalytic motifs (Kumar and Joyce 2003). Mutagenic PCR was also performed at the end of each successive round after the first round to introduce further diversity, resulting in the eventual successful isolation of the bifunctional self-cleaving / ligase ribozymes.

1.4.3.4 Evolution of Natural Ribozymes

While a large number of RNAs catalyzing a wide range of chemistries have been evolved by *in vitro* selection, there are only very few known natural ribozymes catalyzing a limited number of reactions, such as phosphodiester cleavage, ligation and peptide bond formation. So why are natural ribozymes so limited in numbers, and how are these catalytic activities evolved in nature? The answer to the first question may simply be the fact that some of the reactions catalyzed by *in vitro* selected ribozymes are not required for cells, or that protein enzymes performing the same functions are much more efficient than their RNA counterpart. While many of the natural ribozymes may have been remnants of the RNA world, they all serve important functions in different modern cellular
Figure 1-14: Catalytic strategies employed by two different natural ribozymes.

a. The peptidyl transferase center (active site) of *H. Marismortui*. Contacts made by 23S ribosomal RNA nucleotides are in green, water molecule in black, P site tRNA and its 3' tethered peptide chain (Pept) in blue, and the A site tRNA and its tethered 5' amino acid in red. Numbers indicate nucleotide position on the respective RNA and dotted lines indicate possible interactions that are within hydrogen bonding distance. The groups involved in hydrogen bonding are indicated following the nucleotide and position number (Reprinted from [Beringer, M. and M. V. Rodnina (2007). “The Ribosomal Peptidyl Transferase.” *Molecular Cell* 26(3): 311-321] with permission from Elsevier).

b. The active site of the *glmS* ribozyme. Cleavage occurs at the phosphodiester bond between an adenosine A(-1) and the immediate 3' guanosine nucleotide G1. Glucosamine-6-phosphate is shown at the bottom, and has been suggested to assist in catalysis by extracting a proton from the nucleophilic 2' OH of A(-1) through 2 water molecules. A guanosine at the 40th position (G40) has also been proposed as a possible candidate for deprotonating the 2' OH (From [Klein, D. J. and A. R. Ferre-D'Amare (2006). "Structural Basis of *glmS* Ribozyme Activation by Glucosamine-6-Phosphate." *Science* 313(5794): 1752-1756]. Reprinted with permission from AAAS).

pathways. The ribosomal RNAs, for example, are essential for the synthesis of proteins (Nissen, Hansen et al. 2000; Beringer and Rodnina 2007), while other ribozymes such as the group I intron (Cech, Zaug et al. 1981) and the *glmS* ribozyme (Klein and Ferre-D'Amare 2006), are required for mRNA splicing and the regulation of cellular metabolites, respectively. To address how nature
evolves ribozymatic activity, various research groups have performed structural studies on naturally known ribozymes and examined in detail their catalytic mechanisms. The ribosome, for example, has been suggested to catalyze peptide bond synthesis by a substrate-assisted mechanism (Weinger, Parnell et al. 2004). The rRNAs position the tRNAs such that the 2' hydroxyl group of the 3' terminal adenosine of the P site tRNA is in position and may act as a base to abstract a proton from the incoming nucleophilic amino group (Figure 1-14a) (Beringer and Rodnina 2007). Alternatively, it is also possible that the same 2' hydroxyl group can act a general acid and donate its proton to the adjacent 3' oxygen to promote the release of the P site tRNA from the tethered amino acid (Weinger, Parnell et al. 2004). Similarly, the GlmS ribozyme has been suggested to rely on its cognate ligand glucosamine 6-phosphate (GlcN-6-P) to mediate self-cleavage. GlcN-6-P is positioned in the binding site such that the amino group at the 2' position can serve as a base and deprotonate the 2' hydroxyl nucleophile of the transesterification through two bound water molecules (Figure 1-14b) (Klein and Ferre-D'Amare 2006). Since RNAs have a limited catalytic repertoire as compared to proteins, nature may have evolved RNAs to depend on such substrate-assisted mechanisms as the most efficient catalytic approach. This is in contrast to protein enzymes, which have evolved with more complexity and may have substituted the catalytic roles previously supplied by substrates.

In all the examples described above, the binding and positioning of the substrates are critical for catalysis; thus it may be that many RNA aptamers do have intrinsic catalytic potential, but only upon binding of specific substrates. To
illustrate this idea, Brackett and Dieckmann in 2006 reported that a simple 38 nucleotide RNA aptamer targeting the triphenyl dye, malachite green, could be evolved to a catalytic ribozyme without changing the sequence of the RNA (Brackett and Dieckmann 2006). In their work, they manipulated only the ligand such that the hydroxyl group of the phenyl ring was replaced with an acetate ester. Based on the solution structure, the carbonyl carbon was positioned in a strongly negative electrostatic environment, thus favouring carbocation formation in the transition state. From their work, they successfully demonstrated that the 38-nt RNA can similarly bind to the malachite green analogue, and subsequently catalyze the hydrolysis of the ester bond. While the catalytic rate enhancement was only 8.5 fold over that of background, it was interesting that the ribozyme switched the chemistry from what is a dominant base driven hydrolysis pathway (i.e. in the absence of ribozyme) to an acid driven reaction. In comparison to an acidic pH (pH 5), the acid dependent hydrolysis is in fact 1000 fold faster than the uncatalyzed acid driven hydrolysis. This work demonstrated that transition state stabilization by electrostatic forces inside binding pockets may be sufficient to drive RNA mediated catalysis. Therefore, it is possible for an RNA aptamer or catalyst to evolve new catalytic functions simply by coming across the right substrate. This may in fact be how many of the natural ribozymes have evolved their current functions, leading to the phenomenon of ribozyme promiscuity, as will be discussed in Chapter 4.
CHAPTER 2: *IN VITRO* SELECTION OF FAST PURINE NUCLEOTIDE SYNTHASE RIBOZYMES


2.1 Introduction

The RNA World hypothesis suggests that RNA predates protein in evolution (Bartel and Unrau 1999; Joyce 2002). While this parsimonious model is increasingly consistent with our detailed knowledge of metabolism, it is still unclear if ribozymes can manipulate the small substrates required for an RNA-based metabolism. Specifically, RNA replication requires nucleotide monomers that are in turn synthesized from simpler compounds. Modern metabolism uses PRPP in at least 16 different pathways to synthesize pyrimidine nucleotides (OMP, UMP, Fig. 2-1a), purine nucleotides (AMP, GMP, IMP, XMP, Fig. 2-1b), several pyridine nucleotide cofactors and the amino acids histidine and tryptophan (Mathews 1996). As purine and pyrimidine bases are known to be synthesized by prebiotic processes (Joyce 1989; Orgel 1998), a hypothetical RNA-based metabolism could have used purine and pyrimidine bases together with PRPP to synthesize the nucleotide building blocks required to replicate RNA.

The ability of a ribozyme to efficiently perform small molecular chemistry requires both substrate recognition and rate acceleration. At one extreme, natural
and artificial RNA sequences that bind, but do not react with, small metabolically relevant substrates demonstrate that RNA can recognize small molecules with nanomolar affinities and high specificity (Osborne and Ellington 1997; Wilson and Szostak 1999; Winkler, Nahvi et al. 2004). At the other end of the spectrum, ribozymes that recognize their substrates through Watson-Crick base pairs demonstrate that RNA can perform efficient catalysis when substrate recognition is performed tens of angstroms away from the site of chemical modification (Wilson and Szostak 1999). Decreasing the substrate size forces a catalytic system to perform both recognition and rate acceleration within an increasingly small region of space. Observing the catalytic strategies employed by ribozymes performing similar chemistry on a succession of small substrates may therefore
provide clues as to how RNA compromises between these two important catalytic parameters.

Previously, Unrau and Bartel have successfully isolated three families of ribozymes able to perform pyrimidine nucleotide synthesis using tethered PRPP and free 4-thiouracil ($^{4S}\text{Ura}$) (Unrau and Bartel 1998). These ribozymes displayed high specificity for their $^{4S}\text{Ura}$ substrate and had apparent efficiencies ($k_{\text{cat app}}/K_m$) in the 0.7 to 7 M$^{-1}$ min$^{-1}$ range. One family was optimized and shown to perform rate limiting and highly dissociative chemistry likely to involve the charge mediated stabilization of an oxocarbenium-ion intermediate (Unrau and Bartel 1998; Chapple, Bartel et al. 2003). As this small molecule chemistry differs considerably from other known ribozyme reactions in both mechanism and substrate size, an exploration of purine nucleotide synthesis will provide an important context to our initial nucleotide synthesis studies. In this chapter I will describe the isolation of purine nucleotide synthases with a range of binding affinities and having efficiencies dramatically higher than observed for pyrimidine nucleotide synthase ribozymes.

2.2 Methods

2.2.1 APPRPP and APP$^{6S}\text{G}$ Synthesis

The synthesis of adenylated PRPP (APPRPP) was as previously described (Unrau and Bartel 1998). Adenylated 6-thioguanosine 5'-monophosphate (APP$^{6S}\text{G}$) was synthesized by reacting 100 mM $^{6S}\text{GMP}$ with 50 mM adenosine 5'-phosphorimidazolide (Lohrmann and Orgel 1978) in the presence of 100 mM MgCl$_2$ and 200 mM HEPES (pH 7.4) at 50 °C for 3 hours.
followed by HPLC purification (Unrau and Bartel 1998). $^{68}$GMP was synthesized from 6-thioguanosine (Sigma) according to the method of Breter and Mertes (Breter and Mertes 1990) and purified on a A-25 DEAE Sephadex column (Sigma) using a 50 mM to 400 mM triethylammonium bicarbonate (TEAB) pH 7.5 gradient; product eluted at 250-300 mM TEAB.

2.2.2 Designing of Selection Pools

Two DNA pools were synthesized on an ABI 392 DNA synthesizer using 0.2 μmol 2000 Å control pore glass (CPG) columns (Glen Research) using standard cyanoethyl phosphoramidite chemistry. The first random pool had the final sequence 5'-TTCTAATACGACTCATACTAGAGAGACGCCATCAA-N-95-TGCACCGCAGCAAGC (-N-95-, 95 random nucleotides; first 34-nt, 5' primer; last 16-nt, 3' primer; T7 promoter sequence in italics). The second biased pool was created by mixing together equal amounts of two mutagenized subpools each having the general sequence: 5' TTCTAATACGACTCATA(GGAG)CGAAGTGCCCA-N11=(atcc)N(1-3) (gcctatt)N18=(aacga)N(4-6)=(gctgc)N10=(aataggc)N7-9=(gcggtg)N10=-tTCG(CACCGC)A(GCAAGC) (first 34 nt, 5' primer; last 16 nt, 3' primer; N, random nucleotide; lower case, mutagenized either 10% or 20% for each subpool; brackets delimit the five helical components found in the secondary structure of the family A nucleotide synthase ribozyme and are named with Roman numeral subscripts) (Chapple, Bartel et al. 2003). Variable length random regions were synthesized using a split and pool strategy (Zaher and Unrau 2004). Random and mutagenized nucleotide couplings in both pools were
obtained by mixing equiactive phosphoramidite (Applied Biosytems) stocks (dA:dC:dG:dT in the molar ratios 0.28:0.27:0.23:0.22) as described (Zaher and Unrau 2004). Sequencing of random positions revealed the following relative nucleotide frequencies: 25.8% A (92/356), 21.1% G (75/356), 28.7% C (102/356) and 24.4% T (87/356). Large scale PCR was performed as described (Zaher and Unrau 2004) using the following primers: random pool 5’-

TTCTAATACGACTCCTAGTAGGAGACGCCCC and for the biased pool 5’- TTCTAATACGACTCCTAGTAGGAGCGACCAGGTCCC. The 3’ primer was 5’- GCTTGCTGGGTCCGA and was common to both pools. A total of ~2 nmol of DNA from each of the random and mutagenized pools (~3 x 10^{14} unique sequences, given that 4 copies of each unique sequence was expected to be present after PCR amplification) were transcribed into RNA using T7 RNA polymerase (Chapple, Bartel et al. 2003).

### 2.2.3 RNA Ligation Protocol

RNA at 3.33 μM was ligated with 60 μM APPRPP for 4 hours at 23 °C in ligation buffer (50 mM HEPES, 10 mM MgCl₂, 3.3 mM dithiothreitol (DTT), 10 μg/mL BSA, 8.3% v/v glycerol and 15% polyethylene glycol 8000 at pH 8.0) using 0.5 U/μL T4 RNA ligase (Wang and Unrau 2002). The reaction was terminated by addition of EDTA, followed by a phenol-chloroform extraction and ethanol precipitation. Marker RNA 125-nt long and having the 3’ terminal nucleotide sequence of UCAGAAGAAGUCACAUUGC-3’ was derivatized to contain a terminal ^6S G by performing a ligation using 33 μM APP^6S G for 4 hours. Marker
RNA was gel purified through an N-acryloylaminophenyl-mercuric acetate (APM) gel (Igloi 1988) for use in Rounds 1 to 3.

2.2.4 Selection

Pool RNA ligated to PRPP was incubated at 0.24 μM concentration in incubation buffer (50 mM Tris-HCl, 150 mM KCl, 25 mM MgCl₂, pH 7.5) supplemented with 0.26 mM 6S Gua for 15 hours (Round 1-5). Five nanomoles of RNA were used in the first round of selection for each pool (~10 copies of each pool sequence). Ribozyme reactions were stopped by adding one volume of gel loading buffer (90% formamide and 50 mM EDTA). Reactive RNA were separated from non-reactive species using denaturing 6% PAGE gels containing 3.75 μM APM. During Rounds 1-5, the RNA for each incubation was divided in two. One half contained radiolabelled RNA pool to allow detection of reactive ribozymes. The second half contained unlabelled RNA pool mixed with the synthetic marker RNA, which served as an internal control to locate the position of the reactive species in the APM gel. Gel fragments at the position of the marker were excised and eluted (300 mM NaCl, 1 mM DTT) overnight, and recovered by ethanol precipitation. The resulting RNA was reverse transcribed using Superscript II (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 560 μM of each dNTP, 5 μM 3' primer, 10 U/μL enzyme) at 48 °C for 1 hour. RNA was hydrolyzed with 100 mM KOH at 90 °C for 10 minutes. cDNA was neutralized with HCl and PCR amplified before re-entering the next round of selection. From Rounds 6 to 10, the RNA pools were subjected to increasing
time pressure by lowering of incubation time in each round (4 hr, 1 hr, 6 min, 1 min, and 15 seconds respectively).

2.2.5 TLC Analysis

Gel purified ribozyme isolates were ligated with PRPP and then reacted to completion with $^{6S}$Gua (15 hour incubation). Reacted RNA was derivatized with radiolabelled cytidine 5'-[$^{32}$P], 3' bisphosphate ($^*$pCp, $^*$ indicates $^{32}$P labelling) using the ligation protocol above in the presence of 2 $\mu$M ATP. $^*$pCp was synthesized by phosphorylating 3'-cytidine monophosphate (Sigma) with $\gamma$-$^{32}$P ATP using T4 polynucleotide kinase (NEB). The RNA radiolabelled at its 3' end was gel purified using a 6% APM gel in order to isolate the thiol-containing material. The recovered RNA was digested into 3'-mononucleotides using T2 ribonuclease (25 mM sodium citrate, 4 mM DTT, pH 4.5, 0.26 U/$\mu$L T2 (Sigma), for 2 hours at 37 °C). A 6-thioguanosine 3'-mononucleotide ($^{6S}$Gp) control was synthesized by ligation of the marker RNA (derivatized with 3' p$^{6S}$G) with $^*$pCp. This sample was APM gel purified and digested with T2 ribonuclease as described above. Digested mononucleotides were separated using two dimensional thin layer chromatography on 10 cm × 10 cm cellulose TLC plates (J.T. Baker) pre-soaked in 1:10 saturated (NH$_4$)$_2$SO$_4$:H$_2$O. The first dimension was developed with: 80% ethanol, second with: 40:1 saturated (NH$_4$)$_2$SO$_4$:isopropanol, both solvents were supplemented with 100 mM $\beta$-mercaptoethanol (Gray 1974; Unrau and Bartel 1998). Samples were spotted 1 cm in from the corner of the TLC plates.
2.2.6 Kinetic Analysis

Kinetic studies were performed in incubation buffer supplemented with MgCl₂ to a final concentration of 75 mM. Time points (6, 30, 60, 150 and 240 min) were taken and stopped by addition of an equal volume of gel loading buffer. The reaction rate at a given ⁶S-Gua concentration was determined by simultaneously fitting the fraction reacted for at least five independent time courses (determined by phosphorimager analysis on a Molecular Dynamics Storm 820 of the resulting gels) to the equation $F = \beta(1-e^{-k_{obs}t})$ using the program KaleidaGraph (Synergy Software). $F$ being the fraction reacted at time $t$, $k_{obs}$ the apparent first order rate constant and $\beta$ the total fraction able to react. The resulting rates were fit to the Michaelis-Menten equation $k_{obs} = k_{cat\, app}\frac{[{}^{6}\text{S}\text{-Gua}]}{(K_m + [{}^{6}\text{S}\text{-Gua}])}$. A weighted error analysis was performed to obtain $k_{cat\, app}$ and $K_m$.

2.2.7 Analogues

Purine analogues 6-thioguanine $\varepsilon(347 \text{ nm, pH 1}) = 20900 \text{ M}^{-1} \text{ cm}^{-1}$ (Elion and Hitchings 1955), 6-thiopurine $\varepsilon(325 \text{ nm, pH 1}) = 20500 \text{ M}^{-1} \text{ cm}^{-1}$ (Elion 1962), 2-methyl-6-thiopurine assumed: $\varepsilon(330 \text{ nm, pH 1}) = 20000 \text{ M}^{-1} \text{ cm}^{-1}$, 6,8-dithiopurine $\varepsilon(358 \text{ nm, pH 1}) = 27800 \text{ M}^{-1} \text{ cm}^{-1}$ (Robins 1959), 2,6-dithiopurine, 6-thio-9-methylpurine, 6-hydroxy-2-thiopurine, 2-hydroxy-6-thiopurine, 2-thiopurine, 2-amino-9-butyl-6-thiopurine, 2-methylthio-6-thiopurine, 8-methyl-6-thiopurine, 2,6-dithio-7-methylpurine, 6,8-dithio-2-hydroxypurine, 2,6,8-trithiopurine as well as 4-thiouracil were obtained from Sigma-Aldrich and saturated solutions were dissolved in $1.05 \times$ incubation buffer.
2.3 Results

2.3.1 Selection

We were curious to understand the effect a secondary structure bias might have on the outcome of a selection for purine nucleotide synthesis. Could a secondary structure motif previously selected for its ability to perform pyrimidine nucleotide synthesis be beneficial to purine nucleotide synthesis? To address this question we constructed two pools which were subjected to \textit{in vitro} selection in identical fashion. The first consisted of a 95-nt long random sequence pool having a diversity of $\sim 3 \times 10^{14}$ different sequences spread uniformly throughout sequence space. Our second pool, having the same number of different sequences, approximate length (92-98 nt of variable sequence), and an identical 3' primer binding sequence as the first, also contained significant amounts of random sequence (50-56 nt) but was on average only 4-8 mutations away from being able to form the complete secondary structure of a structurally complex pyrimidine nucleotide synthase ribozyme (Figure 2-2) (Chapple, Bartel et al. 2003).

RNA sequences from both random and structurally biased pools were selected for their ability to promote glycosidic bond formation between the 1-pyrophosphorylated ribose at their 3' ends and a free $^{6S}$Gua substrate using an APM gel shift strategy (Figure 2-3) (Unrau and Bartel 1998). The isolation of RNA containing a thiol group was possible due to the slowing of sulphur containing material in a mercury containing gel (Igloi 1988). After 6 rounds of \textit{in
**Figure 2-2: Template design of the mutagenic pool.**

Coloured boxes indicate hypothesized stem and pseudoknot structures formed in the original 7.02 pyrimidine nucleotide synthase ribozyme (total of 5 helical stems). Constant 5’ and 3’ primers are indicated in green and random nucleotides (r-nt) in black. Other colours represent nucleotides mutagenized at 10% or 20% frequency.

*vitro* selection, $^{6S}$Gua dependent ribozymes having the same mobility as a RNA-p$^{6S}$G marker on an APM gel were observed in both pools after 15 hours of incubation. At this point selection pressure was increased for both pools by lowering the incubation time allowed with the $^{6S}$Gua substrate (see methods). By round 10, nearly 1% of each pool had reacted after 15 seconds of incubation.

Both pools of $^{6S}$G synthases had reaction rates (given by the ribozyme pools reactivity per unit substrate concentration), that were 50 to 100 times higher than the efficiency observed for an equivalent pool of pyrimidine nucleotide synthases (Figure 2-4). The $^{6S}$G synthases of both pools seemed to have reached
Figure 2-3: Overall *in vitro* selection scheme for purine nucleotide synthase ribozymes.

**a.** An RNA pool is derivatized with PRPP using APPRPP and T4 RNA ligase. **b.** RNA pools tethered to PRPP are then incubated with $\text{^{85}S}\text{Gua}$. Ribozymes catalyzing the synthesis of $\text{^{85}S}\text{G}$ from $\text{^{85}S}\text{Gua}$ release pyrophosphate (PP$_i$). **c.** An APM gel was used to isolate active from non-reactive RNA (left lane). A radiolabelled marker RNA derivatized with a terminal $\text{^{85}S}\text{G}$ nucleotide was mixed with pool RNA and used to indicate the position of active ribozymes (right lane). **d.** Reactive ribozymes are excised from the gel and eluted. **e.** The enriched RNA pool is then reverse transcribed into DNA, PCR amplified and transcribed back into RNA ready to enter the next round of selection.
Figure 2-4: Purine ribozyme activity as a function of selection round.

The random sequence pool (empty circles) and the biased pool resulting from mutagenizing the family A pyrimidine nucleotide synthase ribozyme (filled circles) have similar purine synthase activities during the course of the selection. Incubation times were initially 15 hours and by Round 10 pools were incubated for only 15 seconds. In contrast to the purine nucleotide synthase pools, a previous selection for pyrimidine nucleotide synthases (solid squares) resulted in ribozyme populations that were 50-100 times less active (Unrau and Bartel 1998). Reaction rates were calculated by dividing the observed first order reaction rates by the substrate concentration.

maximum reaction rates by round 9, slightly faster than the previously selected

4S U synthases that required at least one more round of selection to plateau.
2.3.2 Sequencing Analysis

Sequencing round 10 of the structurally biased pool revealed at least 20 distinct (as judged by primary sequence alignment) families of which 6 contained multiple isolates (Figure 2-5a). One family was isolated eight times; this family was also found as a single isolate in round 6. One family was repeated three times, and four families were isolated twice each. These families are henceforth called MA to MF, where “M” stands for “mutagenized”. The majority of families (12/20) shared a short conserved UCUUU sequence motif (we accepted in this count one C to U containing isolate) that was not found in the family A motif, and 6 of these 12 families appeared to extend this motif by another six residues to AGGCGUUCUUU (refer to Supporting Information Figure 1a). The short motif was found in a random sequence region of the biased pool immediately 5’ to a well-conserved hairpin loop, which was found in 18/20 families. This hairpin, which forms stem V in the original family A RNA motif (see methods for sequence information), contains one arm that was not mutagenized in order to allow the efficient binding of a reverse transcription primer and thus would be expected to be conserved by chance 46% of the time. Only one isolate (MF) was found that hypothetically contained all five helices of the family A motif, although this isolate had a mismatch in the middle of stem III of the family A motif. This would be expected to occur ~2% of the time by chance; roughly consistent with the number of isolates found (1/20). Moreover, folding with the PKNOTs algorithm (Rivas and Eddy 1999) indicated that this sequence was likely to adopt a fold considerably different from that of the family A nucleotide synthase.
Figure 2: Alignments of round 10 isolates from biased and random sequence pools.

a. Families resulting from the biased pool. The sequences have been aligned with respect to the mutagenized helical elements, underlined in blue, of the family A pyrimidine nucleotide synthase ribozyme.

b. Families resulting from the random pool. Even though this pool had an identical 3' primer binding sequence, a common motif of the sort found in the biased pool is not observed. The bracketed number in each of the families resulting from the random pool reflects the number of repeats observed for that isolate. Constant primer regions are underlined in red.
The round 10 random sequence pool was found to contain at least 33 distinct families. In contrast to the biased pool, none of the families appeared to form a terminal hairpin loop and none contained the UCUUU motif near their 3’ end even though both pools contained exactly the same 3’ terminal sequence (Figure 2-4b). Three of the random families, named RA, RB and RC, occurred more than once (two were repeated three times and one was repeated twice out of 38 sequences total). One sequence from each of the three repeating families along with three sequences from the remaining 30 ‘orphaned’ families (called RD, RE and RF) were selected for further analysis. These named sequences together with the functional families resulting from the structurally biased pool have been submitted to GenBank and have accession numbers AY701990-AY702000.

2.3.3 Magnesium Dependence

Two families from each pool (MA & ME and RA & RE) were evaluated for their difference in reaction rate with $^{65}$Gua under varying magnesium level conditions. The reactivity of the isolates appeared to increase with magnesium concentration, with activity saturating or even decreasing above 75 mM MgCl$_2$. This optimal magnesium concentration was therefore used for all further kinetic assays on the remaining named families. The most reactive sequence (highest reaction rate and fraction reacted) from the random pool was a sequence from family RA, and for the mutagenized pool, it was from family MA (of all the tested isolates only the MD isolate was found to be unreactive). The initial reaction
velocities for the random pool families (RB-RF) were, on average, 1.2 to 3 times slower than RA, while the mutagenized pool families (MB-MF, excluding MD) displayed a 5-11 fold decrease relative to MA. Since the fastest families had the highest frequency of occurrence, the selection appears, as desired, to have isolated ribozymes based on their catalytic prowess.

2.3.4 Product Characterization

To characterize further the utilization of $^6$S Gua by the purine synthase ribozymes, RA and MA along with one of the less common isolates from each pool (RE and ME) were analyzed using thin layer chromatography. Ribozymes reacted with $^6$S Gua were labelled with 5'-$[^{32}\text{P}]$ pCp ($^p$Cp, * indicates $^{32}\text{P}$ radiolabel), so as to specifically tag the 3' most nucleotide (Figure 2-6a). Ribozymes containing a thiol modification were purified using an APM gel and digested into mononucleotides that have a 3' phosphate using T2 RNAse (Figure 2-6b, c). A marker RNA derivatized with $^6$S GMP at its 3' end was also labelled with $^p$Cp and used to generate a radiolabelled $^6$S Gp standard. Because the T2 digested standard, when treated with calf intestinal phosphatase (CIP), resulted in the production of radiolabelled inorganic phosphate and the disappearance of a shifted band on a high percentage APM polyacrylamide gel, we concluded that the nuclease was able to cleave the phosphodiester linkage joining the terminal $^6$S G and the $^p$Cp label. Radiolabelled standard $^6$S Gp was mixed with T2 digested RNA radiolabelled during transcription with $\alpha-[^{32}\text{P}]$ UTP. The resulting two-
Figure 2-6: Overall *in vitro* selection scheme for purine nucleotide synthase ribozymes.

Two-dimensional TLC analysis of the reaction products resulting from four different purine nucleotide synthase reactions. **a.** Reacted RNA or an RNA construct synthesized so as to have a terminal 3’ p65G is derivatized with *pCp (32P labelled phosphate denoted by asterisk) using T4 RNA ligase. **b.** The radiolabelled, thiol-containing material is purified on an APM gel. **c.** The recovered material is digested into mononucleotides using ribonuclease T2. **d.** TLC showing a mixture of known 3' radiolabelled mononucleotides. From left to right spots are: 65Gp, Gp, Ap, Up and Cp. **e.** Ribozyme isolates reacted with 65Gua and treated as described in panels **a** to **c** are shown in the left column. A mixture of ribozyme digests and marker mononucleotides are shown in the right column. The first axis is vertical. TLC origins are indicated by black open circles.
dimensional TLC standard revealed five spots, corresponding to Ap, Cp, Gp, Up and $^{\text{6S}}$Gp (Figure 2-6d). The $^{\text{6S}}$Gp spot was not apparent in digestion of unrelated radiolabelled RNA containing no 3' $^{\text{6S}}$G. Two-dimensional TLC showed a single radiolabelled spot for all four ribozymes treated with the same *pCp ligation procedure as used to generate the $^{\text{6S}}$Gp standard (Figure 2-6e). Mixing the ribozyme sample with the five nucleotide reference indicated that the ribozyme dependent spot co-migrated precisely with the $^{\text{6S}}$Gp standard and resulted in a spot ~2 times more radioactive than the initial reference marker. As this TLC system has previously been shown to resolve quite similar nucleotides (for example pseudouridine from uridine) (Gray 1974; Unrau and Bartel 1998), we are reasonably confident that an N-9 linkage is in fact being produced (note also 6-thio-9-methylpurine did not react detectably, see below).

**2.3.5 Kinetics**

Both RA and MA were subjected to more detailed kinetic analysis, since both isolates had the highest frequency of occurrence and initial velocity compared to other sequences found in both pools. $^{\text{6S}}$Gua, which is sparingly soluble in water, was titrated over its solubility range. First order rate constants were extracted as a function of $^{\text{6S}}$Gua by fitting simultaneously at least five independent time courses at a particular substrate concentration (see methods). The resulting rates were then fit to the Michaelis-Menten equation, revealing an apparent $K_m$ of 78 +/- 11 μM and a $k_{\text{cat, app}}$ of 0.018 +/- 0.007 min$^{-1}$ for the RA isolate (Fig. 2-7), with an apparent efficiency ($k_{\text{cat, app}}/K_m$) of 230 M$^{-1}$ min$^{-1}$. In
Figure 2-7: Catalytic rate ($k_{\text{obs}}$) for MA and RA as a function of $6^S$Gua concentration.

Ribozyme MA (indicated in filled circles) had kinetics which fit a straight line, indicating an apparent efficiency of 284 M$^{-1}$min$^{-1}$. Ribozyme RA (indicated in filled triangles) fit well to the Michaelis-Menten equation, which indicated a $K_m$ of 78 µM and $k_{\text{cat app}}$ of 0.02 min$^{-1}$. The linear behaviour of MA suggests that aggregation or some other form of general ribozyme inhibition is not responsible for the RA kinetics.

In contrast, the reaction rate of MA was directly proportional to $6^S$Gua concentration giving an apparent efficiency of 284 M$^{-1}$min$^{-1}$. The linear rate dependence with substrate concentration observed for MA makes it unlikely that the rate plateau
observed for RA at high $^{6S}$Gua concentrations was due to non-specific ribozyme inhibition such as aggregation. The uncatalyzed rate of purine nucleotide synthesis was undetectable when a short radiolabelled RNA (sequence 5'-AAC) derivatized with PRPP was incubated for as long as 8 days with $^{6S}$Gua as judged by APM gel shift (less than 5 parts in $10^4$ detected, uncatalyzed rate calculated to be $< 2 \times 10^{-4} \text{M}^{-1} \text{min}^{-1}$).

### 2.3.6 Substrate Specificity

The substrate preference of both MA and RA was examined using 15 different thiol-containing purine and pyrimidine compounds. Saturated solutions of each compound were prepared in standard incubation buffer and four compounds showed weak activity with either MA or RA as indicated by APM gel shift (6-thiopurine, 2-methyl-6-thiopurine, 2,6-dithiopurine and 6,8-dithiopurine). As many of these compounds were suspected to be contaminated with 6-thiopurine, they were HPLC purified on a C18 column. The reaction efficiency of these compounds was then compared to $^{6S}$Gua by incubating the purified material at a uniform concentration of 0.26 mM with the ribozymes. Only 6-thiopurine reacted at a rate that was even close to that of $^{6S}$Gua with MA displaying an apparent efficiency of ~0.1-0.5 M$^{-1}$min$^{-1}$ (assuming a linear reaction rate, Figure 2-8) while RA was 5-10 fold slower still. MA also reacted with 2-methyl-6-thiopurine and 6,8-dithiopurine, but about 1-2 and 4-10 times slower than with 6-thiopurine, respectively. In contrast, RA did not react detectably with
Figure 2-8: Ribozyme incubation with 6-thioguanine or 6-thiopurine.

Reactions were performed for both random (RA) and biased pool isolates (MA). Substrate concentrations were held at 0.26 mM. Incubation with $^{6S}$Gua for 30 minutes, 4 hours and 18 hours resulted in 13%, 30%, 30% respectively of ribozyme RA reacting and 26%, 30% and 29% respectively for MA. Incubation of both ribozymes with 6-thiopurine for 18 hours resulted in signals of ~0.5% for RA and 3.6% for MA.

either compound. All four product bands had unique APM gel shifts.

Measurements with radiolabelled guanine, though highly desirable, were not attempted due to the very low solubility of guanine (maximum solubility of ~5 μM, which is ~50 fold lower than that of $^{6S}$Gua) and the relatively small amount of ribozyme that could be used to perform such an assay.
2.4 Discussion

How is the structural information encoded in an RNA’s primary sequence able to influence the outcome of an in vitro selection? We observed a bias resulting from imposing the secondary structure of the family A pyrimidine nucleotide synthase onto one of our high diversity pools. While the overall structure never appeared to be preserved, nearly all of the Round 10 biased pool isolates contain the hairpin-stem V of the family A pyrimidine nucleotide synthase (Chapple, Bartel et al. 2003). The conservation of this stem was expected due to chance 46% of the time (excluding wobbles) implying its high conservation in nearly every isolate (18/20) conferred an overall benefit to ribozyme function. This hairpin was found in the fastest most populated families MA and MB, but was also present in 12/18 families which contained a UCUUU motif (of which 6 can be extended to AGGCGUUCUUU) immediately upstream of the helix. The distinctly different sequence of these families in other regions makes it likely that these families are independent representatives of one overarching motif class defined at least partially by a UCUUU-hairpin motif. This motif at the same sequence region was not found in the random pool even though the random pool contained RNA with terminal nucleotide sequences identical to that of the biased pool. This suggests that the motif is a direct consequence of the imposed secondary structure. It is curious in this regard that while the structural bias increased the frequency of occurrence for the UCUUU-hairpin-loop motif, it did not result in ribozymes notably more efficient than found in the random pool (see Figure 2-4 to compare final pool activities).
A selection involving a small substrate such as $^{6S}$Gua represents a significantly different catalytic challenge than presented by selections involving substrates capable of high affinity binding through Watson-Crick pairing (Wilson and Szostak 1999). With small substrates, even though both the substrate concentration and the incubation time can be experimentally varied, at short enough times and low enough substrate concentrations only the ratio between $k_{\text{cat}}$ and $K_m$ directly effect ribozyme survival. It is therefore not unexpected that ribozymes resulting from a given pool should have roughly similar purine nucleotide synthesis efficiencies (but does leave unexplained why separate pools would have the same efficiency). What is interesting is that two ribozymes with similar efficiencies appear to have exploited quite dissimilar catalytic strategies. RA binds $^{6S}$Gua tightly ($K_m \sim 80 \, \mu M$), discriminates well against quite similar compounds and performs a slow chemical step ($k_{\text{cat app}} \sim 0.02 \, \text{min}^{-1}$). In contrast, MA has slightly worse substrate discrimination than RA, binds its substrate with low affinity, and by implication has a $k_{\text{cat app}}$ significantly higher than RA.

The ability of RA and MA to distinguish between many closely related $^{6S}$Gua derivatives indicates that appreciable contacts must be formed with the substrate at some point during the course of glycosidic bond formation. While the MA and RA purine synthases generally did not react detectably with a range of 6-thioguanine derivatives, they were somewhat tolerant of substitutions at the 2 position reacting thousands of times slower with, for example, 6-thiopurine (Fig. 6). It is curious in this regard that the protein enzyme HGPRTase is also unable to discriminate strongly between hypoxanthine and guanine and normally
accepts both substrates (Craig and Eakin 2000). A similar pattern is displayed by
a naturally occurring guanine aptamer found in the xpt-pbuX mRNA (Mandal,
Boese et al. 2003) and an artificially selected guanine aptamer (Kiga, Futamura
et al. 1998); both discriminate weakly against xanthine and hypoxanthine. While
these correlations may be coincidental, the isolation of two artificial ribozymes
that share the same general recognition characteristics found with a variety of
naturally occurring guanine aptamers, as well as protein enzymes that recognize
guanine, suggest that this pattern of recognition may be universal.

The purine nucleotide synthases were considerably more efficient than the
equivalent pyrimidine nucleotide synthase previously isolated. Even after
performing further selections which improved the family A pyrimidine nucleotide
synthase by 35 fold (Chapple, Bartel et al. 2003), the purine synthases we found
directly from both random and biased pools were still two fold faster (50-100
times faster than the initial pyrimidine nucleotide synthase isolates). The most
obvious explanation for this rate difference is that $^{4S}$Ura is simply harder to
recognize than $^{6S}$Gua. As uracil aptamers are currently unknown this assumption
is difficult to access objectively but appears reasonable given the superior
stacking potential expected from purine substrates.

There is however, a second possibility. The chemistry of glycosidic bond
formation is surprisingly and dramatically influenced by nucleobase composition.
Purine nucleotides are much more thermodynamically stable than pyrimidine
nucleotides. The synthesis of OMP by EC 2.4.2.10 has a $\Delta G^0$ (standard state
free energy change at pH 7) of nearly zero (Tavares, Lee et al. 1987; Bhatia,
Vinitsky et al. 1990), while AMP or GMP synthesis by EC 2.4.2.7 or 2.4.2.8 has a \( \Delta G^o \) of \(-7 \) kcal/mol (Kornberg, Lieberman et al. 1955; DeWolf, Emig et al. 1986; Xu, Eads et al. 1997). While the reason for this large free energy difference does not appear to have been well explored it is striking that the hydrolysis of PRPP has a \( \Delta G^o \) of \(-8.4 \) kcal/mol (Frey and Arabshahi 1995); only \( 1.4 \) kcal/mol more negative than observed for purine nucleotide synthesis. Kinetically, in the absence of an enzyme, the acid-catalyzed cleavage of a purine glycosidic bond at low pH can be estimated to occur at a rate \( 10^5 \) to \( 10^6 \) times faster than the equivalent pyrimidine glycosidic linkage (calculated using deoxyribose nucleosides which are much more labile than ribose nucleotides in acid (Zoltewicz, Clark et al. 1970)). These thermodynamic and kinetic statements are consistent with the finding that purine nucleosides, and not pyrimidine nucleosides can be synthesized by dehydration between ribose and nucleobases (Orgel 1998). Taken together, these observations suggest that the uncatalyzed rate of purine nucleotide synthesis starting from PRPP is likely to be much higher than for pyrimidine nucleotide synthesis even though neither rate could be detected (Unrau and Bartel 1998).

The pertinent question may therefore be, not why is purine nucleotide synthesis superior to pyrimidine nucleotide synthesis, but why is it only 50-100 fold better? It appears most likely that RNA has difficulty precisely positioning its limited range of functional groups in a catalytic pocket small enough to optimize both small substrate binding and transition state stabilization simultaneously. This is generally consistent with our observation that both random and
structurally biased pools resulted in ribozymes with very similar efficiencies (Fig. 3), even though the initial amount of structural information differed between the two pools. More specifically, our kinetic data suggests that RA and MA have considerably different substrate recognition strategies and yet have very similar overall efficiencies. It has been hypothesized that a general relationship between informational complexity (the amount of information required to specify an RNA structure) and function may exist (Szostak 2003; Carothers, Oestreich et al. 2004). We suggest that the precise positioning of functional groups required for small molecule catalysis conflicts with the scale of RNA’s relatively large monomers and provides a natural basis to relate structural information to catalytic function. This complexity-function relationship might be expected to saturate or change its character if this conflict is in fact the limiting process that governs the emergence of small molecule RNA catalysts. It will therefore be of considerable interest to study purine and pyrimidine nucleotide synthase ribozymes in greater structural and kinetic detail in order to explore this important aspect of RNA catalysis.
2.5 Contributions

The work described in this chapter was primarily conducted by Matthew Lau, with the assistance of Kelly Cadieux in the synthesis of the compound APP\textsuperscript{6S}G, and Elaine Chiu in the sequencing of the selection pools. This work was conducted under the guidance of Dr. Peter Unrau.
CHAPTER 3: SECONDARY STRUCTURE CHARACTERIZATION OF TWO PURINE NUCLEOTIDE SYNTHASE RIBOZYMES

3.1 Introduction

The potential of RNAs to synthesize ribonucleotides, which are the basic structural units for constructing larger RNA polymers, is an important aspect to the RNA World hypothesis. In the previous chapter, we described the discovery of ribozymes that can mediate purine nucleotide synthesis from tethered 5-phosphoribosyl 1-pyrophosphate (PRPP) and 6-thioguanine (\(6^S\)Gua) (Lau, Cadieux et al. 2004). This work, in conjunction with the study done previously on ribozyme-mediated pyrimidine nucleotide synthesis (Unrau and Bartel 1998), demonstrates that nucleotide synthesis by RNAs is highly plausible in the early world.

Comparing the two classes of nucleotide synthase ribozymes, we noticed that there were several obvious differences (Unrau and Bartel 1998; Lau, Cadieux et al. 2004). Firstly, there were at least 53 different families of purine nucleotide synthase ribozymes that were successfully isolated, as compared to only 3 families of pyrimidine nucleotide synthase ribozymes. Secondly, the more abundant purine nucleotide synthase ribozymes were on average 50-100 times more efficient than their pyrimidine synthase counterpart. Thirdly, it is more thermodynamically favourable to synthesize a purine nucleotide (\(\Delta G^{\circ} \) of ~ -7 kcal/mol) than a pyrimidine nucleotide (\(\Delta G^{\circ} \) of ~ 0 kcal/mol) using PRPP as the...
starting substrate (Tavares, Lee et al. 1987; Xu, Eads et al. 1997). These results, together, support the notion that it is seemingly easier for ribozymes to synthesize a purine than a pyrimidine nucleotide.

To further address this idea, we describe here our work in deducing the minimal core motifs (or minimal sequences elements required for functionality) and secondary structures of purine nucleotide synthase ribozymes in an effort to further understand their catalytic mechanisms. We hypothesize that they adopt simpler secondary topologies for catalysis as compared to their structurally complex pyrimidine synthase counterpart (5 helical stems, 3 pseudoknots, inferred from comparative sequence analysis, site-directed mutagenesis and cross-linking experiments), since there are apparently more solutions to ribozyme-mediated purine nucleotide synthesis (Chapple, Bartel et al. 2003; Lau, Cadieux et al. 2004). By performing a re-selection with $^{6S}$Gua, we were able to successfully isolate truncated purine nucleotide synthase ribozymes with catalytic efficiencies comparable to their parents (Lau, Cadieux et al. 2004). These truncated ribozymes originated from two recombined pools generated with a Non-Homologous Random Recombination technique (NRR), in which sequences from each pool contained randomly reassembled sequence elements of a previously isolated purine synthase ribozyme (Lau, Cadieux et al. 2004; Wang and Unrau 2005). The isolated truncated sequences were on average 2.5 fold shorter than their parents, and represent the minimal core motifs required for catalysis.
To determine the secondary structures from the minimal core motifs, we performed site-directed mutagenesis on the shortest functional truncated sequence from each recombined pool. Interestingly, these truncated ribozymes required only two or three non-sequence specific hairpin loops for catalysis, which is structurally much less complex as compared to the pyrimidine nucleotide synthase ribozymes (Chapple, Bartel et al. 2003). The results from this work further strengthened the idea that ribozyme-mediated nucleotide synthesis, in various aspects, is much simpler for purine than pyrimidine, as well as illustrating the power of the NRR technique in deducing core sequence elements required for catalysis in functional nucleic acids.

3.2 Materials and Methods

3.2.1 Synthesis of Truncated Pools by Non-Homologous Recombination

MA and MF DNA sequences were first amplified by a mutagenic PCR method using 1 × mutagenic PCR buffer (10 mM Tris-HCl, 50 mM KCl, 7 mM MgCl₂, 0.1% gelatin, pH 8.3), 1 × dNTPs (0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP and 1 mM dTTP), 0.05 U/μL TAQ polymerase, along with 0.3 mM of 5' primer (5'-TTC TAA TAC GAC TCA CTA TAG GAG CGA AGT GCC C, italics indicates T7 promoter) and 3' primer (5'-GCT TGC TGC GGT GCG A) (Cadwell and Joyce 1992). The reaction products were purified by extraction with phenol and chloroform to remove the enzyme, followed by ethanol precipitation. The two DNA pools were then mildly digested with deoxyribonuclease I (5 × 10⁻⁴ U/μL) to give fragments of ~40 nucleotides as indicated by separation through a 3%
agarose gel. To randomly reassemble the fragments together, we first performed blunt end extension with T4 DNA polymerase (40 mM Tris-HCl, 4 mM MnCl₂, 4 mM EGTA, 66 mM sodium acetate, 13.3 mM magnesium acetate, 0.5 mM DTT, 0.2 mM dNTPs, 0.01% BSA, 0.1 U/µL T4 DNA polymerase (NEB), pH 7.9) for 30 min at 12 °C, followed by ligation with T4 DNA ligase (50 mM Tris-HCl, 3.3 mM MnCl₂, 3.3 mM MgCl₂, 11 mM magnesium acetate, 3.3 mM EGTA, 55 mM sodium acetate, 0.2 mM dNTPs, 0.01% BSA, 5.9% PEG-8000, 1.2 mM DTT, 1.2 mM ATP, 0.1 U/µL T4 DNA ligase (Invitrogen), pH 7.6) for 20 hours at 23 °C to generate a diverse pool with fragments that are linked in random orientation and order. The two resulting pools were gel purified to exclude short fragments (40 nucleotides or shorter), and PCR amplified to generate a high copy number of each unique sequence (estimation of ~10¹¹ different sequences in each pool).

### 3.2.2 Selection

Both recombined MA and MF DNA pools were separately transcribed and bodylabelled with α-[³²P] UTP using T7 RNA polymerase, followed by purification through a denaturing 6% polyacrylamide gel. A marker sample containing unrelated bodylabelled RNAs of lengths 271, 125, 70 and 40 nucleotides was also loaded in a separate lane. The two RNA pools, at 2 µM RNA concentration, were then ligated with 60 µM APPRPP for 4 hrs at 23 °C in ligation buffer (50 mM HEPES, 10 mM MgCl₂, 15 mM sodium phosphite, 10 µg/mL BSA and 8.3% v/v glycerol at pH 8.0) using 0.5 U/µL T4 RNA ligase (Wang and Unrau 2002). The RNA sequences derivatized with PRPP (RNA-PRPP, dash symbolizes 5’ - 3’ phosphodiester linkage) were extracted with an equal volume of phenol and
chloroform, followed by incubation at 0.25 μM in incubation buffer (50 mM Tris-HCl, 150 mM KCl, 75 mM MgCl₂, pH 7.5) containing 0.25 mM 6S-Gua for 15 hrs at 23°C. The resulting RNA, along with a 6S-G marker sample in a separate lane, was separated through a 6% denaturing PAGE gel containing 3.75 μM APM to isolate reactive (thiol-containing) from non-reactive RNA. The 6S-G marker sample was composed of the 3 transcription marker sequences derivatized with 3’ 6S-G, which were generated by ligation with the compound APP^{6S}G using the ligation conditions described above.

For rounds 1-3, the shifted region between the 125-nt RNA tethered with 6S-G (125-nt-6S-G) and 40-nt-6S-G was excised. The resulting RNAs were reverse transcribed using Superscript II (50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 560 μM of each dNTP, 5 μM 3’ primer, 10 U/μL enzyme, pH 7.5) at 48°C for 1 hr. RNA was hydrolyzed with 100 mM KOH at 90 °C for 10 min, followed by neutralization with HCl. The subsequent cDNAs were subjected to error-prone PCR amplification as described above, before re-entering the next round of selection. Starting from round 3, length selection pressure was applied to both pools during the transcription purification step by excising only the lower one quarter of the dominant RNA band product and all of the shorter RNA populations below. Similarly, during the APM gel step the shifted region between the dominant product band and the 40-nt-6S-G marker was excised and eluted. Time selection pressure was also applied, starting in round 3, by reducing the incubation time of the two RNA pools with 6S-Gua by half during each round (Round 3, 4 ,5, 6, incubation time of 2 hrs, 1 hrs, 30 hr, 15 min, respectively).
3.2.3 Substrate Analogues

All analogues were obtained from Sigma Aldrich and dissolved in 1.05× incubation buffer. 6-thioguanine $\varepsilon$ (347 nm, pH 1) $= 20,900 \text{ M}^{-1} \text{ cm}^{-1}$ (Elion and Hitchings 1955), 6-thiopurine $\varepsilon$ (325 nm, pH 1) $= 20,500 \text{ M}^{-1} \text{ cm}^{-1}$ (Elion 1962), 6,8-dithiopurine $\varepsilon$ (358 nm, pH 1) $= 27,800 \text{ M}^{-1} \text{ cm}^{-1}$ (Robins 1959), 2-methyl-6-thiopurine assumed $\varepsilon$ (330 nm, pH 1) $= 20,000 \text{ M}^{-1} \text{ cm}^{-1}$ were used at 0.25 mM concentration while analogues 6-thioxanthine, 6-thio-9-methylpurine and 2-thiocyotosine were used at saturating concentrations.

3.2.4 UCUUU Randomization and Re-selection

DNA pools of MAa and MFa containing a 5-nt randomized region were synthesized on an ABI 392 DNA synthesizer using 0.2 µMole of 2000 Å control pore glass (CPG) columns and standard cyanoethyl phosphoramidite chemistry. The random nucleotide couplings (positions 36 to 40 for MAa and positions 41 to 45 for MFa) was obtained by mixing of equiactive phosphoramidite stocks (ABI) in the molar ratios of 0.28:0.27:0.23:0.22 (dA:dC:dG:dT, respectively) (Zaher and Unrau 2005). The two pools were subsequently transcribed to RNA and re-selected for activity with $^{6S}$Gua as in the selection described above. To give an estimate of nucleotide distribution in the randomized region, transcribed RNA pools in each round were first removed of the 5’ triphosphates by treatment with 0.5 U/µL of calf intestine phosphatase (Roche, 50 mM Tris-HCl, 0.1 mM EDTA, pH 8.5) at 50°C for 1 hr, followed by 5’ labelling (phosphorylation) with 2 µM $\gamma$-$[^{32}P]ATP$ (70 mM Tris-HCl, 10 mM MgCl$_2$, 5 mM DTT, pH 7.6) using 10 U/µL T4 polynucleotide kinase (NEB) for 40 min at 37°C. The resulting RNA were
digested using 10 U/μL T1 RNAse (Fermentas, 18 mM sodium citrate, 8 M urea, pH 6.0) for 10 min at 23°C and resolved on a 7% denaturing polyacrylamide sequencing gel.

3.2.5 Site Directed Mutagenesis

Mutants of MAa and MFa were generated by PCR amplification of the wildtype DNA with primers containing specific mutations and / or deletions. Primers were synthesized on a DNA synthesizer as described above. The resulting DNA mutants were transcribed into RNA, derivatized with PRPP, incubated with $^{6S}$Gua for 48 hours, and analyzed for activity by mercury gel shift as in the selection. Mutants were identified as active when the quantified shifted products were at least 0.5% above background.

3.2.6 Kinetics

The RNA sequences of MAa, MF, MFa, tMFa-50 and tMFa-46 derivatized with PRPP were incubated with $^{6S}$Gua for 0 min, 2 min, 8 min, 30 min, 50 min and 120 min. Time points were stopped by the addition of an equal volume of gel-loading buffer. The reaction rate at a particular $^{6S}$Gua concentration was deduced by plotting the fraction reacted for at least three independent time courses to the equation $F = \beta(1-e^{-k_{obs}t})$ using the program KaleidaGraph (Synergy Software), F being the fraction reacted at time t, $k_{obs}$ the first-order apparent rate constant, and $\beta$ is the total fraction able to react. The resulting rates were then fitted to the Michaelis-Menten equation $k_{obs} = k_{cat \, app} [^{6S}Gua] / (K_m + [^{6S}Gua])$. 
3.3 Results

3.3.1 Pool Construction by Non-Homologous Recombination

We were interested in understanding the secondary structures of purine synthase ribozymes and deducing their minimal core motifs. We approached this by constructing two pools, each consisting of random recombined components of a previously characterized purine nucleotide synthase ribozyme (Lau, Cadieux et al. 2004). Motifs that are critical for structure and functionality of the ribozyme will be conserved in truncated sequences that are able to react with $^{65}\text{Gua}$. The purine synthase ribozyme isolates selected for secondary structure characterization were MA and MF, both of which were isolated from the mutagenized pool in our initial selection for purine nucleotide synthase ribozymes (Lau, Cadieux et al. 2004). MA was chosen because it was highly efficient ($284 \, \text{M}^{-1} \, \text{min}^{-1}$) and was the fastest isolate out of all of the analyzed purine nucleotide synthase ribozymes (Lau, Cadieux et al. 2004). MF was selected because it was the only sequence which conserved of all the sequence elements hypothesized to be involved in the folding of the previously isolated Family A pyrimidine nucleotide synthase ribozymes (Chapple, Bartel et al. 2003; Lau, Cadieux et al. 2004). This observation maybe misleading, however, as MF originated from a pool that was biased to contain such structural sequences and according to the computational folding algorithm MFOLD, this sequence is more favoured thermodynamically to fold into other structures (Zuker, Matthews et al. 1999). 

The two recombined MA and MF pools were constructed using a technique called Non-Homologous Random Recombination (Figure 3-1, see...
Parental MA and MF sequences were first amplified using a mutagenic PCR method to introduce point mutations and diversity. The two resulting DNA pools were mildly digested with deoxyribonuclease I to generate double stranded fragments with “sticky ends”. The fragments were subsequently extended to blunt end by T4 DNA polymerase, followed by ligation with T4 DNA ligase to generate sequences with fragments that are linked in random orientation and order. The two pools were then gel purified to exclude fragments shorter than 40-nt, and PCR amplified to generate copies of each unique sequence. The two pools were then transcribed to RNA and subjected to re-selection with 6S Gua (Wang and Unrau 2005).

methods for details on pool construction) (Wang and Unrau 2005). First, we amplified the parental MA and MF DNA sequences using a mutagenic PCR method to introduce point mutations and diversity (Cadwell and Joyce 1992). This method employs a higher magnesium concentration and a biased dNTP concentration relative to typical PCR conditions such that the TAQ polymerase would have increased chance of mis-incorporations (see methods). The two DNA
Figure 3-2: In vitro selection scheme for truncated $^{6S}\text{G}$ synthase ribozymes.

a. Recombined MA and MF pool were first tethered with PRPP using APPRPP and T4 RNA ligase. b. The two pools were then incubated with $^{6S}\text{Gua}$. Ribozymes capable of synthesizing $^{6S}\text{G}$ from $^{6S}\text{Gua}$ release pyrophosphate (PP$_i$) as the leaving group. c. An APM gel was used to separate active from non-reactive RNA (left lane). A mixture of radiolabelled marker RNA of different lengths derivatized with a terminal $^{6S}\text{G}$ nucleotide was loaded in an adjacent lane to indicate the position of active ribozymes (see methods for details). d. Active RNA species were excised from the gel and eluted. e. The resulting sequences then reverse transcribed into DNA, amplified and enriched using a mutagenic PCR method (Cadwell and Joyce 1992), and transcribed back into RNA for the next round of selection.
pools were then mildly digested with deoxyribonuclease I to generate random double stranded fragments, followed by treatment with T4 DNA polymerase to remove 5' and 3' overhangs (giving blunt ends). The fragments were then randomly reassembled together by ligation with T4 DNA ligase to generate sequences containing fragments that are linked in random orientation and order. The two resulting recombined pools were subsequently gel purified to exclude short fragments (41 nucleotides or shorter), and PCR amplified to generate a large number of each unique sequence (initial estimation of $\sim 10^{11}$ different sequences in each of the pools). Sequence analysis and alignment of 13 sequences from each of the two starting pools showed minimal sequence conservation relative to their parents. This could be due to random mutations from error-prone PCR, deletions, re-localization of sequence elements and the inverse insertion of fragments.

### 3.3.2 Selection

The two recombined pools were derivatized with PRPP and selected for reactivity with $^{6S}$Gua as described previously (Figure 3-2) (Lau, Cadieux et al. 2004). During the mercury gel (APM gel) selection step in each round, markers of different RNA lengths derivatized with 3' $^{6S}$G were simultaneously loaded in adjacent lanes. For rounds 1-3, the region between 125-nt-$^{6S}$G and 40-nt-$^{6S}$G was excised and eluted (see methods). The purified RNA material were reverse transcribed and amplified using error-prone PCR to introduce additional sequence variation (Cadwell and Joyce 1992) before entry into the next selection
Figure 3-3: Reactivity of round 3 recombined MA and MF pool sequences with $^{65}$Gua.

Round 3 sequences from both recombined MA and MF pools were tethered with PRPP and incubated with $^{65}$Gua (RMA-PRPP + $^{65}$Gua and RMF-PRPP + $^{65}$Gua, respectively) for 30 min and 120 min and analyzed on an APM gel. Parental MA and MF sequences tethered with PRPP were reacted with $^{65}$Gua in parallel as a control. Markers were loaded in the leftmost and rightmost lanes to indicate the gel shift positions of different length RNA derivative with 3' $^{65}$G. $M_1$ = 125-nt RNA tethered with $^{65}$G, $M_2$ = 70-nt RNA tethered with $^{65}$G, $M_3$ = 40-nt RNA tethered with $^{65}$G and $M_{123}$ = mixture of markers $M_1$, $M_2$ and $M_3$ (see methods for details). RMA and RMF stands for round 3 recombined MA and MF pools, respectively.

Interestingly, after 3 rounds, activity was readily detectable for both recombined pools by our mercury gel shift assay (Figure 3-3, see methods). To isolate the fastest and shortest functional sequences, we added both time and length pressures to the two pools in the successive rounds (Table 3-1). Length selection pressure was introduced by biasing the recovery towards shorter RNA during the RNA transcription gel purification step (see methods). Figure 3-4
illustrates the RNA length distribution after each successive round of selection. Time pressure was applied by continuously reducing the incubation time of the two pools with 6S Guo by two fold each round. Ribozyme activity was still readily detectable for both pools after 6 selection rounds in which the incubation time was reduced to 15 minutes. The RNA sequences of both pools from the same round were reverse transcribed, PCR amplified, and subjected to sequencing.

3.3.3 Sequencing Analysis

3.3.3.1 Alignment of Round 6 Recombined MA Pool Sequences

Analysis of the round 6 recombined MA pool revealed 13 different sequences, with two of the sequences occurring more than once (one was repeated twice and one was repeated three times out of 16 sequences total, Figure 3-5a). Sequence alignment revealed, relative to wildtype MA, a 34 to 37 nucleotide deletion region immediately downstream of the 5' constant primer region for all the truncated sequences (see Figure 3-5a). Other than this region, no other deletions were observed, with the remaining sequence elements aligning closely with the parental sequence (6 sequences with one or two single nucleotide mutations). None of the truncated sequences contained fragments in a different position or obvious fragments that were assembled in reverse orientation relative to MA (i.e. MA contains a 5'-ACG UCG UAC-3' motif, no truncated sequences contain 5'-CAU GCU GCA-3'). Five truncated sequences were selected (including one of the duplicates and four randomly selected sequences) for activity analysis, and all five sequences showed activity after
Table 3-1: Length of sequences and reaction time with $^{6S}$Gua for both recombined MA and MF pools in each selection round.

<table>
<thead>
<tr>
<th>Selection Rounds</th>
<th>Recombined MA Pool</th>
<th>Recombined MF Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range of Length Excised (nt)</td>
<td>Incubation Time (Minutes)</td>
</tr>
<tr>
<td>1</td>
<td>125 to 40</td>
<td>900</td>
</tr>
<tr>
<td>2</td>
<td>125 to 40</td>
<td>900</td>
</tr>
<tr>
<td>3</td>
<td>110 to 40</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>90 to 40</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>85 to 40</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>85 to 40</td>
<td>15</td>
</tr>
</tbody>
</table>

incubation with $^{6S}$Gua for 4 hours. To deduce the secondary structures of the truncated sequences, the sequence denoted MAa, 89-nt in length, was chosen for further analysis.

3.3.3.2 Alignment of Round 6 Recombined MF Pool Sequences

Analysis of the round 6 recombined MF pool revealed 46 different sequences, with one of the sequence occurring twice out of 47 total sequences analyzed (Figure 3-5b). Interestingly, sequence alignment showed a similar trend as the recombined MA pool, with all the truncated sequences showing a deletion region immediately after the 5' constant primer region relative to wildtype MF. The number of nucleotides deleted in the region was much more variable, however, ranging from 32 to 55 nucleotides, with the shortest sequence being
Figure 3-4: Length distribution of recombined pools through 6 selection rounds.

Figure 3-5: Sequencing analysis of round 6 recombined MA and MF pools.

a. Alignment of 16 sequences from round 6 recombined MA pool. Truncated MAa sequence and parental MA sequence are indicated. 
b. Alignment of 47 sequences from round 6 recombined MF pool. Truncated MFa sequence and parental MF sequence are indicated. Sequences in both alignments are aligned relative to their parental sequence at the bottom. Shaded regions indicate matches relative to the consensus sequence at the top of each alignment. The colour bars at the very top indicate relative consensus alignment strength, with red being the strongest to dark blue the weakest. Constant primer regions are underlined in red. Sequences are arranged based on their mutagenesis pattern (see text).

68-nt in length (Figure 3-5b). Other than this main deletion region, other minor variations were also observed. The most striking were 5 different mutations or
deletions that were simultaneously observed in 19 sequences (40% of the population), with 1. mutation at T77A (first letter = wildtype nucleotide, number = nucleotide position, last letter = nucleotide mutation), 2. an addition of a T at position 79, 3. deletion of two A residues at positions 89 and 90, 4. a mutation at G87C, and 5. a mutation at C104T (position numbers are relative to the consensus sequence in the recombined MF pool alignment, mutations are based on the parental MF sequence, see Figure 3-5b). Interestingly, 16 out of these 19 sequences also have an additional mutation at A92G. A similar trend was shared by two other sequences with the exception of a C addition instead of a T at position 79, and a mutation at G87A. Other minor variations include the presence of a single mutation at U109G in 17 sequences, 8 sequences with T76C, and 6 sequences with A88G. In general, the remaining sequence elements align very closely to MF. Again, five truncated sequences were selected (including one of the duplicates, 3 randomly selected sequences of different lengths, and the shortest sequence) for activity analysis, and all five sequences showed activity after incubation with $^{68}$Gua for 4 hours. The shortest sequence out of the five analyzed, denoted MFa, 68-nt in length, was selected for further analysis.

### 3.3.4 Site Directed Mutagenesis

Both MAa and MFa sequences were subjected to computational folding using the program SimFold to predict possible RNA secondary structures (Andronescu, Zhang et al. 2005). This software is an upgraded version to the commonly used nucleic acid folding algorithm MFOLD (Zuker, Matthews et al. 1999), and calculates, for a given sequence, the pseudoknot-free secondary
Table 3-2: Mutants synthesized for structural analysis of MAa, MFa and 50-nt truncated version of MFa (tMFa-50).

The first column of each table indicates the positions that were mutated or deleted. For mutations, the first letter represents the wildtype nucleotide, followed by the position number in the RNA sequence, and the last letter corresponds to the mutated nucleotide (see bottom right for position numbering and predicted secondary structure of each sequence). Mut = mutation, C. Mut = complementary mutations, * indicates an addition of a U nucleotide between U27 and G28 in the tMFa-50 ribozyme. The second column indicates the structure or sequence of the ribozyme that is being addressed by the particular mutant sequence. The third column indicates if the mutant is active in catalyzing \textsuperscript{65}G synthesis (+ indicates active, - indicates inactive).

<table>
<thead>
<tr>
<th>a)</th>
<th><strong>MAa Mutagenesis</strong></th>
<th>b)</th>
<th><strong>MFa Mutagenesis</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deletion / Deletion Positions</strong></td>
<td><strong>Structure / Sequence</strong></td>
<td><strong>Activity</strong></td>
<td><strong>Deletion / Deletion Positions</strong></td>
</tr>
<tr>
<td>Deletion 1 to 14</td>
<td>5’ Primer Region</td>
<td>-</td>
<td>Deletion 1 to 14</td>
</tr>
<tr>
<td>Deletion 1 to 3</td>
<td>5’ Primer Region</td>
<td>-</td>
<td>Deletion 1 to 3</td>
</tr>
<tr>
<td>Deletion 36 to 40</td>
<td>UCUUU region</td>
<td>-</td>
<td>Deletion 36 to 40</td>
</tr>
<tr>
<td>Mut U10C, C12G</td>
<td>Stem1</td>
<td>-</td>
<td>Mut U10C, C12G</td>
</tr>
<tr>
<td>Mut G19G, A21G</td>
<td>Stem1</td>
<td>-</td>
<td>Mut G19G, A21G</td>
</tr>
<tr>
<td>C. Mut G19G, A21G, U10C, G12G</td>
<td>Stem2</td>
<td>-</td>
<td>C. Mut G19G, A21G, U10C, G12G</td>
</tr>
<tr>
<td>Mut C14G, A15C, U16G, G17A</td>
<td>Stem1 Loop</td>
<td>+</td>
<td>Mut C14G, A15C, U16G, G17A</td>
</tr>
<tr>
<td>Mut C31G, G34C</td>
<td>Stem2</td>
<td>-</td>
<td>Mut C31G, G34C</td>
</tr>
<tr>
<td>Mut C85G, G88C</td>
<td>Stem2</td>
<td>-</td>
<td>Mut C85G, G88C</td>
</tr>
<tr>
<td>Mut G68G, G70C</td>
<td>Stem3</td>
<td>-</td>
<td>Mut G68G, G70C</td>
</tr>
<tr>
<td>Mut C79G, G81C</td>
<td>Stem3</td>
<td>-</td>
<td>Mut C79G, G81C</td>
</tr>
<tr>
<td>C. Mut U73G, U74C, C75G, G76A</td>
<td>Stem3 Loop</td>
<td>+</td>
<td>C. Mut U73G, U74C, C75G, G76A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c)</th>
<th><strong>tMFa-50 Mutagenesis</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deletion / Deletion Positions</strong></td>
<td><strong>Structure / Sequence</strong></td>
</tr>
<tr>
<td>Mut G11C, C13G</td>
<td>Stem 1</td>
</tr>
<tr>
<td>Mut G18C, C20G</td>
<td>Stem 1</td>
</tr>
<tr>
<td>Mut CBA</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut A9U, U22A</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut G13C, C20G</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut U13G, G19C</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut C13G, G19C</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut U13G, G19C</td>
<td>Stem 1</td>
</tr>
<tr>
<td>C. Mut U13G, A15C, U16G</td>
<td>Stem 1 Loop</td>
</tr>
<tr>
<td>Mut U23G, G31C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>Mut C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>Mut C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>Mut C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>C. Mut U29G, G31C, C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>C. Mut U29G, G31C, C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>C. Mut U29G, G31C, C40G, G42C</td>
<td>Stem 2</td>
</tr>
<tr>
<td>Mut U23G, U35C, G36G, G37A</td>
<td>Stem 2 Loop</td>
</tr>
<tr>
<td>Deletion 27</td>
<td>UCUUU</td>
</tr>
<tr>
<td>*Addition of U after 27</td>
<td>UCUUU</td>
</tr>
<tr>
<td>Mut U23G</td>
<td>UCUUU</td>
</tr>
<tr>
<td>Mut U23A</td>
<td>UCUUU</td>
</tr>
<tr>
<td>Mut U23C</td>
<td>UCUUU</td>
</tr>
</tbody>
</table>

structure that has minimum free energy according to a standard thermodynamic model. The following site-directed mutagenesis experiments were designed
Figure 3-6: Secondary structure predictions of ribozymes MAa, MFa and the 50-nt truncated version of MFa (tMFa-50).

a. The 3 helical stem structure of MAa. b. The 2 helical stem structure of MFa. c. The 2 helical stem structure of tMFa-50 (18 nucleotides deleted from the 5' terminus relative to MFa). The constant primer regions are coloured in blue and the UCUUU motif in each sequence is coloured in dark yellow. Roman numerals indicate the stem number. Arabic numerals indicate the nucleotide position in each sequence.

Based on such predictions. All of the mutations were introduced by PCR amplification with primers containing the mutation and / or deletion sites. A listing of all the mutants tested for MAa, MFa and a further truncated version of MFa with 50-nt in length (tMFa-50) is outlined in Table 3-2. Figure 3-6 illustrates the predicted secondary structural folding of the three ribozyme isolates from our site-directed mutagenesis data.
3.3.4.1 Mutagenesis of the Ribozyme MAa

MAa is the shortest functional truncated sequence from the round 6 recombinant MA pool, with a 14-nt constant 5' primer region and a 16-nt constant 3' primer region (Figure 3-6a). These regions are required for PCR amplification during the selection, but may not be needed for the functionality of the ribozyme. To examine whether these constant regions were required for catalysis, we first constructed a mutant containing a deletion of the entire 14-nt 5' primer region, but no activity was observed (Table 3-2a). We next attempted deletions of only 3 nucleotides from the 5' terminus, but similarly no activity was detected, thus indicating that MAa does require most, if not all, of its 5' primer region for functionality.

To confirm for the presence of the 5' stem structure (stem I), we first generated disruptive mutations at U10C and C12G, which should disrupt the predicted stem structure (generating both C-A mismatch and G-G mismatch, dash indicates base pairing). If the sequence, or structure, of the stem is critical for functionality of the ribozyme, then the mutant sequence should be inactive towards 6S Gua, and indeed that was what we observed (see Figure 3-6a, Table 3-2a). Similarly, no activity was detected when we inserted disruptive mutations on the other side of the predicted stem (G19A, A21U). To restore the stem structure, we again generated RNA sequences containing mutations U10C, C12G, but we also added complementary mutations G19C and A21G. If MAa is structurally but not sequence specific at this stem region, then activity should be restored if U10-A21, C12-G19 were changed to C10-G21, G12-C19, respectively (Watson-Crick base pairing preserved). Indeed, activity was observed with the
complementary mutations as detected by our mercury gel shift assay. To strengthen our claim, we generated complementary mutations to 4 out of 5 possible base pairs predicted to be involved in stem formation (U10C, G11C, C12G, C13G, G18C, G19C, U20G and A21G mutations generated simultaneously), and consistent with our hypothesis, this truncated ribozyme retained activity.

According to the folding program SimFold (Andronescu, Zhang et al. 2005), the next predicted structural motif for MAa is an 8 base pair long stem that forms between residues 24 to 31 and 56 to 64, with an A residue bulge at residue 61 (G24-64C, A25-63U, G26-62C, 61A bulge, C27-G60, G28-C59, U29-A58, G30-C57 and C31-G56). Our mutagenesis data, however, did not show any evidence of this structure forming. Instead, we suggest the presence of a 6 base pair stem (stem II) forming from positions 30 to 35 with the 3' most nucleotides at positions 84 to 89 (Figure 3-6a). This prediction is based on data that we obtained by performing the same mutagenesis experiments as described above, with disruptive mutations at C85G, G88C abolishing activity and complementary mutations of C85G, G88C, C31G and G34 restoring activity (Table 3-2a). Similarly, when we changed the sequence content in 4 out of 6 base pairs involved in stem formation (positions 31 to 34 and 85 to 88) to alternate Watson-Crick base pairs, activity was retained (Figure 3-6a, Table 3-2a).

The final structural motif predicted was a hairpin loop formed from residue position 67 to residue 82, consisting of a 6 base-paired stem with a 4-nt loop (stem III) (Figure 3-6a). The same sequence elements are present in MFa and
this hairpin stem structure is predicted to be present in all of the purine nucleotide synthase ribozyme families of the mutagenized pool from the previous selection (Lau, Cadieux et al. 2004). Our mutagenesis data similarly supports the presence of this hairpin, but showed that the stem region is not sequence specific, as we can replace 4 out of the 6 possible base pairs with other Watson-Crick base pairings (Figure 3-6a, Table 3-2a).

### 3.3.4.2 Mutagenesis of the Ribozyme MFa and tMFa-50

To determine whether the primer regions at both ends of MFa are required for catalysis, we first generated a truncated version of MFa containing deletion of the 14-nt constant 5' primer sequence. Interestingly, this shortened MFa sequence of 54-nt was also active (Table 3-2b). To determine the minimal core motif of MFa, we continued our deletion strategy and removed another 4 nucleotides after the constant primer region. To our surprise, this isolate was similarly functional, with activity similar to that of MFa (Table 3-2b). We denoted this sequence as tMFa-50, which stands for the truncated version of MFa with 50-nt in length. We next generated a third mutant containing another 3-nt deletion from the 5' terminus (up to the 22\textsuperscript{nd} position since the 23\textsuperscript{rd} residue is perfectly conserved in all of the truncated MF sequences), along with the mutation A22G (for transcription initiation by RNA polymerase), but no reactivity was detectable after 15 hours of incubation with $^{6}$S Gua.

Figure 3-6b and c illustrates the secondary structure prediction from our experimental data for MFa and tMFa-50, respectively. We generated mutants only for tMFa-50 because the first 18 nucleotides of MFa were shown, as
described above, to be unnecessary for catalysis. We performed similar mutagenesis experiments as described above for MAa and our data was supportive of a two stem-loop structure (stem I and stem II) formed by MFa and tMFa-50 during catalysis (Figure 3-6b and c, Table 3-2 b and c).

In an effort to determine if functional sequences that are even shorter than tMFa-50 could be found, we attempted further deletion experiments targeting stem II of tMFa-50. When we deleted the first four nucleotides (G28, U29, G42, C43, see Table 3-6c and Figure 3-6c) involved in the formation of stem II, the mutant was rendered inactive. However, when we instead deleted the last four nucleotides involved in formation of the stem region (U32, G33, C38, A39, the base pairs closest to the loop), product synthesis was observed. While this ribozyme isolate, denoted tMFa-46, was functional, it resulted in a much lower yield of synthesized products (very mild activity was observed only after 15 hours of incubation time) relative to MFa (Table 3-3). This signifies the importance of the stem length (number of Watson-Crick base pairs) in contributing to the overall catalytic efficiency of the MFa ribozyme.

**3.3.5 Replacement of Loop Sequences in Predicted Stem-Loop Structures with the Nucleotide Sequence GNRA**

After establishing the potential stems that are formed in MAa and MFa, we next examined the importance of the 4-nt loops within these stem-loop structures. In MAa, there are two stem-loop structures, stem I with loop sequence CAUG, and stem III with loop sequence UUCG. MFa similarly has two stem-loop structures, stem I with loop sequence CAUA and stem II with loop sequence
UUCG. Interestingly, alignments of either truncated MAa or truncated MFa isolates revealed sequence variations at various positions of the hypothesized loop regions (Figure 3-5). To determine the importance of sequence specificity within the loop regions, we substituted each loop with a GNRA sequence (N = any nucleotide, R = purine nucleotide, see Table 3-2a and c). The 4-nt GNRA loop has been shown in literature to contribute to the overall stability enhancement of RNA stem loop structures (Antao, Lai et al. 1991; Heus and Pardi 1991). It is important to note, however, that stem III of MAa and stem II of MFa both consist of highly stable UNCG loops, which are similar in stability as GNRA loops. Interestingly, MAa and MFa mutants with GNRA loop replacements in all the stem-loop structures were active, with reaction rates similar to their parents. These results not only demonstrate the sequence variability of the loop regions, but more importantly further supports the presence of the hypothesized stem structures in both MAa and MFa.

3.3.6 {G}CUUU Motif is Critical for Functionality of MFa, but not MAa

Comparing the overall alignments and predicted secondary structures of the truncated MA and MF sequences, we noticed that not only did sequences from both pools share a similar stem II structural motif, but they also contain a 5-nt sequence motif, UCUUU, that is highly conserved throughout. For MAa, this UCUUU motif is immediately following C36 of stem 3, from residues 36 to 40, whereas for MFa, the motif is found between stem 1 and 2, from residues 27 to 31. Interestingly, this UCUUU motif is also present at the same region (immediately upstream of the 3’ terminal hairpin stem structure) in 12/20 families.
isolated from the mutagenized pool in the initial selection for purine nucleotide synthase ribozymes (Lau, Cadieux et al. 2004). This region, however, was randomized according to the original mutagenized pool template design (see Figure 2-2). Therefore, if the UCUUU motif is not of relevance, it should only appear once in every 1024 sequences (4^5, with 4 possible nucleotides at each of the 5 positions).

To examine the importance of this UCUUU motif, we generated two additional 5-nt randomized MAa and MFa truncated pools, both containing randomized nucleotides at the UCUUU positions (see methods). Both pools were subjected to three rounds of re-selection with 6S Gua using the same selection conditions as described before. For the 5-nt truncated MAa pool, reactivity was observed even before the pool was subjected to selection and did not increase significantly from round to round. For the MFa pool, very mild activity became detectable in round 2, and reactivity dramatically increased by round 3.

We sequenced both the round 0 and round 3 sequences of the 5-nt randomized MAa and MFa pools and examined their nucleotide distribution in each of the five randomized positions. Counting the randomized region of both round 0 pools together, we obtained 29% A, 16% G, 29% C and 26% U out of 105 nucleotides. The ratio of Gs was relatively lower than the other nucleotides, but this can be due to the low number of sequences analyzed and nucleotides counted. More importantly, we did not observe any bias towards a particular nucleotide in any of the five positions.
Analysis of the round 3 MAa 5-nt randomized pool did not reveal any particular pattern in the five random nucleotide region. Counting all five positions together, we obtained a ratio of 24% A, 4% G, 30% C and 52% U out of 55 nucleotides, showing a bias towards Us and a low preference for Gs. However, when we examined each position individually, none of the positions consistently had the same nucleotide and no obvious bias towards any nucleotide was observed (i.e. at position 1, 3 As, 0Gs, 3 Cs and 5 Us). Since this 5-nt sequence can be highly variable, we hypothesize that perhaps it is not required for MAa catalysis. This was not the case, however, as the mutant containing deletion of the UCUUU region was inactive (Table 3-2a). This suggests that perhaps the 5-nt sequence region is required only as a spacer region where its sequence composition is of little significance.

The sequencing data for the round 3 MFa 5-nt randomized pool gave very much the opposite result. Out of 22 sequences analyzed, 18 had the sequence UCUUU in the 5 nucleotide randomized region, while the other 4 had the sequence ACUUU. The chances of any position having the same nucleotide throughout 22 sequences, assuming an equal random distribution of nucleotides, is $\sim 10^{-12}\%$ ($0.25^{22}$), thus it is most likely that the last 4 nucleotides CUUU are critical for functionality of MFa. The first randomized position (position 27) was slightly more variable, with four sequences containing an A instead of a U. To examine the specificity of this nucleotide residue for MFa catalysis, we generated three different mutants with a single nucleotide mutation of A, C or G at position 27 of tMFa-50 and examined their reactivity with $^{68}\text{Gua}$. Interestingly, both
mutants containing A or C mutations were active, while the G mutant was not. The A mutant and the C mutant reacted at rates of about 2 and 7 fold slower, respectively, relative to MFa. In the discussion section, this 5-nt motif in MF and MF related sequences will be denoted {G}CUUU, with {G} indicating that it can be any nucleotide (A, C or U), but not G.

3.3.7 Kinetic Studies

We were interested in comparing the reaction rates and efficiencies of the truncated MA and MF sequences relative to their parents. We performed detailed kinetic analysis on the parental sequence MF, the truncated sequences MAa and MFa, as well as the two further truncated MFa sequences tMFa-50 and tMFa-46 (Figure 3-7). These sequences were also analyzed for magnesium dependence and substrate specificity, and the results are reported in the later sections. MA was not analyzed because it was previously determined to react with an apparent efficiency of 284 M⁻¹ min⁻¹ (Km and kcat could not be determined since we were limited by the maximum saturating concentration of ⁶⁵Gua) (Lau, Cadieux et al. 2004). To determine their apparent efficiencies, the five sequences were first derivatized with PRPP and subsequently reacted with varying concentrations of ⁶⁵Gua (35 μM, 87.5 μM, 175 μM, 262.5 μM, 350 μM and 700 μM). The resulting reaction rates were plotted against substrate concentration (⁶⁵Gua) using the Michaelis-Menten equation and both a Km and kcat values were determined for each sequence (Figure 3-7). For parental MF, we obtained a Km = 50 μM, kcat = 0.027 min⁻¹, and an apparent efficiency of ~540 M⁻¹ min⁻¹ (kcat / Km). For MFa and the further truncated MFa sequences, their Km values were much higher than
Figure 3-7: Kinetic Analysis of MF and 4 other truncated isolates.

Reaction rates were plotted against $^{65}$Gua concentrations for ribozymes a. Parental MF, b. MAa, c. MFa, d. tMFa-50, and e. tMFa-46. f. Table summarizing the $K_m$, $k_{cat}$ and $k_{app}$ efficiency values obtained from fitting each plot to the Michaelis-Menten equation. A $K_m$ and $k_{cat}$ value cannot be extracted from previous kinetic analysis of the MA ribozyme due to the saturation limit of the $^{65}$Gua concentration. The initial slope of the plot was taken as the $k_{app}$ efficiency for MA (Lau, Cadieux et al. 2004).
their parent MF, resulting in a general lowering in catalytic efficiency (MFa = 440 M$^{-1}$ min$^{-1}$, tMFa-50 = 180 M$^{-1}$ min$^{-1}$ and tMFa-46 = 340 M$^{-1}$ min$^{-1}$, see Figure 3-7f). For MAa, we extracted a $K_m$ value of 210 μM, a $k_{cat}$ value of 0.14 min$^{-1}$, and an apparent efficiency of ~670 M$^{-1}$ min$^{-1}$, which is ~2.3 fold higher than MA.

3.3.8 Magnesium Dependence

To examine the metal dependence of the truncated sequences, we performed a magnesium titration (0 mM, 10 mM, 25 mM, 50 mM, 75 mM and 100 mM) for each of the sequences described above. In the selection, sequences were selected to react with $^6S$Gua at 75 mM MgCl$_2$ because MA from previous work was shown to have maximum reactivity at this concentration (Lau, Cadieux et al. 2004). For MAa, it was similarly the most reactive at 75 mM MgCl$_2$. MF, on the other hand, reached its optimal reactivity by 25 mM MgCl$_2$, and the truncated MF sequences showed a similar trend. These results suggest that the deletion of the 32-55 nucleotide region immediately downstream of the constant 5' primer do not seem to interfere with the binding of magnesium for both MA and MF, nor was there an effect on the rate of catalysis.

3.3.9 Substrate Specificity

To give further insight into the active site of the truncated purine synthase ribozymes, we examined their ability to react with different $^6S$Gua analogues containing modifications at different sites on the base. From my previous work described in chapter 2, MA was shown to exhibit catalytic activity with 6-thiopurine (~4 fold less reactive than $^6S$Gua), suggesting that it can tolerate
Figure 3-8: Substrate specificity of MF, MAa and MFa.

Ribozymes a. MF, b. MAa, and c. MFa were tethered with PRPP (MF-PRPP, MAa-PRPP and MFa-PRPP), incubated with various purine substrates for 1 hour or 15 hours, and separated through APM gels for detection of catalytic activity. Analogues tested were: A = $^{85}$Gua (control), B = 6-thiopurine, C = 2-methyl-6-thiopurine, D = 6-thioxanthine, E = 6,8-dithiopurine, F = 6-thio-9-methylpurine.

certain modifications at the 2 position (Lau, Cadieux et al. 2004). To examine if the parental MF ribozyme and the truncated sequences can tolerate similar or additional modifications, we tethered MF, MAa and MFa with PRPP and incubated them with 6-thiopurine, 2-methyl-6-thiopurine, 6-thioxanthine, 6, 8-dithiopurine and 6-thio-9-methylpurine. Interestingly, we observed a similar trend as all three sequences showed reactivity with 6-thiopurine after incubation for 15 hours, but not with any of the other analogues (Figure 3-8).
3.4 Discussion

3.4.1 Structural Comparison Between Purine and Pyrimidine Nucleotide Synthase Ribozymes

In comparing the predicted secondary structures of the purine versus the pyrimidine nucleotide synthase ribozymes, it appears that the pyrimidine nucleotide synthase ribozymes require a much higher level of structural complexity for mediating catalysis. From our mutagenesis data, the ribozyme isolate MFa can be truncated to 50-nt while retaining catalytic activity comparable to its parent, and requires only a 2 stem-loop structure for functionality. This is in direct contrast to the shortest truncated pyrimidine nucleotide synthase ribozyme of 124-nt, with 5 helical regions (3 of which are pseudoknots), 3 loop regions and 7 interhelical joining regions. The likelihood of finding a sequence with the structural motifs predicted (intrinsic probability) of the purine and pyrimidine nucleotide synthase ribozymes can be calculated by the equation \( P = \frac{r}{4^n} \)

where \( P \) is the intrinsic probability, \( r \) is the number of possible sequences of length \( n \) that satisfies the criteria of the specified structure, and \( n \) is the total length of the sequence (Sabeti, Unrau et al. 1997). The intrinsic probability of finding a sequence with the two structural helical stems flanking the single stranded \{G\}CUUU motif observed in MFa would be \(~1\) in \(10^8\) sequences \((r = 4^9 \times 4^8 \times 6 \times 6 \times 3, n = 35, r \) value determined from 9 sequence independent Watson-Crick base pairs, 8 total arbitrary nucleotides in loop regions, 2 sequence independent Watson-Crick base pairs that in addition can tolerate G-U wobble, and the \{G\} position can be either U, A or C, see Figure 3-6c). This is much more probable than encountering a sequence with the five helical stem structure of the
pyrimidine nucleotide synthase ribozyme about once in every $10^{17}$ sequences (calculated from 28 Watson-Crick base pairs in the helical structures). The intrinsic probability of finding a pyrimidine nucleotide synthase should in fact be significantly less, because we did not take into account any of the constant sequences flanking the stem structures and any base pairs in the stem structures that are sequence specific. Both of these would lead to a lower intrinsic probability, but for the ease of calculation they were ignored. These results are consistent with previous discoveries, using similar *in vitro* selection approach, of at least 53 different purine nucleotide synthase ribozyme families as compared to only 3 for pyrimidine synthase ribozymes, suggesting that there are more solutions to ribozyme-mediated purine nucleotide synthesis (Unrau and Bartel 1998; Lau, Cadieux et al. 2004).

### 3.4.2 Sequence Specificity of Truncated MA and MF Isolates

By changing the sequence composition of the stem and loop regions, we have shown that the structures of MAa and MFa are not sequence dependent. In any of the stem regions for both sequences, at least half of the nucleotide complementary base pairs can be changed to alternative Watson-Crick base pairing while still retaining catalytic activity. Similarly, we observed that the sequence within the loop regions of the two stem-loop structures in MAa and MFa are not sequence dependent and can be modified. These results raise an interesting question. Why did we not observe more sequence variation in these structural regions from our sequencing analysis of the round 6 recombined MA
and MF pools? In MAa, for example, the region within the first stem-loop structure was conserved throughout all of 16 sequences that we have examined.

We believe that the limited sequence variation observed within the structural regions were mainly due to two factors. The first is the low mutagenesis rate that was introduced during the mutagenic PCR amplification step at the end of each round. Since non-Homologous recombination generates diversity mainly through reassembly of pre-existing sequence elements, the majority of individual nucleotide mutations and sequence variations would therefore have to come from the mutagenic PCR step. However, on average per nucleotide we are expecting only a 0.66% +/- 0.13% mutation rate over the course of the PCR. This gives, on average, at any residue position, only one mutation for every 151 nucleotides. Secondly, from our kinetic analysis we observed that the truncated isolates containing wildtype sequences in their stem regions generally had higher reaction rates than mutants that form the same stems but using different base pairing sequences (i.e. MAa reactivity is ~0.09 min⁻¹, as compared to ~0.002 min⁻¹ for the same MAa sequence but with complementary mutagenesis at 4 base pairs in the first stem region). The reaction rates of sequences with wildtype loops were also moderately higher than mutant sequences with GNRA loops. Since sequence variations at structural regions seemingly did not confer any benefits to the ribozymes, and together with the fact that very mild point mutagenesis was implemented during the selection, it is not surprising, therefore, that the isolates selected were biased towards sequences with wildtype fragments containing occasional minor mutations.
3.4.3 Importance of the \{G\}CUUU Motif

The sequence motif NCUUU observed in MF and its truncated sequences appears to be critical for their functionality. According to our re-selection and mutagenesis data, the \{G\}CUUU motif is highly conserved in MF related sequences and is not likely participating in the formation of any secondary structural elements. It maybe, however, involved in the formation of the active site or tertiary interactions with other regions of the RNA. To examine this phenomenon, we attempted to perform cross-linking experiments with both MAa-6S\textsuperscript{G} and MFa-6S\textsuperscript{G} in an effort to examine whether the NCUUU motif is interacting with or in close proximity to the terminal 6S\textsuperscript{G} product. This proved to be difficult, however, as we were limited by weak radioactive signal, due primarily to the weak cross-linking exhibited by both MAa-6S\textsuperscript{G} and MFa-6S\textsuperscript{G} (<1% of products cross-link). We believe that the inability to perform cross-linking experiments may be due to the instability of the ribozyme structures, even at high magnesium concentrations. This is supported by our enzymatic digestion and chemical interference results in which we observed similar digestion and modification patterns in both denaturing and native conditions. We are currently exploring alternative methods for both increasing the RNA labelling signal and the cross-linking efficiency.

So is this all important \{G\}CUUU sequence a universal motif required by all purine nucleotide synthase ribozymes? To address this question, we aligned all of the known families of purine nucleotide synthase ribozymes from the initial selection (not including the truncated sequences in this study) and observed that only 45% (24 out of 53 families) contained the \{G\}CUUU sequence (Lau,
The probability of finding a CUUU motif occurring at least once in a stretch of 95 random positions (average length is 125-nt, removing 30-nt of constant primer regions) would be roughly one out of every three sequences ($\frac{92}{4^4} \approx 0.33$). Of the 24 families with the \{G\}CUUU motif, only half of them contained the motif followed immediately by a terminal stem structure. MA and its truncated sequences such as MAa, for example, do possess the string of \{G\}CUUU sequence, but from our UCUUU randomization re-selection we have shown that this region is not sequence specific for MAa and any of the 5 nucleotides can be varied. Together, this shows that while the \{G\}CUUU motif may be critical for functionality in ribozymes such as MF, it is not a sequence element required universally by all purine nucleotide synthase ribozymes.

### 3.4.4 Different Strategies for Ribozyme-Mediated Purine Nucleotide Synthesis

From our study, it is appears that MA and MF are exploiting different strategies for synthesizing $6^S$G. MA has been shown from the previous report to have low binding substrate affinity but performs a fast chemical step, with an overall efficiency of $284 \text{ M}^{-1} \text{ min}^{-1}$. MF, on the other hand, binds its substrate more tightly ($K_m$ of 50 $\mu$M), reacts at a much slower rate ($k_{cat}$ of 0.027 min$^{-1}$), but has an overall efficiency even higher than that of MA ($k_{apparent \; efficiency}$ of 540 M$^{-1}$ min$^{-1}$) (Lau, Cadieux et al. 2004). When we examined the kinetics of the truncated sequences, we found that MAa was slightly more efficient than MA and similarly was very fast and does not strongly recognize its substrate (as indicated by high $K_m$ value). MFa and further truncated MFa sequences, on the other hand,
did not share kinetic characteristics similar to MF. Taking MFa, for example, we noticed that the deletion of the 56 nucleotides immediately after the 5' constant primer region resulted in a slightly lower overall catalytic efficiency. While the catalytic rate of MFa was actually higher than its parent, we noticed its $K_m$ value was similarly higher as well. When we examined the next truncated sequence tMFa-50, we observed a similar trend with the additional deletion of the 5' constant primer region. While the catalytic rate of this truncated isolate remained the same as MFa, we observed an even higher $K_m$ value resulting in a lower efficiency. From these results, it seems that the truncated MF ribozymes are progressively less efficient and lose their substrate binding affinity with increasing number of deletions. Therefore, the deletion fragments observed in truncated MF sequences may play a role in substrate recognition for MF, even though they may not be required for catalysis. This is consistent with our observations that the truncated MF sequences (MFa, tMFa-50 and tMFa-46) are all tolerant to the $^6S\text{Gua}$ analogue 6-thiopurine as a substrate (Figure 3-8).

### 3.4.5 Non-Homologous Random Recombination as a Tool for Secondary Structural Analysis and RNA Evolution

In our study, we designed two starting pools containing recombined fragments of a parental purine nucleotide synthase ribozyme, hypothesizing that sequence elements that are critical for functionality would be conserved and possibly reinserted in other positions. Interestingly, after re-selection for reactivity with $^6S\text{Gua}$, we did not observe any isolates in either recombined pools that contained obvious sequence fragments added in reverse orientation or located in
a different region along the sequence. This was somewhat surprising considering that random fragment insertions were readily observed in the starting recombined pools. This may in part be due to the short total length of the core motif (i.e. the structure is so packed that any movement of structural sequences will disrupt the overall folding of the ribozyme) or the fact that the sequential order of nucleotides in the wildtype sequence is simply the most superior arrangement possible (i.e. fragment reassembly does not result in more efficient catalysts).

From this work, we have shown that Non-Homologous Random Recombination can be a powerful approach in deducing core sequence elements that are required for catalytic RNAs. It also has the potential to evolve new ribozymes of different functionality but with similar structural motifs relative to a pre-existing ribozyme. An example would be the class of ligase ribozymes isolated by in vitro evolution that contained similar tertiary structure as the large domain of the self-splicing group I intron of Tetrahymena from which they were evolved (Jaeger, Wright et al. 1999). Non-Homologous recombination would favour such selections because it biases the pool to conserve various sequence motifs by reassembling sequence fragments as opposed to introducing random point mutations (Cadwell and Joyce 1992).
3.5 Contributions

The work described in this chapter was primarily conducted by Matthew Lau, with the assistance of Danny Shum in both the synthesis of the starting pools and the selection. This work was conducted under the guidance of Dr. Peter Unrau.
CHAPTER 4: A PROMISCUOUS RIBOZYME MEDIATING
A SUGAR ALDEHYDE REACTION AND AN
UNEXPECTED NUCLEOTIDE SYNTHESIS CHEMISTRY

4.1 Introduction

The ability of RNA catalysts (ribozymes) to synthesize nucleotides as the
basic building blocks for the assembly of RNA polymers is an essential aspect of
the RNA World hypothesis (Fuller, Orgel et al. 1972; Gilbert 1986; Orgel 1998;
Joyce 2002). In chapter 2, we described the isolation of PRPP dependent
pyrimidine and purine nucleotide synthase ribozymes, which could synthesize
either tethered $^{4S}$UMP or $^{6S}$GMP respectively (Figure 4-1a) (Unrau and Bartel
1998; Lau, Cadieux et al. 2004). These findings indicate that both pyrimidine and
purine nucleotides could have been synthesized by RNA catalysts using PRPP
dependent pathways early in evolution. However, the transition from prebiotic to
metabolic synthesis of nucleotides remains a fundamental unsolved problem. For
example, were such substrates readily available in the early world? Work by
different researchers, such as John Oro and Leslie Orgel, suggest possible
prebiotic pathways for the synthesis of both purine and pyrimidine bases (Oro
1961; Sanchez and Orgel 1970). The case for PRPP is different, however, as
the activated sugar hydrolyzes readily in aqueous solution (Trembacz and
Jezewska 1990; Dennis, Puskas et al. 2000), and it is difficult to imagine a
prebiotic source for this important molecule.
Figure 4-1: Ribozyme-mediated nucleotide synthesis pathways studied.

a. The family A pyrimidine synthase ribozyme catalyzes the synthesis of $^{4S}$UMP using tethered PRPP and $^{4S}$Ura as substrates, releasing pyrophosphate in the process (Unrau and Bartel 1998). The purine synthase ribozyme MA, synthesizes $^{6S}$GMP from PRPP and $^{6S}$Gua (Lau, Cadieux et al. 2004). b. In this study, we isolate ribozymes that use $^{6S}$Gua and tethered PR as substrates.

Ribose may be a viable alternative, although its accumulation and significance in the prebiotic world has been vigorously challenged due to its instability under a variety of conditions (Shapiro 1988; Larralde, Robertson et al. 1995). However, recent work by the Benner laboratory has demonstrated that ribose can be synthesized under prebiotic like conditions, with high yield, by the addition of borate (Ricardo, Carrigan et al. 2004). In the absence of a catalyst, ribose can be heated with purine bases under dehydrating conditions to generate a low yield of purine nucleotides (Fuller, Orgel et al. 1972). This observation, along with Benner’s work, suggests that RNA monomers could have existed
transiently in localized regions of the early Earth where both nucleobases and ribose had accumulated abiotically. If such localized concentrations of nucleosides could have given rise to evolving biological systems, whose information encoding and catalytic components were comprised of RNA, then it is plausible that high concentrations of ribose and nucleobases, left over from the abiotic synthesis of nucleotides, would have been available to drive early metabolism.

In support of this hypothesis, we describe in this chapter an *in vitro* selected ribozyme that can easily induce ribose 5-phosphate (PR) to react with 6-thioguanine (6S\text{Gua}), demonstrating the ability of RNA to chemically manipulate this important sugar (Figure 4-1b). This ribozyme likely exploits the reactive aldehyde moiety of the ribose, thus providing an example of a simple anabolic type reaction. Strikingly, we find that substituting PR with the activated substrate PRPP causes the ribozyme to produce tethered 6S\text{G}, even though no selective pressure was applied to generate this particular nucleotide. This unexpected second chemistry indicates that ribozymes making use of simple abiotic compounds like ribose could have quickly switched to the utilization of higher energy metabolites such as PRPP by virtue of the intrinsic promiscuity of their RNA based active sites. This suggests that early RNA based life could have rapidly evolved nucleotide synthesis strategies that resemble those used now by modern metabolism, while at the same time weaning themselves from the scarce supply of abiotically synthesized nucleotides.
4.2 Materials and Methods

4.2.1 Design of the RNA Pool

A random sequence DNA pool was synthesized on an ABI 392 DNA synthesizer using 0.2 \( \mu \)Mole of 2000 Å control pore glass (CPG) columns and standard cyanoethyl phosphoramidite chemistry. The DNA pool had the final sequence: 5'\textbf{-TTC TAA TAC GAC TCA CTA TAG GAG CGA AGT GCC C-N-9s\textit{-TCG CAC CGC AGC AAG C}} (95 random nucleotides, N-9s, flanked by a constant 34-nt 5' primer and a 16-nt 3' primer region, in bold letters; T7 promoter sequence in italics). The 95 random nucleotide couplings was obtained by mixing of equiactive phosphoramidite stocks (ABI) in the molar ratios of 0.28:0.27:0.23:0.22 (dA:dC:dG:dT, respectively) (Zaher and Unrau 2005). Sequencing of the synthesized pool revealed nucleotide frequencies of 25.8% A, 21.1% G, 28.7% C, and 24.4% T in the 95-nt random region. Large scale PCR was performed using the 5' primer: 5'-\textbf{TTC TAA TAC GAC TCA CTA TAG GAG CGA AGT GCC C} and the 3' primer 5'-\textit{GCT TGC TGC GGT GCG A}.

4.2.2 Substrate Ligation and Product Marker Synthesis

Synthesis of adenosine 5'-diphosphoribosyl 1-pyrophosphate (APPRPP) and adenosine 5'-diphospho-6-thioguanosine (APP\textsuperscript{6s}G) was as previously described (Lau, Cadieux et al. 2004). Adenosine 5'-diphosphoribose (APPR) was purchased from Sigma and further HPLC purified through a reverse phase C18 column (Agilent Technologies). Ligations were typically performed using 2 \( \mu \)M RNA with 60 \( \mu \)M APPR or APPRPP for 4 hrs at 23 °C in ligation buffer (50 mM
HEPES, 10 mM MgCl\(_2\), 15 mM sodium phosphite, 10 \(\mu\)g/mL BSA and 8.3% v/v glycerol at pH 8.0) using 0.5 U/\(\mu\)L T4 RNA ligase (Wang and Unrau 2002). Reactions were stopped by the addition of 12 mM EDTA, 300 mM NaCl, followed by a phenol-chloroform extraction and ethanol precipitation. Marker RNA was synthesized by ligation of APP\(^{6S}\)G (at 33 \(\mu\)M) to the end of a 125-nt long RNA having the 3' terminal sequence of ...UCA GAA GAC AUC ACA UUG C-3', which shares the same last 4 residues as pR1, so as to generate a marker sequence containing a terminal \(^{6S}\)GMP. The resulting marker was gel purified through an \(N\)-acryloylamino phenylmercuric acetate gel (APM gel) (Igloi 1988) for use through rounds 1 through 6.

### 4.2.3 Selection

In the first selection round, \(~1.67 \text{ nmols of DNA, containing } \sim 2.5 \times 10^{14}\) distinct sequences, was transcribed in the presence of \(\alpha\)-\[^{32}\text{P}\] UTP using T7 RNA polymerase and purified through a 6% denaturing polyacrylamide gel (PAGE). A total of 2.5 nmols of this initial RNA pool was ligated (at 5 \(\mu\)M concentration) with APPR as described above. RNA derivatized with PR at its 3' terminus (RNA-PR, dash symbolizes 5' - 3' phosphodiester linkage) was incubated at 0.25 \(\mu\)M for 18 hrs at 23 \(^\circ\)C in incubation buffer (50 mM Tris-HCl, 150 mM KCl, 75 mM MgCl\(_2\), pH 7.5) containing 0.25 mM \(^{6S}\)Gua and 10 mM ATP. Reactions were terminated by the addition of 75 mM EDTA and then applied to 30 kDa MWCO centrifugal filters (Amicon) to concentrate RNA and remove free \(^{6S}\)Gua. The resulting RNA, along with the marker RNA in an adjacent lane, was separated through a 6%
denaturing PAGE containing 3.75 µM APM to isolate reactive (thiol-containing) from non-reactive RNA. The gel slice at the position of the marker was excised, eluted in salt solution (300 mM NaCl, 1 mM DTT) overnight and recovered by ethanol precipitation. The eluted RNA was reverse transcribed for 1 hr at 48 °C in RT buffer (50 mM Tris-HCl, 25 mM KCl, 3 mM MgCl₂, 10 mM DTT, 560 µM of dNTPs, pH 7.5) together with 5 µM of 3’ primer using 10 U/µL Superscript II (Invitrogen). The RNA was hydrolyzed using 100 mM KOH at 90 °C for 10 min followed by neutralization to pH 8 by addition of HCl. The resulting cDNA was PCR amplified and used for the next round of selection.

4.2.4 Precise Cleavage of the 3’ End of pR1

pR1 DNA was PCR amplified using the 5’ selection primer and a 3’ primer with the sequence: 5’-TGA TCT CCG CTT GCT GCG GTG CGA, adding an additional 8-nt to the 3’ end of pR1. The amplified product was transcribed into RNA and digested using a 38-nt 10-23 DNAzyme having the sequence: 5’-TGA TCT CCG GGC TAG CTA CAA CGA TTG CTG CGG TGC GA (binding arms in bold) (Santoro and Joyce 1997). The DNAzyme anchors to the RNA by binding with one arm to the 9-nt 3’ terminal sequence (5’ CGG AGA UCA) and the other arm to the upstream 14-nt sequence (5’ TCG CAC CGC AGC AA), resulting in the cleavage between G₁₂₆ and C₁₂₇ of the RNA and the removal of the terminal 9 nucleotides. The resulting 126-nt RNA was purified by 6% PAGE and dephosphorylated for 1 hr at 50 °C (50 mM Tris-HCl, 0.1 mM EDTA, pH 8.5) using 0.5 U/µL alkaline phosphatase (Roche). The reaction was terminated by heating at 65 °C for 20 minutes, followed by a phenol-chloroform extraction.
ethanol precipitation and purification by 6% PAGE. The purified RNA was 3' extended with radiolabelled cytidine 5'-[\(^{32}\)P], 3' bisphosphate (*pCp, * indicates \(^{32}\)P radiolabel) in the presence of 2 \(\mu\)M ATP using the ligation protocol described above. *pCp was generated by phosphorylation of 6 \(\mu\)M of cytidine 3' monophosphate (Sigma) for 40 min at 37 °C in kinase buffer (70 mM Tris-HCl, 10 mM MgCl\(_2\), 5 mM DTT, pH 7.6) with 2 \(\mu\)M of \(\gamma\)-[\(^{32}\)P] ATP using 10 U/\(\mu\)L T4 polynucleotide kinase (NEB). The resulting radiolabelled RNA was finally dephosphorylated to remove the terminal 3' phosphate and gel purified prior to ligation with APPR.

4.2.5 Kinetic Analysis of pR1

The pR1 isolate was transcribed and 2 \(\mu\)M of the RNAs were derivatized with either PR or PRPP as described above. These constructs were then incubated with \(^{65}\)Gua and time points (0, 18, 42, and 90 hrs) were taken. Time points were stopped by the addition of an equal volume of gel-loading buffer (formamide, 10 mM EDTA and Xylene Cyanol). The reaction rate at constant 0.25 mM \(^{65}\)Gua was determined by fitting the fraction reacted for four independent time courses to the equation \(F = \beta(1-e^{-k_{obs}t})\) using the program KaleidaGraph (Synergy Software), with \(F\) being the fraction reacted at time \(t\), \(k_{obs}\) the first-order rate constant, and \(\beta\) being the maximum fraction reacted.

4.2.6 Terminal Enzymatic Extension Assays

pR1 isolate was derivatized with PR or PRPP as described above, reacted for 96 hrs with 0.25 mM \(^{65}\)Gua, followed by an APM gel purification. The pR1-PR
and pR1-PRPP products were then subjected to 3' ligation with *pCp in the presence of 2 µM ATP. Extension activity was determined using an APM gel shift assay to detect the amount of labelled 5'-[^32P] pCp incorporated. For Poly(A) extension assays, 2 µM of pR1-PR, pR1-PRPP reaction products were extended with 2 µM of α-[32P] dATP in PAP buffer (2.5 mM Tris-HCl, 50 mM NaCl, 10 µM DTT, 0.1 mM MgCl₂, 2.5 mM MnCl₂, 10 µM EDTA, 5% v/v glycerol, pH 8.0) using 0.2 U/µL *E. coli* Poly(A) polymerase (Ambion) for 2 hours at 37 °C.

**4.2.7 TLC Analysis of PRPP Product**

Both unlabelled pR1-PRPP and internally labelled MA-PRPP were reacted to completion with 6S Gua, ligated to 5'-[^32P] pCp and purified by an APM gel as described above. The recovered RNA products were digested to completion for 15 hours at 23°C (25 mM sodium citrate, 4 mM DTT, pH 4.5) using 0.26 U/µL T2 ribonuclease (Sigma) (Lau, Cadieux et al. 2004). Digested pR1 products and MA products (reference) were analyzed using thin layer chromatography by spotting on a 10 cm × 10 cm cellulose TLC plate (J. T. Baker) pre-soaked in 1:10 saturated (NH₄)₂SO₄ : H₂O and developed in two dimensions (first dimension 80% ethanol, second dimension solvent contains 40:1 saturated (NH₄)₂SO₄ : 2-propanol; both solvents contain 100 µM β-mercaptoethanol) (Lau, Cadieux et al. 2004). Samples were spotted 1 cm in from each edge at the bottom left corner of the TLC plates.
4.2.8 Purine Analogues

All purine analogues were obtained from Sigma Aldrich and dissolved in 1.05× incubation buffer. 6-thioguanine $\varepsilon$(347 nm, pH 1) = 20,900 M$^{-1}$ cm$^{-1}$ (Elion and Hitchings 1955), 6-thiopurine $\varepsilon$(325 nm, pH 1) = 20,500 M$^{-1}$ cm$^{-1}$ (Elion 1962), 6,8-dithiopurine $\varepsilon$(358 nm, pH 1) = 27,800 M$^{-1}$ cm$^{-1}$ (Robins 1959), 2-methyl-6-thiopurine assumed $\varepsilon$(330 nm, pH 1) = 20,000 M$^{-1}$ cm$^{-1}$ were used at 0.25 mM concentration while analogues 6-thioxanthine, 6-thio-9-methylpurine and 2-thiocytosine were used at saturating concentrations.

4.2.9 Re-equilibration of the Two PR Products

pR1 RNA was dephosphorylated as before, phenol-chloroform extracted, ethanol precipitated and then purified by 6% PAGE. Gel purified RNA was ligated to PR as previously described, followed by a phenol-chloroform extraction and ethanol precipitation. The resulting pR1-PR construct was then 5' labelled by phosphorylation with $\gamma$-[$^{32}$P] ATP using T4 polynucleotide kinase, followed by subsequent reaction with $^{63}$Gua and purification through an APM gel. The two shifted bands (top band product and lower band product) were excised separately, and the two recovered RNA samples were then incubated in incubation buffer containing 0.25 mM $^{63}$Gua for two days to allow re-equilibration, after which they were each loaded in separate lanes and purified through an APM gel. This procedure was repeated twice, with the topmost shifted band always being cut out from the lane loaded with the initial top band product, and the lower band always being excised from the lane loaded with the initial lower band product.
4.2.10 Reduction of Ribose-5-phosphate

Sodium borohydride (EMD Chemicals) was freshly dissolved in water and immediately added at 10 mM concentration to pR1-PR or pR1-PRPP. The RNA was reduced for 1 hour (until bubbling of hydrogen gas ceased) before being subjected to incubation with $^{68}$Gua.

4.2.11 Processing and Derivatization of pR1 with PdR

4.2.11.1 Synthesis of Shortened pR1

pR1 DNA was PCR amplified using the 5' selection primer and a 3' primer with the sequence: 5' CTC CGC TTG CTG CGG TGC GA, adding an additional 4-nt to the 3' end of pR1. The amplified product was transcribed into RNA, digested using a 34-nt 10-23 DNAzyme with the sequence: 5' CTC CGC TTG GGC TAG CTA CAA CGA TGC GGT GCG A (binding arms in bold), resulting in cleavage between G$_{122}$ and C$_{123}$ of the RNA (Santoro and Joyce 1997). The 122 nt fragment was gel purified, removed of terminal cyclic phosphates by T4 kinase, and gel purified a second time as described above.

4.2.11.2 Synthesis of $^*$pCAAGC-PdR

RNA oligonucleotide with the sequence CAAGCdG (CAAGC being ribonucleotides, dG = deoxyguanosine) was obtained from Dharmacon. 2 $\mu$M of the RNA oligo was first radiolabelled with 2 $\mu$M of $\gamma$-$[^{32}P]$ ATP as described above, followed by gel purification to remove unreacted $\gamma$-$[^{32}P]$ ATP. The purified RNA was subjected to depurination by heating at 80 $^\circ$C for 2 minutes in the presence of 200 mM HCl, and subsequently neutralized with NaOH. The
resulting sample was separated through a 23% PAGE and the lower band product (\textsuperscript{*}pCAAGC-PdR, dG depurinated to PdR) was excised and purified. A similar RNA oligonucleotide pCAAGCdG (with a 5' phosphate) was obtained by Dharmacon and subjected to the same depurination and gel purification procedures as described. The resulting RNA sample was sent for mass spectrometry analysis.

4.2.11.3 Generation of Full Length pR1-PdR

To obtain the desired pR1 sequence with a terminal tethered PdR, 5 \( \mu \)M of the shortened pR1 RNA was ligated with 5 \( \mu \)M of purified \textsuperscript{*}pCAAGC-PdR (contains radioactive 5' phosphate) in the presence of 2 \( \mu \)M ATP using the ligation protocol described above. The resulting RNA was separated through 8% PAGE, and the product corresponding to full-length pR1 was excised and eluted (with bodylabelled pR1 RNA loaded in an adjacent lane as a marker). To enhance radioactive signal, the RNA (pR1-PdR) was further radiolabelled by phosphorylation with \( \gamma \)-\([\textsuperscript{32}P]\) ATP using T4 kinase and gel purified as described above, before incubation with \( ^{65}\text{Gua} \) under selection conditions.
4.3 Results

4.3.1 Selection of the pR1 Sequence

The initial goal of our work was to explore the potential of ribozymes to catalyze the synthesis of purine nucleotides starting from PR. We envisioned two feasible mechanisms that could give rise to conventional nucleotide synthesis. The first involved a two-step reaction where PR is first activated to PRPP, using a previously activated compound such as ATP, and then, in a second step, forming a nucleotide by reaction with a base. Alternatively and less likely, ribozymes might exist that can catalyze the synthesis of nucleotides directly from PR and a nucleobase by dehydration, similar to the heating reaction described by Fuller and Orgel between D-ribose and purine bases (Fuller, Orgel et al. 1972).

As the thermodynamics of this process is marginal for purine nucleotide synthesis and difficult for pyrimidine nucleotide synthesis (Tavares, Lee et al. 1987; Bhatia, Vinitsky et al. 1990; Frey and Arabshahi 1995; Xu, Eads et al. 1997; Alberty 2006), the first model was initially favoured.

To examine the importance of ribose activation in ribozyme mediated nucleotide synthesis, we implemented an \textit{in vitro} selection using ribose, ATP, and nucleobase as potential substrates. More specifically, we attempted to isolate ribozymes that, when derivatized with PR at their 3' terminus, were capable of reacting with $^{6S}$Gua in the presence of ATP (Figure 4-2). Isolation of reactive RNA species was made possible by utilization of an APM gel shift strategy, in which active ribozymes containing a thiol group from $^{6S}$Gua were
Figure 4-2: *In vitro* selection scheme for isolation of ribozymes that utilize PR and $^{65}$Gua.

**a.** A random RNA pool was derivatized with PR by ligation with APPR using T4 RNA ligase, releasing adenosine 5'-monophosphate (AMP). **b.** RNAs derivatized with PR were then incubated with both $^{65}$Gua and ATP. ATP was added as a potential source of free energy for the ribozyme reaction. Ribozymes able to react with $^{65}$Gua are distinct from the unreactive RNA population by virtue of their thiol tag. **c.** Reactive sequences tethered to $^{65}$Gua were isolated from non-reactive sequences by separation through a thiol sensitive APM gel (Igloi 1988). A radiolabelled marker derivatized with $^{65}$GMP at its 3' terminus was loaded in an adjacent lane to indicate the position of the gel-shifted ribozymes. The region of the gel containing the active species was excised, eluted and recovered by ethanol precipitation. **d.** The enriched catalytic RNAs were then reverse transcribed to DNA, PCR amplified and then transcribed back into RNA for entry into the next round of selection.
retarded in their migration through a mercury containing gel (Igloi 1988; Unrau and Bartel 1998; Lau, Cadieux et al. 2004). For the selection, we constructed a random pool consisting of $\sim 3 \times 10^{14}$ different sequences. The pool contained 95 random nucleotide positions flanked by constant primer regions at both the 5' and 3' ends (see methods). This pool was derivatized with PR at its 3' terminus using T4 RNA ligase (see methods) and then incubated in a buffer containing 0.25 mM $^6$S Gua and 10 mM ATP.

After 6 rounds of selection, catalytic RNAs were readily detectable using our APM gel shift assay. The shifted products of the same round were purified, cloned and sequenced; out of 39 sequences analyzed, 22 sequences were identical. This dominant sequence, called pR1, was 127-nt in length. It did not share any obviously conserved sequence or helical elements with other purine and pyrimidine nucleotide synthase ribozymes that we have previously isolated (Unrau and Bartel 1998; Chapple, Bartel et al. 2003; Unrau and Bartel 2003; Lau, Cadieux et al. 2004) as judged by MFOLD (Zuker, Matthews et al. 1999). Of the remaining 17 sequences, none occurred more than once or contained any conserved sequence elements relative to pR1. Although these sequences were highly variable, 12 of the 17 sequences did share a common sequence motif “GGUGG” immediately upstream of the constant 3' primer, compared to the “ACUGG” found in pR1 at the same position. Due to the low frequency of these 17 sequences, only pR1 was subjected to further study. The pR1 sequence has been submitted to GenBank with accession number EU267931.
**4.3.2 The Synthesis of Two Thiol-containing Products Requires PR, but Not Triphosphates**

In order to examine the reactivity of pR1, we ligated PR onto its 3’ end using APPR, producing the pR1-PR construct (dash indicates 3’ - 5’ phosphodiester linkage between pR1 and PR) and incubated it with $^{6S}$Gua under the same conditions as during the selection. Interestingly, pR1 produced two distinct but closely migrating thiol-containing products as indicated by APM gel shift analysis (Figure 4-3, lanes 1-4). To establish that the tethered PR was used as a substrate in catalysis, we incubated pR1 with $^{6S}$Gua without prior ligation to APPR. We observed no product formation based on our APM gel shift assay, indicating that PR was required for the reaction to proceed. In addition, ATP was not required as no change in product yield or band pattern was observed when ATP concentration was titrated from 0 mM to 100 mM. Therefore, ATP was not added in the rest of our experiments. To eliminate the possibility that the 5’ triphosphates of pR1 were utilized in the reaction, we enzymatically removed these phosphates by treatment with calf alkaline phosphatase (CIP) after transcription. The dephosphorylated RNA, after reaction with $^{6S}$Gua, exhibited the same two shifted bands, thus confirming that triphosphates were not involved in producing these thiol-containing products.

**4.3.3 Two Observed PR Products are Not Caused by Sequence Heterogeneity or Alternative Folding**

During transcription by T7 RNA polymerase, additional untemplated nucleotides are often added to the 3’ end of the newly transcribed RNA (Milligan, Groebe et al. 1987; Zaher and Unrau 2004). To rule out the possibility that the
Figure 4-3: The substrate dependent reactions of the pR1 ribozyme.

Tethering pR1 with PR and reaction with 0.25 mM $^{65}$Gua produced two distinct bands as a function of time (pR1-PR + $^{65}$Gua lanes). Changing the tethered substrate to PRPP and reacting with $^{65}$Gua resulted in single band that migrated slightly faster than the two pR1-PR product bands (pR1-PRPP + $^{65}$Gua lanes), consistent with the appearance of three distinct bands when mixing the two 90 hr reaction mixtures (mix lane).

Two products resulted from transcriptional heterogeneity, we purified the pR1 RNA through an 8% polyacrylamide sequencing gel, which allowed us to resolve RNA species with single nucleotide resolution. The size purified ribozyme, after tethering with PR and incubation with $^{65}$Gua, again produced two shifted products. To positively confirm the sequence of the active ribozyme, we enzymatically digested and regenerated the 3' end of pR1 so as to guarantee the terminal pR1 sequence. This was achieved by first appending additional
sequence to the 3’ end of the pR1 RNA sequence and then cleaving off this tail using the DNAzyme 10-23 (Santoro and Joyce 1997), leaving the full length ribozyme minus its terminal cytidine residue (see methods). This shortened pR1 RNA was then carefully gel purified and treated with alkaline phosphatase to remove the terminal 2’-3’ cyclic phosphate that resulted from the DNAzyme induced cleavage. To regenerate the missing cytidine, the RNA was then ligated with 5’-[\(^{32}\)P] pCp, followed by a second round of phosphatase treatment to remove the 3’ phosphate added by the incorporation of 5’-[\(^{32}\)P] pCp. After labelling the ribozyme in this way, PR was added to its 3’ terminus by ligation and the construct was incubated with \(^{6S}\)Gua. We again observed two APM dependent bands, confirming that the two thiol-containing products were not due to variation in length or terminal sequence of the ribozyme.

As stable alternative folds of the pR1 sequence could possibly account for the two observed gel mobilities, the pR1 products were digested with T1 ribonuclease, which cleaves after guanosine residues, leaving 5’ hydroxyls on the cut fragments (Pace, Heinemann et al. 1991). These fragments were then radiolabelled using T4 polynucleotide kinase and \(\gamma^{-[^{32}\text{P}]}\) ATP, before analysis through a 23% denaturing gel. As expected, a complex spectrum of short radiolabelled RNA fragments was observed. When the sample was run into a 23% APM gel, we observed two bands that separated from this mixture (data not shown). These two bands presumably correspond to the two thiol-dependent products observed in Figure 4-3 and suggest that the two bands are distinct
chemical species and not denaturation-resistant folded states of the pR1 ribozyme.

4.3.4 pR1 Utilizes PRPP in Addition to PR and Specifically Utilizes $^{6\text{S}}\text{Gua}$

To explore the chemistry mediated by the pR1 ribozyme, the importance of the two substrates in the reaction, $^{6\text{S}}\text{Gua}$ and PR, was examined. pR1-PR was shown to be highly specific towards $^{6\text{S}}\text{Gua}$, as no activity was observed when incubated with 6 different thiol-containing purine compounds (6-thiopurine, 6,8-dithiopurine, 2-methyl-6-thiopurine, 6-thioxanthine, 6-thio-9-methylpurine and 2-thiocytosine) even after incubation for up to 7 days. This is in contrast to a previously characterized purine synthase ribozyme that showed limited reactivity with 6-thiopurine in addition to its $^{6\text{S}}\text{Gua}$ substrate (Lau, Cadieux et al. 2004).

We next examined whether substituting PR with PRPP would give rise to reaction products. To our surprise, a new and distinct third product was observed (Figure 4-3, lanes 6-9), which did not co-migrate with either of the two PR products (Figure 4-3, mix lane). Since previously isolated ribozymes from our laboratory also use PRPP as a substrate for $^{6\text{S}}\text{GMP}$ synthesis, it was possible that the PRPP product resulted from cross-contamination with these ribozymes. To eliminate this possibility, pR1-PR products and pR1-PRPP product were purified separately from an APM gel, reverse transcribed, and subjected to sequencing. 13 sequences derived from the pR1-PR products and 13 sequences from the pR1-PRPP product were analyzed and all of them were found to be the pR1 sequence. This indicates that the pR1 sequence was responsible for synthesizing all 3 distinct products.
The unusual ability of pR1 to react with both PR and PRPP prompted us to explore the possibility that previously isolated nucleotide synthase ribozymes might also be able to utilize both substrates. We tethered PR onto the 3’ end of 4 previously isolated purine nucleotide synthase ribozyme isolates (RA, RE, MA and ME), which all use PRPP to synthesize $^6S\text{GMP}$ (Lau, Cadieux et al. 2004), and subjected them to incubation with $^6S\text{Gua}$ for 4 days. No reactivity was detected, demonstrating that the ability of pR1 to catalyze reactions with both PR and PRPP was not a general property of nucleotide synthase ribozymes.

4.3.5 Only the PRPP Product is a Conventional Nucleotide

To examine if any of the three products synthesized by pR1 was $^6S\text{GMP}$, we generated a marker with a terminal 3’ $^6S\text{GMP}$ and compared its mobility with the ribozyme products in an APM gel. The marker was synthesized by ligation of the pR1 RNA with APP$^{6S}\text{G}$ (see related method). Interestingly, the pR1-PRPP product was observed to exactly co-migrate with the marker when separated through an APM gel, suggesting that the PRPP product was $^6S\text{GMP}$ and that the PR products were not. To give further insight into the three tethered products, we implemented an extension assay using $\alpha-[^{32}\text{P}]\text{dATP}$ and Poly(A) polymerase. Poly(A) polymerase requires the 3’ most nucleotide of the acceptor sequence to contain a free 3’ hydroxyl (Cao and Sarkar 1992; Martin and Keller 1998). Analysis by APM gel indicated that the PRPP product was recognized by the enzyme whereas the two PR products were not (Figure 4-4a). As a positive control, the tethered $^6S\text{GMP}$ product of the purine synthase ribozyme MA (Lau, Cadieux et al. 2004) was also shown to be a substrate for Poly(A) polymerase
To further reinforce these results, we performed another extension assay using T4 RNA ligase and 5'-[^32P] pCp. T4 RNA ligase requires the donor substrate to contain a cyclic ribose sugar and also a free 3' hydroxyl (Barrio, Barrio et al. 1978; England and Uhlenbeck 1978; England and Uhlenbeck 1978). This enzyme, in contrast to Poly(A) polymerase, tolerates modifications at the 2' position of the acceptor terminal nucleotide, but at the expense of ligation efficiency. Consistent with our polymerase extension assay, we observed ligation only with the pR1-PRPP product and not the pR1-PR reaction products. These results together indicate that the pR1-PRPP product is likely to be $^{6S}$GMP, where as the pR1-PR products, which are not recognized by enzymes evolved to modify the termini of regular nucleic acids, are unusual reaction products.

4.3.6 Characterization of the PRPP Product

To determine whether the product of the PRPP reaction was indeed $^{6S}$GMP, we utilized a 3-step labelling, digestion and thin layer chromatography (TLC) analysis strategy that has been used previously to characterize the products of nucleotide synthase ribozymes (Gray 1974; Unrau and Bartel 1998; Lau, Cadieux et al. 2004). pR1 RNA was first tethered with PRPP, reacted with $^{6S}$Gua, purified, and subsequently 3' labelled with 5'-[^32P] pCp to tag the 3' terminal nucleotide. The 3' tagged RNA was purified a second time using an APM gel and the resulting RNA was digested into 3' - monophosphate nucleotides with T2 RNAse. Analysis by two-dimensional TLC revealed one spot (Figure 4-4b, bottom left panel). To determine whether the observed spot corresponded to 3'-[^32P] $^{6S}$GMP, a reference was generated containing
Figure 4-4: The pR1-PRPP, but not the pR1-PR, reaction products are enzymatically recognized and co-migrated with a $^{65}\text{GMP}$ standard in two dimensional TLC analysis.

a. pR1-PR, pR1-PRPP and MA-PRPP ribozymes were incubated for 90 hrs with 0.25mM $^{65}\text{Gua}$. RNA was separated and purified through an APM gel, followed by incubation with $\alpha$-[$^{32}\text{P}$] dATP and Poly(A) Polymerase. After four hours these reactions were loaded on an APM gel. Only the region of the gel corresponding to the thiol dependent shift is shown. b. Purified pR1-PRPP reaction products were 3’ end labelled with 5’-[$^{32}\text{P}$] pCp, followed by a second APM gel purification step. The resulting thiol-containing RNA was digested to mononucleotides using ribonuclease T2 and analyzed by two-dimensional TLC (lower left panel, see methods). A reference mixture containing five radiolabelled 3’ monophosphate nucleotides was also created (upper right panel, see methods). This set of standards was then mixed with the pR1-PRPP product digest. The ribozyme product co-migrated with 3’-[$^{32}\text{P}$] $^{65}\text{GMP}$ as judged by this analysis (lower right panel). The black circles indicate the positions of the TLC origins.
radiolabelled AMP, GMP, CMP, UMP and $^{6S}$GMP (Figure 4-4b, top panel) (Lau, Cadieux et al. 2004). When the digested pR1-PRPP product was mixed with the reference, it co-migrated with the $^{6S}$GMP standard as indicated by both 2-D TLC (Figure 4-4b, bottom right panel) and high percentage APM gel (data not shown), further substantiating the claim that pR1-PRPP synthesizes $^{6S}$GMP.

4.3.7 Characterization of the PR Products

To further characterize the pR1-PR reaction, we excised the two product bands separately from an APM gel (Figure 4-5a). Due to their close mobility in the mercury gel, it was difficult to ensure cross-contamination was completely avoided between the two bands. To address this issue, we loaded two lanes and excised RNA from the top half of the top most product (Figure 4-5a, left most panel, left lane) and RNA from the bottom half of the lower product (Figure 4-5a, left most panel, right lane). We reasoned that this should dramatically reduce the amount of cross-contamination between the two bands. Re-exposure of the APM gel showed that indeed, this procedure did an excellent job of retaining only the desired species in each case. Quantification suggested that the maximum contamination from the other product band was no more than ~20%. The resulting top and lower band samples were eluted and incubated in incubation buffer containing $^{6S}$Gua for a further two days. Surprisingly, both the excised upper and lower products re-equilibrated back to two bands. These products were therefore extracted again, and after a total of three successive excisions of the upper half of the upper band and the lower half of the lower band, we were still able to observe re-equilibration. Re-equilibration of the top band resulted in a
Figure 4-5: Interconversion of the two pR1-PR dependent products.

a. The left most panel shows two lanes loaded with the same sample from a pR1-PR reaction with $^{65}$Gua (labelled as “Initial”). The top product was excised from the left lane, and passaged as follows: eluted RNA was incubated at room temperature for 18 or 32 hours in the presence of $^{65}$Gua before loading onto a fresh APM gel ($T_1$ panel). The top band was then excised again and the procedure was repeated two more times ($T_2$ and $T_3$ panels). The same procedure was performed starting from the lower product band excised from the right lane (left most panel and $L_1$, $L_2$ and $L_3$ panels). Each panel shows the incubation time course and a re-exposure after excision of either the upper (T panels) or lower (L panels) reaction products. The lowest band visible in the T and L panels corresponds to material lacking a thiol tag. b. The 32 hour re-equilibration time points for the upper (T series) and lower (L series) are directly compared. Aliquots were stored at $-80^\circ$C prior to loading.
~2 : 1 ratio of top to lower band while the lower band reformed close to equal amounts of both products (Figure 4-5b). We estimated that after three rounds of excision the maximum contamination between bands if there was no re-equilibration would be < 1% (0.2^3). These results therefore indicate that the two products can interconvert on a relatively short time scale, with the suggestion that the upper band is more stable than the lower. This equilibration was accompanied by the time dependent accumulation of an unshifted band (Figure 4-5a right panels, lower most band), which made up a significant fraction of the total RNA material. This band was produced much faster than would be expected for oxidation of the thiol product (Lau, Cadieux et al. 2004), and appears likely to represent a secondary reaction event that separates the thiol tag from the radiolabelled ribozyme body.

4.3.8 Reduction of Ribose Prevents Product Formation

Next, we chemically manipulated the tethered PR substrate and examined the resulting effect on catalysis. Since the ribose sugar of PR is capable of interconverting between a closed ring conformation and an open aldehyde form (>99% in furanose structure and <1% as aldehyde at equilibrium (Angyal 1969)), we hypothesized that perhaps it was the acyclic form that was responsible for the two observed products due to the intrinsic reactivity of aldehydes with nitrogenous compounds. To partially address this hypothesis, we reduced the tethered PR to ribitol 5-phosphate by treatment with the strong reducing agent sodium borohydride (NaBH₄) (Ricardo, Carrigan et al. 2004). Interestingly, reduction of pR1-PR with NaBH₄ prior to reaction with ^6^SGua resulted in no
product formation (Figure 4-6 lanes 1 to 8). As a control, the same experiment was then carried out with pR1-PRPP (Figure 4-6 lanes 9 to 16). Satisfyingly, this reaction was unaffected by pre-treatment with NaBH₄. This can be explained by the fact that PRPP contains a pyrophosphate at the 1 position, which prevents formation of the linear aldehyde form and consequently its reduction. In the pR1-PRPP reactions, with or without NaBH₄ pre-treatment, we observed the formation of a low yield of the two PR dependent products. This is likely due to the fact that PRPP is not particularly stable and hydrolyzes overtime to either PR or 5-phosphoribosyl-1,2 cyclic phosphate (Dennis, Puskas et al. 2000). Since NaBH₄ reduction is a rapid process, it should therefore only reduce the PR initially present in the reaction mixture, but not the PR formed over time from PRPP hydrolysis. Meola and Sandwick in 2003 have shown that PRPP breaks down at an average rate of ~0.001 hr⁻¹ at neutral pH and that this rate increases ~140 fold with the addition of 32 mM Mg²⁺ at neutral pH (Meola, Yamen et al. 2003). In fact, when the Mg²⁺ concentration was raised to 450 mM, 10% of the original PRPP was degraded in 30 hrs. Since our ribozyme was incubated for 96 hours and makes use of 75 mM Mg²⁺, it was expected that a significant portion of the PRPP would degrade to PR, giving rise to the observed PR dependent products. While not perfectly diagnostic, the suppression of the PR reaction by NaBH₄ suggests that the reactive terminal aldehyde of the linear ribose is involved in pR1-PR dependent chemistry but does not preclude the importance of the ribose cyclic form.
Figure 4-6: Reduction of ribose prevents the synthesis of the PR products.

pR1-PR and pR1-PRPP ribozymes were pre-incubated with (+" columns) or without (-" columns) sodium borohydride. These ribozymes were then incubated with 0.25 mM $^{6S}$Gua for 0, 18, 42 and 90 hours. The diagrams below each time course series represent the state of the tethered substrates prior to incubation with $^{6S}$Gua. As pR1-PRPP is not reduced by sodium borohydride, but can slowly hydrolyze to pR1-PR after the reduction treatment, pR1-PR dependent bands are found in both the "+" and "-" columns for the pR1-PRPP sample.

4.3.9 Deoxyribose is an Alternative Sugar Substrate for pR1

Incubation with analogues of $^{6S}$Gua, as described before, demonstrated that pR1 is highly specific towards $^{6S}$Gua, but did not reveal further details.
regarding the mechanism of the catalyzed chemistry involving PR. Since the
$^6$S Gua analogue experiments gave us limited information, we next attempted to
manipulate the other substrate, the tethered ribose. Deoxyribose 5-phosphate
(PdR) was an excellent candidate because it differs from ribose only at the 2'
position, with a hydrogen group instead of a hydroxyl positioning in the equatorial
position below the ring. One of the obvious approaches was to derivatize PdR at
the 3' terminus of PR1 by ligation with adenylated PdR (APPdR). However, this
compound was not commercially available, and was difficult to synthesize in our
hands. Instead, we performed a multi-step experiment in which we first obtained
a PR1 sequence minus the 5 terminal nucleotides at the 3' terminus (Figure 4-
7a). This was performed similarly as described before by first appending the PR1
sequence with additional nucleotides at the 3' terminus, followed by DNAzyme
cleavage and subsequent treatment with T4 kinase to remove the resulting
terminal 2'-3' cyclic phosphate. To generate the tail sequence of PR1 tethered
with PdR, we obtained a custom-made RNA oligonucleotide CAAGCdG, which
contains a deoxyguanosine nucleotide at the 3' terminus. This oligo was
phosphorylated with $\gamma^-[32]$P ATP, followed by depurination in hydrochloric acid.
Since purine deoxyribonucleotides depurinates at a rate of ~520 fold faster than
purine ribonucleotides (Zoltewicz, Clark et al. 1970), only the guanine of dG was
removed upon acid treatment, resulting in $^p$CAAGC-PdR ($^p$ indicates 5'
radioactive phosphate group, and -PdR indicates tethered PdR). This was
supported by both polyacrylamide gel (a new lower product appeared after 2
minutes), and by mass spectrometry analysis (mass of the purified lower product,
1828.25 u, corresponds to pCAAGC-PdR, as compared to 1960.21 u for pCAAGCdG). The shortened pR1 sequence was then ligated with *pCAAGC-PdR, and analysis by polyacrylamide gel revealed a longer product (pR1-PdR) after ligation that migrated to the same position as full-length pR1 (Figure 4-7a). Interestingly, when we incubated purified pR1-PdR with $^{6S}$Gua we observed the emergence of a mercury-dependent product overtime, which reached a maximum yield of ~3.8%. The PdR product appeared to migrate closer to the bottom PR band than the top PR band when analyzed through an APM gel (compare the lane of 90 hrs for pR1-PR + $^{6S}$Gua (-) to the lanes 18 hrs and 90 hrs for pR1-PRPP + $^{6S}$Gua (-) in Figure 4-7b).

Since PdR can similarly open its sugar ring into an open chain aldehyde conformation like PR, we hypothesize that if the product synthesized involves an aldehyde, it should similarly be shut down by reduction. To examine this phenomenon, we performed the same experiment by pre-treating pR1-PdR with sodium borohydride prior to incubation with $^{6S}$Gua (Figure 4-7b). While the PdR pathway was not completely inhibited, due likely to the lower stability of PdR in the open chain form (i.e. PdR opens its sugar ring less readily than PR) (Munanairi, O'Banion et al. 2007), we did observe a drastic 20 fold reduction in product synthesis relative to untreated pR1-PdR after a 4 day incubation with $^{6S}$Gua.
Figure 4-7: pR1 reaction with $^{65}\text{Gua}$ using PR or PdR as tethered substrates.

a. Tethering pR1 RNA with deoxyribose 5-phosphate. pR1 DNA (with last 14-nt shown) was first 3’ extended by an additional 3-nucleotide through amplification with an extended PCR primer, followed by transcription to RNA. The resulting extended RNA was then digested with a DNAzyme 10-23 designed to base pair with 8-nucleotides on each side of the RNA flanking the cleavage site between G$_{122}$ and C$_{123}$ (see methods). The cleaved RNA was subsequently dephosphorylated to remove the 2’, 3’ cyclic phosphate group remaining on G$_{122}$. To regenerate the full length pR1 RNA with a tethered terminal deoxyribose, the shortened pR1 RNA was subjected to ligation with the RNA oligonucleotide P-CAAGC-PdR, which resulted from acid depurination of P-CAAGC-dG in a prior step. pR1 DNA is indicated in green, pR1 RNA in cyan, DNAzyme 10-23 in red, terminal deoxyguanosine and its depurinated form in blue. P = phosphate group, dG = deoxyguanosine and dash mark indicates tethering.

b. pR1-PR and pR1-PdR ribozymes were pre-incubated with (+’ columns) or without (-’ columns) sodium borohydride. These ribozymes were then incubated with 0.25 mM $^{65}\text{Gua}$ for 0, 18 and 90 hours. The shifted products are shown in this APM gel.
4.3.10 Kinetics

To examine the catalytic rate of pR1, we performed separate time course experiments for pR1-PR and pR1-PRPP. The first order rate constants were obtained at 0.25 mM $^{6S}$Gua concentration by plotting the fraction reacted versus time for 4 independent time courses (see methods). The measured first order rate of synthesis for pR1-PR ($\sim 0.03 \text{ hr}^{-1}$ to 0.07 hr$^{-1}$) was ~2 fold slower than the synthesis rate for pR1-PRPP ($\sim 0.06 \text{ hr}^{-1}$ to 0.1 hr$^{-1}$). To examine the uncatalyzed rates, a 5-nt and a 125-nt RNA were ligated to either APPR or APPRPP. These RNAs were then incubated with $^{6S}$Gua for 7 days. Assuming that we could detect a peak twice the background intensity upon exposure to a phosphorimager screen, the complete absence of an APM dependent gel shift implied an uncatalyzed rate of less than $2.5 \times 10^{-6} \text{ hr}^{-1}$ at 0.25 mM $^{6S}$Gua concentration. The catalytic rate enhancement by pR1 for either the PR or PRPP dependent chemistry is therefore is at least $10^4$ fold faster than the uncatalyzed rate. We next examined the magnesium requirement for pR1 by performing a Mg$^{2+}$ titration experiment ranging from 0 mM to 75 mM (Figure 4-8). Interestingly, the PR products were detectable at 1 mM Mg$^{2+}$ concentration, whereas 10 mM was required for the formation of the PRPP product. The finding that pR1-PR was less Mg$^{2+}$ dependent than pR1-PRPP suggests the possibility that metal ions are utilized differently for each reaction. In both cases, however, the optimal Mg$^{2+}$ concentration appears to plateau at 10 mM Mg$^{2+}$ and increasing Mg$^{2+}$ beyond 10 mM did not significantly enhance the rate or the fraction of products synthesized for either reaction.
Figure 4-8: Magnesium dependence of pR1 ribozyme.

a. pR1-PR and b. pR1-PRPP reactions with $^{65}$Gua held at 0.25 mM. Time course for 0, 18 and 90 hours were conducted with magnesium concentrations at 0, 1, 10, 25, 50 and 75 mM.
4.4 Discussion

4.4.1 Catalysis of Two Sequential Distinct Chemical Steps is Difficult for *In vitro* Selected Ribozymes

Nucleoside formation from ribose and a purine or pyrimidine base in a single step is thermodynamically unfavourable, having a $\Delta G^\circ$ of $\sim 7$ kJ mol$^{-1}$ (Alberty 2006) or $\sim 28$ kJ mol$^{-1}$ (Tavares, Lee et al. 1987; Bhatia, Vinitsky et al. 1990) respectively. Understanding the difficulty of synthesizing $6^S$G from an unactivated sugar, we added ATP during our selection, hypothesizing that a thermodynamically favourable two-step reaction might be selected where PR is first activated to PRPP by ATP and then in a second step converted to $6^S$GMP when incubated with $6^S$Gua (overall $\Delta G^\circ$ of $\sim -38$ kJ mol$^{-1}$) (Frey and Arabshahi 1995; Xu, Eads et al. 1997). While thermodynamically feasible, this strategy requires two distinct catalytic activities to be contained within a single RNA sequence. As ribozymes and DNAzymes catalyzing sequential chemical steps have been previously isolated (Pan and Uhlenbeck 1992; Huang and Yarus 1998; Brown, Li et al. 2003; Zaher and Unrau 2006), we believe that a two-step purine nucleotide synthase ribozyme might not be impossible to discover in random sequence RNA pools, depending on the information required for specification (Szostak 2003). Each motif would act sequentially to produce $6^S$GMP in two distinct chemical steps. These two RNA motifs are most easily imagined as distinct non-overlapping entities (Soukup and Breaker 2000; Kumar and Joyce 2003), but could possibly consist of alternative folds of the same primary sequence (Schultes and Bartel 2000). In either case, the inability of the pR1 ribozyme to utilize ATP as an intermediate in nucleotide synthesis illustrates
the difficulty of isolating a ribozyme, from a random RNA pool, that can mediate two distinctly different and metabolically relevant chemical steps.

### 4.4.2 Ribozyme Promiscuity

Alternatively, as has been argued for the evolution of new protein enzyme function (O’Brien and Herschlag 1999), catalytic promiscuity might have played an important role in the evolution of ribozymes mediating small molecule chemistries. While many ribozymes exhibit good substrate discrimination, there are small molecule ribozymes that are remarkably promiscuous in their substrate requirements. For example, two structurally distinct capping ribozymes both appear to utilize a covalent intermediate for their phosphate dependent chemistry (Huang and Yarus 1997; Zaher and Unrau 2006; Zaher, Watkins et al. 2006). As a consequence, both ribozymes accept a broad range of nucleotide substrates equally well (Huang and Yarus 1997; Zaher, Watkins et al. 2006). In addition, some, but not all, nucleotide synthase ribozymes exhibit a tolerance for reacting with alternative nucleobases. The family A \(^{4S}\)UMP synthase reacts \(~10,000\) fold slower with uracil than its \(^{4S}\)Ura selection substrate and a \(^{6S}\)GMP synthase reacts \(~1,000\) times more slowly with 6-thiopurine than \(^{6S}\)Gua (Unrau and Bartel 1998; Lau, Cadieux et al. 2004). This suggests that it is possible to isolate ribozymes that evolve in substrate preference, while retaining a common tertiary structure and mechanism. However, while attempts to select a \(^{6S}\)GMP synthase from a RNA pool biased towards a \(^{4S}\)UMP synthase ribozyme structure were successful, no evidence was found for secondary structure conservation between the \(^{6S}\)GMP synthases and their progenitor (Lau, Cadieux et al. 2004). Likewise, pR1 did not
show any sequence or secondary structure conservation relative to any of the previously characterized nucleotide synthase ribozymes. This is not surprising, however, as pR1 was selected to utilize a ribose substrate with different chemical composition than PRPP for catalysis of an alternative PR dependent chemistry.

A more commonly observed form of catalytic promiscuity involves hydrolytic reactions that are typically viewed as undesirable side effects of in vitro selection. Many ribozymes catalyzing small molecule chemistries are also known to accelerate the hydrolysis of their activated substrates (Piccirilli, McConnell et al. 1992; Huang and Yarus 1998; Unrau and Bartel 1998; Zaher and Unrau 2006). Similarly, both the lead-dependent ribozyme (termed leadzyme) (Pan and Uhlenbeck 1992) and the 17E variant of the 8-17 DNAzyme, can efficiently catalyze the hydrolysis of the 2', 3'-cyclic phosphate resulting from the initial phosphodiester bond cleavage reaction (Brown, Li et al. 2003). A notable exception in this regard is the finding that an RNA aptamer selected to bind the dye malachite green also has the intrinsic ability to accelerate the acid driven hydrolysis of an acetoxy derivative of the dye by ~1000 fold (Brackett and Dieckmann 2006). The pR1 ribozyme, as compared to the catalytic nucleic acids described above, is unique in that it can mediate an alternative metabolic relevant chemistry more efficiently than the distinctly different activity it was selected to perform.

4.4.3 Promiscuity of the pR1 Ribozyme

The unanticipated PRPP dependent nucleotide synthesis chemistry mediated by pR1 is surprisingly efficient (~5 M\(^{-1}\) min\(^{-1}\)) and is ~2 fold faster than
PR dependent chemistry that enabled it to survive *in vitro* selection. This nucleotide synthesis rate is very similar to the activity of pyrimidine nucleotide synthase ribozymes selected from random sequence (~4 M\(^{-1}\) min\(^{-1}\)) (Unrau and Bartel 1998) and is only ~50 times slower than PRPP dependent purine nucleotide synthase ribozymes selected directly for \(^{6}\)GMP synthesis (~250 M\(^{-1}\) min\(^{-1}\)) (Lau, Cadieux et al. 2004). This similarity in nucleotide synthesis rates is quite unexpected and indicates that the mechanism of the PR dependent chemistry in the pR1 ribozyme must be such that the PRPP secondary reaction can proceed smoothly.

The promiscuity demonstrated by pR1 towards different sugars prompted us to explore other closely related PR analogues, such as deoxyribose 5-phosphate. Interestingly pR1 not only can use PdR as a substrate, but the catalyzed reaction also exhibited ~10-fold higher product yield than the PR reaction. This result, however, may be misleading because the pR1-PdR sequences were synthesized from the ligation of short pR1 sequences with pCAAGC-PdR, and after gel purification, all of the resulting pR1 RNAs would be derivatized with PdR. Synthesis of pR1-PR, on the other hand, required the ligation with APPR using T4 RNA ligase, and the efficiency of PR incorporation in this step has not been fully characterized. It is estimated to be less than 30%, as demonstrated in ligation of RNAs with adenylated constructs, under our ligation conditions. There are many other factors, however, that affect the ligation process, including the terminal nucleotide of the donor sequences, as well as the properties of the acceptor substrate (England and Uhlenbeck 1978).
Nonetheless, the ability of pR1 to use PdR for catalysis further demonstrates that this ribozyme, while demonstrating \textsuperscript{65}Gua specificity, is highly promiscuous towards the sugar substrate.

\textbf{4.4.4 Likely Mechanism of pR1-PR, pR1-PdR and pR1-PRPP Chemistries}

The two thiol-containing reaction products formed by pR1-PR are consistent with the Schiff base chemistry involving the open chain aldehyde form of ribose (Figure 4-10a). Different reducing sugars, including glucose, ribose and PR, are well known to react with amino acids and nucleotides in the Maillard reaction, a process commonly observed in the browning of foods (Nissl, Ochs et al. 1996; van Boekel 2006; Munanairi, O'Banion et al. 2007). The first step is a condensation reaction, in which an amine group acts as a nucleophile, reacting with a sugar carbonyl to form a carbon-nitrogen bond (Sandwick, Johanson et al. 2005; van Boekel 2006). The Schiff base of the resulting Maillard product then undergoes an isomerization called the Amadori rearrangement to form ketosamines, which overtime degrade to various glycation end products (Smith, Taneda et al. 1994).

In our re-equilibration experiment, we observed that the lower-band reaction product (Figure 4-5a) is less stable than the upper-band product and appears to convert preferentially to the upper-band product over time. This could imply that the initial Schiff base (lower band) converts to the Amadori product (top band), which over time degrades, resulting in the accumulation of material lacking a thiol tag (Figure 4-5a, extreme bottom band). This interpretation is in line with the observation that upon reduction by sodium borohydride of the
Figure 4-9: Reduction of the lower PR band product.

The left most panel shows 3 lanes loaded with the same sample from a pR1-PR reaction with $^6$S-Gua (labelled as “Before Cut, BC”). The bottom product was excised from all 3 lanes, and passaged as follows: eluted RNA was reduced by sodium borohydride for 30 minutes at room temperature, incubated for 32 hours in the presence of $^6$S-Gua, ethanol precipitated to remove any remaining reducing agent and loaded onto a fresh APM gel (middle panel). The bottom band was then excised again and the procedure was repeated, but without reduction by sodium borohydride. The resulting sample was loaded through a third APM gel for analysis (right panel). For the left and middle panel, the gel was re-exposed and shown on the right after excision of the bottom band product ("After Cut", AC).

bottom band product after purification from an APM gel (which would lock it down in the Schiff base conformation), a more dominant bottom band product is observed when separated through a second APM gel (Figure 4-9, middle panel). By cutting the bottom band product again and analysis through a third APM gel, we once again see an increase in the ratio of the bottom to top band products.
(Figure 4-9, rightmost panel). Obtaining clear results from this experiment has been difficult, due to the lost of RNA (radioactive signal) over the multiple reduction and purification steps (i.e. after reduction the sample must be purified in a series of steps to remove any remaining reducing agent prior to loading onto a fresh mercury gel), and the close separation between the two band products. While we were unable to see complete inhibition of the upper band in our mercury gel analysis, due likely to the contamination of the top product during gel excision caused by smearing, we did observe an accumulation of the lower band over a series of reduction and re-equilibration iterations. This suggests that by reducing the Schiff base (bottom band), we have prevented re-equilibration to the Amadori product (top band). This reaction mechanism is in distinct contrast to the reaction with PRPP, which generates a nucleotide glycosidic linkage based on our TLC and enzymatic recognition data (Figure 4-4b).

The reactivity of pR1 using PdR as a tethered substrate is also consistent with the predicted Maillard reaction. PdR, which contains a proton instead of a hydroxyl group at the 2 position of the sugar, can react in similar fashion with N-9 of $^{65}$Gua upon opening of its sugar ring (Figure 4-10b). The resulting Schiff base, however, cannot undergo further rearrangement to the more stable Amadori compound due to the absence of the hydroxyl at the second carbon (Munanairi, O’Banion et al. 2007), resulting in the synthesis of a single product, as indicated from our mercury gel shift assay. We also observed similar inhibition effects upon pre-treatment of pR1-PdR with sodium borohydride, which strongly suggests that an aldehyde chemistry is involved in the synthesis of both PdR and PR products.
(Figure 4-7). Comparing the relative shifted positions in a mercury gel, the PdR product appears to migrate in closer proximity to the lower PR band than the upper band, suggesting that the PdR product may also be the consequence of a Schiff base formation (Figure 4-7b). We attempted to mix the PdR product with those formed from the PR reaction together to examine for possible co-migration, but due to the close separation of the band products, we were unable to draw any clear conclusions. It is important to note, however, that we do not expect exact co-migration of the PdR product with any of the PR bands, due to the difference in chemical composition of the products.

Mechanistically, the simplest hypothesis consistent with our data is that the pR1 ribozyme uses charge stabilization to enhance the PR and PRPP reactions (Figure 4-10). The PR reaction likely involves the reactive carbonyl of the open chain ribose, consistent with the complete absence of activity observed upon its reduction. As charge stabilization is known to enhance Maillard chemistry, with PR reacting appreciably faster than ribose (Sandwick, Johanson et al. 2005), it is plausible that the pR1 ribozyme utilizes the negative charge available from its phosphodiester backbone to stabilize the positively polarized charge at the carbonyl carbon, so as to encourage the acyclic reaction (Sandwick, Johanson et al. 2005) with a nucleophile on the $^{6S}$Gua substrate (Figure 4-10a). Moreover, since the ribozyme utilizes the N-9 of $^{6S}$Gua for its PRPP dependent chemistry, and if we assume that the ribozyme does not reorganize its active site when either PR or PRPP is tethered, it appears likely that the N-9 of $^{6S}$Gua is involved in either chemistry (as suggested in Figure 4-
Figure 4-10: Proposed reaction mechanisms for the synthesis of different pR1 products.

a. In the pR1-PR reaction, the negative charge present in the phosphate group from the phosphodiester backbone can stabilize the carbonyl present in the acyclic form of ribose. This would encourage nucleophilic attack from the N-9 of $^{65}$Gua to the carbonyl group, resulting in the formation of an imine (Schiff base). The resulting product can then undergo an Amadori rearrangement to give a ketosamine product. b. In the pR1-PdR reaction, the deoxyribose sugar can similarly undergo acyclization and react with $^{65}$Gua to form a Schiff base. However, since PdR does not have a 2' hydroxyl group, it cannot form the enaminol intermediate and undergo isomerization to give the more stable Amadori product. c. In the pR1-PRPP reaction, the phosphate group can similarly stabilize the carbocation of the ribose that is formed upon leaving of the pyrophosphate, resulting in the nucleophilic attack and formation of the glycosidic bond with the N-9 of $^{65}$Gua. $\delta^+$ indicates partial positive charge.
10). As charge stabilization of the reactive oxocarbenium ion intermediate found at the anomeric carbon of the furanose ring is well known to be utilized by glycosidic bond cleaving enzymes (Dinner, Blackburn et al. 2001) and is likely to be involved in ribozyme mediated pyrimidine nucleotide synthesis (Figure 4-10c) (Unrau and Bartel 2003), we propose that the catalytic promiscuity achieved by the bifunctional pR1 ribozyme results from a specific pattern of negative charges that can stabilize both reactions. This active site, which accelerates both PR and PRPP dependent reactions, is sensitive to the additional negative charge from the pyrophosphate of PRPP and requires increased levels of magnesium to perform the nucleotide synthesis chemistry maximally (Figure 4-10c). It is interesting to note that examples of both in vitro selected DNAzymes and ribozymes have similarly been shown to use charge stabilization for catalysis (Li and Sen 1996; Travascio, Li et al. 1998; Brackett and Dieckmann 2006).

4.4.5 Evolutionary Significance of Ribozyme Small Molecule Substrate Promiscuity

The promiscuous chemistry of the pR1 ribozyme demonstrates how a ribozyme can immediately switch its substrate utilization from a simple metabolite such as PR to that of PRPP without a single change to its primary sequence. As supplies of nucleotides on the early Earth were likely to have been extremely limiting, ribose and its simple derivatives could have played an important role in early metabolism (Ricardo, Carrigan et al. 2004). If early RNA based replicative systems were able to efficiently utilize ribose to react with abiotic nucleobases, not only would they have been able to generate compounds of metabolic
importance, but they could also have provided a means for increasing the local concentration of both ribose and nucleobase substrates. Such a process, combined with the assumption that early life was encapsulated, would enhance the rate of chemical reactions within an early cell and provide a means to transiently store important metabolites (Bartel and Unrau 1999; Szostak, Bartel et al. 2001). If such a metabolism, which has not been shown, can naturally lead to the synthesis of PRPP, then the synthesis of nucleotides, as demonstrated here by the surprising catalytic flexibility of the pR1 ribozyme, could have immediately emerged as an intrinsic property of an early RNA based catalytic system.
4.5 Contributions

The work described in this chapter was conducted by Matthew Lau under the guidance of Dr. Peter Unrau.
CHAPTER 5: CONCLUDING REMARKS AND SIGNIFICANCE OF WORK

5.1 Discovery of Ribozymes Mediating Efficient Purine Nucleotide Synthesis by \textit{In vitro} Selection

We have successfully isolated ribozymes that are capable of promoting the synthesis of a purine nucleotide ($^{6S}$G) from tethered PRPP and $^{6S}$Gua using an \textit{in vitro} selection technique. The two most proficient purine synthases, annotated as RA and MA, have apparent efficiencies of 230 and 284 M$^{-1}$min$^{-1}$, respectively. Compared to previously selected pyrimidine nucleotide synthase ribozymes, these ribozymes are 50-100 times more efficient. Interestingly, RA has a $K_m$ of $\sim$80 $\mu$M for 6-thioguanosine, whereas MA has no detectable affinity for this molecule, indicating that considerably different forms of substrate recognition are utilized by the two ribozymes. Both ribozymes showed good substrate discrimination against 6-thioguanosine analogues, with only a slight tolerance for substitutions at the 2 position. It is of interest that the protein enzyme HGPRTase, and a naturally occurring aptamer found in the \textit{xpt-pbuX} mRNA also discriminate weakly against guanine substrates varying at the 2 position. Analysis by TLC showed that the fastest ribozyme isolates, when incubated with $^{6S}$Gua, do indeed produce tethered 6-thioguanosine.

Our work here demonstrates the important trade-off parameter of substrate recognition and catalytic rate, as observed in ribozyme MA and RA, which must be considered for ribozymes mediating small molecular chemistry. To
synthesize nucleotides most efficiently in an early environment rich with $^6S$Gua, the fast MA-like ribozymes would be favoured since substrate specificity is not important in this case. On the other hand, if the environment was abundant in other competing substrates (i.e. 6-thiopurine) and scarce in $^6S$Gua, the substrate specific RA-like ribozymes would be favoured since the synthesis of the correct nucleotides are more important in this case.

The success of this project has resulted in the discovery of the first known purine nucleotide synthase ribozymes, which mimic the cellular protein phosphoribosyltransferases involved in critical PRPP-dependent nucleotide synthesis pathways. The work presented in this chapter, in conjunction with previous work on pyrimidine nucleotide synthase ribozymes, demonstrates the ability of ribozymes to synthesize both purine and pyrimidine nucleotides. This strongly suggests that ribonucleotides and larger RNA molecules could have preceded proteins in a hypothetical RNA World.

### 5.2 Characterizing Purine Nucleotide Synthase Ribozymes by Non-Homologous Random Recombination

To gain further insight into the nucleotide synthesis chemistries mediated by RNAs, we deduced the minimal core motifs and secondary structures of two purine synthase ribozymes, and compared their structural complexity with the previously isolated pyrimidine synthase ribozymes. This was performed by Non-Homologous Random Recombination on two previously isolated purine nucleotide synthase ribozyme isolates, MA and MF, and subsequent re-selection for activity with $^6S$Gua. The recombined MA and MF pools were constructed such
that each pool contained randomly recombined sequence elements relative to its parent. The two pools were selected for highly truncated sequences that were able to promote $6S\text{G}$ synthesis. The selection resulted in the isolation of short reactive species from both pools containing a 32 to 55-nt deletion region relative to each of the two 126 and 124-nt long parents. The shortest sequence observed was further truncated with the result being a final 50-nt sequence that still had activity comparable to its parent. By site-directed mutagenesis, we have determined that 3 stem loop regions for MA, and 2 stem loop regions for MF, are critical for functionality.

Our results from this work illustrate the overall simplicity of both sequence and secondary structural requirements for purine nucleotide synthase ribozymes. Compared to a previously isolated pyrimidine nucleotide synthase ribozyme, we observed a drastic difference in structural complexity. For the purine synthase ribozymes, only 2 stem loop motifs were required for functionality, as compared to 5 helical regions (3 of which are pseudoknots) for their pyrimidine synthase counterpart. The intrinsic probability of finding a sequence with minimal purine synthase ribozyme structure was correspondingly higher ($\sim 1$ in $10^8$ sequences) as compared to the pyrimidine synthase ribozymes ($\sim 1$ in $10^{17}$ sequences).

Together, along with observations described in chapter 2, these results suggest that ribozyme-mediated purine nucleotide synthesis is much simpler and far more superior in efficacy as compared to pyrimidine nucleotide synthesis.

From our selection and further truncation experiments, we deduced a functional sequence that was only 50-nt in length, while retaining similar catalytic
efficiencies relative to its parental MF ribozyme. The size of this particular ribozyme is comparable to some of the shortest known ribozymes, such as the 43-nt full length hammerhead ribozyme (Martick and Scott 2006). The truncated ribozymes isolated from this work, because of their short length, may be an excellent starting point for the evolution of ribozymes with alternative functions, or for the engineering of multi-functional catalytic nucleic acids. Their sequence simplicity may also be advantageous for future crystallography studies, which will provide further details as to their mechanism of action and the factors contributing to their catalytic potential in mediating purine nucleotide synthesis.

5.3 Isolation of a Promiscuous Ribozyme Catalyzing Both Purine Nucleotide Synthesis and an Unusual Sugar Chemistry

To address the ability of RNAs to catalyze a reaction between an unactivated sugar and a purine base, we implemented a selection using a high diversity RNA pool tethered to PR at its 3' terminus and successfully isolated the ribozyme pR1. Interestingly, this ribozyme isolate did not require ATP as an external energy source for catalysis and can produce 3 different products: two when tethered with PR, and a third product when tethered with PRPP. The third product from PRPP was completely unexpected and judged by gel electrophoresis and TLC analysis to correspond to 6-thioguanosine. Extension assays with Poly(A) polymerase and T4 RNA ligase revealed that only the PRPP product, but not the two PR products, was capable of 3' extension. Since 3' extension by both enzymes require a 3' terminal nucleotide, this suggested that the PR products formed were likely not conventional nucleotides. Further
characterization revealed that the two PR derived products could reassert between each other, and upon reduction of the ribose prior to reaction with $^6$S Gua, neither of these two products form. Interestingly, pR1 can also use PdR as an alternative substrate to form one product, and a similar inhibitory effect was observed upon prior reduction of PdR. This indicates the involvement of the sugar aldehyde in its open chain conformation in the reaction. Our results suggest that pR1 catalyzes a Maillard chemistry through the PR dependent pathway, forming a Schiff base as one of the products, which can subsequently undergo an Amadori rearrangement to generate the second product.

The pR1 ribozyme is the first example of an in vitro selected bifunctional ribozyme that shows both substrate and catalytic promiscuity in its active site. pR1 faces the challenge of committing to either one of the two chemistries, and its fate is ultimately dependent upon which of the substrates, PR or PRPP, is present in the active site at the time of the reaction. Catalytic promiscuity, as we propose, is achieved by charge stabilization of either the carbonyl of the open ribose chain to promote the aldehyde chemistry, or with the oxocarbenium ion intermediate to promote the glycosidic linkage between the furanose and the base. As this ribozyme was only selected for using PR as a substrate, its ability to also utilize PRPP in a different chemistry illustrates how latent functions in ribozymes may be of importance, and as such, provides a new paradigm for the study of ribozyme evolution.
BIBLIOGRAPHY


