RNA-MEDIATED CHEMISTRIES: A CASE OF REPLICATION AND CAPPING

by

Hani Zaher

B.Sc., Simon Fraser University, Canada, 2002

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

DOCTOR OF PHILOSOPHY

in the Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Spring 2007

Approval

Name	Hani Zaher
Degree	Doctor of Philosophy (Science)
Title of thesis	RNA-mediated chemistries: A case of replication and capping
Examining committee	Chair: Dr. Edgar Young

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

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

Dr. Michel R. Leroux
Associate Professor of Molecular Biology and Biochemistry
Committee Member

Dr. David J. Vocadlo
Assistant Professor of Chemistry
Internal Examiner

Dr. Steven A. Benner
Director and Distinguished Fellow
Foundation for Applied Molecular Evolution
External Examiner

Date Defended: March 22, 2007

Abstract

Our current understanding of biology suggests that early life relied predominantly on RNA for both catalysis and the storage of genetic information. Tremendous efforts have been undertaken to confirm that the catalytic abilities of RNA can sustain this "RNA World". However, the ability of ribozymes to utilize certain chemistries known to be critical for present-day biology, and hence likely to be important for ribo-organisms, has yet to be demonstrated.

Numerous natural and artificial ribozymes have been shown to facilitate reactions that invert stereochemistry at phosphorous reaction centres. I have isolated and characterized an RNA-capping ribozyme that retains stereochemistry during the synthesis of a 5′-5′ RNA cap. Stereochemical data determined using thio-phosphate modifications, together with an early rate-limiting step, suggest that this ribozyme utilizes two distinct inverting chemical steps - proceeding via a ribozyme-covalent intermediate - during catalysis.

A ribo-organism, like contemporary organisms, would have required ribozymes to replicate an RNA genome within some sort of primordial compartment. To provide support for this requirement, I have isolated an improved RNA polymerase ribozyme, referred to as B6.61, from a mutagenized pool containing ~9×10^{14} different sequences using a novel large-scale *in vitro* compartmentalization system. B6.61 polymerized all tested primer-template (PT) complexes faster than its parent, the Round-18 polymerase ribozyme. For one PT complex a rate enhancement of more than 80 fold was observed for extensions longer than one helical turn. The new variant also exhibited improved

fidelity on a number of templates. Most interestingly, B6.61 was found to copy one PT complex by almost two complete turns of an RNA helix.

To further study RNA-mediated replication, I investigated the role of nucleic acid structure during the course of oligonucleotide extension by T7 RNA polymerase. In addition to normal transcription, the enzyme can produce anomalous transcripts in the absence of a promoter. I have found oligonucleotides that are able to form transient unimolecular loop structures closed by as little as one base-pair to be viable substrates. This intermittent extension process was found to be quite efficient, and adds to the understanding of viral RNA replicative strategies.

Keywords: RNA world; ribozyme; capping; stereochemistry; retention; covalent intermediate; replication; polymerization; extension; fidelity; viral replication; viroid; *in vitro* selection; compartmentalization

To Nabil, of course

Acknowledgments

I would like to express my sincere gratitude to my senior supervisor Dr. Peter Unrau for his guidance, inspiration, patience, and trust during my graduate career. I am also grateful to my committee members, Dr. Dipankar Sen and Dr. Michel Leroux, for the encouragement, helpful advice, and useful comments on the thesis and other manuscripts.

I also thank the internal examiner, Dr. David Vocadlo, for the very useful help and suggestion on enzyme mechanisms. I appreciate the time taken by the external examiner, Dr. Steven Benner, out of his busy schedule to attend my defence.

I am also thankful for the friendship and support of Edward Leung, and also for the useful comments and advice on earlier versions of this thesis. I am grateful to Tim Le Fevre for his suggestions and comments on the thesis, to R. Ammon Watkins for his assistance in the capping project, and all other members of the Unrau and Sen laboratories, past and present, for their help and advice. I also wish to thank Sameh Al-Natour, Ali Ali, Hasan Cavusoglu, and Vibhore Sharma for their friendship and support. I thank the staff in the Department of Molecular Biology and Biochemistry for their help.

I am very grateful for the endless support by Carrie during the past two years, and her help in putting this thesis together.

Finally, needless to say, my family for everything.

The work presented in this thesis was supported by grants and fellowships from the following agencies, Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Michael Smith Foundation for Health Research, and Simon Fraser University.

Table of Contents

List of Tables

List of Figures

Abbreviations

^{4S}Upppp$_s$G:	4-thiouridine(5′)α-thiotetraphospho-(5′)guanosine
^{4S}UTP:	4-thiouridine triphosphate
APM:	N-acryloyl-aminophenyl-mercuric acid
AppppG:	Adenosine(5′)tetraphospho-(5′)guanosine
Apppp$_s$G:	Adenosine(5′)α-thiotetraphospho-(5′)guanosine
ATP:	Adenosine 5′- triphosphate
bp:	Base-pair
CIP:	Calf intestinal phosphatase
CTP:	Cytidine 5′- triphosphate
DEPC:	Diethyl pyrocarbonate
DNA:	Deoxyribonucleic acid
dNTP:	Deoxynucleotide 5′- triphosphate
dsDNA:	Double-stranded DNA
EDTA:	Ethylenediaminetetraacetic acid.
GTP:	Guanosine 5′- triphosphate
HPLC:	High performance liquid chromatography
Imp ^{4S}U:	4-thiouridine 5′- phosphorimidazolide
Imp A:	Adenosine 5′- phosphorimidazolide
IVC:	*In vitro* compartmentalization
NDP:	Nucleotide 5′- diphosphate
NMP:	Nucleotide 5′- monophosphate

nt:	Nucleotide
NTP:	Nucleotide triphosphate
PAGE:	Polyacrylamide gel electrophoresis
PCR:	Polymerase chain reaction
PNK:	Polynucleotide kinase
PT:	Primer Template
Rib:	Ribozyme
RNA:	Ribonucleic acid
ssDNA:	Single-stranded DNA
T7 RNAP:	T7 RNA polymerase
Tris:	Tris(hydroxymethyl)methyl amine
TSA:	Transition-state analogue
UTP:	Uridine 5′- triphosphate
WT:	Wild type

CHAPTER 1:
Introduction; the "RNA world" and ribozyme chemistry

"... But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, lights, heat, electricity, etc. present, that a protein compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed. ..."

Charles Darwin Letter to J.D. Hooker, February 1, 1871

1.1 Setting the stage:

Throughout mankind's history, philosophers and scientists have struggled in formulating a definite hypothesis that could explain the emergence of life on our planet. All known organisms, past and present, use the same polymer (DNA) for storing genetic information and employ the same class of catalysts (largely protein enzymes) to ensure their survival. This observation strongly suggests that all life forms descended from a common ancestor that utilized the very same polymers. In all domains of life these two polymers are dependent on each other; the DNA nucleotide sequence encodes the amino acid content of the proteins and hence their function, while proteins are required for the maintenance and replication of the DNA genome. The communication between DNA and proteins is actively carried out by RNA, which was once thought to be its only role (Gesteland et al., 2006).

The DNA/RNA/protein central dogma of molecular biology is too complicated to have arisen *de novo*, as it requires both nucleic acid and proteins to be formed simultaneously from simple abiotic organic precursors (Orgel, 2004b). This prompted researchers to ask which arose first- functional proteins or informational nucleic acid- that is, a "chicken and egg" paradox. In the 1960s it was well established that the

evolution of all living organisms occurred by natural selection through mutations and replication (Woese, 1967). Thus, the contest between nucleic acid and proteins was simplified by asking which polymer is capable of sustaining a self-replicating system. With the structure of the DNA double helix solved (Watson & Crick, 1953a, b), it was well understood how nucleic acid could direct the synthesis of its complement from mononucleotides through Watson-Crick base-pairing, while a general mechanism for replicating a polypeptide was, and still is, unknown. Therefore, the answer to the "chicken and egg" dilemma was obvious: nucleic acid preceded proteins (Woese, 1967).

The ability of RNA to serve as a genetic carrier was accepted because it is very similar to DNA, and some viruses use it specifically for this purpose (Ortin & Parra, 2006). Additionally, in 1965 the secondary structure of alanine tRNA from yeast was reported (Holley, 1965) showing that RNA, like proteins, could adopt a secondary structure. With these two factors in mind, Orgel and Crick boldly made the controversial suggestion that RNA could act functionally as a catalyst and hence was once the sole polymer in living organisms (Crick, 1968; Orgel, 1968). They used the argument that nucleotide coenzymes used by protein enzymes were molecular fossils from an era where RNA functioned independently of protein (White, 1976). Francis Crick went on to suggest that the original ribosome was composed solely of RNA (Crick, 1968). It was agreed that the existence of such a ribo-organism would only be possible if RNA could assume some of the functions currently performed by protein enzymes. However, the authors of all these papers failed to recognize that RNA catalysts are still present in modern metabolism, and assumed that the superior protein enzymes have entirely replaced them.

The surprising discovery by Cech and Altman that RNA, like protein enzymes, can catalyze reactions (Kruger et al., 1982; Guerrier-Takada et al., 1983) bolstered those previous assumptions and sparked interest in RNA-mediated catalysis. The role of catalytic RNAs, ribozymes, in contemporary biology was further substantiated by discoveries showing that RNA was able to catalyze a diverse set of reactions, most are critical for viral replication or genomic organization (Buzayan et al., 1986; Peebles et al., 1986; Forster & Symons, 1987b, a; Uhlenbeck, 1987; Kuo et al., 1988; Saville & Collins, 1990). The hypothesis that RNA preceded the DNA/RNA/protein world was no longer ignored, and was eventually coined by Gilbert as the "RNA World" hypothesis (Gilbert, 1986). The hypothesis was given a major boost by the recent demonstration that protein synthesis is an RNA-catalyzed reaction as revealed by recent crystal structures of the ribosome (Ban et al., 2000; Nissen et al., 2000).

The origin of life has traditionally been studied using two approaches; looking backward through phylogenetic or palaeontologic evidence, or moving forward starting with small molecules and conditions believed to be present on early earth (Benner et al., 2006). Because approaching the problem from the first window is beyond the scope of this thesis and has proven to be difficult (Bartel & Unrau, 1999), it will not be discussed any further.

Supporters of the "RNA World" hypothesis are divided into two different classes (Benner et al., 2006). While both groups agree that RNA was once the sole polymer responsible for the storage of genetic information and catalysis, one group is more stringent in that they state that a ribo-organism was the first form of life, while the other group believes that some RNA-like polymer may have preceded RNA. The purpose of

the following sections is to provide a summary of the advances made in the field of prebiotic chemistry in support of the "RNA first" hypothesis. We will conclude that RNA was probably not the first polymer to support life, and put forward a few plausible alternative polymers that may have reigned in a pre-RNA era.

1.2 Moving forward "RNA first" perspective:

For the "RNA first" hypothesis to be sound, there must be a pathway that could explain the *de novo* emergence of RNA from inanimate matter. The dawn of RNA on early earth has been of interest to many prebiotic chemists and has been subdivided into three main problems: 1) prebiotic synthesis of nucleotides, 2) prebiotic synthesis of polynucleotides, and 3) nonenzymatic replication of RNA. Having solved these problems, natural selection of functional ribozymes in the prebiotic soup could lead to the first ribo-organism (Orgel, 2004b).

1.2.1 Prebiotic synthesis of nucleosides.

In prebiotic chemistry, conditions that are likely to have been prevalent on the primitive earth are used during the course of the synthesis of the desired material. The term "prebiotic" is not well defined and somewhat elastic, since no one is in a position to claim what the early earth looked like (Lazcano & Miller, 1996). However, a few general restrictions have been agreed upon by the community. First, the starting material, besides being plausibly prebiotic to the researcher, must be present in adequate amounts. Second, water must be used as the solvent if applicable, and third the yield must be respectable (Orgel, 2004b).

1.2.1.1 Ribose synthesis, the formose reaction.

The formose reaction, used to synthesize ribose, is the most prebiotically plausible of the reactions required to generate RNA, and is the best understood (Mizuno & Weiss, 1974). In 1861, Butlerow generated a sweet mixture when formaldehyde was incubated with calcium hydroxide (Butlerow, 1861). The mixture was later characterized to contain ribose and other sugars (Butlerow, 1861; Breslow, 1959; Zubay, 1998). Intriguingly, the reaction is catalyzed by an impurity in the formaldehyde that Butlerow fortunately failed to purify. Glycolaldehyde, which is made from the polymerization of two formaldehyde molecules, acts as an initiator in the autocatalytic cycle. The cycle proceeds with forward and reverse aldol reactions, where tautomerization plays a role in the interconversion between aldehydes and ketones (Benner et al., 2006).

In the presence of hydroxide ions, glycolaldehyde loses a proton to form an enediolate. The carbon atom of the enediolate then attacks a molecule of formaldehyde to form the three-carbon glyceraldehyde, which in turn can undergo the same reaction with formaldehyde to form a four-carbon carbohydrate. This molecule can then undergo cleavage, after isomerization, to form two molecules of glycolaldehyde resulting in an auto-catalytic cycle (Figure 1-1a). The enediolate of glycolaldehyde can attack the carbonyl carbon of glyceraldehyde to form a variety of pentose sugars including ribose (Figure 1-1b).

The formose reaction has one major shortcoming in that it fails to produce significant amounts of ribose (Decker et al., 1982). Under alkaline conditions, ribose loses a proton to form an enediolate that can react with another ribose, formaldehyde, or glyceraldehyde. The formose reaction, unless modified, usually ends up as an

uncharacterized brown mixture (Benner et al., 2006). Two methods have been recently used to stabilize ribose and bias the formose reactions towards the synthesis of ribose. In the first method, glycolaldehyde is substituted by glycolaldehyde phosphate, changing the reactive species to phosphate esters (Mueller et al., 1990). Unlike the hydroxycarbonyl group of sugar, the resulting pentose-2,4- diphosphate cannot carry out a nucleophilic attack. In a second surprising report, borate minerals were found to stabilize the cyclic form of ribose by complexing with the 2′, 3′ hydroxyl groups rendering it unreactive (Ricardo et al., 2004).

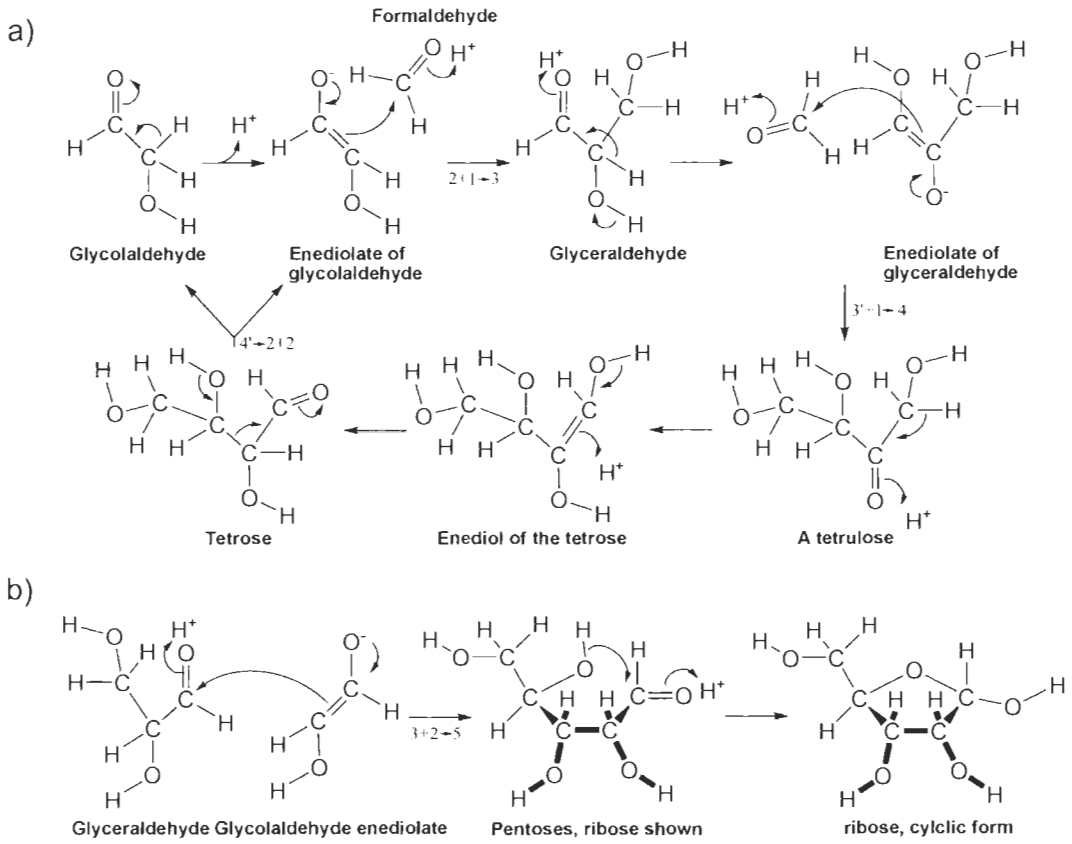


Figure 1-1: Proposed mechanism of the formose reaction. a) The reaction is catalyzed by glycolaldehyde, and results in an autocatalytic cycle. In the presence of base, glycolaldehyde loses a proton to form its enediolate, which then attacks one molecule of formaldehyde to produce the 3-carbon glyceraldehyde. The glyceraldehyde in turn loses a proton to form an enediolate that can attack formaldehyde generating a 4-carbon sugar, tetrulose. This sugar undergoes rearrangements through an enediol to give a tetrose, which can fragment to produce two molecules of glycolaldehyde. **b)** The enediolate of glycolaldehyde can attack glyceraldehyde to generate a pentose sugar. The resulting pentoses are ribose, arabinose, xylose, or lyxose. One of the hydroxyl groups can attack the electrophilic carbonyl group producing the cyclic form of the sugar (Benner et al., 2006).

1.2.1.2 Synthesis of bases.

One of the most notable reactions in the history of prebiotic chemistry is the synthesis of the purine base, adenine, from hydrogen cyanide (HCN) and ammonia reported almost 50 years ago (Oro, 1961; Oro & Kimball, 1961, 1962). The reaction goes through a set of HCN polymerization reactions that lead to the accumulation of a dark precipitate that releases appreciable amounts of adenine and trace amounts of guanine upon hydrolysis (Miyakawa et al., 2000; Miyakawa et al., 2002b, a). The HCN

polymerization proceeds through the formation of an HCN tetramer, the only known stable intermediate, followed by a complex and currently not understood series of reactions. Several proposed mechanisms have been put forth that could explain the synthesis of adenine from the HCN tetramer (Orgel, 2004a). Orgel and co-workers have shown that the addition of ammonia to HCN produces formamidine, which when reacted with the HCN tetramer gives 4-amino-5-cyano-imidazole (AICN). AICN then reacts with one molecule of formamidine to produce adenine (Ferris & Orgel, 1965, 1966a; Sanchez et al., 1967, 1968), (Figure 1-2). Adenine has been also produced by the direct addition of HCN to AICN, or by heating formamide (Saladino et al., 2001). The production of AICN from HCN through formamidine requires high concentrations of ammonia. The presence of ammonia on early earth is highly contentious, making these syntheses controversial. Ferris and Orgel have shown that AICN can be produced photochemically from HCN in quantitative yields (Ferris & Orgel, 1966b), and it has been regarded as a potential prebiotic route to adenine.

The most plausible prebiotic route to pyrimidines is through the reaction of cyanoacetylene with urea to produce cytosine (Figure 1-3), which can hydrolyze to form uracil. The tenability of this synthesis stems from the observation that cyanoacetylene is readily formed when an electric discharge passes through a mixture of nitrogen and methane (Ferris et al., 1968). Furthermore, the synthesis of pyrimidines through this route can proceed in parallel to that of adenine from HCN (Orgel, 2004b).

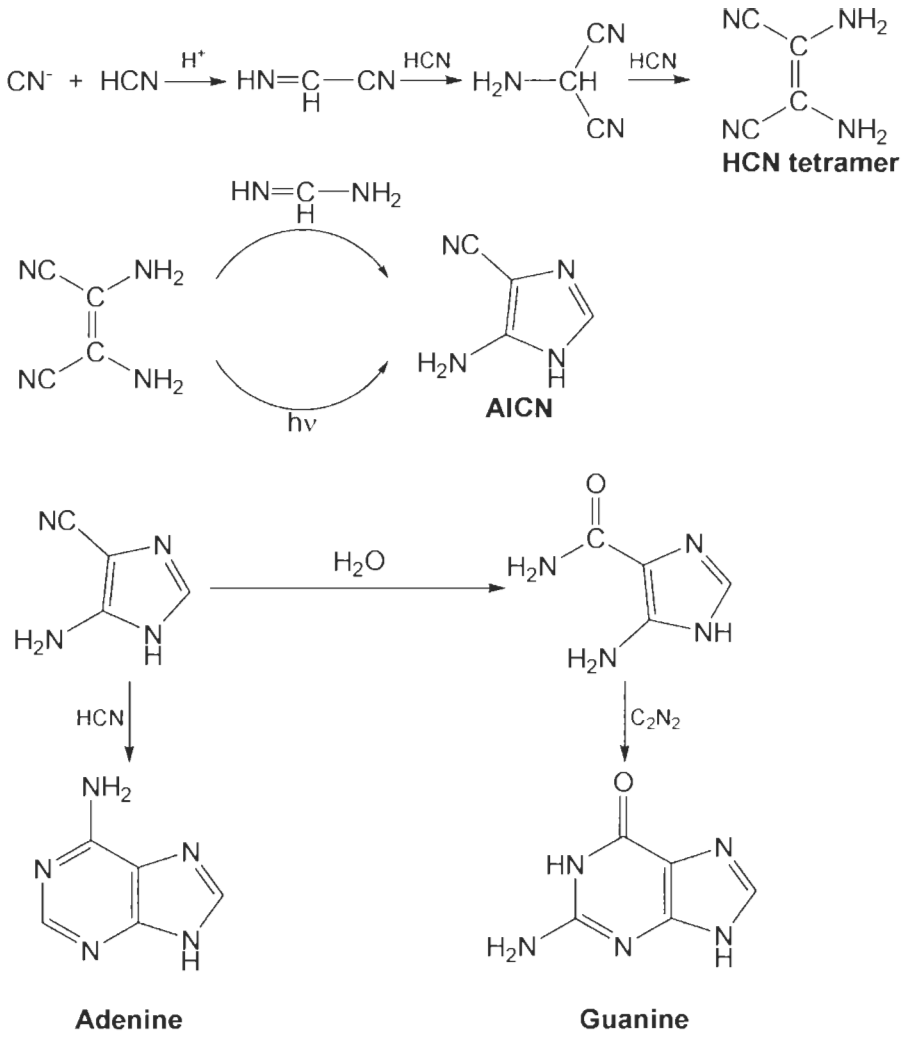


Figure 1-2: A possible prebiotic route to adenine and guanine from HCN. HCN undergoes a series of polymerization reactions to form a tetramer. The tetramer, through a reaction with formamidine or by photoisomerization, produces 4-amino-5-cyano-imidazole (AICN). AICN has been shown to react with HCN to give adenine (Saladino et al., 2001). The hydrolysis product of AICN, 4-amino-imidazole-5-carboxamide, can react with cyanogen to produce guanine.

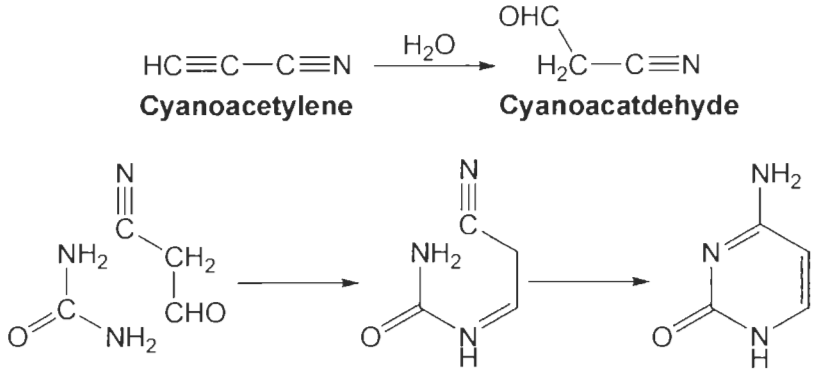


Figure 1-3: Plausible prebiotic synthesis of cytosine. The hydrolysis of cyanoacetylene gives cyanoacetaldehyde, which reacts with urea to produce cytosine (Ferris et al., 1968).

1.2.1.3 Nucleoside synthesis.

The synthesis of nucleosides from ribose and bases is the most problematic reaction in the chain of reactions required for the prebiotic synthesis of RNA (Muller, 2006). Nevertheless, if D-ribose is heated together with guanine in seawater in dry phases, ~5% of the reaction products is authentic β-guanosine. The equivalent reaction for adenine is not as promising, as the major reaction takes place at the amino group (N6) of the base. However, 3% of β-D-adenosine can be produced if the reaction products are mildly hydrolyzed (Fuller et al., 1972).

Cytidine and uridine syntheses are even more problematic and to date no direct synthesis for either has been reported (Joyce, 2002). α-cytidine has been synthesized by incubating a mixture of ribose, cyanamide, and cyanoacetylene in aqueous solution. When ribose is substituted by ribose-5-phosphate, α-cytidine-5′-phosphate is obtained with a yield of 40%; the nucleotide can be photo-anomerized to the β-anomer but only in 5% yield (Sanchez & Orgel, 1970). In a recent report by the Sutherland group, α-D-cytidine-5′-phosphate was synthesized by first reacting 2-aminooxazole, a condensation product of glycolaldehyde and cyanamide, with D-glyceraldehyde-3-phosphate to give pentose-aminooxazoline-5′-phosphates (Figure 1-4). In the second step the nucleotide is formed by the treatment with cyanoacetylene, similar to the previous synthesis (Anastasi et al., 2007).

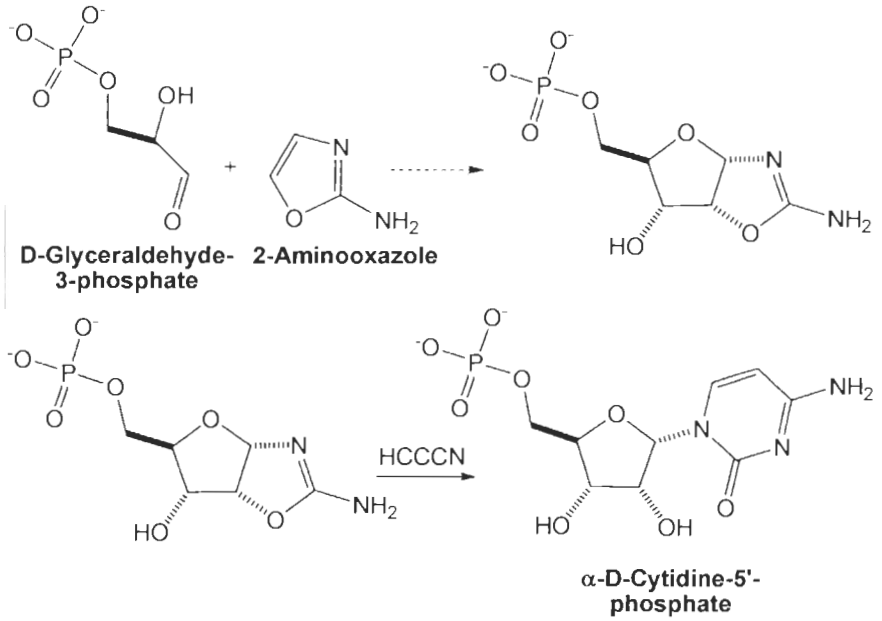


Figure 1-4: Possible prebiotic synthesis of the nucleotide α-D-cytidine-5′-monophosphate. D-glyceraldehyde reacts with 2-aminooxazole, a condensation product of glycolaldehyde and cyanamide to give pentose aminooxazolines (the ribose product is shown). In the second step, ribose aminooxazoline reacts with cyanoacetylene producing the nucleotide (Anastasi et al., 2007).

1.2.2 Prebiotic synthesis of polynucleotides.

1.2.2.1 Activation of nucleotides.

The polymerization of nucleotides into oligonucleotides requires the nucleotides to be first activated. Two modes for the activation of nucleotides have been proposed, phosphorylation and phosphorimidazolides (Figure 1-5, discussed in the next section).

Early attempts to phosphorylate nucleotides using inorganic phosphates employed cyanamide or cyanate as condensing agents. Although these methods proceed with appreciable yields of phosphorylated nucleotides, they require the reaction to be conducted in the absence of water as it effectively competes for the activated phosphate intermediate trimetaphosphate (Lohrmann & Orgel, 1968). The first successful attempt to phosphorylate nucleotide under aqueous conditions utilized ammonium phosphate and urea as a catalyst, and heating at moderate temperatures of 50 to 70°C (Lohrmann &

Orgel, 1971). This procedure results in ~70% of the starting nucleotide converted to a mixture of 2′, 3′ and 5′ phosphorylated product. This protocol has been altered to allow for the biased synthesis of 5′-activated nucleotides (Reimann & Zubay, 1999). However, phosphorous is mainly present as insoluble calcium phosphates on the present day earth, which is presumed to have also been true on early earth (Orgel, 2004b). When ammonium phosphate is substituted by calcium phosphate, the above urea-catalyzed reaction proceeds much slower. Nonetheless, conducting the reaction with hydroxylapatite at an elevated temperature of 100°C has been reported to yield 20% of the phosphorylated nucleotide (Lohrmann & Orgel, 1971).

1.2.2.2 Polymerization of nucleotides.

Phosphate activated nucleotides react extremely slowly under moderate conditions of pH and temperature (with a $t_{1/2}$ of many years) that makes it practically impossible to study their polymerization in the laboratory (Bartel & Szostak, 1993; Muller, 2006). Nucleotides activated as phosphoramidates (Figure 1-5) such as nucleotide 5′-phosphorimidazolides have been extensively used instead of their phosphate counterparts (Orgel, 2004b). These nucleotides can be obtained by reacting nucleotide-5′-monophosphates with imidazole, and hence are considered prebiotic (Lohrmann, 1977). One drawback usually associated with these compounds is their limited relative availability on early earth due to their ability to hydrolyze readily. Therefore they have been used only as a rough example of a class of reactions that may have been responsible for the polymerization of activated nucleotides (Orgel, 2003).

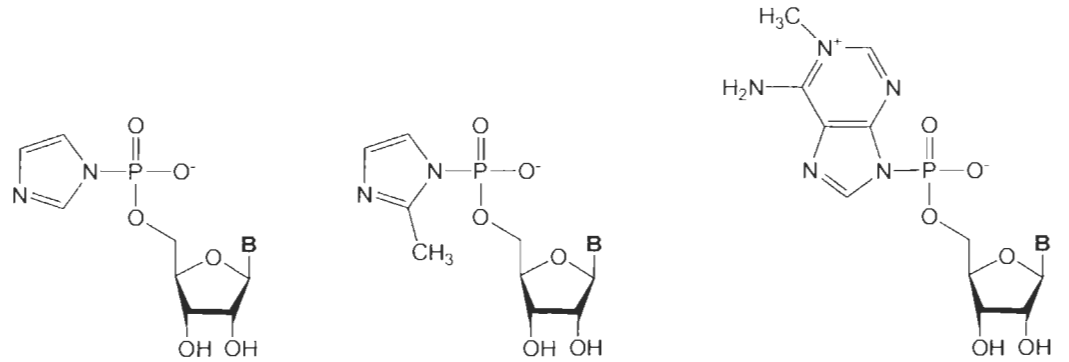


Figure 1-5: Commonly used phosphoramidates in nonenzymatic polymerization reactions.

Nucleotide 5′-phosphorimidazolides oligomerize in aqueous conditions to form short linear and cyclic polymers in the presence of Pb^{2+} (Sawai & Orgel, 1975; Sleeper et al., 1979; Sleeper & Orgel, 1979). The synthesis can be directed to long linear oligomers by performing the reaction under eutectic conditions (Monnard et al., 2003). The resulting oligonucleotides contain up to 40% 3′-5′ linkages. In striking recent reports, the abundant clay mineral, montomorillonite, has been used as a catalyst in the synthesis of long oligonucleotides from dilute concentrations of nucleotide 5′-phosphorimidazolides (Prabahar & Ferris, 1997; Huang & Ferris, 2003; Ferris, 2006). Some oligonucleotides can reach up to 40-nt, if the nucleotide phosphoramidates are added as 1-methyladenine instead of the imidazolides (Huang & Ferris, 2003). The regioselectivity for the 3′-5′ phosphodiester linkage can reach up to 80%, but it is variable and highly dependent on the base and the activating group (Ferris, 2006).

1.2.2.3 Nonenzymatic replication of oligonucleotides.

The replication of RNA in the absence of protein or RNA catalysts has been proposed to have likely played a role in the origin of the "RNA World" (Orgel, 2003).

The nonenzymatic templated addition of mononucleotides has been exclusively carried out using nucleotide phosphoramidates. The first efficient example of regioselective polymerization was the synthesis of poly(G) oligonucleotide across a poly(C) template using guanosine-5′-phosphorimidazolide (ImpG). The reaction required the presence of Pb^{2+} or Zn^{2+} divalent metals in addition to Mg^{2+} (Lohrmann & Orgel, 1980). Remarkably, the regioselectivity of the reaction is highly dependent on the choice of the metal; in the presence of Mg^{2+} and Pb^{2+}, the linkage is almost entirely 2′-5′. When Zn^{2+} is added instead of Pb^{2+}, highly pure 3′-5′ linkage is obtained. However, this reaction could not be extended to the addition of nucleotides other than G across homopolymers. Surprisingly, when the activated nucleotides are supplied as 2-methyl imidazolides instead of imidazolides, the reaction was found not to require any divalent metals apart from Mg^{2+}. Moreover, the resulting phosphodiester linkages were almost exclusively 3′-5′ (Inoue & Orgel, 1981). Perhaps the greatest advantage of this reaction is that all four bases are capable of being incorporated provided that the template has at least 60% of its residues as cytidines. What follows from this constraint, however, is that the resulting oligonucleotides would contain cytidines at a maximum frequency of 40% preventing repeated rounds of replication (Joyce, 1987). The reaction is also inhibited by the L-enantiomers of the activated nucleotides (Joyce et al., 1984), which are present as frequently as the D-enantiomers in any prebiotic synthesis.

1.3 "Pre RNA" world hypothesis:

The short overview of nucleotide synthesis and polymerization should convince the reader of the inevitable conclusion that at present the prebiotic synthesis of RNA is

not at all feasible. Although many individual steps have been demonstrated to proceed under prebiotic conditions, few result in substantial yield and those usually tend to produce a clutter of compounds. Furthermore, the collective syntheses require a range of conditions that are not compatible with each other (Joyce, 2002; Orgel, 2003).

These obstacles have led to suggestions that the first life form did not utilize RNA but rather some other polymer. Just as RNA has been proposed to cause the emergence of DNA and protein, this system would have been responsible for the birth of RNA (Joyce, 2002). This approach has attracted many researchers and has been regarded as being more plausible than the alternative "RNA first" hypothesis. Despite obstacles associated with the nature of this polymer- there is no guidance from any known biology- the field of pre-RNA world has made decent progress in recent years. Systematic investigations of alternative bases, sugars and linkages have identified some remarkable potential systems (Eschenmoser, 1999). Chief among them is threose nucleic acid (TNA); a nucleic acid-like polymer based on α-L-threofuranosyl units that are linked through 2′-3′ phosphodiester bonds (Schoning et al., 2000), (Figure 1-6). Like RNA, TNA has the ability to pair with itself and RNA through Watson-Crick base-pairs. Unlike RNA, TNA can overcome some of the complexities wrought by prebiotic syntheses. Threose is a much simpler molecule than ribose, it is one of only two four-carbon sugars and can only be joined at two positions (2′ and 3′). Given the ability of TNA to pair with RNA, it is not difficult to imagine how a TNA world could have evolved into the "RNA World".

Peptide nucleic acid (PNA) has been also suggested as an alternative candidate for the pre-RNA era. PNA consists of N-(2-aminoethyl) glycine residues having bases

attached through methylenecarbonyl group with a peptide-like backbone (Nielsen et al., 1991), (Figure 1-6). A prebiotic route to an entire PNA monomer is currently unknown, but aminoethylglycine has been observed in spark discharge reactions from nitrogen, ammonia, methane and water (Nelson et al., 2000). PNA can form a duplex with itself and RNA through Watson-Crick base-pairs. The advantages of PNA over RNA are its relative stability and lack of chirality. On the other hand, PNA monomers have the ability to undergo an intramolecular N-acyl transfer reaction that prevents them from undergoing any standard replication mechanisms (Eriksson et al., 1998).

The catalytic potential of TNA and PNA has yet to be explored due to difficulties associated with the amplification of these polymers that is required during the course of *in vitro* evolution (see section 1.5). Nonetheless, recent reports have shown that certain family B archaeal DNA polymerases are able to synthesize short stretches of TNA, up to 50-nt, on a DNA template (Chaput et al., 2003; Chaput & Szostak, 2003; Kempeneers et al., 2003). As there are currently no enzymes that could reverse transcribe TNA back to DNA, the Szostak lab has cleverly established a display method where a single-stranded TNA molecule is attached to its encoding duplex DNA (Ichida et al., 2005), similar to the mRNA-display method used for protein selections (Roberts & Szostak, 1997). Thus, selection of functional TNA sequences can be accomplished by recovery and subsequent amplification of the attached DNA. Once the potential of TNA for forming catalytically active structures is established, its role as predecessor of RNA can be assessed; above all its ability to synthesize RNA.

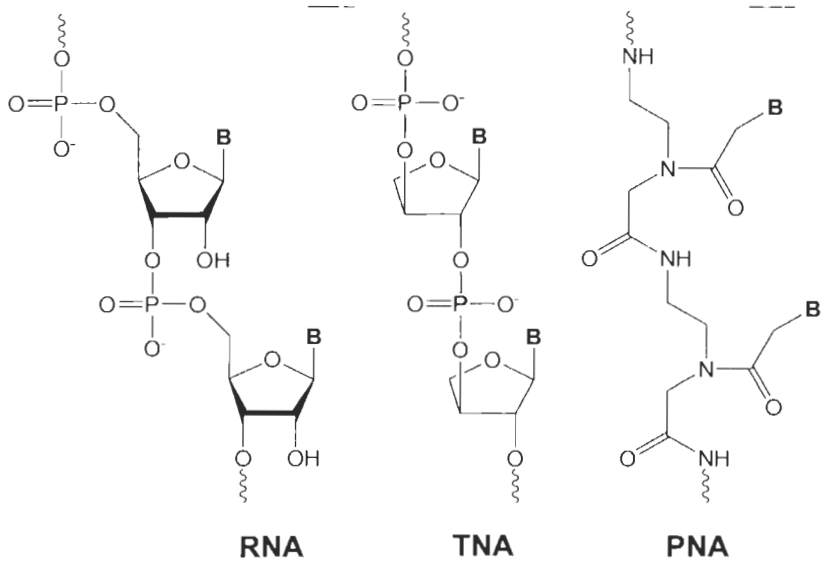


Figure 1-6: Potential pre-RNA world polymers.

1.4 Ribozymes:

The discovery of ribozymes has unequivocally established the "RNA World" as one of the most credible explanations for the embarkation of contemporary biology (Gilbert, 1986). The catalytic ability of RNA was also surprising due to the limited functional groups available for ribozymes; four compared to the twenty exploited by its protein counterpart. To exacerbate this, the functional bases of RNA are largely placed to the interior of duplexes, whereas the amino acid side chains of proteins are arranged optimally on the outside of α-helices and β-sheets to provide tertiary contacts (Doherty & Doudna, 2000). Additionally, the high charge and flexibility of the nucleic acid backbone were once expected to limit the functional capabilities of RNA (Narlikar & Herschlag, 1997). Nonetheless, ribozymes- both naturally occurring and *in vitro* selected- that accelerate reactions up to 10^{11} over the background rate have been found (Doudna & Lorsch, 2005).

1.4.1 Naturally occurring ribozymes.

All of the known naturally occurring ribozymes to date, apart from the ribosome, catalyze reactions at phosphoryl centres that encompass transesterification and phosphate hydrolysis reactions (Fedor & Williamson, 2005). These ribozymes are further divided into two main groups based on their sizes, the small ribozymes (50-150 nt in length) and the large ribozymes (several hundred nucleotides in length). The small ribozymes are found in viral, virusoid, or satellite RNA genomes, and include the hammerhead, hepatitis delta virus (HDV), hairpin, and *Neurospora* Varkud satellite (VS). All of these ribozymes share a similar mechanism where the 2′ hydroxyl nucleophile attacks the adjacent phosphorous in the RNA phosphodiester linkage, resulting in cleavage products having 2′,3′-cyclic phosphate and a 5′ hydroxyl groups (Doherty & Doudna, 2000). This reaction is required during the processing of the rolling circle replication products into genome-length strands. The newly discovered metabolite-responsive (riboswitch) nucleolytic ribozyme *glmS* also belongs to this family of ribozymes. This sequence is present in the mRNA of the *glmS* gene of *Bacillus subtilis*, which encodes glucosamine fructose-6-phosphate aminotransferase. Cleavage by the ribozyme is activated by binding to its effecter molecule, glucosamine-6-phosphate, which results in a product terminated by a cyclic 2′3′ phosphate (Winkler et al., 2004). As a result of this reaction, the initiation of translation of the downstream gene is inhibited.

The large ribozymes are more structurally complex RNAs, and include group I and II introns and the universally conserved ribonuclease P (RNase P). Group I and II introns catalyze two transesterification reactions that result in self-splicing, while RNase P cleaves precursor RNAs during their maturation to generate functional 5′ termini

(Kazantsev & Pace, 2006). In these reactions, the nucleophile and the phosphate reaction centre do not reside closely in sequence or are located on different molecules. As a result, the complexity associated with these ribozymes is to provide the proper orientation of the nucleophile and the labile phosphate required for precise cleavage or self-splicing. Interestingly, a new ribozyme discovered in the human β-globin mRNA, which undergoes self-cleavage in the presence of GTP and Mg^{2+} to promote transcriptional termination, may employ a mechanism similar to that of RNase P (Teixeira et al., 2004).

1.5 *In vitro* selection techniques:

In vitro selection is a powerful technique used to isolate rare functional RNA or DNA sequences from nucleic acid libraries having a diversity of up to 10^{16} different molecules (Wilson & Szostak, 1999). Originally developed by the labs of Gold and Szostak to isolate RNA ligands for small molecules (aptamers), it has since been expanded to isolate exotic ribozymes and deoxyribozymes that have deeply expanded the functional capabilities of nucleic acids (Ellington & Szostak, 1990; Tuerk & Gold, 1990). By providing catalysts of absolute importance to a ribo-organism, like nucleotide synthases and RNA polymerases, the technique has boosted the plausibility of the "RNA World" (Unrau & Bartel, 1998; Johnston et al., 2001; Lau et al., 2004). *In vitro* selection methods were made possible due to three key innovations in molecular biology- advances in automated DNA synthesis, the invention of PCR, and the isolation of reverse transcriptase (Wilson & Szostak, 1999). There are currently three *in vitro* selection approaches for the isolation of nucleic acid catalysts, namely conventional (self-modifying) techniques, transition-state analogue binding methods, and *in vitro*

compartmentalization (IVC). In all of these cases, the selection starts with a pool of sequences (Zaher & Unrau, 2005)- discussed in detail in appendix I- which have elements that are completely random or mutagenized. The selection is carried out by subjecting this pool to repeated rounds of selective amplification steps that depend on the physical or biochemical properties of the available sequences.

1.5.1 Conventional techniques.

The majority of the selections that have been performed to date have utilized the "self-modifying" method because of its relative ease of applicability (Joyce, 2004). This methodology takes advantage of the fact that most reactions involve bond-forming or breaking by tethering one of the reactive groups to the candidate enzymes and supplying the other substrate, if applicable, free in solution. In selections where bond-forming is considered, the free substrate is chemically tagged, and hence successful catalysts get decorated with the tag allowing them to be enriched (Figure 1-7). On the other hand, in experiments where bond-breaking catalysts are being isolated, the chemical tag is attached to the tethered substrate so as to enrich the sequences that dispose of this tag.

Perhaps one of the most commonly used tags by conventional *in vitro* selection is a simple oligonucleotide where the gain or loss of this element confers active sequences different gel mobilities or alternative hybridization abilities. The stringency of such selections can be greatly improved by adding a biotin tag to the oligonucleotide substrate. Thus, bond-forming catalysts become themselves biotinylated allowing them to bind to streptavidin-coated beads, whereas bond-breaking results in functional sequences being released off the beads. The oligonucleotide tag methods have been commonly used in the

isolation of ribozymes performing nucleic acid cleavage and ligation reactions (Joyce, 2004).

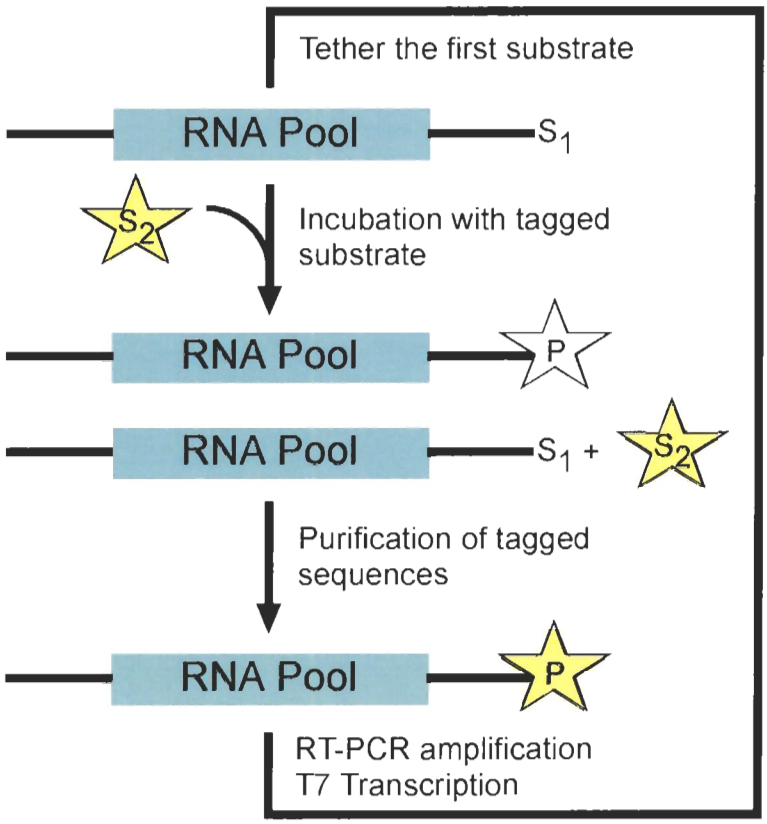


Figure 1-7: Conventional *in vitro* selection. The population of RNA sequences (solid lines indicate fixed sequence, required for amplification) is modified with the first substrate (S_1), and then allowed to incubate with the second substrate (S_2) that is chemically tagged. Successful sequences catalyze the reaction at their ends (P), hence tagging themselves. Tagged molecules are purified away from the inactive population, and subjected to PCR-amplification. This is followed by transcription to produce the progeny RNAs that then undergo a subsequent round of selection.

Reactions that involve smaller substrates are typically more difficult to design, largely due to the steric clash induced by the tag once attached to the small substrate, especially if the tag is larger than the substrate itself (Johnston et al., 2001). Nevertheless, biotin has been effectively used as a tag in many selections such as those employed in the isolation of Diel-Alderase and acyl transferase ribozymes (Lohse & Szostak, 1996; Tarasow et al., 1997). Alternatively, modification of substrates at the atomic level can be used as an effective handle during the selection process. A notable

example is the thio-modified bases/nucleotides, which have been used to isolate both nucleotide synthases and RNA polymerases (Unrau & Bartel, 1998; Johnston et al., 2001; Lau et al., 2004). This substitution of only one oxygen atom for a sulphur in a potential ribozyme results in this molecule having a slower mobility in gels containing mercury (Igloi, 1988), and hence allows it to be isolated. The methodology has also been used to isolate kinase, alcohol dehydrogenase and capping ribozymes, the latter being the subject of chapter 2 (Lorsch & Szostak, 1994; Tsukiji et al., 2004; Zaher et al., 2006).

The subset of the pool that survives the selection step must then undergo an amplification process. RNA molecules are first reverse transcribed to cDNA using reverse transcriptase in the presence of a 3′ DNA primer and dNTPs. The cDNA is then amplified by PCR that employs a thermostable DNA polymerase and two DNA primers. The 3′ primer is the same as used during reverse transcription, while the 5′ primer is used to make the sense strand and introduces a promoter sequence for the DNA-dependent RNA polymerase, typically T7 RNA polymerase (T7 RNAP). The double-stranded PCR products are then transcribed to generate the progeny RNA molecules for the next round of selection. The PCR step can also be used to introduce diversity by employing mutagenic conditions; for *Taq* it involves biased dNTPs and slightly elevated divalent metal concentrations (Cadwell & Joyce, 1994). As the selection proceeds, conditions are usually altered to favour molecules with enhanced properties- generally the reaction time is decreased to bias the development of sequences with faster reaction rates and the substrate concentration is progressively lowered to promote an improvement in K_m. Once the desired enrichment is achieved- typically assessed using activity- the remaining RNA molecules are cloned, sequenced and evaluated for activity.

True catalysts must accelerate the reaction rate without being modified or consumed during the process. As a result, RNAs resulting from conventional methods are not true enzymes. Nevertheless, it is possible for a self-modifying RNA to be converted into a true catalyst by separating it into two modules- one acting as the catalyst, while the other as the substrate. The ease of implementation of such an approach depends critically on the type of interaction between the two modules, and in most cases it is accomplished when base-pairing between the catalyst and the substrate is required. Turn-over is then achieved through dissociation of the product and the subsequent docking of a new substrate. However, this approach has not worked reliably for ribozymes catalyzing small molecule chemistries as the tether tends to play a critical role in the recognition and stabilization of the substrate complex (Unrau & Bartel, 1998).

1.5.2 Transition-state analogue binding methods.

This technology takes advantage of the notion that a catalyst accelerates the reaction rate by stabilizing the transition-state of that reaction. Transition-state is defined as the state corresponding to the highest energy in a reaction coordinate and hence it is highly unstable (Jencks, 1963). Stable analogues of transition-states for many reactions have been commonly synthesized by organic chemists to understand enzymes mechanisms, and in 1986 they were used to isolate the first catalytic antibodies (Pollack et al., 1986; Tramontano et al., 1986). In this strategy, a pool of antibodies was screened for their ability to bind tightly to a transition-state analogue (TSA) of a peptide hydrolysis reaction. As hoped, some of the isolated sequences were found to catalyze the desired reaction. The feasibility of this method in the isolation of functional nucleic acid was evident by the selection of two novel ribozymes and one deoxyribozyme (Prudent et al.,

1994; Conn et al., 1996; Li & Sen, 1996). The ribozyme isolated by Prudent et al. catalyzed the non-covalent interconversion of a bridged biphenyl substrate, while Conn et al. isolated a ribozyme for porphyrin metallation. Li and Sen isolated the only deoxyribozyme using this methodology; also for porphyrin metallation.

The advantage of the TSA method resides in the fact that isolated catalysts always perform multiple-turnover chemistry. On the other hand, aptamers that bind the TSA tightly are often incapable of catalyzing the desired reaction (Gold et al., 1995; Lorsch & Szostak, 1996), and to date, the three instances described above are the only examples of successful TSA-based attempts. The shortcomings of this approach have been attributed to the approximation to the actual transition-state exploited by researchers during the design of TSAs. Furthermore, even in the three successful attempts, the catalytic rate enhancements have been moderate because the critical contacts required to accomplish this are not necessarily those required for tight binding to the TSA.

1.5.3 *In vitro* compartmentalization.

All evolutionary systems, natural or artificial, require that the genotype (typically DNA) and the encoded phenotype (RNA or protein) are linked to each other. Conventional selection techniques achieve this requirement by subscribing the RNA molecules the role of acting as both the genotype and phenotype. In contrast, natural selection correlates functional polypeptides and informational polynucleotides using cellular compartmentalization. As a consequence, the survival of each cellular unit depends on the proper functioning of the cellular machinery expressed by its genes. IVC mimics natural selection where the genotype and phenotype, unlike conventional

selections, are separated and linked through the compartmentalization of both within water-in-oil droplets (Tawfik & Griffiths, 1998).

Typically in IVC ribozyme selections, one of the substrates is physically linked to the pool DNA. The resulting tethered genes are subsequently encapsulated along with T7 RNAP and NTPs required for the transcription of the phenotypic RNAs. A result of compartmentalization, functional RNA sequences catalyze the formation of the product on the corresponding coding-genomic DNA molecules (Figure 1-8). Ribozyme-encoding genes, with the product attached, are separated away from genes that encoded inactive RNA sequences and as result still carry the unmodified substrate (Griffiths & Tawfik, 2006; Miller et al., 2006). This subset of DNA genes is then amplified using PCR before the next round of encapsulated selection is carried out. The applicability of this technique has been verified in a number of protein selections (Tawfik & Griffiths, 1998; Griffiths & Tawfik, 2003; Cohen et al., 2004), and recently it has been adopted in high throughput sequencing (Margulies et al., 2005). As for ribozyme selections, IVC has been used in the isolation of Diels-Alderases, RNA ligases and RNA polymerases (Agresti et al., 2005; Levy et al., 2005); the latter is the subject of chapter 4.

The advantage of IVC over conventional methods is three-fold. First, enzymes isolated by IVC are true catalysts performing multiple turn-over chemistry. Second, active RNAs tend to fold into highly ordered structures, and are as such difficult to fully reverse-transcribe and often exhibit significant stalls. Many efficient ribozymes, therefore, are likely to have been lost in conventional selections. IVC bypasses this problem by directly amplifying the DNA sequences. Lastly, the requirement of the

substrate to be tethered to functional RNAs in the "self-modifying" methods, but not in IVC, limits the number of ways a catalytic motif can form.

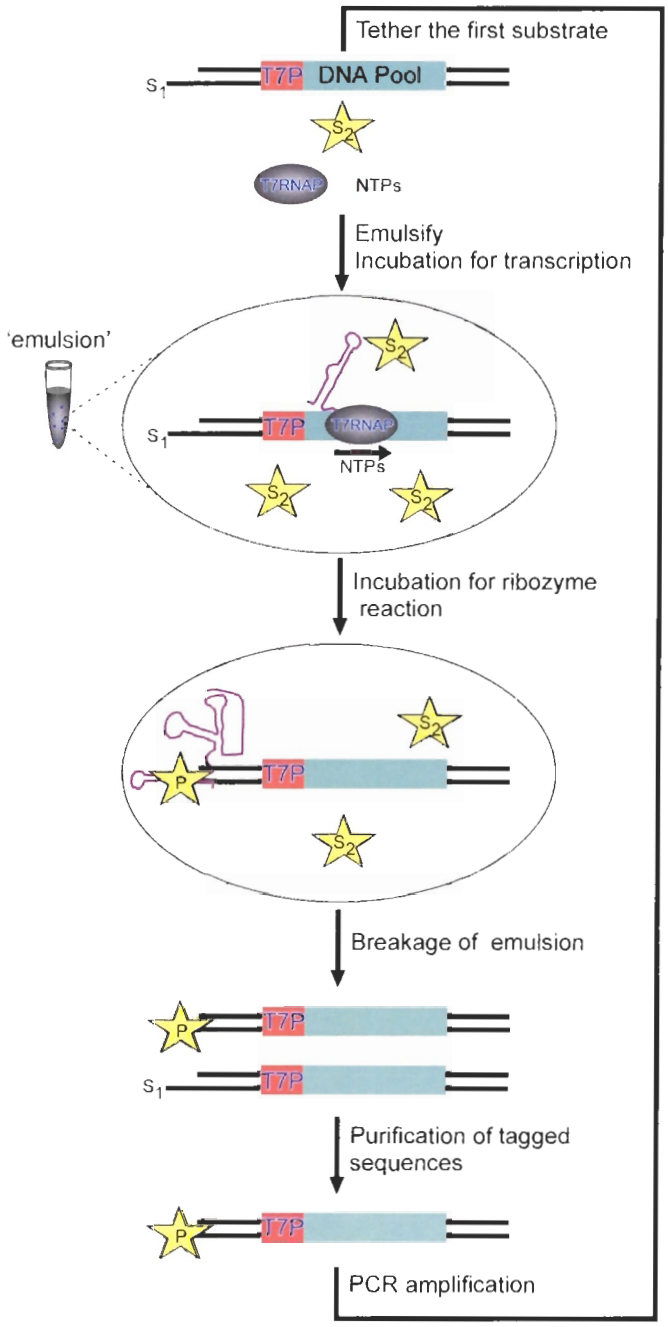


Figure 1-8: ***In vitro*** **compartmentalization methodology**. One of the substrates (S_1) is first attached to the DNA pool (genotype), which contains fixed sequence elements and a T7 RNAP promoter (T7P). The resulting modified pool is subsequently emulsified together with tagged substrates (S_2), T7 RNAP and NTPs. T7-mediated transcription results in the RNA phenotype. Potential catalysts recognize the substrate tethered to the corresponding ribozyme-encoding genes (as a result of compartmentalization), and catalyze the formation of the DNA-tethered product (P). Genes that are modified are purified from ones carrying the initial substrate, and amplified before the next round of selection is carried out.

1.6 Ribozyme mechanisms:

Enzymes employ a number of strategies to stabilize the transition-state between substrate and product during catalysis. For instance, catalysts may add or abstract protons during the reaction, orient substrates so that they are optimally positioned to react, and use binding interactions away from the reaction centre to introduce an unfavourable contact that does not exist in the transition-state (Doudna & Cech, 2002). The lack of diverse functional groups in RNA and its propensity to bind tightly to divalent metal-ions such as magnesium led early suggestions to classify ribozymes as a distinct class of metalloenzymes (Pyle, 1993). That is, all ribozymes require correctly positioned divalent metal-ions to have a direct role in catalysis. While these hypotheses appear to be true for the large ribozymes, the small nucleolytic ribozymes seem to employ a different strategy and it is almost certain that none of them are metalloenzymes.

1.6.1 Stereochemistry.

The stereochemical course of many protein-catalyzed reactions has classically been used to unveil the type of mechanism utilized by the enzyme (Frey et al., 1982). In particular, protein enzymes that catalyze reactions similar to those accelerated by natural ribozymes (phosphodiester bond cleavage and ligation reactions) have been studied using phosphorthioate substituted substrates (Eckstein, 1985). In this approach, one of the nonbridging oxygens in the phosphodiester linkage, which naturally is not chiral, is replaced by a sulphur atom. As such phosphorous reaction centre is chiral, the stereochemical course of the reaction can be determined. The same approach has been used to monitor the stereochemical course of the reactions catalyzed by naturally-occurring ribozymes, and in all cases tested, the reactions proceed with an inversion of

configuration at the phosphorous centre (McSwiggen & Cech, 1989; Rajagopal et al., 1989; Slim & Gait, 1991; Padgett et al., 1994), (Figure 1-9). These findings were used to propose the in-line displacement mechanism for ribozyme-catalyzed reactions, which proceeds via a trigonal bipyramidal transition-state (Figure 1-9). In chapter 3, I discuss the stereochemical course of a capping ribozyme that retains stereo-configuration during the course of the reaction.

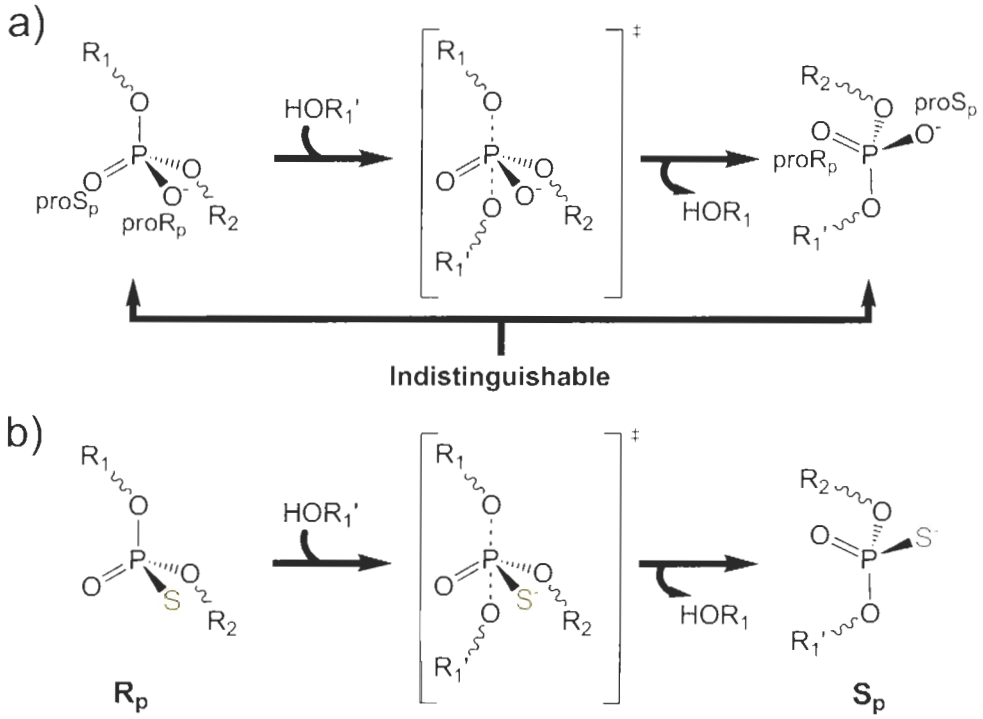


Figure 1-9: Phosphorthioate substitutions reveal ribozyme mechanism. **a)** All naturally-occurring ribozymes, which catalyze reactions at a phosphodiester linkage, utilize an in-line displacement mechanism. The reaction proceeds through a trigonal bipyramidal (penta-coordinated) transition state that leads to an inversion of stereo-configuration that is not detectable; the phosphorous reaction centre is not chiral. **b)** Atomic substitution of a non-bridging oxygen by a sulphur (shown is the Rp stereoisomer) can be used to probe the stereochemical course of ribozyme-catalyzed reactions, as the phosphorous is made chiral in this case. This approach has been used to verify the in-line displacement mechanism, which leads to inversion around the reactive phosphorous (shown is the Rp→Sp inversion).

1.6.2 Group I introns.

The self-splicing group I intron of *Tetrahymena* was the first ribozyme to be discovered (Kruger et al., 1982). Since then, group I introns have been found in

precursors of ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA) of many organisms, including the nuclei of protozoa, the mitochondria of fungi, the chloroplasts of algae, and bacteria and their phages (Doudna & Cech, 2002). They are classified as such based on their secondary structures and common mechanism of action. The self-splicing reaction proceeds via two transesterification steps, where an exogenous guanosine or GTP molecule acts as the nucleophile (Figure 1-10).

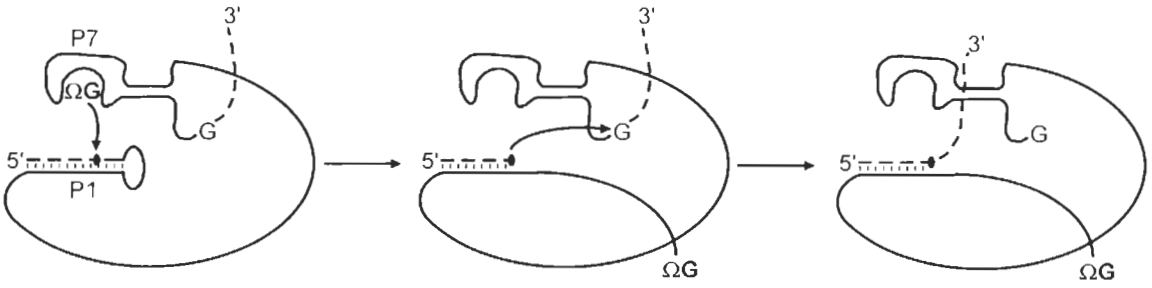


Figure 1-10: Schematic of the group I intron self-splicing reaction. Exons are shown as dashed lines, while the intron is drawn in a solid line. In the first step, the 3′ hydroxyl of a bound guanosine (ΩG) attacks the 5′ splice site (within the P1 stem), cleaving it, and becoming attached to the 5′ end of the intron. In the second step, the intron- still held in the P1 region- attacks the 3′ splice site, resulting in the joining of the two exons and the release of the intron.

Naturally, these ribozymes do not perform multiple turnover; but as the excised intron preserves the active site, it can be re-engineered to catalyze splicing in *trans* giving a true catalyst. The *trans*-acting ribozyme performs the first step of self-splicing with multiple turnover using an RNA oligonucleotide that mimics the 5′ exon. This version of the *terahymena* group I intron has been used almost exclusively to dissect each step of the reaction pathway. The oligonucleotide substrate first base-pairs with an internal guide sequence in the ribozyme, rescuing the P1 helix. Specific 2′ hydroxyl groups along the minor groove of this P1 element in turn promote the presentation of the substrate to the active site. The 3′ hydroxyl of the guanosine nucleophile, after binding to the G site within P7, attacks the splice site phosphate cleaving the substrate. Finally, the cleaved

product is released; this last step is rate-limiting for multiple turnover due to base-pairing and tertiary interactions that slow the diffusion of the product from the active site (McConnell et al., 1993).

Similar to many protein enzymes that catalyze analogous reactions, the group I introns employ divalent metals directly for catalysis (Narlikar & Herschlag, 1997). Early investigations of the locations and the number of catalytic metals utilized biochemical means where individual phosphate or ribose hydroxyl groups were substituted with "softer" thiophosphate or amino groups and assessed for changes in metal specificity, that is rescue by a thiophilic metal like Mn^{2+} or Cd^{2+}. These procedures resulted in a three-metal model; the first metal (M_A) is used to stabilize the negative charge developed in the transition-state, the second metal (M_B) helps in the deprotonation of the 3′-OH of the guanosine nucleophile, and the third metal (M_C) was proposed to assist M_A in orienting the substrates properly and the stabilization of the trigonal bipyramidal transition-state (Shan et al., 2001), (Figure 1-11). While recent crystal structures of the *Azoarcus* group I intron includes all the ligands for this three-metal model, the structure shows only two Mg^{2+} ions are bound in the active site. M_1 is equivalent to M_A, while M_2 is coordinated to both the 3′ and 2′ hydroxyl groups of the guanosine nucleophile and hence takes the position of M_B and M_C (Stahley & Strobel, 2005).

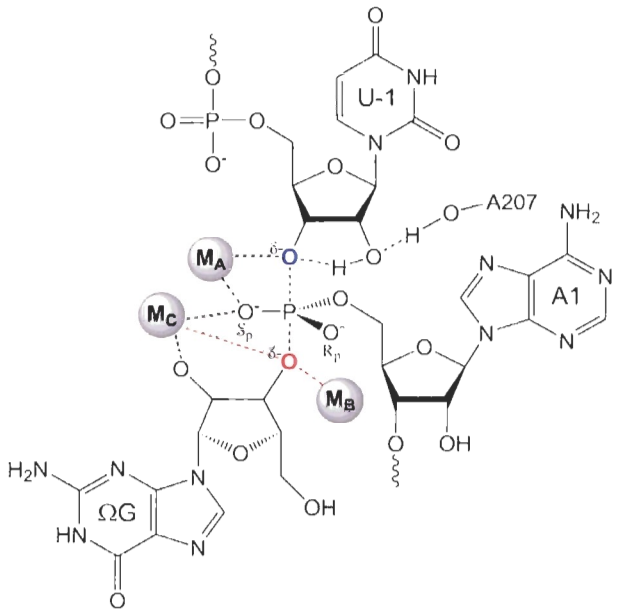


Figure 1-11: The three and two divalent metal-ion models of step one of the self-splicing reaction by the *Tetrahymena* group I intron. In the three metal-ion mechanism (Shan et al., 2001), the first ion M_A stabilizes the leaving group oxyanion (shown in blue) by neutralizing the negative charge that develops during the cleavage reaction. The same metal also interacts with the pro-S_p nonbridging oxygen, which is also coordinated to M_C. In addition, M_C stabilizes the 2′ OH of the ΩG nucleophile. M_B activates the incoming 3′ hydroxyl nucleophile- the interaction is depicted in red as it is absent in the two metal-ion mechanism. In this model, M_C is responsible for the activation of the nucleophilic oxygen (Stahley & Strobel, 2005).

1.6.3 Group II introns.

Found in bacteria and organellar genes, group II introns catalyze the specific self-excision and ligation of the flanking RNA sequences to produce mature transcript (Doudna & Cech, 2002). As in the case with the group I introns, the reaction catalyzed by the group II introns involves two transesterification steps. In contrast to group I introns, the nucleophile resides internally within the intron as an unpaired adenosine close to the 3′-end of the intron. In the first step, the 2′-OH of this adenosine attacks the 5′ splice-site phosphate, resulting in a 5′ free exon and a branched lariat structure similar to the one produced by the splicesome. In the second step, the 5′ exon attacks the 3′

splice site, leading to the ligation of the two exons and the release of the intron as a lariat (Michel & Ferat, 1995), (Figure 1-12).

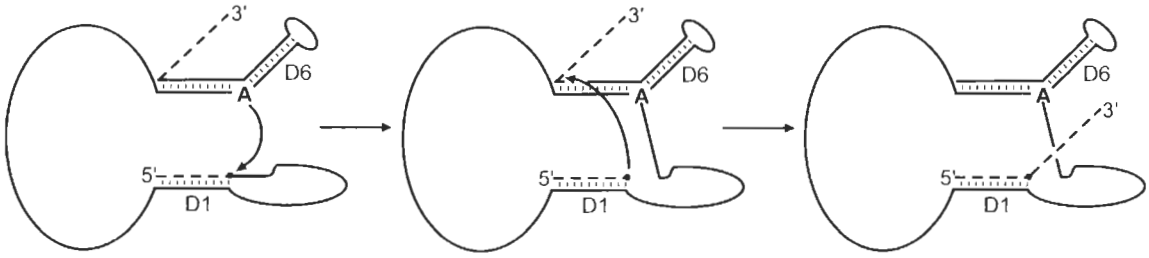


Figure 1-12: Self-splicing pathway by the group II introns. In step 1, the 2′ hydroxyl of an adenosine residue within domain 6 of the intron attacks the 5′ splice site, annealed to domain 1 of the intron, resulting in the formation of a lariat. In step 2, the 5′ splice site attacks the 3′ splice site, joining the two exons and releasing the intron as a lariat.

Like the group I introns, group II introns have an absolute requirement for divalent metal-ions for catalysis. Substitutions of the 3′ leaving group oxygens for sulphur inhibit both the first and the second step of the self-splicing reaction. Thiophilic metal ions restore activity, suggesting that a divalent metal-ion is required to stabilize the 3′-oxyanion leaving group (Gordon et al., 2000), (Figure 1-13). Additionally, both steps are inhibited by S_p sulphur substitution, however the reactions are not rescued by thiophilic metal ions (Padgett et al., 1994). As a result, no metal ions interactions with the two phosphate reaction centres have been demonstrated conclusively (Fedor, 2002). Phosphorthioate substitution at regions near the active site have also identified possible ligands for divalent metal ions, but it is currently unclear whether these metal ions mediate catalysis directly or merely stabilize the active site structure (Sigel et al., 2000; Hertweck & Mueller, 2001). The lack of conclusive biochemical data and high resolution structures have made it difficult to assess how many metal ions are directly involved in catalysis, but the common agreement is that it is likely to be two (Fedor, 2002).

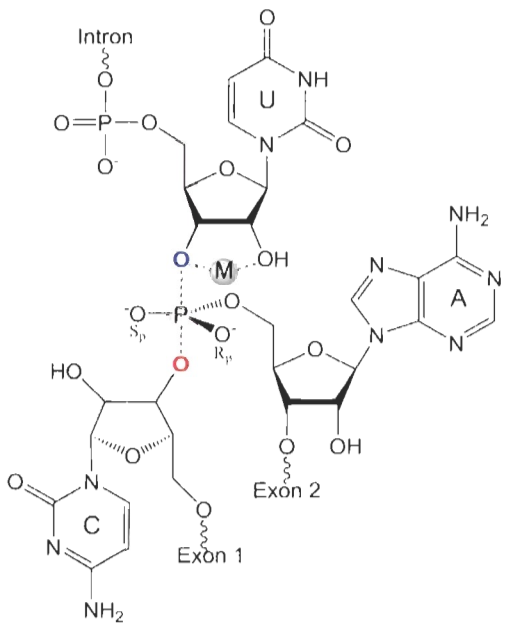


Figure 1-13: Model for metal-ion interactions during step 2 of the reaction catalyzed by group II introns. A single metal ion has been proposed to stabilize the leaving oxyanion (shown in blue), while at the same time interacting with the 2′ hydroxyl group of the intron (Padgett et al., 1994).

1.6.4 RNase P.

RNase P is a ribonucleoprotein complex that catalyses the maturation of the 5′ end of tRNA in all three domains of life (Kazantsev & Pace, 2006). The enzyme has been also found to process alternative substrates such as transfer messenger RNA (tmRNA) and 4.5S RNA; like tRNA, both of these RNAs are also involved in protein synthesis. Bacterial RNase P contains one small basic protein, while the archaeal and eukaryal complexes contain 4-5 and 10 different protein subunits respectively (Hartmann & Hartmann, 2003). In all three cases, the protein component is essential for activity *in vivo* (Schedl & Primak, 1973; Apirion, 1980; Waugh & Pace, 1990; Chamberlain et al., 1998). Nonetheless, the RNA component of the bacterial RNase P has been known to be catalytic by itself *in vitro* for more than two decades, and hence the enzyme is classified as a ribozyme (Guerrier-Takada et al., 1983). Interestingly, at the time of writing this

thesis, the RNA component of the human RNase P was also shown to be functional by itself *in vitro*; albeit very slowly (Kikovska et al., 2007).

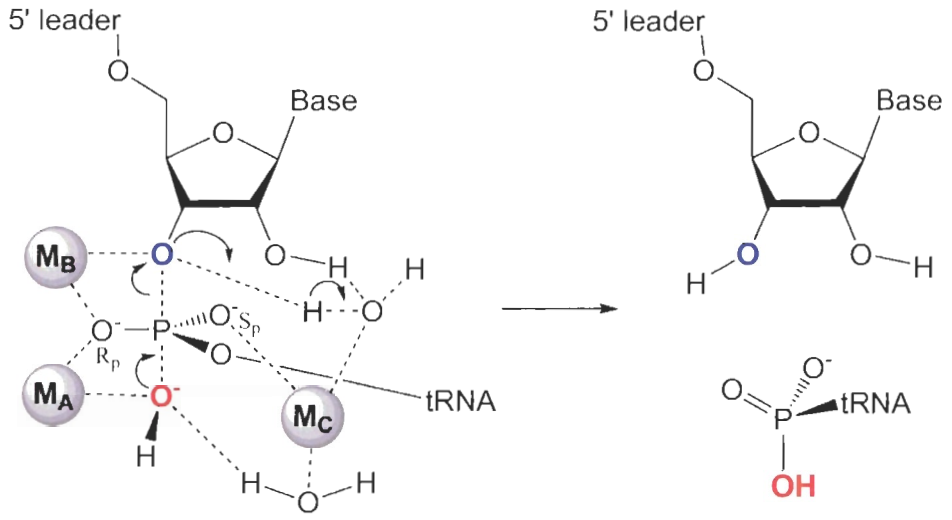


Figure 1-14: The proposed transition state of the RNase P reaction. The pro-R_p oxygen of the scissile phosphate is coordinated to two metals M_A and M_B. M_A also activates the incoming hydroxide nucleophile (drawn in red) and positions it for an in-line attack. M_B is responsible for neutralization of the negative charge being developed on the 3′ leaving oxygen group (depicted in blue). The third metal M_C, through an inner sphere water molecule, interacts with the 2′ hydroxyl group of the leader strand. This results in a reduction in the pK_a of the associated water and, as a result, assists in the transfer of a proton to the 3′ oxygen. The same metal is thought to interact directly with the pro S_p oxygen and with the nucleophilic oxygen through an inner sphere water molecule (Kazantsev & Pace, 2006).

In contrast to group I and II introns, RNase P performs multiple turnover *in vivo*. RNase P catalyzes the hydrolysis of a phosphodiester linkage, generating a mature tRNA with a 5′ phosphate terminal and a leader sequence containing a 2′-3′-cis-glycol terminal (Figure 1-14). The reaction has an absolute dependence on divalent metal ions, typically Mg^{2+} or Mn^{2+}. Hill analyses of the metal dependence together with thiophilic rescue of phosphorthioate substitutions suggest a three-metal-ion mechanism for the catalytic action by RNase P (Smith & Pace, 1993; Warnecke et al., 1996). In this model (shown in figure 1-14), the Mg^{2+}-hydrate coordinate complexes surround the scissile phosphodiester

bond to form a trigonal bipyramidal transition-state. The 2′-OH of the leader sequence is thought to facilitate the protonation of the 3′-oxygen leaving group (Cassano et al., 2004).

1.6.5 Hammerhead.

Hammerhead ribozymes are the smallest naturally occurring ribozymes and are probably the most studied. They were originally discovered in viroids and satellite RNAs of plant viruses where they catalyze a specific phosphodiester bond isomerization during the process of rolling-circle replication (Prody et al., 1986; Uhlenbeck, 1987; Symons, 1997). Recently they have been identified in non-viral species genomes such as schistomes (Martick & Scott, 2006). The hammerhead motif is likely to be the most efficient self-cleaving sequence that can be isolated from random libraries (Salehi-Ashtiani & Szostak, 2001). As a result, it has been suggested that the ribozyme may have arisen multiple times during the evolution of functional RNA sequences.

Since its discovery, the hammerhead has been the subject of major controversy (Fedor & Williamson, 2005). Early biochemical and structural investigations suggested two mutually exclusive mechanisms regarding the extent of conformational change required to activate catalysis in the ribozyme. For example, phosphorthioate substitutions identified a critical divalent metal ion that is coordinated to the pro-R_p nonbridging oxygens of both the A9 residue and the cleavage site C17 (Peracchi et al., 1997), (Figure 1-15a), while the crystal structures of a minimal hammerhead showed that these residues are 20 Å apart (Pley et al., 1994), (Figure 1-15a). Furthermore, kinetic analysis of the ribozyme-catalyzed reaction in the presence of high monovalent ions concentrations showed that divalent metal ions are not essential for the catalytic step but play a role in

the stabilization of the active ribozyme structure (Curtis & Bartel, 2001; O'Rear et al., 2001).

Recent crystal structures of a full-length hammerhead seem to unify these previously irreconcilable differences where tertiary contacts not found in the minimal constructs were found to prime the ribozyme for catalysis (Martick & Scott, 2006), (Figure 1-15b). Similar to previous structures, no Mg^{2+} ions were observed in the new crystal structure, and it suggests a general acid-base strategy for catalysis by the ribozyme. The deprotonation of N1 of G12 and its close proximity to the 2′-OH of the cleavage site C17 in the structure (3.5 Å, Figure 1-15c) corroborates previous suggestion of the general base role played by this residue (Han & Burke, 2005). Previous biochemical data have also implicated the 2′-OH of G8 in catalysis (McKay, 1996), the structure suggests that this hydroxyl group is the active functional group for the general acid, i.e. responsible for the protonation of the 5′ oxygen leaving group. A proposed model of the transition-state configuration is shown in figure 1-15d.

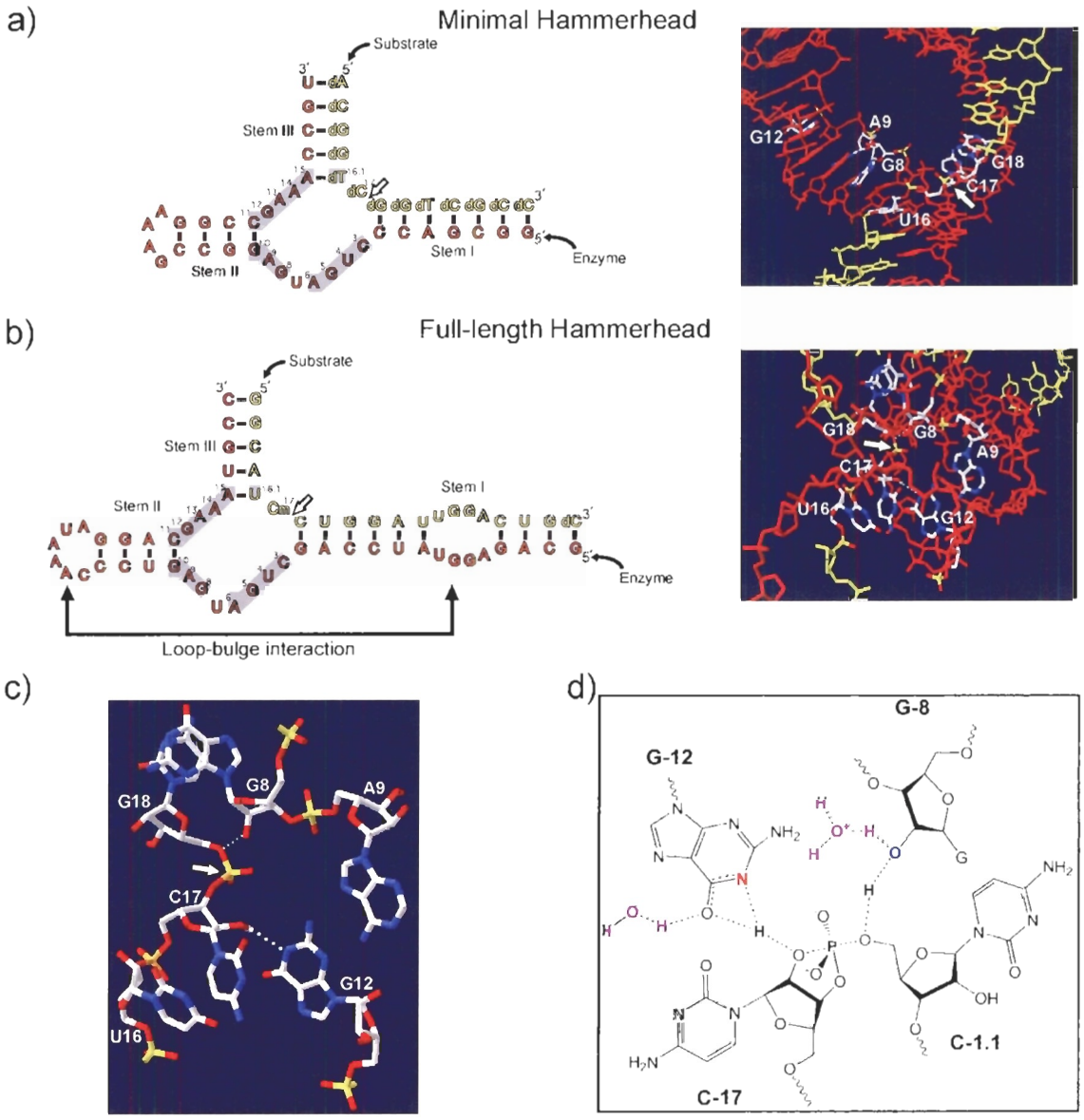


Figure 1-15: The structure and mechanism of the hammerhead ribozyme. **a)** Secondary (left) and crystal (right) structures of a minimal hammerhead (Pley et al., 1994). The substrate is shown in yellow, and is made up of deoxynucleotides so as to prevent cleavage by the ribozyme. The enzyme is drawn in red with the conserved residues displayed on a grey background. All residues are numbered using standard convention. The block arrow indicates the scissile phosphate. **b)** Structure of the full-length hammerhead (Martick & Scott, 2006). The residue Cm17 signifies a 2′-O methyl modification to prevent cleavage. The active site residues, inferred from biochemical data, in both tertiary structures are shown in CPK. Dramatic changes, due to loop-bulge interactions, are observed that reshape the active site in the full-length construct relative to the minimal one. **c)** The active site of the full-length hammerhead ribozyme. **d)** A proposed transition state of the cleavage reaction (Martick & Scott, 2006). G-12 acts as a general base whereby the deprotonated N1 (in red) abstracts a proton from the 2′ hydroxyl nucleophile activating it for the in-line displacement. The 2′ hydroxyl (blue) of G-8 acts as a general base by donating a proton to the 5′ leaving oxygen. Both specific acid and base by water molecules (magenta) may also play a role during catalysis. All molecular models were made using the DeepView/Swiss-PdbViewer v3.7 software package using PDB ID numbers 1HMH (minimal hammerhead) and 2GOZ (full-length hammerhead).

1.6.6 HDV.

The HDV ribozyme catalyzes the same reaction as that of the hammerhead, and it is required during the processing of the viral replication products. Both the genomic and antigenomic strands of the virus harbour self-cleaving ribozymes that have identical secondary structures (Kuo et al., 1988). A recent report from the Szostak lab has shown that the human *CPEB3* gene contains an HDV-like sequence suggesting that the ribozyme may have arisen from the human transcriptome (Salehi-Ashtiani et al., 2006).

In 1998, the Doudna group solved the first crystal structure of a genomic HDV ribozyme (Ferre-D'Amare et al., 1998). The structure revealed that the ribozyme forms an enclosed cleft where the strand scission takes place with no divalent metal ions near the active site. Instead, a set of conserved essential nucleotides surrounds the scissile phosphate, suggesting the possibility of their direct role in catalysis. The crystal structure suggested, for the first time, a possible role for a nucleobase as general acid, namely C75. Biochemical investigations that followed went on to support this hypothesis (Fedor & Williamson, 2005). In a prominent experiment, an inactive ribozyme that had C75 substituted for an abasic residue was rescued in the presence of imidazole (Nakano et al., 2000). The biochemical and biophysical data support a model (figure 1-16) where C75 acts as a general acid by donating a proton to the 5′ bridging oxygen leaving group, while a hydrated metal ion abstracts a proton from the 2′-OH nucleophile (Das & Piccirilli, 2005).

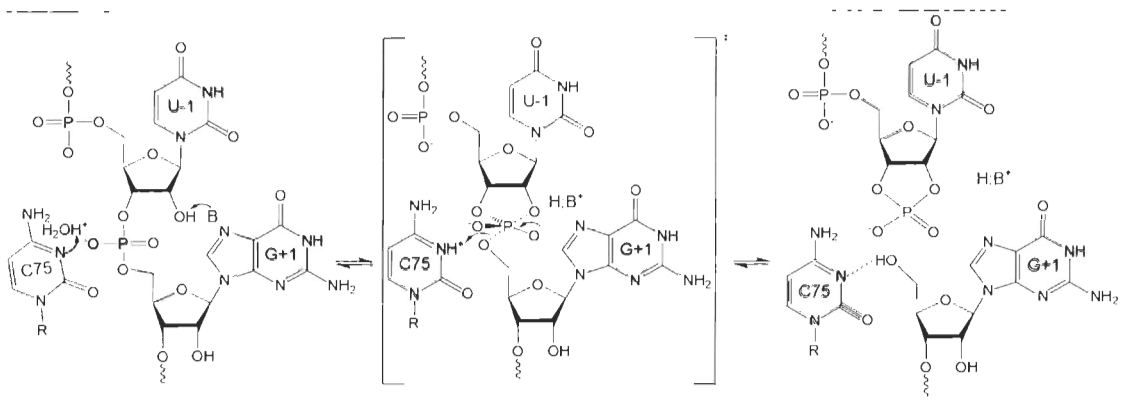


Figure 1-16: General acid catalysis by the HDV ribozyme. A protonated form of the C75 residue serves as a general acid by donating a proton to the 5′ leaving oxygen. An unknown general base, presumably a divalent metal-ion, removes a proton from the nucleophilic 2′ hydroxyl group activating it for the in-line attack (Das & Piccirilli, 2005).

1.6.7 Ribosome.

The ribosome is a large biological complex that is responsible for protein synthesis in all domains of life. It is made up of two subunits that are in turn composed of ribosomal RNA and many proteins. Earlier studies where ribosomal proteins were excluded one at a time suggested that the reaction is catalyzed by the 23S RNA (Noller et al., 1992). The first crystal structures of the large subunit, which is responsible for the peptidyl transfer reaction, revealed that there are no proteins within 20 Å of the active site- labelled by the transition-state analogue CCdAp-Puromycin- confirming that the ribosome is indeed a ribozyme (Ban et al., 2000; Nissen et al., 2000).

In the structure, N3 of the universally conserved A2451 (*E. coli* notation) lied close to the transition-state analogue. The authors proposed a general base role for this residue in removing a proton from the A-site bond aminoacyl-tRNA making it susceptible for an attack by the carbonyl group of the growing peptide (Muth et al., 2000). However, this model was later refuted as the ribosome was found to undergo major conformational switches during catalysis and the original structure captured an

inactive state (Schmeing et al., 2005). Mutational studies by the Green lab indicated that this residue is not critical for the peptidyl transfer reaction, but is required for peptide release (Youngman et al., 2004). As a result of these findings and others, the emerging model is that the ribosome achieves catalysis by mere orientation and proximity of substrates (Fedor & Williamson, 2005). However, recent investigations suggest a substrate-assisted mechanism; in these experiments the rate of peptide bond formation was observed to fall 10^6 fold when the 2′-OH of the 3′-terminal adenosine of the P-site tRNA was removed. Although these findings do not exclude the proximity model, the authors argue that the hydroxyl group might be acting as a general base by abstracting a proton from the amino group (Weinger et al., 2004). The mechanism of the ribosome, to date, remains unresolved and is the subject of hot debates within the community (Rodnina et al., 2006).

1.6.8 Mechanism of an artificial nucleotide synthase.

Although less studied, mechanisms of *in vitro* selected ribozymes offer a back-drop into the range of catalytic strategies available for RNA. Isolated by Unrau and Bartel from a random pool of 10^{15} different sequences, the a15 ribozyme catalyzes the tethered formation of ^{4S}uridine nucleotide from attached 5-phosphoribosyl 1-pyrophosphate (pRpp) and free ^{4S}uracil (^{4S}Ura) with an efficiency that is 10^8 times greater than the uncatalyzed reaction (Unrau & Bartel, 1998). Protein enzymes that catalyze the analogous reaction exploit a number of mechanistically distinct strategies. Probing the transition-state of the a15 ribozyme using kinetic isotope effect indicated that the nucleotide synthase stabilizes a highly dissociative reaction, quite different from the S_N2-

type catalyzed by natural ribozymes, involving an oxocarbenium-ion intermediate (Unrau & Bartel, 2003), (Figure 1-17). Unexpectedly, this mechanism resembles that of hydrolytic protein enzymes, such as uracil DNA glycosylase, which uses the negative charge in its DNA substrate to stabilize the oxocarbenium transition-state.

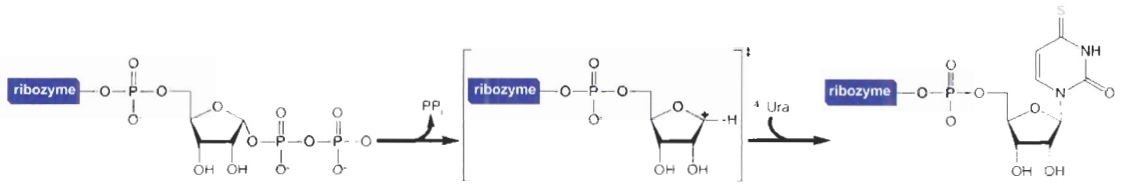


Figure 1-17: Proposed mechanism of the a15 nucleotide synthase ribozyme. The reaction goes through a stepwise dissociative mechanism that involves the parting of the pyrophosphate group to form an oxocarbenium transition state. This is followed by the nucleophilic addition of ^{4S}uracil to form ^{4S}uridine (Unrau & Bartel, 2003).

1.7 RNA-catalyzed RNA replication:

The "RNA world" hypothesis relies on the premise that there exist certain ribozymes that are capable of replicating a ribo-organism genome. Inspired by the mechanisms of modern protein polymerases, researchers have been investigating RNA's capability of polymerizing an RNA primer-template complex as a first step in the creation of a true replicase ribozyme (Bartel & Unrau, 1999; Szostak et al., 2001; McGinness & Joyce, 2003). Conventional *in vitro* evolution techniques have resulted in a number of ribozymes that extend a primer in a template-dependent fashion from nucleotide triphosphates (Been & Cech, 1988; Bartel et al., 1991; Chowrira et al., 1993; Johnston et al., 2001; McGinness et al., 2002; Lawrence & Bartel, 2005). The most advanced being the 189-nt Round-18 polymerase, which can polymerize one primer template complex by more than a helical turn (Johnston et al., 2001). Although this polymerase possesses many of the attributes that are usually associated with protein

enzymes, its activity must be greatly improved before auto-replication can be considered. I have recently used IVC to isolate a superior polymerase ribozyme, which is discussed and compared to other polymerase ribozymes in chapter 4.

1.8 Sub-viral replication strategies:

An RNA replicase, if designed as above, would require two bouts of polymerization reactions to produce one copy of the original strand. The first bout would result in a complementary strand that in all likelihood would form a highly stable duplex with the original strand. This duplex in turn would prevent the synthesis of the desired strand from the complementary strand, unless a strategy to overcome this problem is developed. Contemporary DNA enzymes overcome this obstacle by utilizing the energy from the formation of phosphodiester bonds to perform strand displacement. Although it is conceivable that a ribozyme could be designed to perform strand displacement, no such strategy for an RNA genome exists in modern metabolism. It seems as if nature has evolved a mechanism that is largely different from that of DNA replication to avoid the formation of more stable double stranded RNAs. Replication of sub-viral genomes such as those of hepatitis delta virus and viroids is a case for this point (Symons, 1997; Flores et al., 2005).

Viroids are small single stranded RNAs, 200-400 nt in length, that infect a variety of plant species. The genomes are circular with high degree of self-complementarity (Gross et al., 1978), and do not code for any proteins, and hence viroids are the smallest known pathogens (Flores et al., 2004). Viroid replication proceeds via a rolling-circle mechanism, thought to be catalyzed by the host RNA polymerase II (Muhlbach &

Sanger, 1979). The ability of these RNAs to hijack the active site of an enzyme that normally accepts DNA is currently not understood. The replication process results in two strands, a long linear strand containing many genome-lengths, and a long linear positive strand that is then processed into the circular genome (Daros et al., 1994). The mechanism of autocatalytic amplification of an RNA called X by T7 RNAP is thought to be analogous to that of viroids (Konarska & Sharp, 1989, 1990). RNA X is a 64-nt linear sequence that is almost completely self-complementary, and also has a four-fold symmetry allowing it to adopt a clover-leaf structure (Figure 1-18). The replication of RNA X, like those of viroids, requires single stranded templates, and both the template and replica are released as single stranded. I have investigated the mechanism of this process by probing the sequence and secondary structure requirement for T7 RNAP-mediated extension of oligonucleotides (Zaher & Unrau, 2004), see chapter 5.

It is highly conceivable that primordial RNA replicases utilized strategies similar to present-day mechanisms. In this case, the template and replica structures would have contributed significantly to strand displacement. Finally, the observation that single stranded RNA genomes are amplified by DNA-dependent RNA polymerases suggest that the ability of single strands of RNA to replicate may be a more general property of RNA (Wettich & Biebricher, 2001).

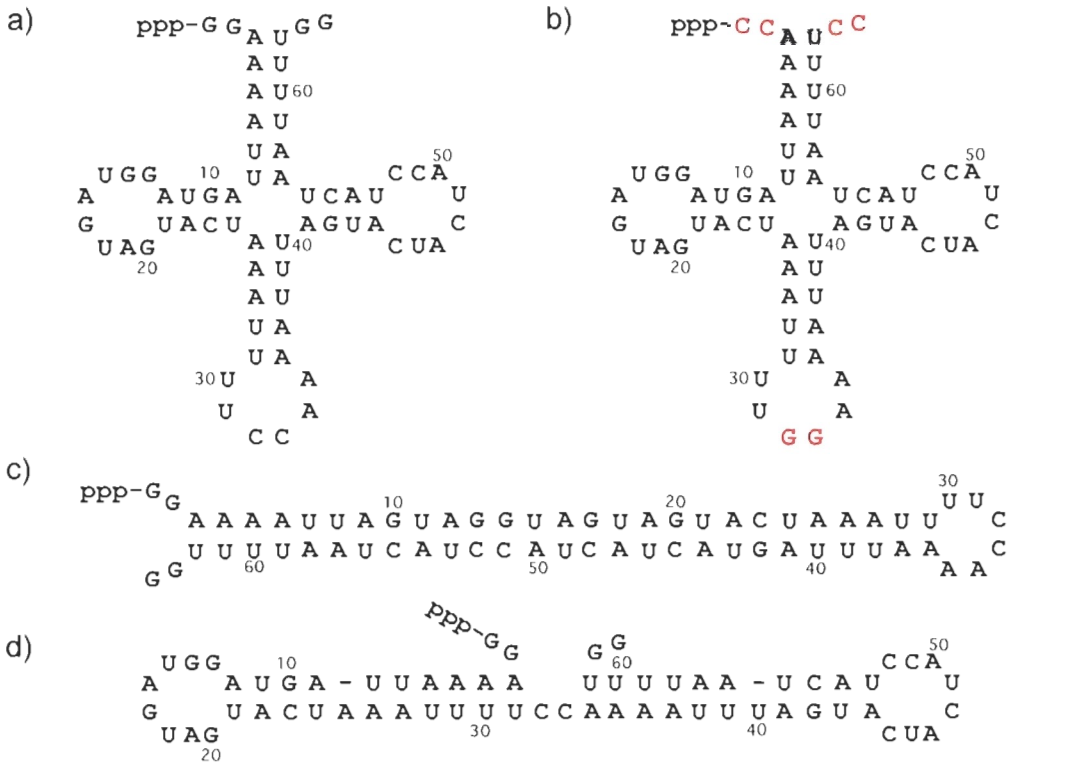


Figure 1-18: RNA X adopts alternative secondary structures. The clover-leaf structure of the '-' (G) strand, **a)**, and '+' (C) strand, **b)**. The '+' strand differs from the '-' one by only having 2 cytidine residues instead of guanosines at each of its 5′ terminus and 3′ terminus, and 2 guanosine residues instead of cytidines in the middle loop (The different residues are displayed in red**)**. Alternatively, RNA X has the potential to form a stable rod-like structure, **c)**, or a dumbbell structure, **d)**. In both cases only the '-' strand is shown.

CHAPTER 2:
Two independently selected capping ribozymes share similar substrate requirements

This chapter is largely based on the manuscript; Zaher HS, Watkins RA, Unrau PJ. 2006. "Two independently selected capping ribozymes share similar substrate requirements". *RNA* **12**:1949-1958.

2.1 Introduction:

Enzymes are noteworthy in that they efficiently perform specific chemistry. But how did these catalysts evolve and what were the fundamental constraints that governed their early emergence? These questions are difficult to address with protein enzymes, as billions of years of evolution have often erased the information required to understand the evolutionary progression from an emergent to a fully competent catalyst (Bartel & Unrau, 1999). The emergent catalytic properties of RNA are much easier to study than that of protein, as RNA catalysts (ribozymes) can be selected artificially from pools of random sequence in the laboratory (Wilson & Szostak, 1999; Joyce, 2002; Fiammengo & Jaschke, 2005). The emergent properties of these RNAs may be of particular importance as catalytic RNAs are critical to modern metabolism (Ban et al., 2000; Winkler & Breaker, 2005) and were likely to have played a dominant role early in evolution prior to the emergence of protein catalysis (Gilbert, 1986; Joyce, 2002).

Artificially selected ribozymes, like protein enzymes, have generally been shown to exhibit good substrate discrimination over a range of chemistries (Cech et al., 1981; Unrau & Bartel, 1998; Seelig & Jaschke, 1999; Wilson & Szostak, 1999; Lau et al., 2004; Tsukiji et al., 2004; Fiammengo & Jaschke, 2005). In some cases the basis for substrate discrimination may be obvious; such is the case with self-cleaving ribozymes, where their

base-pairing properties make possible highly specific phosphodiester bond cleavage reactions (Doherty & Doudna, 2000). In other cases the mechanism behind substrate discrimination and its importance to catalysis is harder to resolve. For example, we have found purine and pyrimidine nucleotide synthase ribozymes that exhibit good substrate discrimination between small substrates having molecular weights less than 170 Da (Unrau & Bartel, 1998; Lau et al., 2004). This substrate specificity is achieved even though these ribozymes exploit a range of catalytic strategies – some exhibit good substrate binding affinity, while others have high intrinsic chemical rates but poor substrate affinity (Lau et al., 2004). This indicates that, at least for nucleotide synthesis ribozymes, a range of catalytic strategies is available. A detailed investigation of the catalytic strategies utilized by other artificially selected ribozymes performing chemistries with a range of small molecule substrates will help to further delimit the catalytic properties of newly emergent ribozymes, such as those possibly encountered early in evolution.

The chemistry of capping provides a unique window into the importance of substrate discrimination and RNA catalysis. Capping involves the exchange of nearly equivalent phosphate groups so as to form a nucleotide 5′-5′ cap on an RNA strand using a small nucleotide substrate (Equation 2-1). During the course of enzyme-mediated capping, the phosphate moieties of the nucleotide are required for chemistry, while the distal base and sugar of this nucleotide are utilized to provide substrate recognition and discrimination. As nucleotide substrates are, in principle, easily recognized by RNA it might be expected that capping ribozymes like their protein counterparts, would also distinguish between nucleotide substrates. An exploration of the catalytic and substrate

utilization of capping ribozymes should therefore provide insight into how the requirements imposed by chemistry regulate RNA-mediated catalysis.

Equation 2-1

Previously Huang and Yarus isolated a general RNA capping ribozyme called Iso6. Iso6 accelerates the formation of a variety of 5′-5′ RNA cap structures and releases pyrophosphate in the presence of Ca^{2+} ions (Huang & Yarus, 1997a). Notably the ribozyme is very promiscuous in its substrate requirements. Iso6 only requires a terminal unblocked phosphate-containing nucleophile to attack the α-phosphate of the ribozyme's 5′-triphosphate (Huang & Yarus, 1997c). Additionally, the ribozyme possesses pyrophosphatase, decapping and exchange activities (Huang & Yarus, 1997b). The ribozyme displays a similar ~10 μM affinity for both guanosine triphosphate and diphosphate, but exhibits a substrate affinity towards guanosine monophosphate that is ~15 fold lower (Huang & Yarus, 1998). In this chapter, we report the isolation of a second capping ribozyme, called 6.17, from an independent selection and using a different selective approach. Intriguingly, this ribozyme has similar kinetic and substrate recognition properties to Iso6 – even though both ribozymes have distinct metal ion preferences and are likely to adopt different secondary structures. This suggests that, in contrast to the majority of small molecule reactions mediated by RNA, the chemistry of capping is incompatible with facile substrate discrimination; an interesting finding given

the proven ability of RNA to distinguish between different nucleobases (Noeske et al., 2005).

2.2 Materials and methods:

2.2.1 Pool construction and isolation of capping ribozyme 6.17.

A high diversity DNA pool having the sequence *TTCTAATACGACTTATA*GGAC CGAGAAGTTACCC-$N_{(76)}$-CCTTGG-$N_{(76)}$-GGCACC-$N_{(76)}$-ACGCACATCGCAGCAA C (italics indicate the T7 promoter sequence, N indicates a random nucleotide position) was made from three synthetic single-stranded pools, as described previously (Unrau & Bartel, 1998; Zaher & Unrau, 2005). RNA transcripts were made by *in vitro* T7 transcription (40 mM Tris-HCl pH 7.9, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 4 mM ATP, 4 mM CTP and 2 mM UTP and 4 U/µl T7 RNAP incubated for 1 h at 37°C), (Zaher & Unrau, 2004). A total of ~ 1.5×10^{15} different RNA sequences were used with approximately 5 to 6 copies of each sequence in the initial pool (a total of 15 nMoles of RNA). Initially the selection was designed to isolate RNA polymerase ribozymes. This was carried out by incubating the RNA with an RNA template designed to hybridize to the 3′-end of the pool and to provide a template for ^{4S}UMP incorporation using the following incubation conditions: 0.3-0.5 µM RNA, 1.0 µM template, 2 mM ^{4S}UTP, 50 mM Tris-HCl pH 7.6, 25 mM $MgCl_2$, 150 mM KCl at 22°C for 20 hr. The crude ^{4S}UTP used in Rounds 1 and 2 of the selection was synthesized by incubating 10 mM ^{4S}UDP and 20 mM ATP with 50 mM Tris-HCl pH 7.8, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 5mM DTT and 1mM spermidine together with 0.15 U/µl of nucleoside-5′-diphosphate kinase (Sigma) for 30 min at room temperature. This

preparation was then crudely enriched and used in the pool incubation. The final composition of the synthesis was approximately 62% ^{4S}UTP, 12% ADP and 26% ATP as determined by analytical HPLC. Further rounds used ^{4S}UTP from USB. After incubation with ^{4S}UTP, the pool was dehybridized from the template by the addition a competing oligonucleotide able to hybridize to the template and EDTA to chelate Mg^{2+}. The pool was then gel purified. Reactive RNA able to incorporate a ^{4S}U residue were then purified on a 5 % PAGE gel containing APM, and visualized using a Storm 820 phosphorimager. Round four of this pool was found to contain both template-dependent and template-independent sequences (the former to be discussed elsewhere). A template-independent selection was carried out by excluding the template from future rounds in order to isolate primer independent self-tagging ribozymes. In rounds 5 and 6 of the selection the ^{4S}UTP concentration was lowered to 1 mM. Incubation times during the selection for rounds 1 to 6 were 24, 24, 24, 23, 1 and 1 h respectively. By round 6, the selection was stopped; the RNA was cloned and 25 isolates sequenced. After screening for activity, isolate 6.17 was studied in detail due to its faster kinetics. The truncated *cis* sequence c6.17 has been submitted to GenBank and has accession number DQ371299.

2.2.2 Characterization of *cis* acting capping ribozyme 6.17.

The site of nucleotide addition was mapped by incubating the ribozyme with [α-^{32}P]-UTP and subjecting the resulting radiolabelled RNA to partial alkaline hydrolysis (5 μM RNA in 50 mM $NaHCO_3$ at 90°C for 30 min in the presence of 0.5 μg/μl tRNA). The reaction was stopped by adjusting the pH to 7.0 using 1 M Tris-HCl. A G-ladder was produced by partial T1 digestion (5 μM RNA, 0.5 μg/μl tRNA, 6 M Urea, 0.02 U/μl

T1 (Fermentas) at 22°C for 2 minutes). The reaction was stopped by the addition of 2 × formamide loading dye (95% formamide, 10 mM EDTA, 0.025 % Xylene Cyanol and 0.025 % bromophenol blue) and storing the sample on dry ice before loading onto a sequencing gel.

The phosphates of the incorporated nucleotide were shown to be protected from phosphatase by subjecting the ribozyme that had been capped with [α-^{32}P]-UTP to calf intestinal phosphatase (CIP) (New England Biolabs); 50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT supplemented with 1 U/µl of enzyme and incubating for 1 h at 37°C.

Kinetic reactions were carried out under selection conditions using 0.5 µM ^{32}P-body-labelled RNA. To calculate the K_m of isolate c6.17, time-courses with nucleotide concentrations ranging from 17 µM to 3 mM were performed. Initial rates were extracted from time course data and plotted against ^{4S}UTP, ^{4S}UDP, and ^{4S}UMP concentrations and fit to the Michaelis-Menten equation.

2.2.3 t6.17 capping assay.

Isolate 6.17 was truncated and separated into two modules called A and B to facilitate analysis. Module A consisted of residues 150 to 222 of the 6.17 isolate, while module B consisted of residues 1 to 33 and contains the reactive 5' triphosphate that is capped. Module B was 3'-end labelled by incubating 5 µM RNA with 2 µM [^{32}P]-pCp in 50 mM HEPES pH 8.3, 10 mM DTT, 10 µg/ml BSA, 8.3 % v/v glycerol, 1 mM ATP, 1 U/µl T4 RNA ligase for 2 h at 22°C (Wang & Unrau, 2002). The labelled RNA was then gel purified by 12% PAGE. [^{32}P]-pCp was synthesized by incubating 30 µM Cp (Sigma)

in polynucleotide kinase buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6) supplemented with 5 μCi/μl of [γ-^{32}P]-ATP (3000 Ci/mmol, NEN) and 0.5 U/μl of T4 polynucleotide kinase (Invitrogen) at 37°C for 30 minutes. The enzyme was then inactivated by heating at 65°C for 15 minutes.

Capping reactions were carried out, in single turnover conditions, by annealing module A (2 μM) and module B (0.2 μM) by heating at 80°C for 5 minutes and gradually cooling to 22°C before the addition of the selection buffer. Reactions were started by the addition of nucleotide at a concentration of 1 mM. Time points were taken at 3 h and 24 h and samples were then resolved using 15 % sequencing PAGE.

2.2.4 Decapping and exchange activity.

Module B was first capped with UTP and carefully gel purified on a 12 % gel. The resulting capped construct was then incubated with a five-fold excess of module A under varying concentrations of ^{4S}UTP. Time-points were taken and resolved using 15 % APM gels in order to separate ^{4S}UTP caps from the initial UTP capped substrate and hydrolyzed products.

2.2.5 Capping rate as a function of pH.

[^{32}P]-pCp labelled module B was annealed to module A as before and incubated in 50 mM Mes (pH 5.7), Mes (pH 6.0), Mes (pH 6.5), Bes (pH 7.0), Bes (pH 7.5), Tris (pH 7.6), Epps (pH 8.5) and Ches (pH 9.5) buffers, in the presence of 20 mM $MgCl_2$, 150 mM KCl and 1 mM ^{4S}UTP. Aliquots were taken at 4, 8, 24 and 48 h and separated on a 10 % PAGE containing 10 μM APM. The time-courses from three independent

experiments were fit to first order kinetics, and apparent capping rates were then extracted and plotted against the pH values.

2.2.6 Metal requirement and Mg^{2+} titrations.

[^{32}P]-pCp labelled module B was annealed to module A as before and incubated in 50 mM Tris pH 7.6, 150 mM KCl and 1 mM ^{4S}UTP in the presence of one of the following metal ions; Mg^{2+}, Mn^{2+}, Ca^{2+}, $Co^{3+}(NH_3)_6$, Zn^{2+} and Ni^{2+} at 5 mM. Time points were taken and resolved on an APM gel. Dependence of capping on Mg^{2+} was carried out by performing capping reactions in the presence of Mg^{2+} concentrations spanning 1.25 to 20 mM. Aliquots were taken at 4, 8, 24 and 48 h and separated on a 10 % PAGE containing 10 μM APM. The apparent capping rates from three different experiments, resulting from a first-order fit, were plotted against the Mg^{2+} concentrations and fit to the Hill equation.

2.2.7 Secondary structure determination.

The protection of A and G residues within either module A or B were probed using chemical and enzymatic methods under three different incubation conditions: denaturing, native and assembled. Denaturing reactions were carried out in 25 mM sodium cacodylate pH 7.5, 150 mM KCl and 5 mM EDTA and at a temperature of 90°C, while the native reactions were performed in 25 mM sodium cacodylate pH 7.5, 150 mM KCl and 25 mM $MgCl_2$ at 22°C. Probing of hybridized module A and module B were performed in native buffer. Modifications of the N7 position of adenosine residues were carried out by incubating 200 μl of RNA sample with 1 μl DEPC (Sigma) at 90°C for 15

minutes (denaturing conditions) or 10 µl DEPC at 22°C for 1 h (native and assembled conditions). The RNA was then ethanol precipitated twice. Cleavage was initiated by resuspending the pellet in 20 µl of 1 M aniline-acetate pH 4.5 at 60°C and incubating for 15 minutes in the dark. The aniline was removed by speed-vacuuming for two hrs. The nucleic acid was resuspended in formamide loading dye. Partial enzymatic T1 digestion of G residues were carried out for the three incubation conditions. The cleavage products from the chemical modifications and T1 digestion were resolved by 15 % PAGE alongside a base-hydrolysis ladder of the appropriate end-labelled module.

2.2.8 Photo-crosslinking of ^{4S}U cap structures to the ribozyme.

Module A was dephosphorylated using CIP and the resulting RNA 5′-labelled with [γ-^{32}P]-ATP (3000 Ci/mmole, NEN) at 2 µCi/µl in the presence of T4 polynucleotide kinase at 0.5 U/µl in polynucleotide kinase buffer for 30 minutes at 37°C. This was followed by gel purification on a 10 % gel. A small amount of the resulting radio-labelled module A was annealed to an excess of module B and a capping reaction initiated by the addition of ^{4S}UTP or ^{4S}UMP at a concentration of 1 mM. After 24 hrs of incubation the sample was exposed to UV light through a ELISA plate (selecting > 300 nm) from a UV transilluminator (Fotodyne, Bio/Can Scientific) for 30 min. The cross-linked RNA was gel purified using 10 % PAGE. The site of the cross-link was mapped using T1 digestion in urea and base hydrolysis as described earlier, and the resulting fragments were resolved on a 12 % sequencing gel. The same analysis was also performed for 3′-labelled module B (using [α-^{32}P]-pCp).

2.3 Results:

The selection was initially designed to isolate possible RNA polymerase ribozymes based on their ability to incorporate ^{4S}UMP onto their 3′-ends when annealed to a poly(A) template and incubated with 1-2 mM concentrations of ^{4S}UTP. Such candidates were enriched based on their ability to be super-shifted on a gel containing N-acryloyl-aminophenyl-mercuric acid (APM) (Igloi, 1988). Due to the large incubation volumes of Round 1 and 2, crude ^{4S}UTP containing significant amounts of ADP and ATP was used, while pure ^{4S}UTP was used in Rounds 3 to 6. By Round 6, 2.5% of the pool RNA was retained on an APM gel after 4 hours of incubation with no sign that activity increased in further rounds of selection. Twenty-five isolates were therefore cloned from Round 6 and screened for activity. Clone 6.17 showed the fastest kinetics, with ~25% reacting after 4 hours, and was selected for further study. Analysis of the reaction mediated by this ribozyme indicated that this sequence did not perform a polymerase type activity at the 3′-terminus but rather formed a nucleotide cap at its 5′- terminus. This 272 nt long ribozyme was truncated from its 3′-end as well as internally resulting in a fully active 94 nt *cis* construct referred to hereafter as c6.17. Nevertheless, the 94 nt size of the ribozyme still made it difficult to study capping with nucleotides lacking thiol tags. To facilitate analysis, a two component *trans* ribozyme (called t6.17, shown in Figure 2-1a) was designed by cutting c6.17 in the loop closing helix I. A short 33 nt long sequence called module B, possessing a 17 nt long tail that was not important for activity but that increased transcriptional yield, became capped when incubated with a 73 nt long module A (residues 150 to 222 of the initial ribozyme). The t6.17 construct retained nearly all of

the activity of the *cis* construct and had only a slightly lower rate constant and substrate binding affinity than its c6.17 parent under single turnover conditions.

2.3.1 Ribozyme 6.17 synthesizes 5′-5′ caps.

Trans mediated capping of module B was confirmed by the following set of experiments. 1. Removing the triphosphate from the module B transcript and replacing it with a 5′ hydroxyl or 5′ monophosphate inactivated the ribozyme. 2. Unlabelled ribozyme incubated with [α-^{32}P]-UTP became radiolabelled and did not lose this ^{32}P label upon treatment with Calf intestinal phosphatase (CIP) suggesting that the α-phosphate of the incorporated nucleotide was not exposed. Furthermore, the location of this modification was mapped by alkaline hydrolysis to the very 5′ end of module B. 3. When module B was labelled with [γ-^{32}P]-GTP by transcription and then incubated in the presence of module A, radioactive pyrophosphate was released. 4. Unlabelled ribozymes became radiolabelled with no appreciable accumulation of radiolabelled phosphate or pyrophosphate when capped with [γ-^{32}P]-ATP as judged by gel electrophoresis containing phosphate and pyrophosphate standards. The simplest interpretation of these four independent pieces of data is that the ribozyme mediates a reaction where the α-phosphate of the 5′ triphosphate found on module B is joined to the terminal phosphate found on the nucleotide substrate.

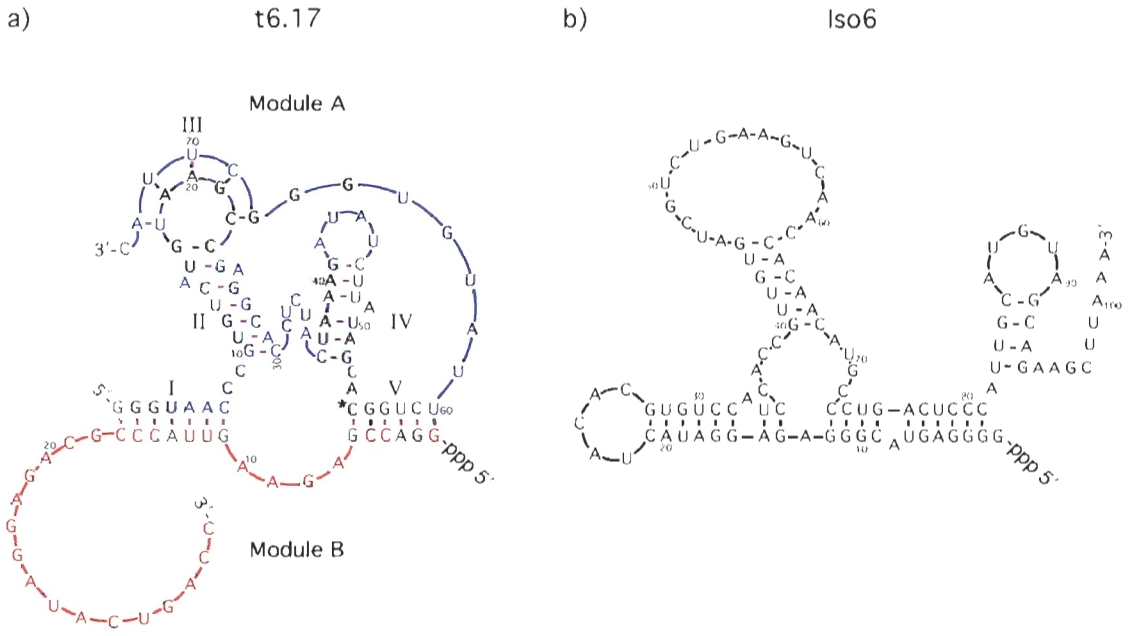


Figure 2-1: Proposed secondary structures of capping ribozymes 6.17 and Iso6. a) The secondary structure of t6.17, module A is in blue, module B is in light red. RNA substrate recognition (module B) is accomplished by helices I and V. In addition, module A has two hairpins, II and IV. The loop of hairpin II forms a pseudoknot with the 3′ terminus of module A, forming helix III. Black residues are protected from T1 nuclease and DEPC modification in the absence of the other module. Green residues are protected only in the presence of the other module. Pink bars represent proposed base-pairing interactions, while black bars indicate base-pairing confirmed by manual co-variation experiments. The black oval indicates the importance of residue U_{A60} on module A in forming a wobble base-pair with G_{B1} of module B. The asterisk indicates where the ^{4S}U cap quantitatively cross-links when exposed to UV light. **b)** The secondary structure of Iso6 proposed by Huang and Yarus. Adapted from (Huang & Yarus, 1997b).

2.3.2 Secondary structure of the ribozyme.

The bimolecular t6.17 capping ribozyme was particularly amenable to structural probing. Diethyl pyrocarbonate (DEPC) probing of the N7 position of adenosine residues and T1 nuclease partial digestion at unstructured G residues were particularly informative. Module A contained many residues that were enzymatically or chemically protected even in the absence of module B. Guanosine residues G_{A17}, G_{A21}, G_{A56}, G_{A57}, G_{A66}, G_{A67}, and G_{A68} were found to be protected from T1 in native conditions (Figure 2-2a, left panel). Residues A_{A19}, A_{A20}, A_{A38}, A_{A39}, A_{A40}, A_{A51} and A_{A54} were protected from DEPC modification, and were also independent of the presence or absence of

module B (Figure 2-2a, right panel). The protection of A_{A19}, A_{A20}, G_{A21} and G_{A68} residues suggested the possibility of a pseudoknot interaction between residues U_{A18}-C_{A22} and G_{A68}-A_{A72} resulting in helix III (this helix can in principle range in length from 5 to 7 bp). Point mutants designed to destabilize this helix were constructed and a complete co-variation of the base pair G_{A21}:C_{A69} to U_{A21}:A_{A69} was found not to affect catalytic rate, while the individual point mutants had a significant impact on the ribozyme's ability to promote capping (Figure 2-2.b).

Guanosine residues G_{B2}, G_{B6} and G_{B11} of module B were enzymatically protected only in the presence of module A (Figure 2-2c, left panel). DEPC modification of module B also showed the specific protection of adenosine residues A_{B3} and A_{B14} in the presence of module A (Figure 2-2c, right panel). Based on the T1 and DEPC protection patterns for module B, we propose that module A and module B recognize each other through two base-paired regions: helix I consisting of nucleotides G_{A1}-G_{A7} and nucleotides G_{B11}-C_{B17}, and helix V consisting of nucleotides C_{A55}-U_{A60} and nucleotides G_{B1}-G_{B6}. Point mutants that disrupted a proposed base-pair in helix V (G_{A56}:C_{B5}) completely inhibited activity. Consistent with the existence of a helix, changing the G_{A56}:C_{B5} pair to a U_{A56}:A_{B5} pair, restored capping activity (Figure 2-2d).

A model of the t6.17 ribozyme, consisting of five helical elements, is shown in Figure 2-1a. Our ribozyme structure is consistent with DEPC and T1 probing as well as co-variation data obtained for two of these helical elements. Module B recognition is achieved via helices I and V. In addition, module A has two helical elements, II and IV, with the pentaloop protruding from helix II interacting with the 3′- tail of the ribozyme to form a pseudoknot, labelled helix III. Interestingly, this helix can be extended to form a

7 bp helix and the exact interplay between helix II and III will require further study to resolve. The bulged helixes II and IV are predicted to exist based on thermodynamic folding (mfold V3.1.2; (Zuker, 1999)). Consistent with this model, the two bulged adenosines in helix II were sensitive to DEPC treatment in native conditions. Helix IV was supported by the existence of protected residues A_{A38}, A_{A39}, A_{A40}, and A_{A51}. This relatively complex RNA motif consists of 31 bp and approximately 22 nt of interhelical sequence. The t6.17 ribozyme appears likely to adopt a secondary structure different from Iso6. Using thermodynamic folding, S1 nuclease protection and lead cleavage mapping, Huang and Yarus predicted that the secondary structure for Iso6 consists of four helical elements (Figure 2-1b; (Huang & Yarus, 1997b)). Iso6 may potentially form a pseudoknot-like structure by the pairing of residues A_{54}-U_{57} with residues G_{96}-U_{99}. However, the resulting structures would remain topologically distinct from that of 6.17; in particular helix IV of 6.17 is located 3′ to the reactive 5′-terminal α-phosphate of the ribozyme, while the corresponding 3′-end hairpin of Iso6 would be 5′ to the reaction centre.

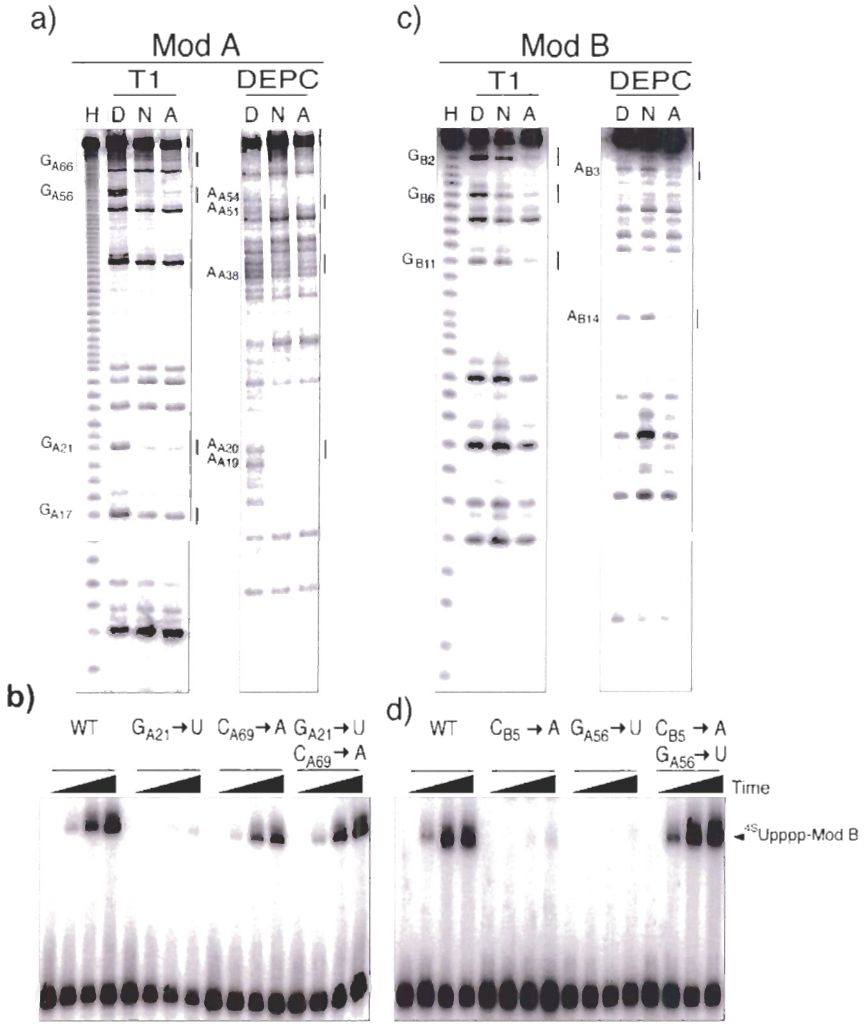


Figure 2-2: Secondary structure probing and manual co-variation experiments. a) T1 partial hydrolysis reactions (left panel) and DEPC reactions (right panel) of [^{33}P]-5′-end labelled module A. T1 hydrolysis showed the protection of module A guanosine residues G_{A17}, G_{A21}, G_{A56}, G_{A57}, G_{A66}, G_{A67} and G_{A68} in native conditions that were independent of the presence of module B. DEPC treatment indicated that adenosine residues A_{A19}, A_{A20}, A_{A38}, A_{A39}, A_{A40}, A_{A51} and A_{A54} of module A were involved in secondary structure formation and were again independent of module B. In all cases H indicates hydrolysis ladder, D indicates denatured complex, N indicates native conditions in the absence of the other module, while A indicates probing of the fully assembled construct under native conditions. **b)** A predicted pseudoknot in module A was confirmed by constructing point mutants in G_{A21} and C_{A69}. While mutating G_{A21} to U completely inhibited activity, altering C_{A69} to an A residue lowered the capping rate to a 1/3 that of the wild-type. When the double mutation was assayed, capping rate was restored to wild-type levels. **c)** T1 partial hydrolysis reactions (left panel) and DEPC reactions (right panel) of [^{32}P]-3′-end labelled module B. Residues G_{B2}, G_{B6} and G_{B11} of module B were protected from T1 only in the presence of module A. DEPC treatment of module B showed the protection of adenosine residues A_{B3} and A_{B14} only in the presence of module A. **d)** Kinetics of point mutants designed to test the interaction between modules A and B. Residues G_{A56} and C_{B5} were mutated so as to disrupt a predicted base pair. Only the double mutant that changed the base pair from a C:G to a A:U was as functional as the wild type sequence supporting the predicted secondary structure.

2.3.3 Cross-linking of capped products.

^{4S}UTP is an efficient photo-induced crosslinking reagent when excited by ~310 nm light. We found that when module B was capped with either ^{4S}UTP or ^{4S}UMP, a quantitative crosslink formed with module A. When these cross-linked RNA species were gel purified and subjected to base hydrolysis and T1 digestion, the cross-links were mapped in each case to residue C_{A55}, even though the length of the phosphodiester linkage varied from 2 to 4 phosphates (Figure 2-3a). As expected, labelling module B at its 3′ end with [^{32}P]-pCp produced a crosslink with identical mobility that mapped to the extreme 5′ end of module B. The location of the crosslink, C_{A55}, is at the opposite end of helix V, and is 6-bp away from the triphosphate required for capping activity (Figure 2-3a). We hypothesized that if helix V were flexible near the site of capping, it could allow the observed cross-link. This idea was tested by constructing point mutants that either stabilized or destabilized the triphosphate end of helix V. Replacing the terminal U_{A60}:G_{B1} wobble pair with C_{A60}:G_{B1} destroyed capping activity, while replacing the wobble pair with A_{A60}:G_{B1} lowered capping efficiency by ~6 to 7 fold (Figure 2-3b). This data indicates that helix V needs flexibility near the cap site, and that the wobble U_{A60}:G_{B1} base pair is required for optimal activity.

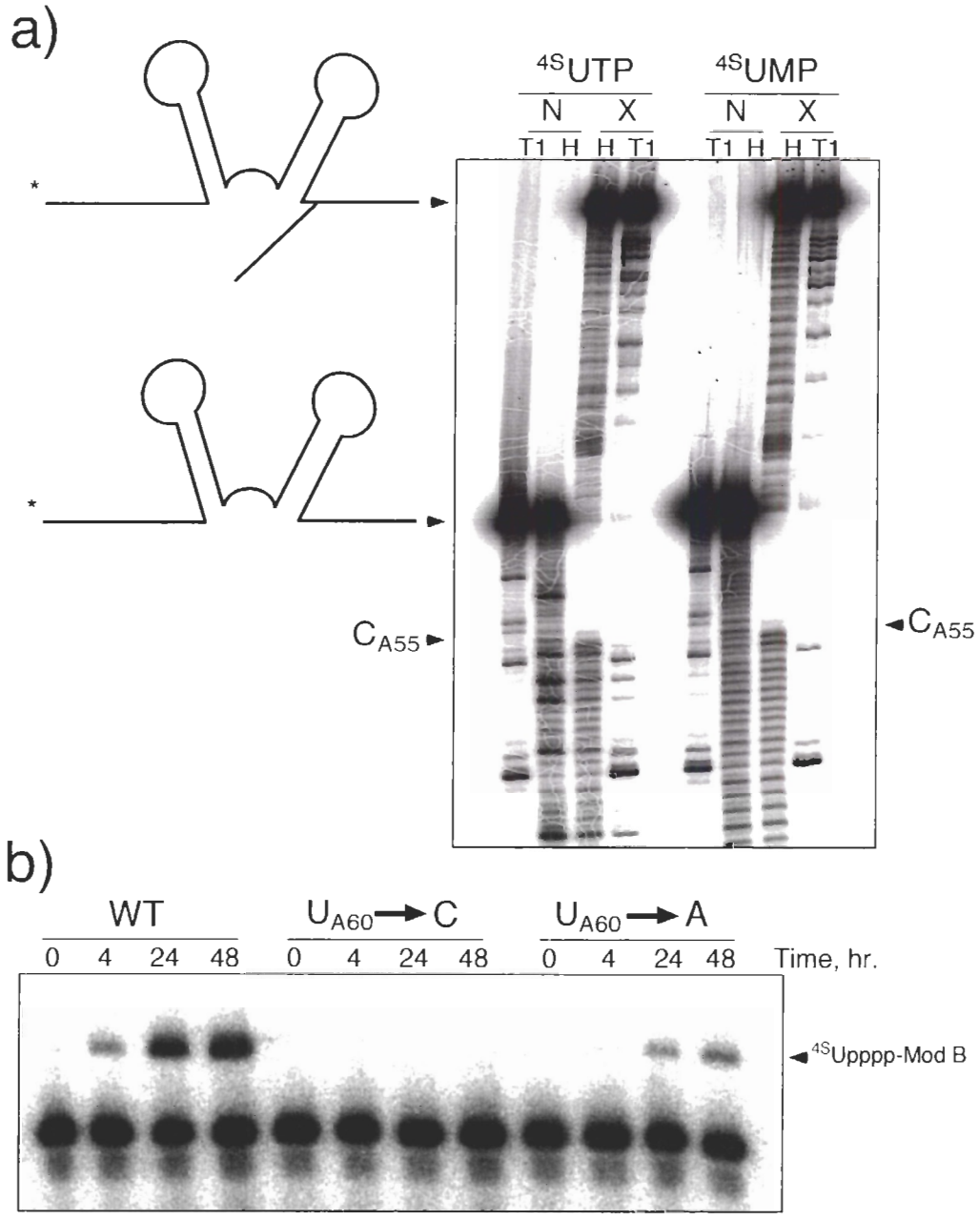


Figure 2-3: Cross-linking of capping products and flexibility of helix V. a) Base hydrolysis (H) and T1 partial digestion (T1) of ^{4S}UTP-capped (left lanes) or ^{4S}UMP-capped (right lanes) module B cross-linked to a [^{32}P]-5′-end labelled module A (X). The cross-link was formed by exposing a 24 h time-point of a capping reaction to UV of wavelength > 300 nm for 30 minutes, nearly 100% of the ^{4S}U capped product formed a single cross-linked species. Unreacted ribozyme that did not show a mobility shift on an APM gel did not cross-link when exposed to UV light (N) and was also subjected to base hydrolysis and partial T1 digestion. For both cap types, the cross-link was assigned to C_{A55}. **b)** The importance of the U_{A60}:G_{B1} wobble base-pair, was assessed by constructing point mutants. The first of which altered U_{A60} to C and stabilized the base-pair, but abolished capping. Destabilizing the end of the helix by mutating the U_{A60} to an A residue was tolerated but capping activity was lowered 6 to 7 fold relative to the wild-type.

2.3.4 Metal ion dependence and magnesium binding of 6.17 ribozyme.

To test the metal ion requirements of 6.17, t6.17 was incubated in the presence of Mg^{2+}, Ca^{2+} and Mn^{2+}, $Co^{3+}(NH_3)_6$, Zn^{2+} and Ni^{2+}. Capping took place only in the presence of Mg^{2+}, Ca^{2+} and Mn^{2+}; however, the capping rate slowed ~2 fold when Mn^{2+} was substituted for Mg^{2+} and by a factor of ~7 fold when only Ca^{2+} was present (Figure 2-4a). Supplementing a reaction containing 20 mM Mg^{2+} with 5 mM Ca^{2+} or 5 mM Mn^{2+} had no effect on the capping rate, suggesting that Mg^{2+} was the preferred metal ion for catalysis and that Ca^{2+} and Mn^{2+} do not effectively compete for binding. A Hill analysis indicated that several ion binding sites were important for catalysis or folding and that magnesium was bound loosely by the ribozyme (see Figure 2-4b). In contrast, the capping ribozyme Iso6 isolated by Huang and Yarus (Huang & Yarus, 1997a) exhibited a strong preference for Ca^{2+} and showed no activity even when incubated with 100 mM Mg^{2+}. In fact, both magnesium and strontium were found to be competitive inhibitors for the calcium dependent capping activity of the Iso6 ribozyme (Huang & Yarus, 1997b).

2.3.5 Capping is dependent on nucleotide phosphate chain length but not on base or nucleotide sugar composition.

Substrates containing 5′ phosphate groups were required for catalysis although the length of this phosphate group could be varied. The general substrate requirements exhibited by 6.17 are similar to those displayed by Iso6, which also reacts with a range of substrates that contain at least one exposed phosphate group (Huang & Yarus, 1997c). The first panel of Figure 2-5 shows t6.17 reacting with ^{4S}UTP, ^{4S}UDP, ^{4S}UMP but not ^{4S}U. Capping reactions were observed with a broad range of ribose, deoxyribose and dideoxyribose nucleotide triphosphates and nucleotide monophosphates, indicating that

neither the sugar nor the base composition was essential for activity (Figure 2-5, panels 2, 3, 4 and 5). Dinucleotides containing only internal phosphates were found not to react and implicate the importance of a 5′ nucleotide phosphate of variable length in the ribozyme reaction (Figure 2-5, panel 6).

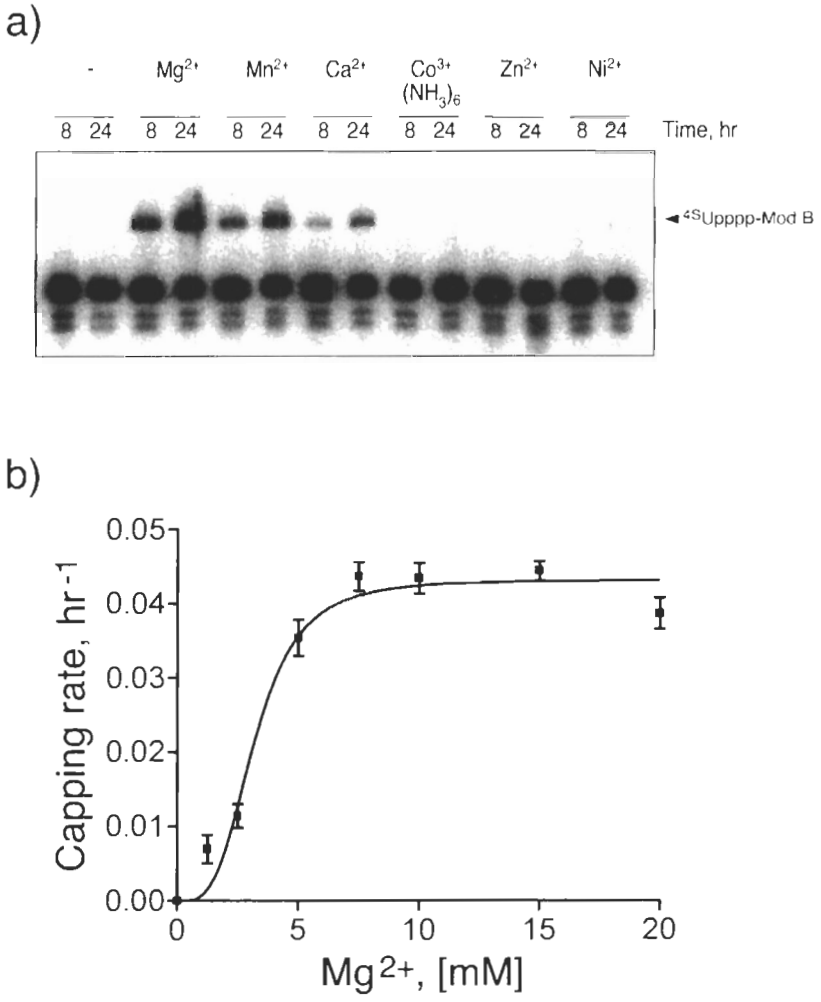


Figure 2-4: Metal-ion requirements and Mg^{2+} dependence. a) Standard ^{4S}UTP capping reactions (pH 7.6) were performed in the presence of 5 mM of the indicated metal ions. Samples were taken at the indicated times and run on a 10% APM gel. Mn^{2+} supported capping with a rate half of that observed in the presence of only Mg^{2+}. Capping also tolerated Ca^{2+}, but with a rate one seventh of that observed with Mg^{2+}. **b)** A cooperative binding analysis using a fit to three independent magnesium titration experiments. ^{4S}UTP capping rates were determined and plotted against Mg^{2+} concentration. A Hill-coefficient of 3.5 ± 0.6, with a binding constant of 3.2 ± 2 mM was measured (assuming equivalent binding sites).

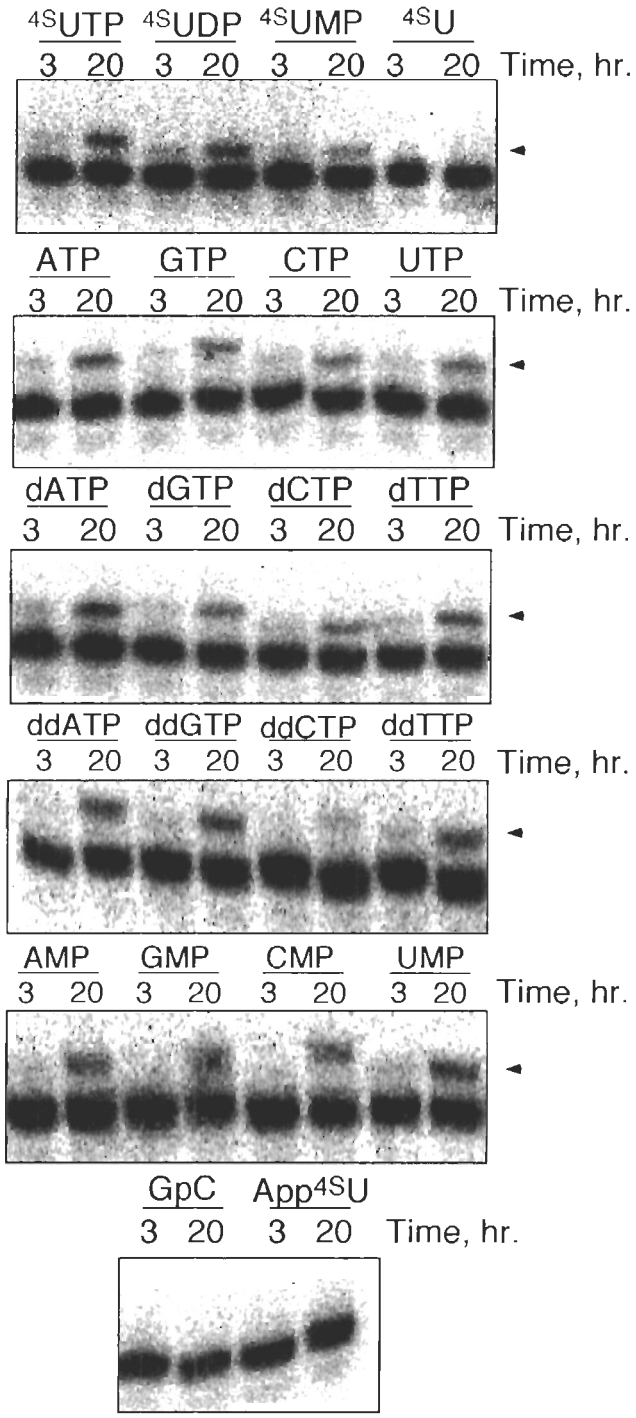


Figure 2-5: General substrate utilization by the capping ribozyme. Capping takes place with nucleotides having at least one 5′-phosphate group, while substrates having internal phosphates (GpC and App^{4S}U) do not support capping. Module B was [^{32}P]-pCp labelled and annealed to a 10 fold excess of module A to form a reactive ribozyme complex. Reactions were started by the addition of 1 mM of the indicated substrate to the ribozyme incubation. Time-points were taken at 3 and 20 h and resolved using 15% PAGE.

Decreasing the phosphate chain length of the nucleotide substrate affected substrate binding affinity but left the ribozyme reaction rate invariant. Nucleotide titrations were performed and initial rates extracted for the c6.17 ribozyme. Fitting these rates to the Michaelis-Menten equation resulted in $k_{cat(app)}$ values of 0.079 ± 0.002 h^{-1}, 0.080 ± 0.002 h^{-1} and 0.078 ± 0.005 h^{-1} for the nucleotides ^{4S}UTP, ^{4S}UDP and ^{4S}UMP respectively. The K_m values for the same series were 24 ± 3 μM, 13 ± 2 μM and 535 ± 100 μM (Figure 2-6). Table 2-1 summarizes these parameters and the relative second order rate constants for 6.17 and Iso6. Both ribozymes exhibit large decreases in $k_{cat(app)}/K_m$ when reacted with a nucleotide monophosphate. Notably the $k_{cat(app)}$ for both ribozymes is nearly completely independent of substrate choice, averaging 0.079 h^{-1} for 6.17 and 0.083 min^{-1} for Iso6. This suggests that the chemical mechanism of both ribozymes is independent of phosphate chain length, and that NMPs in contrast to NDPs or NTPs are poorly recognized.

Table 2-1: Kinetic parameters of Iso6 and c6.17 and their relative efficiencies.

Iso6[a]				c6.17[b]			
Nucl.	$k_{cat(app)}$ (min^{-1})	K_M (μM)	Rel. $k_{cat(app)}/K_M$[c]	Nucl.	$k_{cat(app)}$ (h^{-1})	K_M (μM)	Rel. $k_{cat(app)}/K_M$[c]
GTP	0.084 ± 0.004	13 ± 2	1.0	^{4S}UTP	0.079 ± 0.002	24 ± 3	1.0
GDP	0.089 ± 0.005	11 ± 2	1.3	^{4S}UDP	0.080 ± 0.002	13 ± 2	1.9
GMP	0.077 ± 0.004	200 ± 40	0.060	^{4S}UMP	0.078 ± 0.005	535 ± 100	0.044

[a] Taken from (Huang & Yarus, 1998). Reaction conditions: 20 mM Mes (pH 5.5), 20 mM Ca^{2+}, 37°C.
[b] Reaction conditions: 50 mM Tris (pH 7.6), 20 mM Mg^{2+}, 150 mM K^+, 22°C.
[c] Rel. $k_{cat(app)}/K_M$, is given as the ratio of the $k_{cat(app)}/K_M$ of the appropriate nucleotide to that of the triphosphate one.

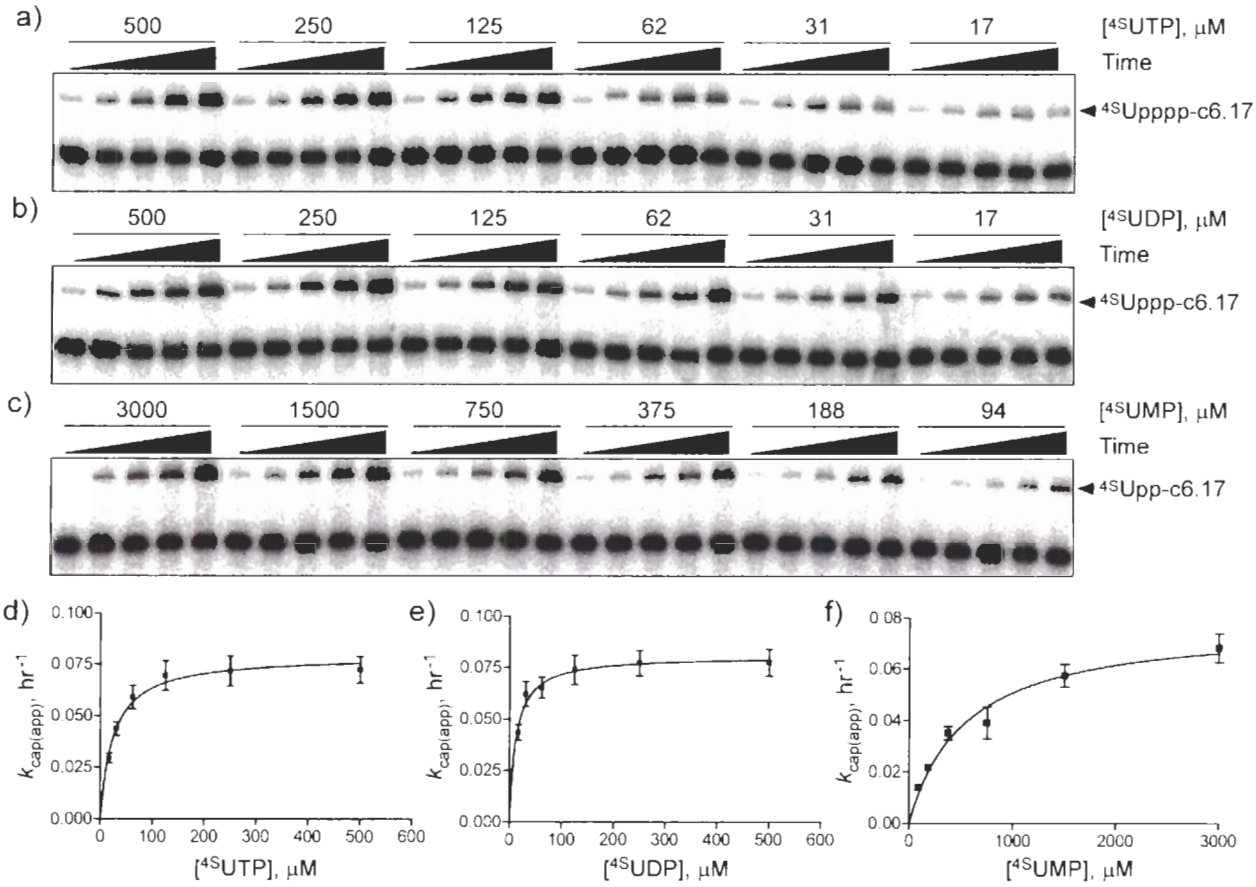


Figure 2-6: Determining the kinetic parameters of capping - dependence on 5′-nucleotide phosphate chain length. Internally labelled c6.17 was incubated with **a)** ^{4S}UTP, **b)** ^{4S}UDP and **c)** ^{4S}UMP at the indicated concentrations. Time points were taken at 1, 2, 4, 8 and 24 hours and resolved on an 8% APM gel. Initial rates were then extracted from each time-course and plotted against the nucleotide concentration. Michaelis-Menten kinetics fit well to the data with **d)** ^{4S}UTP: $k_{cat(app)} = 0.079 \pm 0.002$ h^{-1}, $K_m = 24 \pm 3$ μM, **e)** ^{4S}UDP: $k_{cat(app)} = 0.080 \pm 0.002$ h^{-1}, $K_m = 13 \pm 2$ μM and **f)** ^{4S}UMP: $k_{cat(app)} = 0.078 \pm 0.005$ h^{-1}, $K_m = 535 \pm 100$ μM

.

2.3.6 Decapping and exchange reactions.

In addition to being able to react when activated with a triphosphate, 6.17 was also able to perform cap exchange and decapping. t6.17 was first incubated in the presence of 1 mM UTP and the resulting uridine tetraphosphate cap product (Upppp-Mod B) was carefully gel purified away from unreacted or hydrolyzed module B RNA. The Upppp-Mod B construct was then re-annealed to module A and incubated with varying concentrations (31 to 500 μM) of ^{4S}UTP (Figure 2-7). Capping by exchange of UTP with

^{4S}UTP had a saturated rate of capping of 0.070 ± 0.006 h^{-1}– a value very similar to that observed when the ribozyme was initially activated with a triphosphate (Figure 2-6a). These observations indicate that capped RNA is a good substrate for cap exchange and implies that neither the pyrophosphate leaving group of the triphosphate activated ribozyme nor the nucleotide released during exchange strongly influence catalytic rate. Again this property is shared with Iso6; using a guanine triphosphate cap, Huang and Yarus demonstrated that Gppp-Iso6 when incubated with Gpppp, efficiently performed cap exchange to form Gppppp-Iso6 (Huang & Yarus, 1997b).

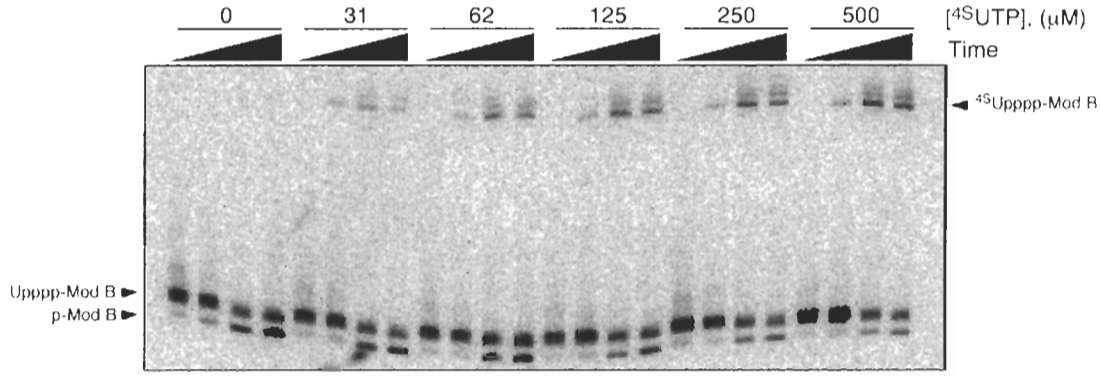


Figure 2-7: Decapping and exchange activity. A [^{32}P]-3′-end labelled module B was capped with UTP (Upppp-Mod B), carefully gel purified and annealed to module A, ^{4S}UTP was titrated from 0 to 500 μM. Time-points were taken at 0, 2, 24 and 48 hours and run on a 15% APM gel. In the absence of ^{4S}UTP, Upppp-Mod B is hydrolyzed to a monophosphate-module B (p-Mod. B). As the ^{4S}UTP concentration is elevated; a competing exchange activity, where ^{4S}UTP attacks the α-phosphate releasing UTP and forming a ^{4S}UTP cap (^{4S}Upppp-Mod B), is observed.

2.3.7 pH profile.

t6.17 showed an increase in reactivity with decreasing pH. The effect of pH on the capping rate was measured from 5.7 to 9.5 using three replicates (Figure 2-8). The pH-rate profile is consistent with two apparent pKa's, the first below six and the second at ~ 8 and will require a more detailed investigation to fully understand. Generally, the

pH dependence of 6.17 is similar to that observed for Iso6, where maximal activity is also observed at low pH (Huang & Yarus, 1997a). Both capping ribozymes are in striking contrast to ribozymes that perform phosphodiester chemistry where chemical rate increases with pH (Bergman et al., 2000).

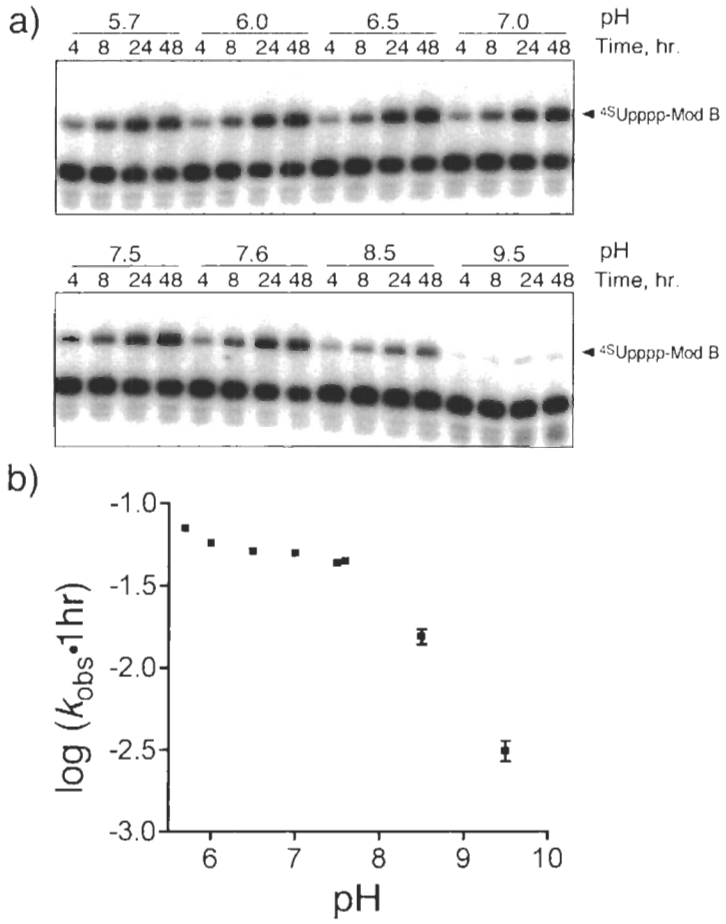


Figure 2-8: pH profile of the ^{4S}UTP capping reaction. . **a)** ^{4S}UTP capping reactions were carried out under varying pH conditions from 5.7 to 9.5, in the presence of 20 mM Mg^{2+}. Aliquots were taken at the indicated times and resolved on an APM gel. **b)** Log capping rates from three different experiments are plotted against pH. Average capping rates (S.D. of error shown) decreased by about 0.02 h^{-1} for every increase in pH unit.

2.4 Discussion:

Why do two capping ribozymes selected from two different laboratories both show a near complete lack of nucleotide substrate discrimination? The procedure used in

the early rounds of selection to isolate Iso6 involved the circularization of RNAs that were able to generate a 5′ terminal monophosphate by liberating a pyrophosphate from their 5′ termini. Only in later rounds of selection were ribozymes enriched based on their ability to produce cap structures when incubated with UTP-agarose (Huang & Yarus, 1997a). The selection resulting in the isolation of Iso6 therefore favoured ribozymes with the general ability to activate the 5′ terminal α-phosphate irrespective of substrate. A completely different approach was utilized to isolate 6.17. Ribozymes able to cap themselves with ^{4S}UTP were selected in early rounds (using thiol sensitive APM gels) after having been incubated with a solution containing significant levels of ATP and ADP in addition to ^{4S}UTP. As a consequence a mild selective pressure in favour of ^{4S}UTP dependent capping should have existed during our selection. Even without this pressure, ribozymes with good levels of substrate discrimination are typically selected by incubation of RNA pools with pure small molecule substrates (Unrau & Bartel, 1998; Seelig & Jaschke, 1999; Wilson & Szostak, 1999; Lau et al., 2004; Tsukiji et al., 2004; Fiammengo & Jaschke, 2005). It therefore appears that selection conditions were not directly responsible for the inability of 6.17 to differentiate between nucleotide substrates and that the chemistry of capping may have in some way prevented the isolation of ribozymes able to utilize a specific substrate.

The primary sequences of 6.17 and Iso6 bear no obvious sequence similarities when compared using a dot plot or manual alignment. Furthermore, the two ribozymes are likely to adopt globally different secondary structures (Figure 2-1) as indicated by structural probing and co-variation experiments (Figure 2-2). However, the possibility of the two ribozymes adopting similar tertiary structures cannot be completely excluded.

Intriguingly, 6.17 and Iso6 are both predicted to have helices that terminate at the site of capping. This helix in 6.17 (labelled helix V in Figure 2-1a) must be quite distorted as the capped product quantitatively cross-links to C_{A55} at the distal end of this helix (Figure 2-3). Consistent with this interpretation, stabilizing helix V by converting the wobble base-pair found at its terminus to a Watson-Crick pair inhibited activity (Figure 2-3b). A similar situation may exist for Iso6 as the nucleotide immediately proximal to the site of capping is predicted to lack a base-pairing partner (Huang & Yarus, 1997b). The helical element, which at least for 6.17 appears likely to be in a distorted geometry, may play an important role in stabilizing the phosphate chemistry that takes place at the reactive terminal α-phosphate for both ribozymes.

Iso6 and 6.17 display an unusual pH dependence, with capping rates increasing as the pH is decreased (Figure 2-8; (Huang & Yarus, 1997a)). Ribozymes that mediate phosphodiester cleavage and ligation reactions exhibit reaction rates with opposite behaviour (Dahm et al., 1993; Pyle & Green, 1994; Bergman et al., 2000). The pH profile of the class I ligase, for example, is explained by the fact that the attacking 3′ hydroxyl group must be deprotonated to facilitate its attack onto the α-phosphate of the triphosphate on the incoming RNA strand so as to form a 5′-3′ phosphodiester linkage (Bergman et al., 2000). The pH dependence of capping can be explained by assuming that protonation of the triphosphate chain is required to facilitate catalysis. In support of this argument, the hydrolysis of ATP, where water attacks the α- or β-phosphate to yield AMP or ADP respectively, is increased at lower pH (Ramirez et al., 1980). In this instance, the neutralization of the γ- or β-phosphate oxyanions increases the electrophilicity of the reaction centre, stabilizing a proposed oxyphosphorane

intermediate. This reaction is accelerated by a factor of 10 to 50 in the presence of divalent metal ions such as Mg^{2+} or Ca^{2+} at pH values > 3 (Ramirez et al., 1980) suggesting that both 6.17 and Iso6 utilize a similar divalent metal ion dependent process during capping. Consistent with this hypothesis, both ribozymes are unreactive in the presence of cobalt hexammine. Like other magnesium-dependent ribozymes, 6.17 tolerates Mn^{2+} and Ca^{2+} (Figure 2-4a; (Fedor, 2002)); however Iso6 prefers Ca^{2+} and only reacts slowly with Mn^{2+} (Huang & Yarus, 1997b). Ribozyme 6.17 has a relatively loose affinity for Mg^{2+} and may as a consequence tolerate ions with slightly larger diameters (the ionic radii of Mg^{2+}, Mn^{2+} and Ca^{2+} are 0.72 Å, 0.83 Å and 1.00 Å respectively). Iso6 in contrast has a fairly high affinity for Ca^{2+}, possibly explaining its ability to tolerate Mn^{2+} but not Mg^{2+} (Huang & Yarus, 1998).

Although 6.17 and Iso6 ribozymes fail to recognize either the sugar or base of their substrates, they both exhibit a high affinity for the phosphate moiety of the capping nucleotide. Both ribozymes have affinities ranging from 11 to 24 μM for nucleotides diphosphates and triphosphates. This affinity decreases by a factor of 15 to 22 when the ribozymes are given nucleotide monophosphates. Intriguingly, chemistry is performed in such a way that even when binding interactions are improved by providing a substrate with two or more phosphate groups, the maximum rate of capping remains invariant (Table 2-1). This data suggests that both capping ribozymes possess complicated active sites that can activate the α-phosphate towards chemistry independent of the attack of the ultimate substrate. Using thio-phosphate modifications to track the stereochemistry of the 6.17 capping reaction, we have recently shown that capping in fact proceeds via a two step mechanism that proceeds via a covalent intermediate (Zaher & Unrau, 2006), see

chapter 3 - closely paralleling the reaction mechanism of protein enzymes. This mechanism implies that the preferential recognition afforded substrates containing at least two phosphate groups is intertwined with the displacement of the pyrophosphate leaving group that initiated the process of cap formation.

By reasons of symmetry, the initial state of either ribozyme when activated with a triphosphate or a nucleotide cap looks chemically indistinguishable from the product of the ribozyme reaction – provided that the ribozymes do not interact with the sugar or base of the nucleotide at the distal end of the cap. Breaking this symmetry would allow the ribozyme to distinguish reactant from product, and would be expected to reduce the promiscuous nature of either ribozyme. The finding that neither Iso6 nor 6.17 break this symmetry has two implications: First as we have argued through the finding that 6.17 and Iso6 cannot distinguish different nucleobases and sugars (Figure 2-5), both ribozymes must focus their catalytic attention in the immediate vicinity of the phosphate groups involved in capping chemistry. Second, the amount of informational complexity required to specify the three dimensional structure of the active site is sufficiently high to preclude the easy isolation of capping ribozymes that can recognize particular nucleotides. This is unusual, as in principle the recognition of a nucleobase via Watson-Crick hydrogen bond interactions is easily achieved by RNA, as recently demonstrated by the solution structures of the adenine and guanine riboswitches (Noeske et al., 2005). We hope that the future evolution of these capping ribozymes using selection conditions where substrate discrimination is required will allow the isolation of RNA capping ribozymes that utilize a specific nucleotide substrate. Such a ribozyme would be predicted not to only utilize a substrate with a defined base, but (as a consequence of the

formation of a nucleotide binding pocket) would also be expected to react only with a nucleotide of defined phosphate chain length. The additional sequence information required to specify this substrate discrimination would provide important clues as to the evolutionary potential of RNA and extend the ideas of Szostak concerning the information content of functional RNA sequences (Szostak, 2003; Carothers et al., 2004).

2.5 Contributions:

R. Ammon Watkins and Peter Unrau carried out the selection, while I conducted all the characterization reported in this chapter.

CHAPTER 3:
A general RNA-capping ribozyme retains stereochemistry during cap exchange

This chapter is largely based on the manuscript; Zaher HS, Unrau PJ. 2006. "A general RNA capping ribozymes retains stereochemistry during cap exchange". *JACS* **128**: 13894-900.

3.1 Introduction:

Ever since the discovery that RNA is capable of catalysis, there has been a fascination with ribozyme mechanism and stereochemistry. Early in the history of RNA catalysis, the divalent metal dependent *Tetrahymena* self-splicing intron, the hammerhead ribozyme, and the group II intron were shown, via the use of phosphothioate modifications (section 1.6.1), to invert stereochemistry during phosphodiester bond formation and cleavage reactions (Figure 3-1a) (McSwiggen & Cech, 1989; Rajagopal et al., 1989; Slim & Gait, 1991; Padgett et al., 1994). The in-line attack model that was developed based on these results has since been validated and extended to ribozymes that mediate chemistry without the direct catalytic participation of magnesium ions – most notably the hairpin ribozyme which was recently crystallized using a vanadate pentacoordinate transition-state mimic (Hampel & Cowan, 1997); (Rupert et al., 2002).

The perception that all ribozyme reactions involve single-step inverting reactions is deeply entrenched within the field of RNA biochemistry. It is useful however, to consider other catalytic strategies that RNA can exploit in order to mediate reactions at phosphorous centres. Shortly after showing the catalytic ability of the *Tetrahymena* self-splicing intron, Zaug and Cech demonstrated that it could be engineered to mimic both an alkaline phosphatase and a phosphoryl transferase – enzymes that utilize covalent

intermediates (Zaug & Cech, 1986). However, questions concerning the stereochemistry and metal ion specificity of these constructs could not be addressed, because the reactive centre within these riboyzmes involves a phosphomonoester and not a phosphodiester. This made it practically impossible to track the reactions stereochemically by using atomic substitution of oxygen with sulphur. Thus, despite nearly two decades of research into RNA catalysis and the isolation of many exotic catalytic activities (Wilson & Szostak, 1999; Fiammengo & Jaschke, 2005), a ribozyme that utilizes a retaining mechanism has not been stereochemically characterized.

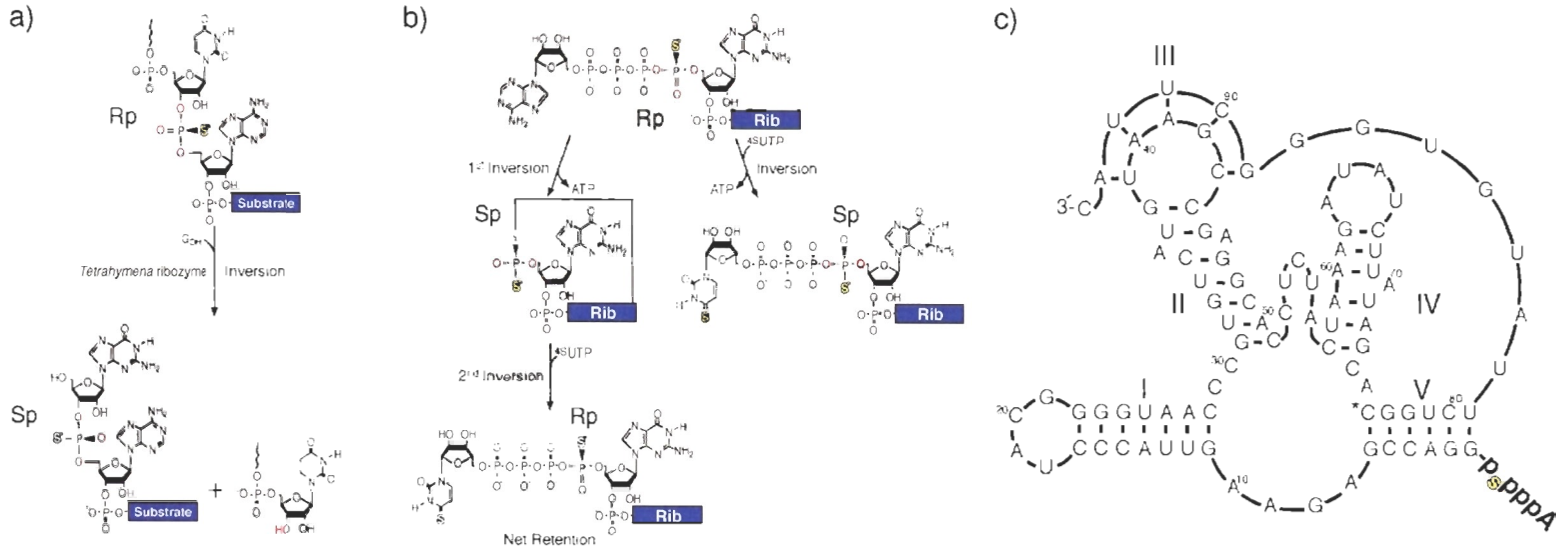
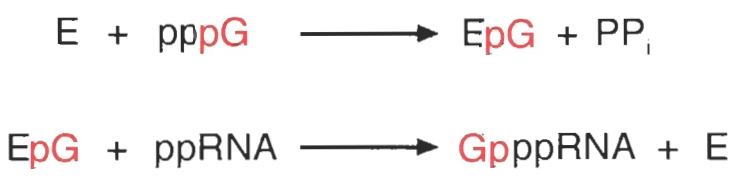


Figure 3-1: Tracking capping stereochemistry using R_p or S_p thiophosphate modified ribozymes. **a)** Classically the endonuclease activity of the *Tetrahymena* ribozyme was shown to invert stereochemistry when an R_p thiophosphodiester was converted to an S_p form after attack by a guanosine nucleoside. **b)** The possible stereochemical outcomes of a cap exchange reaction where an R_p capped ribozyme, (R)-Apppp$_s$-Rib, is exchanged for a 4-thiouridine cap. **c)** Proposed secondary structure of the c6.17 (shown activated with adenosine thiotetraphosphate cap) ribozyme (Zaher et al., 2006).

Capping a nucleic acid produces a 5′-5′ phosphate bridge between the nucleic acid and the capping nucleotide, and has several important biological functions. Capping increases the lifetime of mRNA in eukaryotes, and enzymes such as T4 DNA and RNA ligase activate monophosphates to adenylates so as to facilitate the formation of phosphodiester linkages (Fresco & Buratowski, 1994; Doherty & Suh, 2000; Ho et al., 2004). RNA capping uses GTP:RNA guanylyl transferase to transfer a GMP onto the 5′-diphosphate end of an mRNA in order to form a G(5′)ppp(5′)-RNA cap. This enzyme-catalyzed reaction proceeds via a covalent enzyme-(lysyl-N)-GMP phosphoramidate intermediate that is the result of a nucleophilic attack by a lysine residue onto the α-phosphate of GTP. The cap is formed when the β-phosphate of the mRNA 5′-diphosphate attacks the α-phosphate of the GMP-enzyme intermediate, releasing it from the enzyme by displacement of the lysine residue (Equation 3-1), with an overall retention of stereo-configuration (Fresco & Buratowski, 1994). A similar mechanism is utilized by the DNA and RNA ligases to form a pyrophosphate bridge.



Equation 3-1

Here we characterize the kinetics and reaction stereochemistry of the 6.17 capping ribozyme (chapter 2); (Zaher et al., 2006). The 272 nt long construct was truncated into a fully functional 94 nt long *cis* construct shown in Figure 3-1c (called c6.17). A *trans*-

acting construct (called t6.17, chapter 2), created by opening the loop of helix I, separates the ribozyme into two modules called A and B, with capping taking place at the 5′-end of module B. In this chapter, we show that 6.17 performs capping with unusual rate-limiting kinetics that retains stereo-configuration.

3.2 Materials and methods:

3.2.1 Ribozyme constructs.

RNA transcripts were made by *in vitro* T7 transcription [40 mM Tris-HCl pH 7.9, 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 4 mM ATP, 4 mM CTP and 2 mM UTP and 4 U/μl T7 RNA T7 RNAP incubated for 1 h at 37°C] (Zaher & Unrau, 2004). The resulting transcripts were then gel purified using PAGE. The truncated *cis*-acting ribozyme (c6.17) has the following sequence (5′-gga ccg aga agu uac ccu acg ggg uaa ccc gug uca ugu aag ccg agg cac cuc uac uaa aga uau cuu aua gca cgg ucu uau gug ggc uua c), and has the GenBank accession number DQ371299. Module A of the *trans* construct has the following sequence (5′-ggg uaa ccc gug uca ugu aag ccg agg cac cuc uac uaa aga uau cuu aua gca cgg ucu uau gug ggc uua c) and module B has the following sequence (5′-gga ccg aga agu uac ccg cag agg aua cug acc).

3.2.2 Pyrophosphate release kinetics.

Module B was transcribed in the presence of 2.5 μCi/μl [γ-^{32}P]-GTP (3000 Ci/mmole, NEN) and gel purified. This module was then hybridized to Module A as previously described (Zaher et al., 2006) in the standard incubation buffer (50 mM Tris-

HCl pH 7.6, 150 mM KCl) supplemented with 20 mM $MgCl_2$, and varying concentrations of ^{4S}UTP at 22°C. Control reactions were performed in the absence of module A or ^{4S}UTP. The products were resolved on a 23% sequencing gel together with ^{32}P-inorganic phosphate and pyrophosphate markers. Inorganic phosphate was obtained by dephosphorylating [γ-^{32}P]-ATP (NEN) using calf intestinal phosphatase (CIP, New England Biolabs); 50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT supplemented with 1 U/μl of enzyme and incubating for 1 h at 37°C. Pyrophosphate was made by saving an aliquot of a transcription reaction performed in the presence of [γ-^{32}P]-GTP.

To correlate the capping rate with the pyrophosphate release rate, an [α-^{32}P]-UTP (NEN) body-labelled module B was mixed with [γ-^{32}P]-GTP labelled module B with a specific activity ratio of 1:1 (by scintillation counting). This ratio was confirmed by phosphorimager quantification of a CIP treated RNA sample of the mixture that was run on PAGE (found 51:49; Figure 3-2, CIP lane). The dual-labelled module B was then incubated with module A under varying concentrations of ^{4S}UTP. Time-points were taken at 0, 2, 24 and 48 hr, and resolved using 23% APM gels (Igloi, 1988).

3.2.3 Capping ribozyme kinetics.

Module B was labelled on its 3′ terminus with [^{32}P]-pCp. This construct was then annealed to module A and incubated with varying concentrations of ^{4S}UTP in the standard incubation buffer supplemented with 20 mM $MgCl_2$. Aliquots were taken at 8, 16, 24, 48, 96 and 192 h and were separated on 10% APM gels. Fits were carried out as described in the text using the GraphPad Prism version 4.00 analysis package.

3.2.4 Synthesis of caps.

Guanosine α-thiotriphosphate (GTPαS) was purchased from TriLink Biotechnologies and separated into its S_p and R_p stereoisomers using reverse phase HPLC (ZORBAX XDB-C18 column on an Agilent 1100 series HPLC). After equilibrating the column with 50 mM triethylammonium acetate pH 7.0 for 5 minutes at a flow rate of 1 ml/minute, GTPαS was eluted using a linear gradient increasing by 1% Acetonitrile per minute for 15 minutes. The Sp form eluted at ~7.3% acetonitrile, while the R_p form eluted at ~8.1%. The stereochemistry was assigned based on two evidences, the S_p isomer elutes before the R_p and that T7 RNAP is only able to utilize the S_p isomer during transcription.(Griffiths et al., 1987) Adenosine 5′-phosphorimidazolide (ImpA) and 4-thiouridine 5′-phosphorimidazolide ($Imp^{4S}U$) were synthesized as previously described (Lohrmann & Orgel, 1978). The tetra-phosphates caps [AppppG, (S)-$App_{ppp_s}G$, (R)-$App_{ppp_s}G$, (S)-$^{4S}U_{ppp_s}G$ and (R)-$^{4S}U_{ppp_s}G$] were synthesized by incubating the appropriate NTP at 50 mM with 200 mM ImpA or $Imp^{4S}U$ in 50 mM HEPES pH 7.4 in the presence of 50 mM $MgCl_2$ at 55°C for 4 hr. The reaction was stopped by quenching the magnesium with excess EDTA, caps were purified using the HPLC conditions described earlier. Peaks that occurred as a result of the reaction were collected. The identity of the caps was confirmed by its UV-absorbance and mass spectrometry. The caps were dried using a speed-vac and resuspended in water.

3.2.5 Synthesis of modified ribozymes and capping/exchange reactions - APM analysis.

Capped ribozymes were made by *in vitro* T7 RNAP transcription. GTP concentration was reduced to 1 mM to encourage initiation with the dinucleotide caps,

which were included at 1 mM in the reaction. The resulting RNAs were then gel purified using PAGE. Ribozymes capped with ^{4S}U [(S)-^{4S}Upppp$_s$-Rib and (R)-^{4S}Upppp$_s$-Rib] were further purified using an APM gel and used as stereochemical markers. Capping or exchange reactions were carried out by incubating [α-^{32}P]-UTP internally-labelled ribozymes in the standard incubation buffer in the presence of 10 mM $MgCl_2$, 5 mM $MnCl_2$ and 1 mM ^{4S}UTP. Time-points were taken at 0, 3, 24 and 48 h and resolved using APM gels. RNAs were treated with CIP to remove the triphosphate from ribozymes that initiated with GTP rather than dinucleotide caps during the transcription reactions. Reactions with these CIP-treated ribozymes were conducted as before.

3.2.6 HPLC confirmation of the capping products stereochemistry.

Unlabelled ppp-Rib and (R)-Apppp$_s$-Rib that had been CIP-treated after transcription were reacted at 2 μM concentration with 5 μCi/μl of [γ-^{32}P]-ATP in the same buffer used for the APM analysis. Aliquots were taken at 0, 0.5, 3 and 20 h and gel purified. These samples were then digested with 100 U/μl T1 (Fermentas) and 1 U/μl T2 (Fermentas) at 37°C for 15 minutes, followed by treatment with 1 U/μl CIP for 1 h to remove the 3′-monophosphates resulting from the digestion. Unlabelled AppppG, (S)-Apppp$_s$G and (R)-Apppp$_s$G standards (see cap synthesis section) were added to these tubes and fractionated by HPLC, using the same conditions mentioned earlier. Fractions were collected every 2 minutes- between 0 and 12.5 minutes- every 6 seconds- between 12.5 and 15.1 minutes- and every 2 minutes- between 15.1 and 20 minutes. The fractions were then counted on Beckman LS600SC scintillation counter for 9 minutes in a total

fluid volume of 7 ml (2 ml aqueous sample mixed with 5 ml Formula-989 scintillation fluid).

3.3 Results:

3.3.1 A rate-limiting step precedes capping.

A simple kinetic model requiring an early rate-limiting kinetic intermediate fit a broad range of data. This model was initially suggested by the finding that the rate of pyrophosphate release was fixed and independent of the overall rate of capping (Figure 3-2). The model assumes that both the triphosphate and the capped forms of module B substrates have equivalent kinetics and react to form an early rate-limiting kinetic intermediate (Figure 3-3a). This intermediate was found to partition into either a capped (f_{cap}) or an inert hydrolyzed product ($f_{hyd} = 1 - f_{cap}$). Consistent with this model, a UTP capped module B was able to be recapped with ^{4S}UTP with the rate of this exchange increasing with ^{4S}UTP concentration. At the same time the rate of hydrolysis was observed to fall in direct correspondence (Zaher et al., 2006), (section 2.3.6). Together the two reaction schemes shown in Figure 3-3a define a differential equation that predicts the amount of ^{4S}UTP capped module B as a function of time:

$$F = f\,(e^{-k_{hyd(effective)}t} - e^{-(k_{cap(effective)} + k_{hyd(effective)})t})$$

f being the maximum fraction able to react (and was set to 0.5 based on an analysis of pyrophosphate release shown in Figure 3-2), $k_{cap(effective)} = f_{cap} \cdot k_{lim}$ and $k_{hyd(effective)} = f_{hyd} \cdot k_{lim}$. A representative set of time dependent data and the resulting fits are shown in Figure 3-3b and 3c. The effective capping and hydrolysis rates were then averaged from five independent data sets and plotted as a function of ^{4S}UTP concentration (Figure 3-3d).

The effective rate of capping was found to fit well to a Michaelis-Menten type equation where $k_{cap(effective)} = k_{lim} \cdot [^{4S}UTP]/(K_m + [^{4S}UTP])$. The effective hydrolysis rate was allowed one additional degree of freedom and fit to $k_{hyd(effective)} = k_{lim} \cdot [K_m]/(K_m + [^{4S}UTP]) + k_{hyd(min)}$ (Figure 3-3d). A non-zero value for $k_{hyd(min)}$ would imply that f_{hyd} did not equal $1 - f_{cap}$ and contradict our model. The simultaneous fits agreed very well with the data with a k_{lim} of 0.052 ± 0.002 h^{-1}, a K_m of 62 ± 16 μM and a $k_{hyd(min)}$ of 0.001 ± 0.002 h^{-1}. As expected from this model the value of $k_{cap(effective)} + k_{hyd(effective)}$ nearly equalled the rate of pyrophosphate release determined independently at the same ^{4S}UTP concentration (Figure 3-3d, triangles and diamonds respectively). Indeed, the average rate of pyrophosphate release was 0.050 ± 0.003 h^{-1}, equalling k_{lim} to within error. This simple kinetic model explains a broad range of capping data and has the unusual feature that the rate of pyrophosphate release is invariant and not a function of substrate concentration as is typical for most ribozymes that liberate pyrophosphate during catalysis.

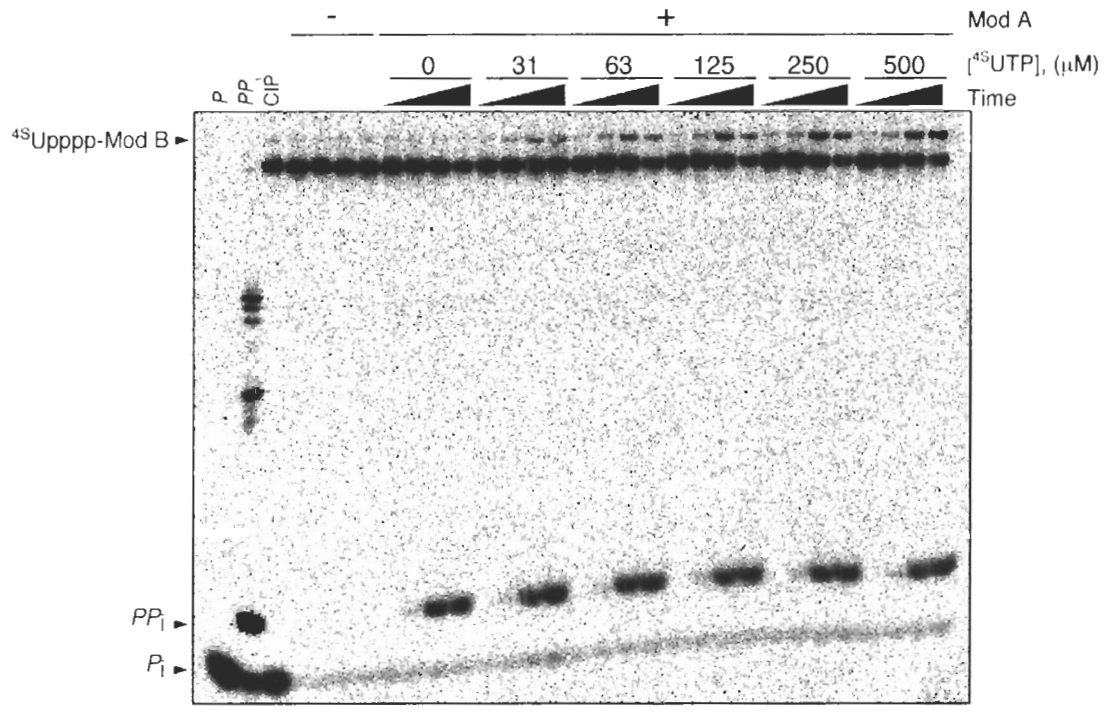


Figure 3-2: Rate of pyrophosphate release is independent of nucleotide substrate concentration. No pyrophosphate was released in the absence of module A, but when module A was present, pyrophosphate was released from module B with a rate that was independent of ^{4S}UTP concentration. During the ^{4S}UTP titration, the rate of capping increased and approached the rate of pyrophosphate release.

The kinetics of t6.17 bear an uncanny resemblance to those of a calcium-dependent capping ribozyme called Iso6 (Huang & Yarus, 1997a). Both ribozymes share the property that the rate of pyrophosphate release is independent of substrate concentration (Figure 3-2 and 3 for t6.17), (Huang & Yarus, 1998). Further, decreasing the phosphate chain length of the nucleotide substrate to a monophosphate has a marked effect on binding affinity for both ribozymes, but leaves the saturated rate of capping unchanged for both ribozymes (Huang & Yarus, 1998; Zaher et al., 2006). Additionally, both tolerate a broad range of nucleotide substrates and have similar pH dependence (Huang & Yarus, 1997b, c, a, 1998; Zaher et al., 2006) even though the ribozymes have different metal ion requirements and secondary structures (Huang & Yarus, 1997c; Zaher et al., 2006). To our knowledge, Iso6 and 6.17 are the only ribozymes that share such a

detailed range of properties. The simplest interpretation of these similarities is that a covalent intermediate is formed by both ribozymes during capping. Although a lariat-type structure of the form suggested in Figure 3-1b was not detected for either ribozyme despite careful gel based analysis, its presence can be inferred by probing the stereochemistry of the reaction. A covalent intermediate would require two sequential nucleophilic attacks, with each inverting the stereochemistry of the α-phosphate found on the 5′ terminus of the ribozyme, resulting in a net retention of stereo-configuration, while a single-step reaction would be expected to invert stereochemistry.

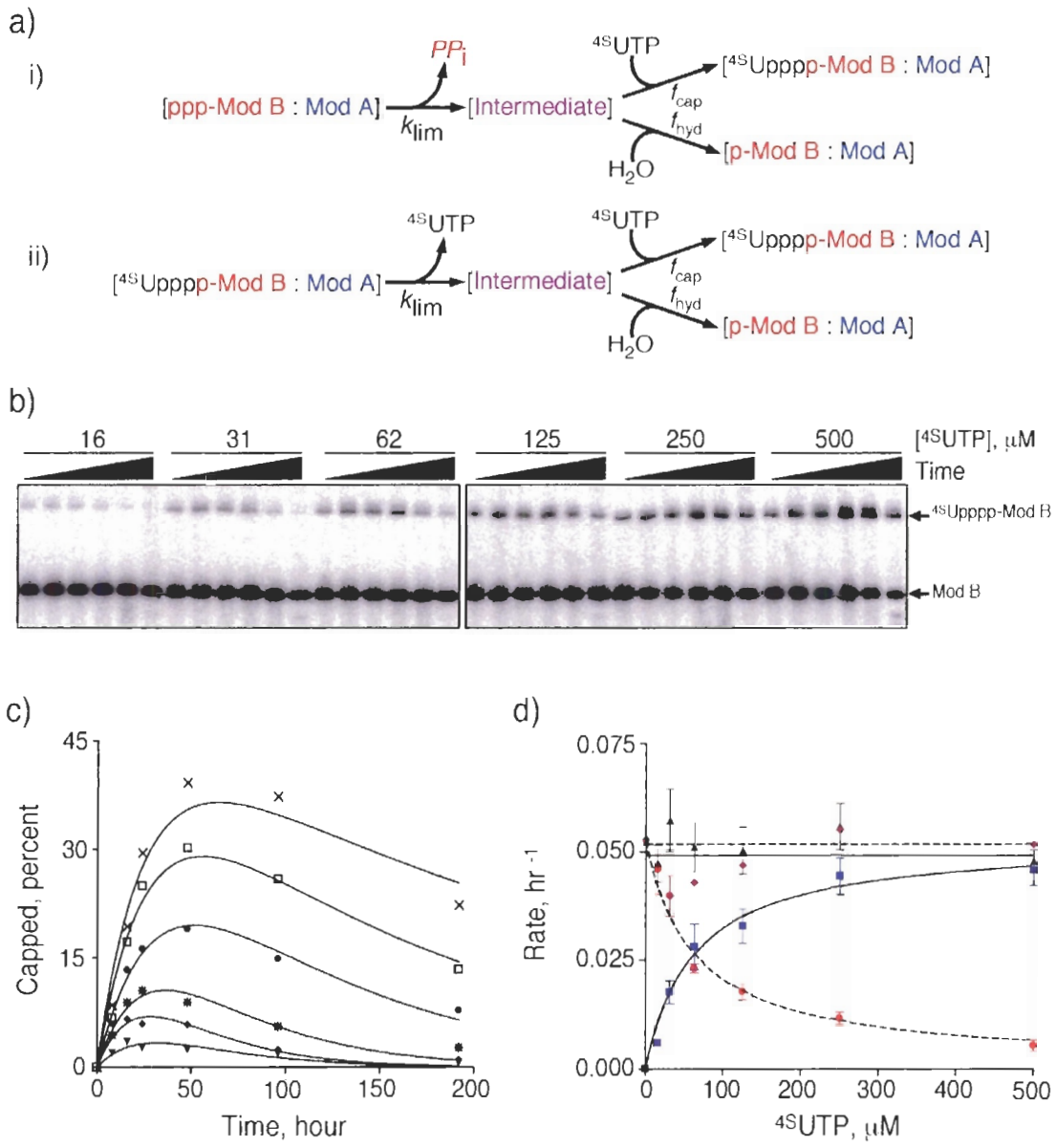


Figure 3-3. The kinetics of capping using ^{4S}UTP as a substrate. **a)** A kinetic model for the trans form of the ribozyme. **b)** APM-polyacrylamide gel showing one of five replicates used to determine capping kinetics, the lower band on the gel being unreacted or hydrolyzed module B. **c)** The ^{4S}UTP dependent temporal fit to the data in b). **d)** The capping rates fit well to a simple Michaelis-Menten type curve while hydrolysis rates agreed well with the model in a).

3.3.2 Ribozyme c6.17 performs capping with net retention of stereo-configuration around the α-phosphate.

Stereochemistry was monitored by rendering the reactive 5′ α-phosphorous of each ribozyme chiral by substituting either of its non-bridging oxygens with a sulphur. Two sets of dinucleotide cap stereoisomers were synthesized (Figure 3-4a). The first, S_p and R_p adenosine(5′)α-thiotetraphospho-(5′)guanosine [(S)- and (R)-Appppp$_s$G, respectively] were used to prepare ribozymes with a stereochemically modified cap and were also used as absolute references. The second set, S_p and R_p 4-thiouridine(5′)α-thiotetraphospho-(5′)guanosine [(S)- and (R)-^{4S}Upppp$_s$G, respectively] were used to generate stereochemical gel mobility markers. Caps are typically incorporated during the initiation of transcription by T7 RNAP (Grudzien et al., 2004), and as we show, allow the synthesis of ribozyme transcripts containing either S_p or R_p (5′) α–thiotetraphosphate caps. Initiation with these caps rather than GαSTP prevented the otherwise unavoidable incorporation of S_p GαSMP into the backbone of the RNA transcripts. The caps were synthesized by first purifying S_p and R_p GαSTP [(S)-ppp$_s$G and (R)-ppp$_s$G respectively], and their absolute stereochemistry was confirmed by their enzymatic utilization and relative HPLC retention times (Griffiths et al., 1987). Each stereoisomer was then reacted with adenosine 5′-phosphorimidazolide (ImpA) or 4-thiouridine 5′-phosphorimidazolide (Imp^{4S}U) (Lohrmann & Orgel, 1978) so as to produce S_p or R_p dinucleotide α-thiotetraphosphate caps that were subsequently HPLC purified (Figure 3-4a). Mass spectrometry and UV spectra together with the relative HPLC retention times of each pair confirmed the expected products. Transcription in the presence of these caps resulted in their incorporation on the 5′ termini of RNA transcripts. (S)- and (R)-

^{4S}Upppp$_s$-RNA shifted significantly higher on an APM gel (as a result of the gels thiol-mercury dependent interactions) than ^{4S}Upppp-RNA, consistent with the two sulphur groups present in the (S)- and (R)-^{4S}Upppp$_s$-RNA constructs. Conveniently, the (R)-^{4S}Upppp$_s$-RNA shifted higher than (S)-^{4S}Upppp$_s$-RNA making it possible to monitor capping ribozyme stereochemistry by gel mobility (Figure 3-4b left two lanes, indicated by M / S and R).

Only the (R)-Apppp$_s$-Rib(c6.17) ribozyme construct was found to perform cap exchange, even though the (S)- and (R)-Apppp$_s$-Rib constructs for Iso6 and c6.17 were tested with a broad range of thiophilic metal ions (Mn, Cd, Pb, Zn, Ni, Fe and Cu). (S)- and (R)-Apppp$_s$-Rib and an Apppp-Rib control were treated with calf intestinal phosphatase (CIP) after transcription and gel purification so as to remove the 5′ triphosphates that inevitably result from initiation with GTP rather than the desired dinucleotide cap. CIP was highly effective in removing this undesired 5′ triphosphate, resulting in the nearly complete suppression of ^{4S}UTP capping in a control reaction (Compare Figure 3-4b CIP Rib lanes to Figure 3-5 ppp-Rib lanes). Apppp-Rib activity was preserved as expected, given the inability of CIP to dephosphorylate a cap structure (Figure 3-4b, Apppp-Rib lanes; Figure 3-5). When (S)-Apppp$_s$-Rib was incubated with ^{4S}UTP, no capping was observed under all tested conditions (Figure 3-4b, (S)-Apppp$_s$-Rib). However, the (R)-Apppp$_s$-Rib(c6.17) construct [Figure 3-4b, (R)-Apppp$_s$-Rib] was active at approximately one tenth the rate of the unmodified construct in optimal conditions using 10 mM magnesium and 5 mM manganese. Removing manganese from the reaction resulted in a modest two-fold decrease in overall rate (data not shown). The reaction product, when mixed with a control (R)-^{4S}Upppp$_s$-Rib transcript, resulted in a

single-band; but two distinct bands were observed when it was mixed with the (S)-^{4S}Upppp$_s$-Rib marker (Figure 3-4b – Co-mig. lanes R and S, respectively). This strongly suggests that the (R)-Apppp$_s$-Rib(c6.17) construct retained stereochemistry during the cap exchange reaction.

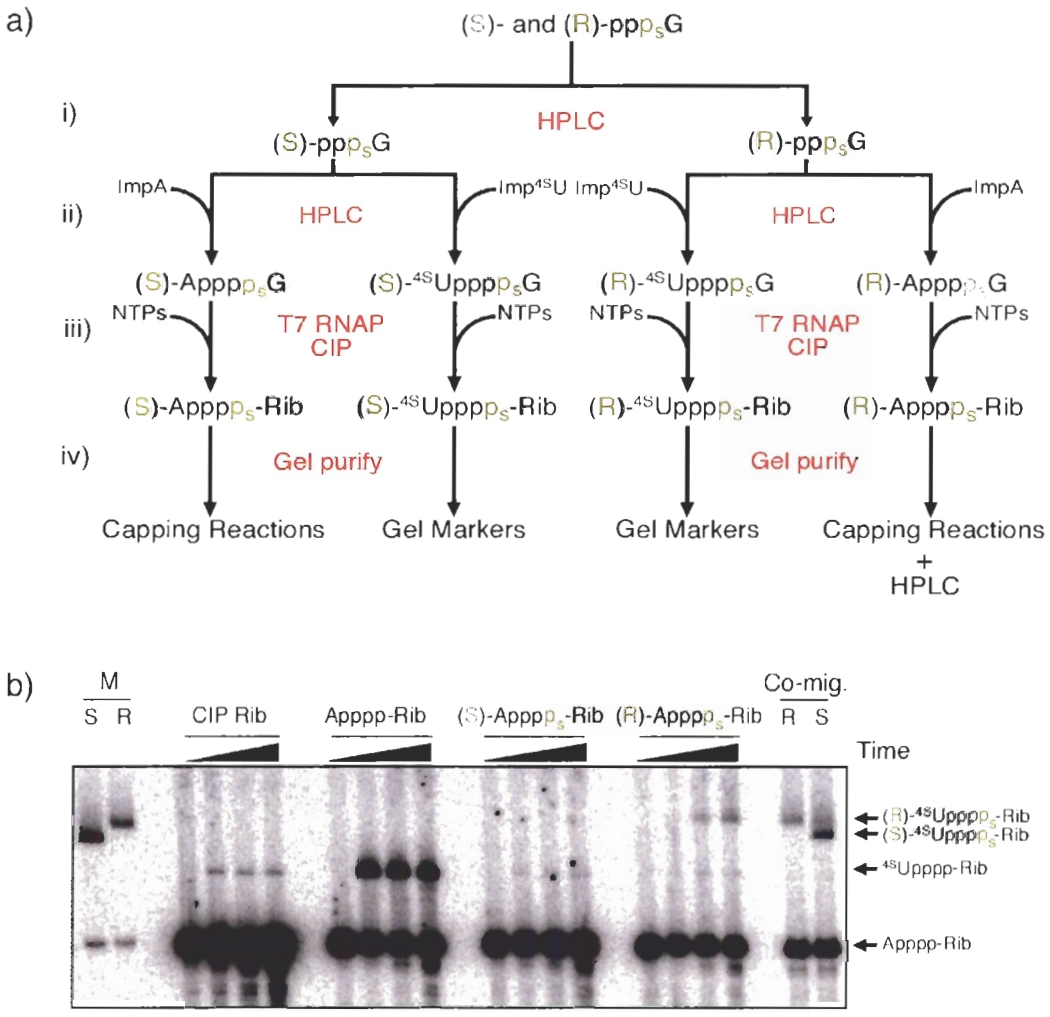


Figure 3-4: Capping proceeds with retention at the reactive (5′) α-phosphate centre - APM analysis. **a)** Synthesis flowchart leading to gel purified R_p or S_p capped ribozymes possessing either adenosine or 4-thiouridine caps. **b)** Demonstration that ribozyme c6.17 forms caps that are retained relative to the initial substrate stereochemistry.

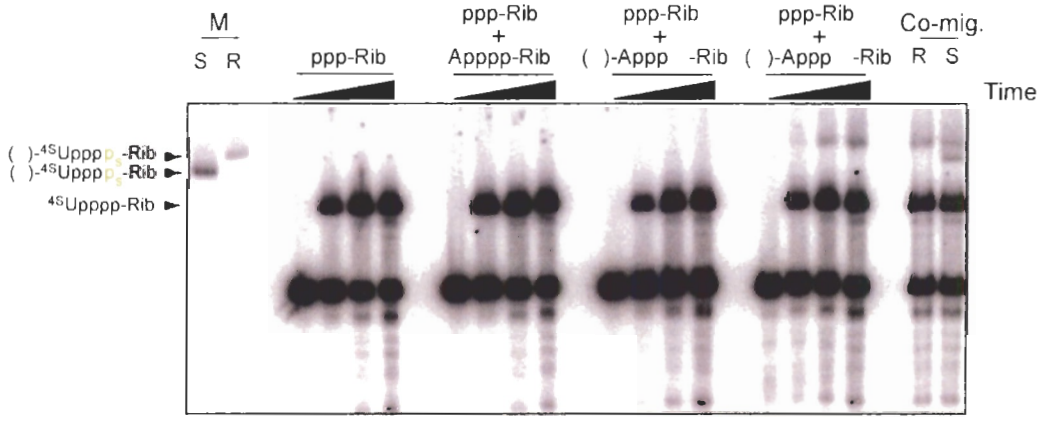


Figure 3-5: Exchange activity of capped c6.17 ribozymes that were not dephosphorylated with CIP after transcription.

A second independent experiment was performed to confirm this finding. (S)- or (R)-Apppp$_s$-Rib(c6.17) together with the equivalent Iso6 constructs were transcribed in the absence of a radiolabel, gel purified and CIP treated to remove triphosphate as described (i.e. Figure 3-4a). As expected, when incubated with [γ-^{32}P]-ATP, only the (R)-Apppp$_s$-Rib(c6.17) stereoisomer was able to exchange its cap for a radio-labelled nucleotide. As a control, ppp-Rib(c6.17) (transcribed in the absence of an initial cap) became capped with [γ-^{32}P]-ATP (Figure 3-6a, Figure 3-7a.ii and b.ii gel inserts,). The resulting radio-labelled materials from the control and the (R)-Apppp$_s$-Rib(c6.17) incubations were carefully gel purified and digested to completion with a mixture of T1 and T2. CIP was then added to remove any external phosphates (Figure 3-6b). Again as expected, digestion of the control reaction (Figure 3-7a.i) left a radio-labelled cap with the same HPLC retention time as the AppppG standard (Figure 3-7a.ii, the solid line shows the UV trace of the dinucleotide cap standards while the dotted lines indicate the eluted radioactivity in cpm). Digestion and CIP treatment of the (R)-Apppp$_s$-Rib(c6.17) product yielded a radio-labelled cap that co-cluted with the (R)-Apppp$_s$G dinucleotide standard - that was used initially to make the construct (Figure 3-7b.ii). Thus in two separate experiments; one involving exchange with ^{4S}UTP, and the second with [γ-^{32}P]-ATP, the (R)-Apppp$_s$-Rib(c6.17) construct was shown to retain stereochemistry.

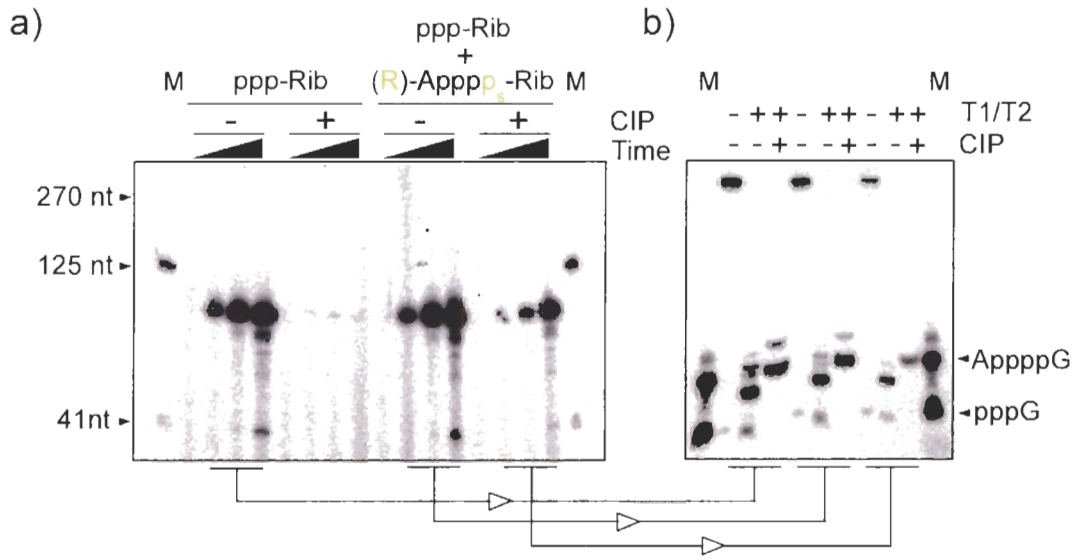


Figure 3-6: Reactions of ribozymes with [γ-^{32}P]-ATP and preparation of samples for HPLC injections. **a)** Unlabelled c6.17 ribozymes, ppp-Rib and a mixture of ppp-Rib and (R)-Apppp$_s$-Rib, were either treated with CIP or kept as is after transcription. These samples were reacted with [γ-^{32}P]-ATP, time-points were taken at 0, 0.5, 3 and 20 h and separated on PAGE. **b)** Radio-labelled ribozymes from (a) were gel purified, digested with T1 and T2 and then treated with CIP.

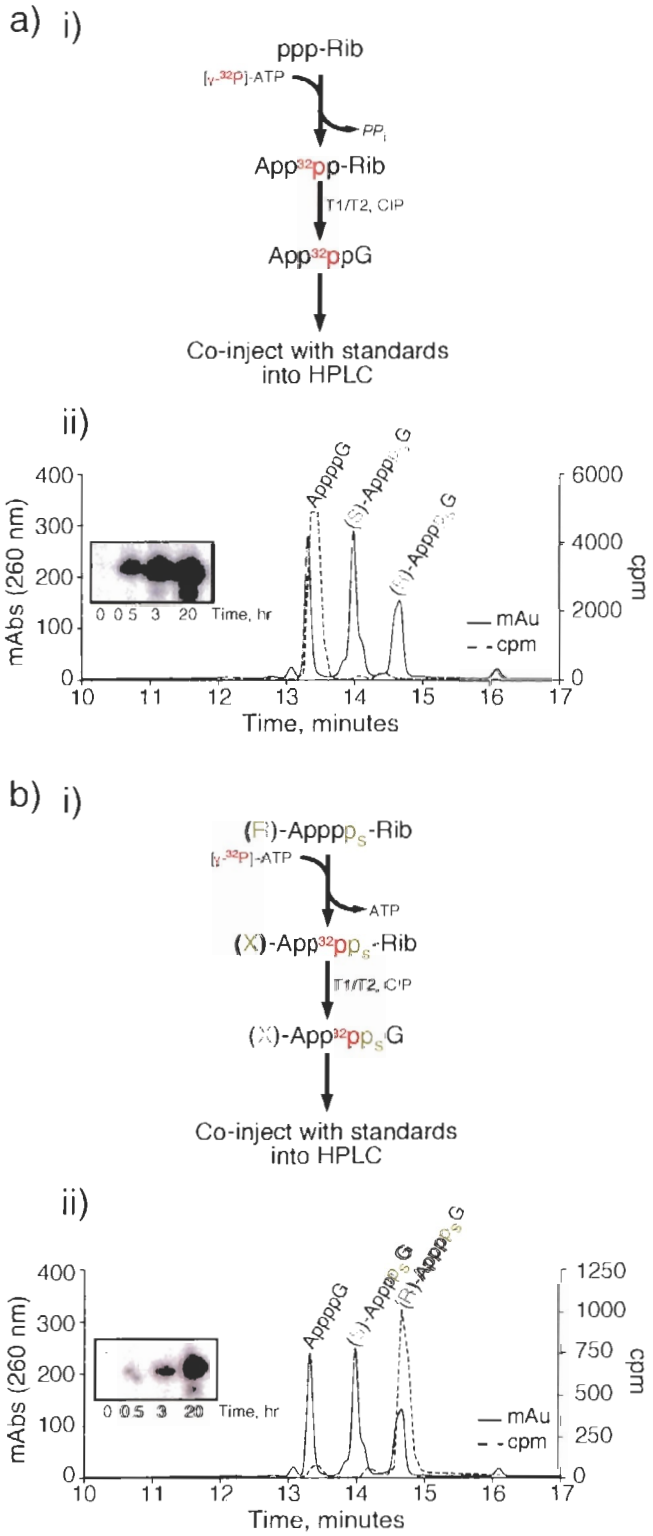


Figure 3-7: (R)-Apppp$_s$-Rib(c6.17) produces a radio-labelled dinucleotide cap when reacted with [γ-^{32}P]-ATP that has the same HPLC retention time as an (R)-Apppp$_s$G cap standard. a.i) Protocol used to manufacture a control radio-labelled dinucleotide cap synthesized by a triphosphate-containing ribozyme. **a.ii)** HPLC chromatogram showing that the radio-labelled cap elutes with the AppppG standard as expected. **b.i)** The same protocol was used for (R)-Apppp$_s$-Rib(c6.17). **b.ii)** This time the radio-labelled cap was found to elute at the same time as the (R)-Apppp$_s$-G standard.

3.4 Discussion:

The retaining mechanism of capping has unusual metal ion dependence. Ribozymes that perform inverting phosphodiester chemistry fall into two general categories, depending on whether or not divalent metal ions directly participate in catalysis. The metal-dependent *Tetrahymena*, hammerhead and group II intron ribozymes tolerate substitution by only one thiophosphate stereoisomer and not the other (McSwiggen & Cech, 1989; Slim & Gait, 1991; Padgett et al., 1994), and are inactive in the presence of cobalt hexammine (Horton & DeRose, 2000; Zaher et al., 2006). The inactive stereoisomer, in the case of the hammerhead, can be rescued by the addition of soft metals such as manganese and cadmium (Scott & Uhlenbeck, 1999). The hairpin ribozyme in contrast is active in the presence of cobalt hexammine (Hampel & Cowan, 1997), and mediates reactions with both R_p and S_p thiophosphates. The finding that our retaining capping ribozyme reacts with only an R_p α-thiophosphate substitution and not an S_p substitution is similar to the stereochemical signatures of the metal dependent *Tetrahymena*, hammerhead and group II intron ribozymes which invert stereochemistry. However, while our retaining ribozyme fails to react in the presence of cobalt hexammine, we find no evidence that soft metals rescue the (S)-Apppp$_s$-Rib(c6.17) construct. These observations indicate that the ribozyme makes critical contacts with the *pro*-S_p oxygen of the α-phosphate during the course of chemistry and hint that a divalent metal ion may be directly involved in capping chemistry by coordinating the *pro*-S oxygen on the α-phosphate of the ribozyme during catalysis.

The retention of stereochemistry strongly suggests that the c6.17 ribozyme utilizes a two-step mechanism. Each step would invert stereochemistry and be separated

by a transient covalent intermediate. Entirely consistent with this mechanism, is the unusual finding that pyrophosphate release is independent of nucleotide substrate concentration during capping. A rate-limiting chemical step can easily be interpreted by a two-step model if the first step is slow and the second step is rapid and partitions the reactive covalent intermediate into either a cap or a hydrolysis product (Figure 3-1b and 3-2). While we could not isolate this covalent intermediate, its steady state concentration would be expected to be low given the available kinetics. We are therefore confident, given the two independent data sets that the retention of stereochemistry observed for c6.17 results from the formation of a transient covalent intermediate during capping.

Formally, this covalent intermediate could involve a ribose 2′-hydroxyl forming a phosphodiester intermediate, or a nitrogen from a base generating a phosphoramidate intermediate. Resolving these two possibilities is a difficult task, which our lab is currently exploring. A candidate nucleotide responsible for forming the intermediate is A_{75} in the *cis*-acting construct (Figure 3-1c). Mutating this residue to a G completely inactivates the ribozyme while substituting it for a thiophosphate modified deoxyadenosine does not significantly affect catalytic rate (data not shown). Moreover, this residue is adjacent to the location of a quantitative crosslink (indicated by an asterisk in Figure 3-1c) between the 4-thiouridine cap product and the equivalent residue in the *trans* form of the ribozyme (Zaher et al., 2006), (section 2.3.3). The inhibition of capping observed when this residue is mutated suggests that a nitrogen in the base is involved in forming the capping intermediate and not a hydroxyl from the backbone sugar. If true, exciting mechanistic parallels between 6.17 and a capping DNAzyme selected by Yi and Breaker (Li et al., 2000) should exist. This DNAzyme is likely to utilize a

phosphoramidate type intermediate - although the detailed mechanism for this DNAzyme is currently unknown. The formation of a reactive phosphoramidate intermediate during capping is consistent with a short lived kinetic intermediate and would be analogous to the utilization of the ε-N of lysine residues seen in protein-catalyzed capping reactions (Shuman & Lima, 2004).

Why might capping utilize a retaining mechanism? Capping involves two charged phosphate groups that must be exchanged in a highly symmetrical reaction. This differs fundamentally from previously stereochemically characterized ribozymes that promote reactions at a phosphorus centre and that use hydroxyl groups as nucleophiles or nucleofuges. Performing capping in a single-step would require the simultaneous stabilization of two highly negatively charged phosphate groups that are in turn proximal to a negatively charged reactive centre – a strategy unlikely to be optimal for either a protein or an RNA-based catalyst. A two-step reaction involving the sequential displacement of a charged nucleofuge followed by the attack of a nearly identical nucleophile requires a more complicated reaction landscape, but at the same time the two transition-states encountered during the reaction would be expected to have dramatically lower activation energies than a single-step reaction. Equally important, by exploiting a two-step mechanism, where the reaction is symmetrical with respect to the reaction coordinate, only one substrate-binding site, that recognizes both the initial nucleofuge and the incoming nucleophile, would be required for catalysis. Thus in contrast to the vast majority of biological reactions that invert stereochemistry, the evolution of capping biochemistry may have been driven inexorably to select a retaining mechanism irrespective of whether nucleic acid or protein mediates the reaction.

CHAPTER 4:
Isolation of an improved RNA polymerase ribozyme by compartmentalization

4.1 Introduction:

The 'RNA world' hypothesis suggests that modern metabolism evolved from a system where RNA served to store genomic information and provide catalysis (Gilbert, 1986). The plausibility of this model has been strengthened by the finding that RNA can catalyze a diverse range of chemical reactions (Wilson & Szostak, 1999; Joyce, 2002; Fiammengo & Jaschke, 2005) and that functional RNAs play critical roles in modern metabolism, most notably the ribosome, RNaseP and the riboswitches (Guerrier-Takada et al., 1983; Ban et al., 2000; Winkler & Breaker, 2005). Although RNA machines like the ribosome are believed to be the relics of an RNA-based past, the current dominance of protein catalysts prevents us from learning how such an RNA system would have replicated. Nevertheless, one can imagine reconstructing a minimal RNA-based system in a test-tube where a small set of ribozymes is capable of auto-replication. The activities of these trans-acting ribozymes would be correlated by their confinement within a common compartment, in analogy to a modern cell (Bartel & Unrau, 1999).

Arguably the greatest limitation of the 'RNA world' hypothesis is the lack of an RNA replicase ribozyme, which has been coined 'the holy grail' in the ribozyme community (Szostak et al., 2001; McGinness & Joyce, 2003). Modern protein polymerases, like E. coli DNA polymerase III, facilitate genomic replication by performing highly processive template-based polymerization of dNTPs initiating from a short RNA primer hybridized to a DNA template (Benkovic et al., 2001). Similarly, the

selection of a polymerase ribozyme able to extend an all RNA primer-template (PT) complex with NTPs would be a key first step in the creation of a true replicase ribozyme. Initial attempts to find such a polymerase focused on expanding the functional capabilities of the naturally-occurring group I self-splicing ribozymes (Been & Cech, 1988; Doudna & Szostak, 1989; Bartel et al., 1991; Chowrira et al., 1993; Doudna et al., 1993; McGinness & Joyce, 2002). Later attention turned towards derivatives of the artificially selected class I ligase (Bartel & Szostak, 1993; Ekland & Bartel, 1995; Ekland et al., 1995). Ekland and Bartel found that some derivatives of this ligase could extend a primer by 6 nucleotides in the presence of nucleotide triphosphates, but required the 5'-end of the template to be annealed to the ribozyme preventing general and processive polymerization (Ekland & Bartel, 1996). This ligase was later turned into a more general polymerase by appending a 75-random nucleotide sequence, and subjecting the resulting pool to a conventional in vitro selection process in the hope that this auxiliary module would help in promoting general primer-template recognition (Johnston et al., 2001). This selection relied on the ability of active ribozymes to extend a PT complex tethered to their 5' ends in the presence of ^{4S}UTP. RNA sequences capable of incorporating ^{4S}UMP onto the tethered primer were then enriched based on their reduced mobility in APM polyacrylamide gels (Igloi, 1988). The selection resulted in the 189-nt long Round-18 ribozyme, which can extend an RNA primer by up to 14-nt in a template-directed manner after 24 h of incubation. However the ability of this ribozyme to polymerize more than one RNA helical turn is limited; more typically the Round-18 polymerase adds only a few nucleotides to a given PT complex. This poor polymerization ability has been attributed to weak and highly variable PT recognition (Lawrence & Bartel, 2003), which

in turn is related in a complex fashion to the fidelity of the enzyme. Although the fidelity is high if the polymerase successfully extends a template by as much as 11 nucleotides (Johnston et al., 2001), this measurement does not take into account aborted extension products that terminate due to a mis-incorporation of a nucleotide. Therefore, the efficiency of this polymerase must be greatly improved in order to create an RNA system with true auto-replicative potential. Achieving this goal will require a complex combination of improved primer-template recognition, processivity, fidelity and NTP utilization. As these polymerase properties are interdependent, it will be increasingly difficult to optimize all of these properties simultaneously using standard methods (see section 1.5.3). Indeed, recent attempts to isolate superior polymerase ribozymes from the same pool that was used to isolate the Round-18 polymerase using conventional in vitro selection techniques have not proven fruitful (Lawrence & Bartel, 2005).

Darwinian evolution has solved the problem of correlating genotype to phenotype by utilizing cellular membranes to ensure that the benefits of an advantageous gene are not lost to the external environment. Although cells have been used for the selection of nucleic acids (Ferber & Maher, 1998; Soukup & Maher, 1998; Zimmerman & Maher, 2002; Buskirk et al., 2003; Wadhwa et al., 2004), they offer little or no control over selection conditions and the diversity of the starting library is very limited. Recently, *in vitro* compartmentalization (IVC) has been used as an alternative for conventional *in vitro* selection in the direct isolation of *trans*-acting ribozymes performing multiple-turnover chemistries (Agresti et al., 2005; Levy et al., 2005). The methodology, which was first described for protein selections (Tawfik & Griffiths, 1998; Miller et al., 2006), allows the coupling of genotype (DNA rather than RNA) and phenotype by

compartmentalizing both within water-in-oil droplets. However, the nucleic acid diversities possible with IVC have only been slightly higher than that achieved with cellular systems, and are 10^4 –10^6 times less than with conventional *in vitro* selection, where diversities of 10^{15} can be sampled (Agresti et al., 2005; Fiammengo & Jaschke, 2005; Levy et al., 2005).

In this study we have modified the IVC system to select for an improved RNA polymerase ribozyme from a doped library having a diversity of ~9×10^{14} sequences. The new system was shown to be robust and effective in isolating polymerase ribozymes after six to seven Rounds of selection. One of the pool isolates was shown to have extension activities superior to the wild-type Round 18 ribozyme demonstrating the power of this new selection technique.

4.2 Materials and methods:

4.2.1 Oligonucleotides.

Deoxy-oligonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer using Expedite chemistry. Ribo-oligonucleotides were purchased from Dharmacon and deprotected according to the company's protocol. The selection DNA/RNA primer (D11P12), having the sequence 5'-d(CAC GAG CCA CG) CUG CCA CCG UG, was made using T4 DNA ligase and a DNA splint with the sequence 5'-d(CAG CGT GGC TCG TG).

4.2.2 Pool construction.

Two oligonucleotides corresponding to the reverse complement of the ligase core having the sequence 5'-CAA CTA G**CT CTT CA**G TTg aga acg ttg gcg cta tcg cgc cac

cgg agg ctg cct ctg cat ccg aag atg ttc tca agc tct gag ggc aga ttt gtc ttt tt $(NN)_{0-5}$ C *CTA TAG TGA GTC GTA T* (T7 promoter antisense sequence italicized and the *Ear* I site bolded leaving the underlined sequence recessed after digestion) were synthesized with the lower case residues mutated at a frequency of 0 (L_W) or 3% (L_M). A split and pool strategy was used to synthesize the $(NN)_{0-5}$ region (Zaher & Unrau, 2005), (Figure 4-1a). The two sequences were PCR amplified with the following 5' PCR primer 5'-CTC GCA CCT G**CC GTG** **G**GA GAC GTT CAT CCT TTA *ATA CGA CTC ACT ATA GG* (introducing the bolded *Btg* I recognition site, which leaves the underlined overhang after digestion), and the following 3' PCR primer 5'-CAA CTA G**CT CTT C**AG TT (Bolded is the *Ear* I recognition site). Two oligonucleotides corresponding to the accessory domain sequence 5'-GAC TGC A**CT CTT C**CA ACa ggc gcc caa tac tcc cgc ttc ggc ggg tgg gga taa cac ctg acg aaa agg cga tgt tag aca cgc caa ggt cat aat cCC CGG AGC TTC GGC TCC (the *Ear* I recognition site is bolded, which leaves the underlined overhang sequence after digestion) were also synthesized with one sequence having the lower case residues unmutagenized (A_W) and the other having them mutagenized at a frequency of 10% (A_M). These oligonucleotides were made double-stranded by PCR amplification with the following 5' PCR primer 5'-GAC TGC A**CT CTT C**CA AC and the following 3' PCR primer 5'-GGA GCC GAA GCA CCG GG. The double-stranded constructs for both the ligase core and the accessory domains were then digested to completion with *Ear* I and heated at 72°C for 15 minutes to encourage the short digested pieces to anneal to the PCR primers. The digested L_M was mixed with either A_M or A_W at a ratio of 1:1 before adding T4 DNA ligase to form the full length pool; the third ligation reaction contained L_W and A_M. Half of the initial selection pool (0.75 nmole) was composed of

the L_MA_M construct, and a quarter (0.375 nmole) of each of the L_WA_M and L_MA_W constructs (Figure 4-1b). The pool was then annealed to a 2-fold excess of the selection template T21 (Table 4-1) to produce a pool attached to the PT complex (called the PT-pool, Figure 4-1c).

4.2.3 Polymerization assay in the context of selection.

The wild type polymerase and the starting pool DNA sequences were digested with *Btg* I and ligated to a ^{32}P 5'-end labelled P23 primer using T4 DNA ligase. These constructs were then annealed to a 2 fold excess of the T21 template and incubated in the transcription buffer in the presence of either gel-purified polymerase ribozyme at 2 μM or T7 RNAP at 10 U/μl. The samples were incubated at 37°C for 1 h followed by heating at 65°C for 10 minutes to inactivate T7 RNAP before incubation at 22°C. Time-points were taken at 0, 1, 2, 4 and 20 h, ethanol precipitated and digested with *Hinf* I to improve extended product resolution. The samples were ethanol precipitated and resuspended in 2 × formamide loading dye (95% formamide, 10 mM EDTA, 0.025 % Xylene Cyanol and 0.025 % bromophenol blue) in the presence of 20 fold excess of competing DNA oligonucleotides and loaded onto 15% sequencing PAGE.

4.2.4 Emulsion procedure.

The emulsion protocol, which was originally described by Tawfik and Griffith (Tawfik & Griffiths, 1998) for protein selections, was modified to allow for the generation of litres of emulsion. A 2.5 ml T7 RNAP transcription reaction containing 10 nM PT-pool template in the selection buffer; 40 mM Tris-HCl pH 7.9, 2.5 mM

spermidine, 50 mM $MgCl_2$, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 2 mM ATP, 2 mM CTP and 2 mM UTP; was prepared in a 50 ml falcon tube (Figure 4-2a) and incubated on ice for at least 10 minutes before the addition of T7 RNAP at 10 U/μl. This was followed by the immediate addition of 47.5 ml of a cold (stored at 4°C) oil surfactant mixture composed of 0.5% Tween 80 (Sigma) and 4.5% Span 80 (Sigma) in heavy mineral oil (Paraffin oil white, Anachemia). The mixture was emulsified by adding DEPC-treated 6 mm glass beads (to bring the total volume to full in the falcon tube), taping the tube to a mini vortexer (VWR), and vortexing at full speed for 5 minutes (Figure 2a). For later Rounds, where the volume of the emulsion was reduced to 1 ml, the emulsion was prepared in an eppendorff tube using a dispersing tool (IKA WORKS). The resulting emulsion from both procedures was incubated at 37°C for 3 h to drive T7 RNAP-dependent transcription followed by heat inactivation of T7 RNAP at 65°C for 15 minutes. After incubation, the aqueous phase was recovered by centrifuging the emulsion at 12,000 × g for 20 minutes, which resulted in a white pellet and clear oil supernatant. After discarding the oil supernatant, a two-fold excess of 30 mM EDTA was added to the pellet and vortexed vigorously to resuspend it. Excess oil was removed by extraction with an equivalent volume of diethyl ether to that of the oil mixture. The recovered aqueous phase was ethanol precipitated, phenol/chloroform extracted to remove T7 RNAP and ethanol precipitated again before resuspension in water. To assess the transcriptional yield, a trace amount of [α-^{32}P]-UTP (NEN) was added in the transcription reaction and aliquot of the material recovered after the ether extraction was run on PAGE.

4.2.5 Selection.

A total of 1.5 nmoles (9×10^{14} different sequences) of pool was used for the first Round of selection. This translated to a total of 3 L of emulsion volume, i.e. sixty 50 ml falcon tubes. The emulsion was made as described and incubated at room temperature for ~18 h after heat-inactivation of T7 RNAP. The recovered nucleic acid was then dehybridized by adding KOH to a final concentration of 50 mM. DNA genotypes that were extended by polymerase ribozymes were hybridized to a 10-fold excess of biotinylated DNA oligonucleotide, whose sequence was complementary to the extended sequence, before prompt neutralization with Tris-HCl. The material was then applied to streptavidin magnetic beads (NEB) and washed 3 times with SSC 0.5 × buffer. Hybridized nucleic acid was eluted by applying 50 mM KOH to the beads, which was then neutralized with Tris-HCl and ethanol precipitated. The DNA was then PCR amplified, *Btg* I digested and ligated to the selection DNA/RNA primer before starting the next Round. After three Rounds of selection the selection was split into two different branches A and B. Branch A was conducted as before with no modifications, while branch B included another selection step that included 0.5 mM ^{4S}UTP in the selection buffer. DNA genomes that were successfully extended by ^{4S}UMP were isolated based on their slower mobility in APM polyacrylamide gels (Igloi, 1988). Material eluted from these gel fragments in the presence of 1 mM DTT was precipitated and subjected to the hybridization-based selection protocol.

4.2.6 Assessment of size distribution using microscopy.

A 50 ml emulsion containing 2.5 ml of 5 μM RNA in the transcription buffer and a 5 μl of SYBR green II (Invitrogen) was made as described earlier. Images of the emulsion were taken with a QIMAGING digital camera mounted on a ZEISS Axioskop equipped with epifluorescence.

4.2.7 Cloning.

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

4.2.8 Polymerization assay.

5'-radiolabed primer (0.1 μM final concentration) was annealed to a 5-fold excess of template and added to either the selection buffer or optimum buffer (200 mM $MgCl_2$, 100 mM Tris-HCl pH 8.5 and 4 mM of each NTP). Gel purified ribozymes (5 μM final concentration or otherwise indicated) were added to the reaction and incubated at 22°C. Reactions were stopped by adding a 10-fold excess of formamide loading dye (95% formamide, 10 mM EDTA, 0.025% Xylene Cyanol and 0.025% bromophenol blue) in the presence of ~20-fold excess of a competing oligonucleotide, products were resolved on 20% sequencing PAGE. In reactions where P2 oligonucleotide was included, it was added at 7.5 μM.

4.2.9 Kinetic analysis.

The fractional amount of radioactivity in each extension product was determined using Image Quant v5.1. These extension values were then summed, plotted against time and analyzed using the GraphPad prism v4.00 to determine the overall consumption rate of the initial PT complex. The rate of synthesis of a particular product was determined by quantifying the fraction of the radioactivity corresponding to that product relative to the total radioactivity in that lane and plotting as a function of time.

4.3 Results:

4.3.1 Pool construction.

The Round-18 polymerase, called WT hereafter, is the only ribozyme that has been shown to exhibit general RNA-templated primer extension ability. In the presence of a particular PT complex, the WT can polymerize a full turn of an RNA helix at 200 mM Mg^{2+} and pH 8.5 after 24 h of incubation (Johnston et al., 2001). However, its polymerization activity is extremely sensitive to the PT sequence, with the addition of a single nucleotide to either the primer or the template sequence often having a dramatic and unpredictable effect on activity (Lawrence & Bartel, 2003). With the aim of improving ribozyme processivity and fidelity we decided to select for improved variants of the WT from a high diversity doped pool (Figure 4-1a) using a large-scale compartmentalized selection scheme.

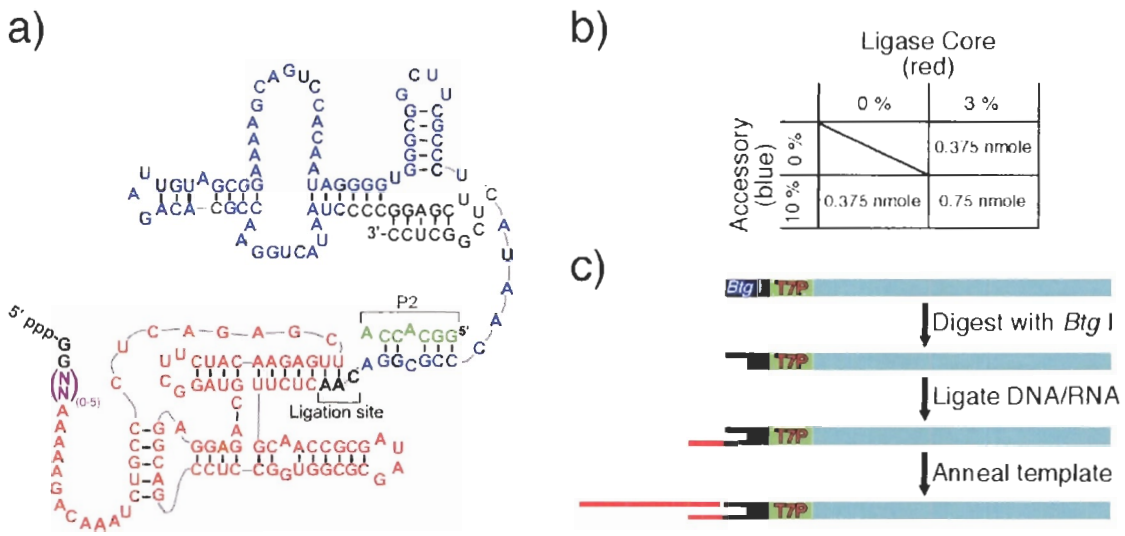


Figure 4-1: Pool design and construction. **a)** Secondary structure of the Round 18 polymerase, which was used to construct the starting pool. The ligase core is shown in red, while the accessory domain is depicted in blue, and the P2 oligonucleotide, found to be important for the WT polymerase ribozyme, is drawn in green. Black residues indicate fixed sequence that was required for transcriptional initiation, PCR amplification or ligation of the two domains after *Ear* I digestion. A random insert of variable length was also introduced, $(NN)_{0\text{-}5}$, 3' to the initiating guanosine residues (magenta). **b)** The starting pool was made from three ligation reactions using the molar amounts shown. **c)** Pool processing for the selection. The DNA pool was digested with *Btg* I leaving a 4-nt overhang that was used to attach the single-stranded selection primer using T4 DNA ligase. This DNA-RNA hybrid (RNA in red) was then annealed to an RNA template in preparation for the selection process (T7P- T7 promoter).

The ligase core of the ribozyme has been previously shown to play an integral role in phosphodiester bond formation and has been found to contain highly conserved sequence elements (Johnston et al., 2001). As a result, two sequences corresponding to the ligase core having no mutagenesis (L_W) or a mutagenesis level of 3% (L_M) were synthesized and converted into double-stranded DNA using large-scale PCR (Zaher & Unrau, 2005). A split-pool synthesis was utilized for both constructs so as to incorporate random nucleotides of the form $(NN)_{0\text{-}5}$ at the 5'-end of the ribozyme (Figure 4-1a) (Zaher & Unrau, 2005). This random insertion provided sequence diversity in a region of the polymerase that had remained fixed ever since the initial characterization of the class I ligase (Ekland & Bartel, 1995)- largely due to limitations of the conventional selection scheme. In contrast to the ligase core, much less is known about the role of the accessory

domain (the auxiliary domain appended to the ligase core during the polymerase selection that confers polymerization ability to the ligase core). Therefore, two accessory double-stranded DNA sequences having mutagenesis levels of either 0% (A_W) or 10% (A_M) were synthesized (Zaher & Unrau, 2005). The ligase core sequences were designed to contain a *Btg* I restriction site upstream of a T7 promoter and an *Ear* I site at their 3'-end. The accessory pool elements contained an *Ear* I site at their 5'-end allowing them to be ligated to the ligase core pool elements with only 3-nt of fixed sequence (Figure 4-1a) after restriction digestion of both with *Ear* I. Pool fragments were ligated to produce full-length pool using 0.75 nmole of L_MA_M, 0.375 nmole of L_WA_M and 0.375 nmole of L_MA_W (Figure 4-1b). The final pool was then digested with *Btg* I and ligated to the D11P12 DNA/RNA primer element (Table 4-1) using T4 DNA ligase. The pool was then annealed to a 2-fold excess of the selection template T21 (Table 4-1) to produce the initial PT-DNA pool (Figure 4-1c). A WT DNA sequence was prepared by cloning the ligation products of the unmutagenized ligase and accessory sequences to obtain a pure sequence.

Table 4-1: PT sequences used in extension assays.

Oligo* Name	Sequence
D11P12	5'd(CAC GAG CCA CG) CUG CCA ACC GUG 3'
D4P12	5'd(CAC G) CUG CCA ACC GUG 3'
P7	5'CUG CCA A 3'
T21	3'GAC GGU UGG CAC GCU UCG CAG 5'
T31	3'GAC GGU UGG CAC GCU UCG CAG CCC CCC CCG G 5'
T26	3'GAC GGU UGG CAC GCC AUC GGA CUA UG 5'
*D, P and T denote deoxynucleotide, primer and template respectively.	

4.3.2 Large-scale emulsion.

IVC has been previously used to isolate RNA ligase and Diels-Alderase ribozymes from libraries having diversities of $\sim 10^9$ and 10^{11} respectively using emulsion volumes of 0.5 to 1 ml (Agresti et al., 2005; Levy et al., 2005). We increased the emulsion volume to 3,000 ml, and by having 10-30 genomes per compartment (weighted by volume distribution), the total pool diversity was expanded to $\sim 9 \times 10^{14}$. The large-scale emulsion was made in 50-ml falcon tube batches, using glass beads and a vortexer (Figure 4-2a). This procedure resulted in compartments having radii typically from 0.5 to 1.5 μm (volumes of 5×10^{-16} to 1.5×10^{-14} L), with a weighted average volume of 2 to 5 $\times 10^{-15}$ L as judged by microscopy (Figure 4-2b) and sedimentation experiments (data not shown). Under these conditions T7 RNAP was functional with an activity 3-fold lower than that of a nonemulsified reaction (Figure 4-2c). The free Mg^{2+} concentration during the incubation was optimized to respect the requirements of both T7 RNAP, which is inhibited by high magnesium levels, and ribozyme activity, which in this case is optimal at high magnesium concentrations. We found that when the free Mg^{2+} concentration (excess over the total NTPs concentration of 14 mM) was set at 36 mM, emulsified transcription yielded ~200 RNA transcripts per DNA sequence (with DNA substrate below the enzymes ~16 nM affinity for the T7 promoter (Zaher & Unrau, 2004), 3 h incubation, 10 U/μl T7 RNAP). Under the same conditions, the WT polymerase was able to extend the untethered selection PT (D11P12:T21) complex by up to 6-nt. Overall these conditions imply that if a single functional DNA sequence were in a compartment, the transcribed ribozymes resulting from this sequence (~200) would be 7 to 20 fold in excess of all PT-DNA substrates found in that particular compartment (10 to 30).

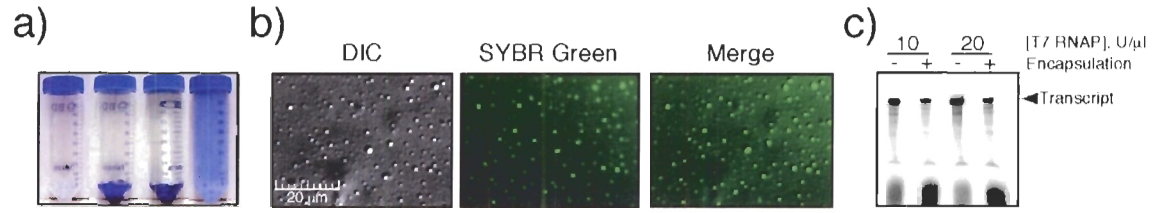


Figure 4-2: Large-scale emulsion manufacture. a) Photographs of the steps used to make 50-ml of emulsion. First (left to right), DEPC-treated glass beads are added into a 50 - ml Falcon tube; second, the aqueous phase (bromophenol blue is included for visualization) is added; third, the oil-phase is poured on top of the aqueous phase before the whole mixture is vortexed to make the water-in-oil droplets. **b)** Microscope images of emulsion containing RNA (180 nt long transcripts at 5 µM) that was stained with SYBR Green. Fluorescence resulting from 450-490 nm excitation co-localized with the DIC image. The droplets are 1-3 µm in diameter. **c)** A denaturing polyacrylamide gel showing that emulsified transcription takes place with a yield 1/3 that of the normal.

4.3.3 Proof of principle and selection design.

Our selection scheme relied on the ability of functional polymerases, which are actively transcribed by T7 RNAP, to extend their own PT-DNA sequences in an entirely *trans* fashion as a consequence of being confined within the same compartment. We have previously shown that T7 RNAP can efficiently extend unstable oligonucleotides in the absence of a promoter (Zaher & Unrau, 2004). This activity could prove highly detrimental to the entire selection process. Successful constructs that are resistant to T7 RNAP-mediated extension require ~ 10-bp of helix between the primer and template. At the same time they must be viable substrates for the WT ribozyme. We found that the selection PT complex having a total of 12-bp of hybridization could not be extended by T7 RNAP but was efficiently polymerized by the WT ribozyme using selective conditions (Figure 4-3). Furthermore, when this primer was ligated to the WT DNA sequence (PT-WT), it was extended by 4-nt when incubated with either gel-purified WT ribozyme or T7 RNAP (Figure 4-3a). When the same primer was ligated to the initial

DNA pool, extension was observed in the presence of gel-purified WT ribozyme, but not when T7 RNAP was added- demonstrating that functional ribozyme polymerases resulting from transcription were required to mediate the extension of the PT-DNA complex.

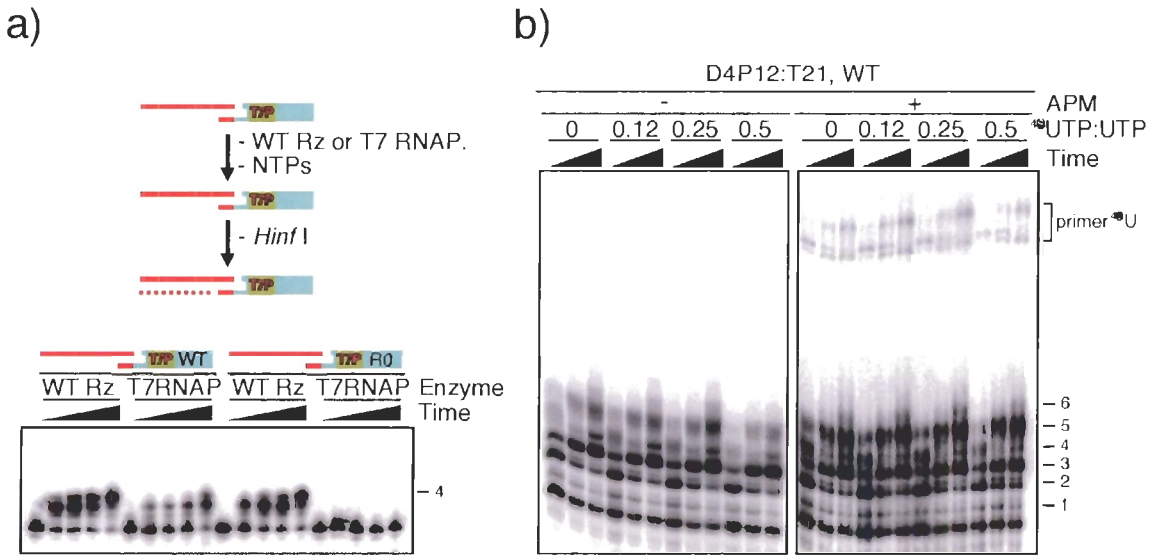


Figure 4-3: The WT polymerase is active in the selection context. a) The PT-WT or the PT-R0 genome constructs were incubated with gel-purified WT (2 μM final concentration) ribozyme polymerase or T7 RNAP, and incubated in the selection buffer. Time-points were taken at 0, 1, 2, 4 and 20 h and the DNA genomes subsequently digested with *Hinf* I to improve gel resolution. Both constructs were extended when the WT ribozyme was added, but only the PT-WT was extended in the presence of T7 RNAP indicating that transcripts produced by T7 RNAP were catalytically active. **b)** The WT ribozyme was transcribed in the presence of varying ^{4S}UTP concentrations (relative to 2 mM UTP which was always present) and gel purified. The four resulting ribozymes were then assayed for their ability to extend PT D4P12:T21 in the selection buffer supplemented with 0.5 mM ^{4S}UTP. Time-points were taken at 0.5 h, 4.5 h and 24 h and resolved on denaturing PAGE plus or minus APM. Ribozymes transcribed in the presence of 1 mM ^{4S}UTP and 2 mM UTP were almost as active as ribozymes that were transcribed in the absence of ^{4S}UTP. All ribozymes were capable of incorporating ^{4S}UMP as indicated by the mercury-dependent gel shift shown in the right panel (marked as primer ^{4S}U in the figure).

Two selective steps were designed to enrich in DNA genomes corresponding to functional polymerase ribozymes. The first used hybridization to purify correctly extended PT-DNA complexes. After incubating the emulsion, the aqueous fraction was recovered, the nucleic acid denatured and an excess of a biotinylated capture probe (having the same sequence as the template used in the PT complex) added (Table 4-2).

Upon renaturation this material was applied to streptavidin magnetic beads and washed so as to retain only those DNA sequences whose RNA primers had been correctly extended by a polymerase ribozyme. These retained DNA genotypes were then eluted and amplified by PCR. This new pool was subsequently *Btg* I digested and ligated to the selection DNA/RNA primer (Table 4-1) before undergoing another Round of selection. A purification factor of 100 could easily be obtained by this hybridization approach (Figure 4-4).

Table 4-2: Selection conditions.

Selection PT:
(D11P12) $^{5'}$d(CAC GAG CCA CG) CUG CCA ACC GUG $^{3'}$
(T21) $^{3'}$GAC GGU UGG CAC GCU UCG CAG $^{5'}$

Round	Capture Probe	^{4S}UTP	P2	[Genome] nM*	Emulsion volume	Inc. t, h
I	d(TTB$_T$ GGA CGC TTC GCA CGG TTG)	-	+	10	3 L	O/N
II	d(TTB$_T$ GGA CGC TTC GCA CGG TTG)	-	+	10	300 ml	O/N
III	d(TTB$_T$ GGA CGC TTC GCA CGG TTG)	-	+	10	10 ml	O/N
AIV	d(TTB$_T$ GGA CGC TTC GCA CG)	-	+	10	1 ml	O/N
AV	d(TTB$_T$ GGA CGC TTC GCA CG)	-	+	5	1 ml	O/N
AVI	d(TTB$_T$ GGA CGC TTC GCA C)	-	+	5	1 ml	O/N
AVII	d(TTB$_T$ GGA CGC TTC GCA C)	-	+	5	1 ml	O/N
BIV	d(TTB$_T$ GGA CGC TTC GCA CG)	+	-	10	1 ml	O/N
BV	d(TTB$_T$ GGA CGC TTC GCA CG)	+	-	5	1 ml	O/N
BVI	d(TTB$_T$ GGA CGC TTC GCA C)	+	-	5	1 ml	4

* 10 nM corresponds to a weighted average, by volume, of 10 to 30 DNA genomes per compartment.
B_T indicates a biotinylated thymidine residue.

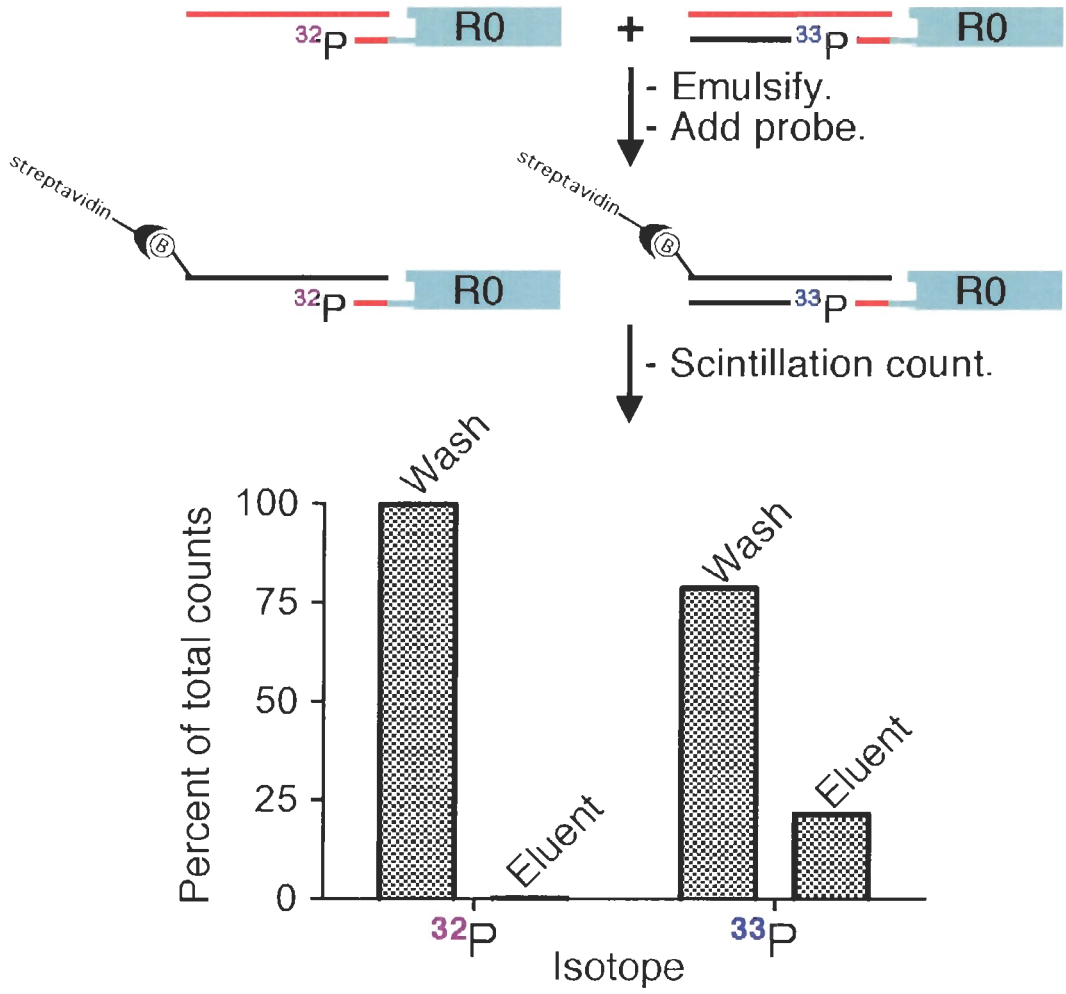


Figure 4-4: The hybridization approach is efficient in isolating extended DNA constructs. A ^{33}P-end labelled D11P12 primer was extended by dNTPs in the presence of T21 using superscript reverse transcriptase (Invitrogen), gel purified and ligated to the DNA pool. This material was then mixed with an equal amount of DNA that had been ligated to a ^{32}P-end labelled unextended D11P12 primer. A mock round of selection recovered 21.4 ± 0.7% of the ^{33}P-labelled material (mock extended) while only 0.21 ± 0.03% of the ^{32}P-labelled construct (unextended) was retained, as determined by spectrum gated scintillation counting.

The stringency of the selection was further increased in some later Rounds of selection by requiring extended PT-DNA genome complexes to contain ^{4S}UMP at sites where UMP was normally incorporated, in addition to having the correctly extended sequence (Figure 4-5 and Table 4-2). This was achieved by adding ^{4S}UTP to the transcription mix prior to emulsification. As a consequence transcribed polymerase

ribozymes must tolerate having ^{4S}UMP in their backbone. Surprisingly, we found WT ribozymes transcribed in the presence of 1 mM ^{4S}UTP and 2 mM UTP were almost as active as ribozymes transcribed in the absence of ^{4S}UTP. Further, ribozymes containing ^{4S}UMP were able to catalyze polymerization of a PT complex using ^{4S}UTP as a substrate (Figure 4-3b) demonstrating that this APM-dependent gel-shift could be effectively used as a purification step.

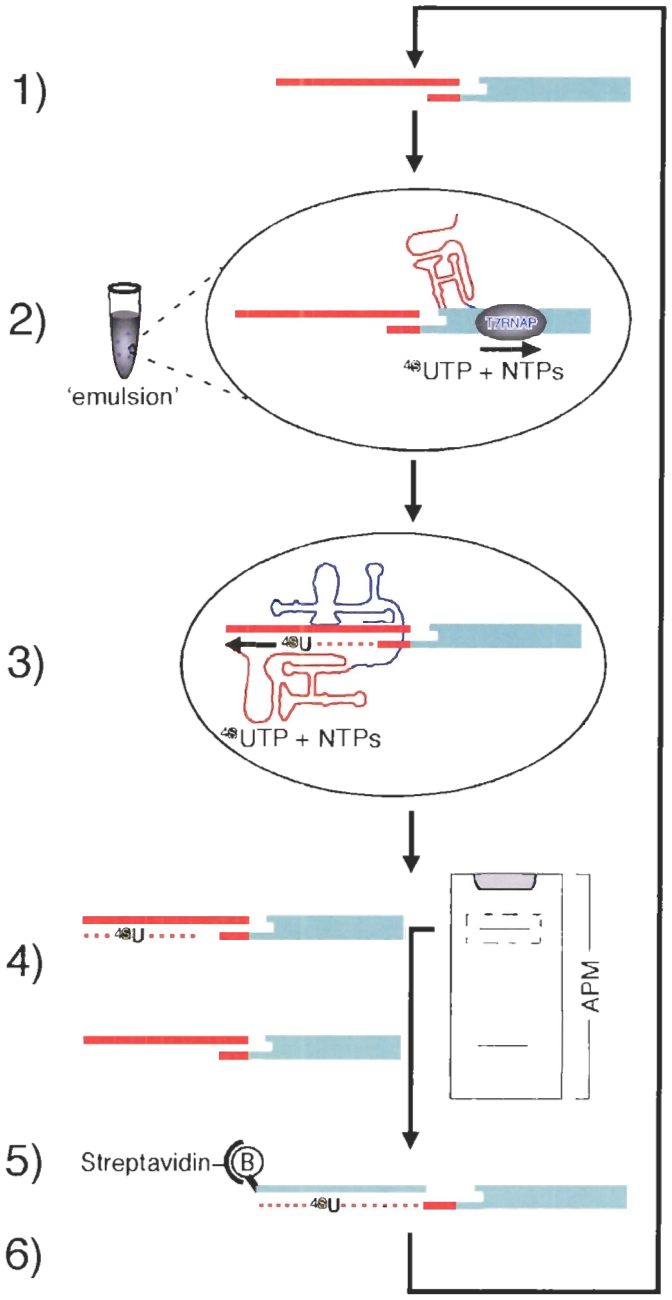


Figure 4-5: Branch B selection scheme. (1) DNA constructs, which were ligated to an RNA primer, are annealed to the selection template. (2) This PT-pool is then compartmentalized within the water-in-oil droplets together with NTPs, ^{4S}UTP and T7 RNAP. Incubation results in T7 RNAP-mediated transcription of the RNA phenotype. (3) Active polymerase ribozymes recognize the PT complex that was attached to their parent DNA genotype and extend it in a template-directed manner while also incorporating ^{4S}UMP. (4) The emulsion is broken open and the recovered DNA genomes separated on an APM gel to enrich in those modified with ^{4S}UMP. (5) The recovered material is then hybridized to a biotinylated probe having a sequence complementary to the extended primer, and applied to streptavidin magnetic beads. (6) The purified pool is eluted from the beads, amplified, digested with *Btg* I and ligated to the selection DNA/RNA primer ready for the next Round of selection.

4.3.4 Selection outcome.

The selection was conducted as described above using a total of 1.5 nmole of DNA genome for the first Round of selection. Based on the pool mutagenesis levels, we estimated the WT sequence to be present with a frequency of 0.3% in the starting pool. Two distinct branches of selection were performed and are summarized in Table 4-2. Branch A of the selection was performed using only hybridization as a selective step and included the P2 oligonucleotide (a 7-nt RNA having the sequence 5′-GGC ACC A added to restore a stem that was present in the parental class ligase I, and that has been reported to be critical for function (Johnston et al., 2001), (Figure 4-1a). After seven Rounds of the branch A selection, the extension activity of the pool was enriched significantly (Figure 4-6). Further Rounds of selection did not seem to enhance activity, instead a competing parasitic activity, characterized by a smear above the radiolabelled primer during PAGE analysis of extension assays, was observed. Sixty clones from Round 7 were assessed for polymerization activity, 12 of which had extension activity comparable to the WT ribozyme, 23 exhibited the parasitic behaviour while the remaining 25 were inactive. Subsequent sequencing of these isolates revealed that 10 of the active isolates were 0 to 2 mutations away from the WT sequence. The remaining two active variants, referred to as A7.15 and A7.29, were significantly different from the WT ribozyme and contained many mutations in the accessory domain (Figure 4-1a). Furthermore, both contained mutations in the critical ligase core. Interestingly, replacing these mutations with the WT ligase core sequence resulted in no appreciable increase in extension rate. Sequencing of the 23 parasitic DNA sequences revealed that they had managed to insert from 1 to 6 consecutive copies of a sequence containing the biotinylated capture probe

sequence (5'-*TTT GGA CGC TTC GCA CGG TTG* CAG D, capture probe sequence in italics) by a currently unknown mechanism. Presumably the orientation of the insertion was important for the survival of these parasites. The RNA transcribed from these sequences could not hybridize to the capture probe, preventing its saturation, while the bottom DNA strands which contained sequence complementary to the biotinylated probe would have been enriched each Round of selection. We attempted to get rid of these parasites by switching the PT complex sequence from Round to Round while altering the biotinylated probe sequence in correspondence. This approach proved successful in removing the unwanted parasitic sequences, but at the same time eliminated polymerization activity even after 14 Rounds of selection. Furthermore, the P2 element was discovered to have inhibitory effects on the WT polymerase (Figure 4-7) suggesting that its omission might have a beneficial effect on the selection.

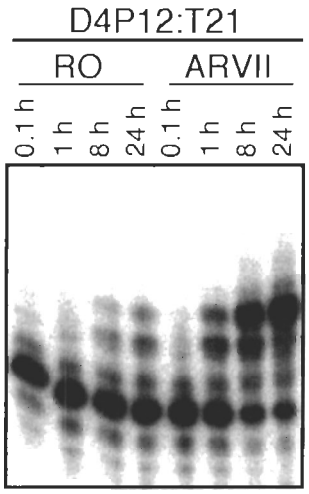


Figure 4-6: Progress of the branch A selection. Gel-purified RNA from the indicated round was added at a final concentration of 0.2 μM to an optimum polymerization reaction. By Round seven, polymerization of D4P12:T21 by 3-nt increased by a factor of ~50.

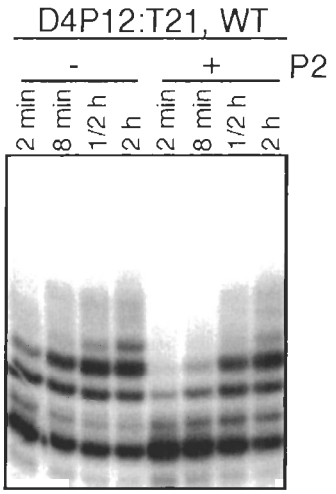


Figure 4-7: P2 oligonucleotide suppresses polymerization by the WT ribozyme. Extension-reactions of D4P12:T21 complex by the WT ribozyme were performed in the presence or absence of the P2 element under selection conditions, and time-points taken at the indicated intervals. In the presence of the P2 RNA the overall rate of polymerization was inhibited by a factor of ~7.

Hoping that the pool contained rare improved polymerases that were lost in the branch A selection due to the overwhelming presence of the parasitic sequences and the effect of the P2 element, we went back to Round 3 and included an APM-gel purification step by adding ^{4S}UTP in the selection buffer and excluding the P2 element from the selection (Branch B, Figure 4-5). The inclusion of the APM step before the hybridization capture step indeed prevented the amplification of the undesired sequences. Branch B was conducted for three Rounds bringing the total number of selection Rounds to 6 with the stringency of the selection being increased by lowering the DNA pool concentration and incubation times in the later Rounds (see Table 4-2). The effect of the APM-purification step was apparent after only two Rounds of selection, and by Round 6 polymerization of PT D4P12:T21 complex by 4-nt had saturated and was ~60 fold higher than the starting pool (Figure 4-8a). Screening 60 isolates from this Round identified a total of 10 active and 50 inactive isolates with one clone, called B6.61, having an increased polymerization activity relative to the WT polymerase. The relative polymerization enhancement was measured by first quantifying the amount of nucleotides incorporated beyond a particular extension product as a function of time and then dividing it by the equivalent WT extension rate. B6.61 was 3 to 4 fold faster at polymerizing six or more nucleotides in the presence of the PT D4P12:T21 complex than either the WT ribozyme or the best isolates from the branch A selection (Figure 4-8b). Interestingly, the B6.61 sequence differed form the WT sequence by a 4-nt insertion between G_2 and A_3 combined with an A_3 to C mutation (5'-GG**A AUA C**A..). This insertion was identical to that found in the A7.15 variant and was similar to that of the 6-nt insertion found in the A7.29 isolate (5'-GG**C AUA CC**A A..). In addition, all three

variants acquired an A_{170} to G mutation in the accessory domain (Figure 4-8c) suggesting a convergent functional requirement for both sequence modifications.

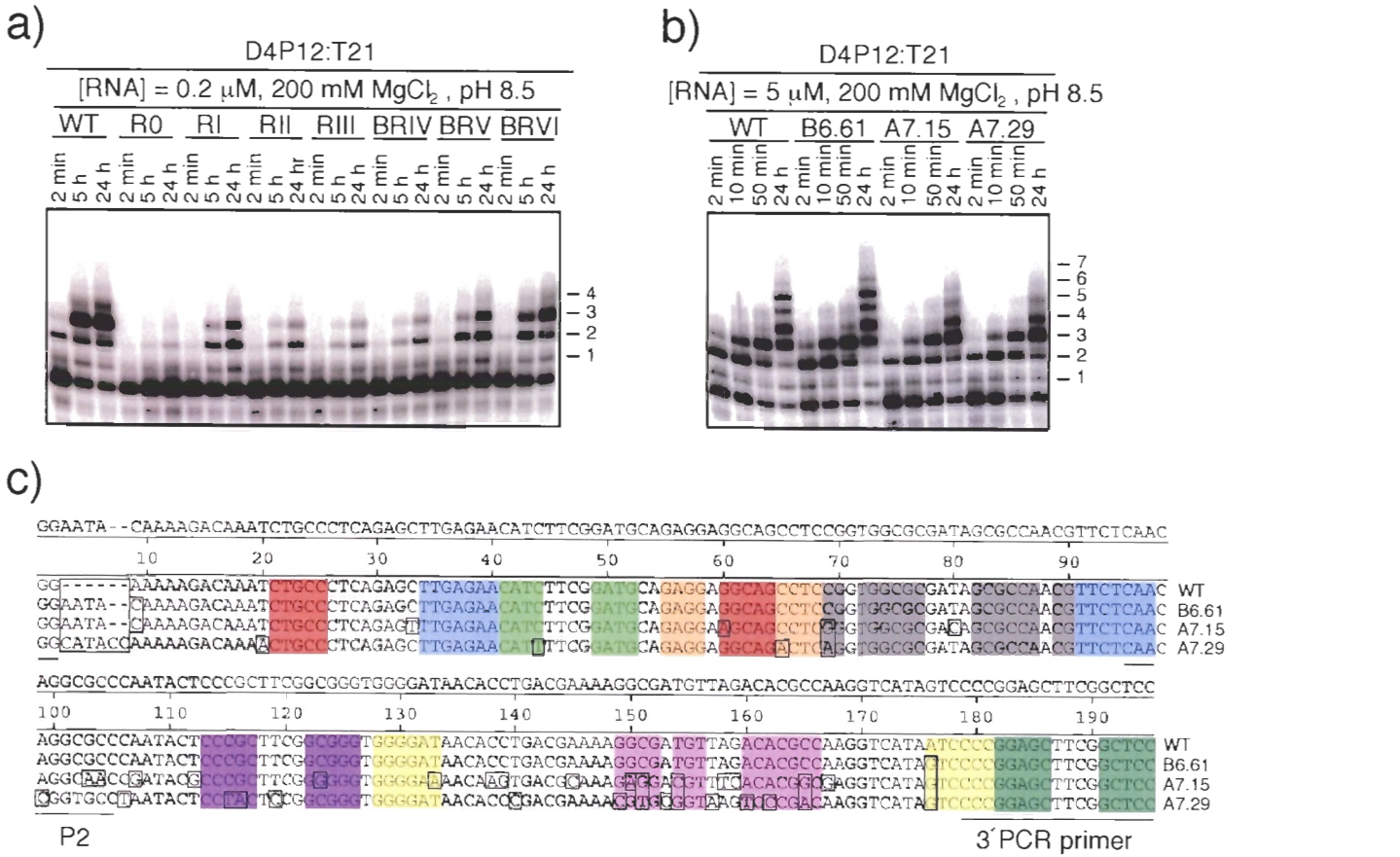


Figure 4-8: Progress and outcome of the selection. **a)** Polymerization activity of gel-purified selection Rounds under the conditions indicated. Aliquots were taken at the indicated time intervals and extension products separated on 20% PAGE. Length of extension products are shown on the right. After six Rounds of the branch B selection, extension activity was enriched by a factor of 60. **b)** Extension of PT D4P12:T21 complex by variants from branches A and B of the selection. B6.61 extended more than 6-nt of the PT complex 3-fold faster than the WT, while A7.15 and A7.29 were slightly slower. **c)** Sequence alignment of the selected isolates relative to the WT sequence. Elements highlighted in the same colour are proposed to form secondary structure with each other based on the previous secondary structure determination of the class I ligase and WT polymerase (Ekland & Bartel, 1995), (Johnston et al., 2001). Residues that are different from the WT sequence are boxed. The initiating guanosine residues are underlined, while the underlined CAA stretch indicates the fixed sequence that was required to ligate the two modules together to make the starting pool.

Given that our selection was performed at substantially lower magnesium and hydroxide ions concentrations than are known to be optimal for the WT ribozyme, we were curious whether the improved activity observed for B6.61 was due to better utilization of either of these ions. Consequently we performed extension reactions for both ribozymes under varying concentrations of Mg^{2+} and different pH values. Both B6.61 and the WT polymerases exhibited the same relative decrease in activity at lower Mg^{2+} and OH^- ions concentrations, and hence the enhancement of activity could not be attributed simply to either of these two factors (data not shown). As our selection was also carried out in an entirely *trans*-fashion, we expected that the selected ribozymes might exhibit a better affinity for the PT complex. Titration of the PT complex P7:T31 up to 15 μM (ribozyme at 5 μM) did not saturate binding for either polymerase as both exhibited a linear increase in extension rate as a function of PT concentration (data not shown). As a result, we could only conclude that the efficiency of polymerization by B6.61 had improved during the IVC selection. B6.61 may therefore have improved its binding affinity for the PT substrate or the mutations observed may be responsible for an increased catalytic rate- perhaps by subtly reorganizing the ribozyme catalytic core in the vicinity of the PT interface.

4.3.5 B6.61 extends beyond the 14-nt limit.

The faster kinetics exhibited by B6.61 were also apparent in the presence of a shorter primer P7 (Table 4-1) when annealed to the T21 template. Although both the WT and the B6.61 polymerases extended the P7:T21 PT by 14-nt, B6.61 polymerized extensions of more than 9-nt 3-fold faster than the WT (Figure 4-9). Arguably most

interesting was the finding that in the presence of template T31, where 8 cytosine and 2 guanosine residues were appended to the 3'-end of T21, B6.61 extended the primer P7 beyond the 14-nt limit, reaching a total extension of more than 20-nt (Figure 4-9), while the WT ribozyme failed to extend more than ~11-nt. This marked increase in extension could be attributed to a more processive polymerization by B6.61. Although the rate of consumption of the P7:T31 PT by B6.61 was only 2-fold faster than found for the WT, the rate of polymerization of RNA longer than 10-nt was at least 80-fold faster, accounting for a total of 7.5% of the primer after 24 h of incubation while the WT had only extended about 0.1%. The relative rate enhancement for longer extension products was even higher but could not be quantified.

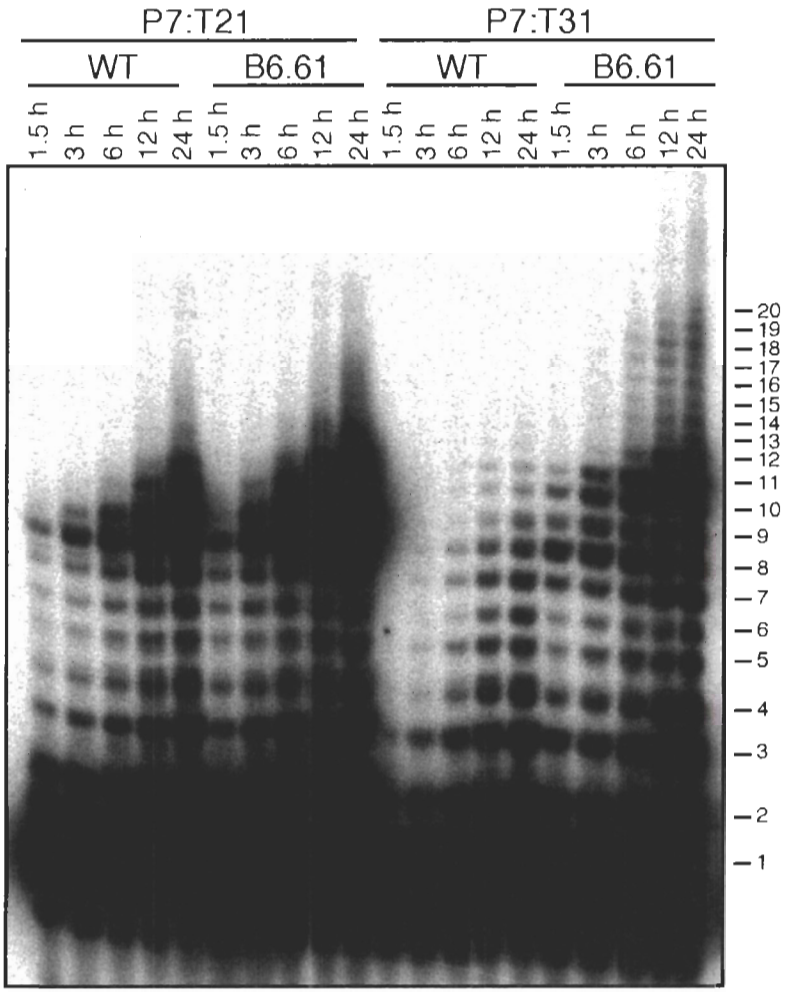


Figure 4-9: B6.61 polymerizes a PT complex by at least 20-nt. The indicated PT complex was incubated in the presence of WT or B6.61 ribozymes under optimum conditions. Reactions were stopped at the specified time-intervals and resolved using 20% PAGE. Although both polymerases extended P7:T21 by 14-nt, B6.61 was 3-fold faster in synthesizing > 9-nt extension products (left panel). In the presence of P7:T31, B6.61 was able to incorporate at least 20-nt in contrast to the 11-nt of extension detectable by the WT ribozyme (right panel).

4.3.6 B6.61 is more accurate.

Why was B6.61 so much better at extending the P7:T31 PT than the WT? To explore this question further both B6.61 and the WT polymerase were incubated with the D4P12 primer, which is able to hybridize to 12-nt of either the T21 or the T31 template. Again as previously observed (Figure 4-8b), B6.61 was faster at extending both the D4P12:T21 and the D4P12:T31 PT complexes than the WT (Figure 4-10a). The

nucleotides incorporated by either ribozyme when given the D4P12:T21 complex appeared to be the correct sequence as judged by the following criteria: (1) Hybridizing the D4P12:T21 extension reaction to the biotinylated probe used in the first Round of selection and subsequent binding to streptavidin magnetic beads resulted in 60-70% of the radiolabelled primer being retained on the beads. (2) The extended products were found to be viable substrates for reverse transcriptase in the presence of T21. (3) Sequential addition of NTPs to the polymerization reaction showed an incorporation of NMPs that correlated with the template sequence. This fidelity was important to confirm as unexpectedly the T31 template when annealed to the D4P12 primer resulted in the very fast incorporation of nucleotides that were not simply encoded by the template sequence (marked by asterisks in Figure 4-10a). Since T31 shares nucleotides 1-21 with T21, the same extension pattern up to the 9-nt addition point on either template had been expected. Nevertheless, B6.61 synthesized more template-encoded product in the 4 to 8-nt addition range and more rapidly than for the WT. Simultaneously, B6.61 produced less of the incorrect extension products strongly suggesting that it has an improved polymerization fidelity.

The observation that B6.61 had a superior fidelity relative to the WT was confirmed by a second extension assay. Polymerization of PT P7:T26 complex by the WT polymerase produced a mis-incorporation after the first extension product (marked by an asterisk in the "-PT WT lanes" Figure 4-10b). This error was later characterized as an incorporation of a uridine rather than a cytosine forming a U:G wobble pair (Figure 4-11). This mis-incorporation was suppressed by the addition of the P2 element to the WT ribozyme but at the cost of introducing a second incorrect incorporation after the 2-nt

product (again a uridine was incorporated, "+PT WT lanes" Figure 4-10b). Remarkably, B6.61 hardly accumulated any incorrect product at either location, producing ~ 3 fold less of the first mis-incorporation in the absence of the P2 element. Only addition of the P2 element caused appreciable mis-incorporation that strongly resembled the WT profile. Thus not only does the new sequence found in B6.61 confer greater polymerization efficiency, but this sequence in some way improves polymerization fidelity by offsetting the importance of the P2 element.

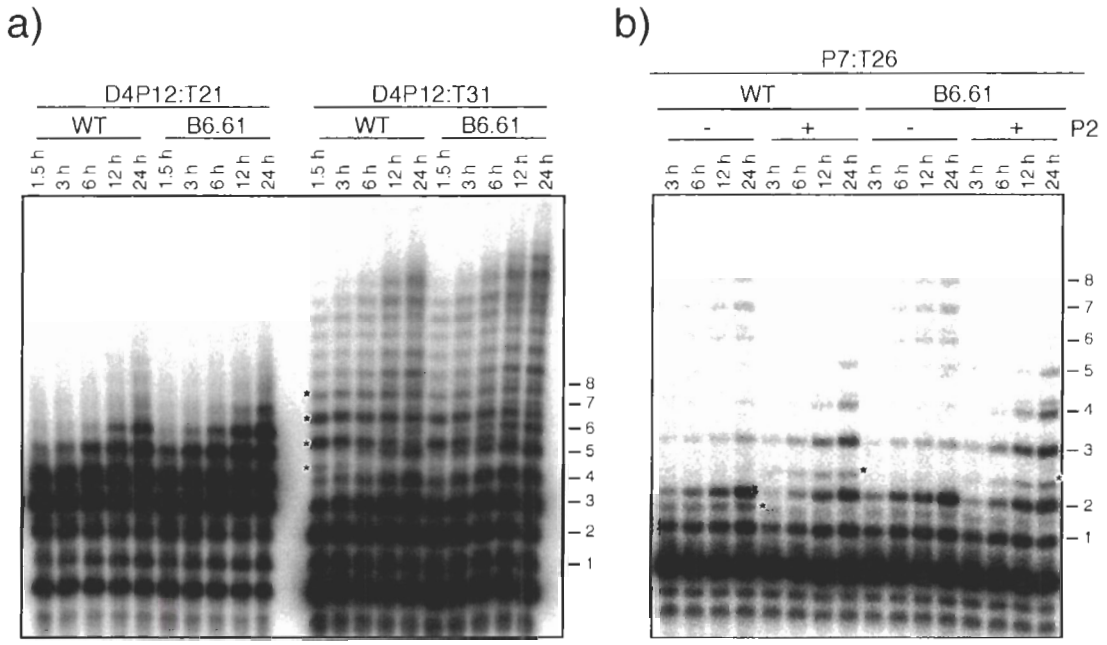


Figure 4-10: The fidelity of B6.61 is superior. **a)** The extension pattern of PT D4P12:T31 by both polymerases was compared to that of PT D4P12:T21. Reactions were conducted under optimum conditions, and aliquots taken at the indicated time-points. Expected extension products are numbered, while incorrect ones are marked by asterisks. Note that B6.61 produces ~3-fold more of the expected products relative to the WT. Products beyond the 8-nt addition were not labelled as there was no standard to determine whether they were correct additions or not. **b)** Effect of P2 on ribozyme fidelity. Ribozyme-catalyzed extension of PT P7:T26 were conducted in the absence or presence of the P2 oligonucleotide under optimum conditions. In the absence of the P2 element, the WT polymerase adds an incorrect product after the 1st addition (marked by an asterisk), while B6.61 produces 3 fold less of the same anomalous product. In the presence of P2, the accumulation of this product by the WT ribozyme was inhibited, but the addition of an incorrect nucleotide after the 2nd addition was enhanced by both polymerases.

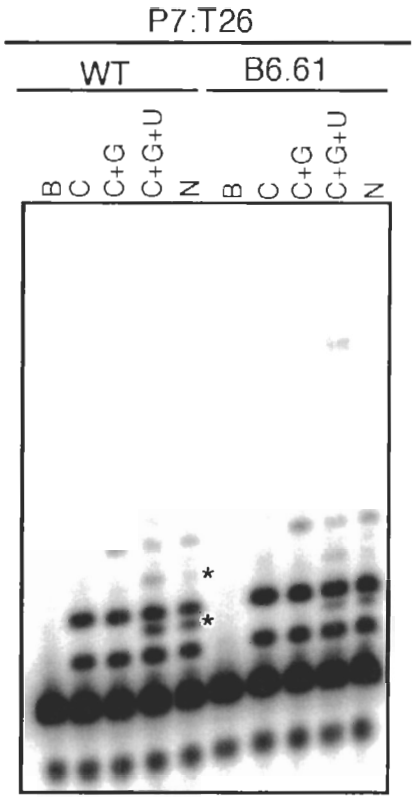


Figure 4-11: The WT polymerase utilizes wobble base-pairing during extension more often than B6.61. Polymerization of P7:T26 by both ribozyme was conducted in the absence of NTPs (B), CTP (C), CTP and GTP (C+G), CTP, GTP and UTP (C+G+U), and all 4 NTPs (N). Only in the presence of UTP is the mis-incorporation observed, where the WT ribozyme incorporates UMP across a guanosine residue ~3 fold more than the B6.61 ribozyme.

4.4 Discussion:

For the first time IVC has been used to select catalytic RNAs from a pool having a diversity comparable to the largest used by conventional *in vitro* selection techniques (~10^{15}) (Bartel & Szostak, 1993). By using emulsion on the 3 L scale we have isolated, from a pool of ~9×10^{14} Round-18 polymerase variants, a superior polymerase ribozyme (B6.61), demonstrating the power of this new large-scale selective technique. B6.61 exhibited better polymerization kinetics, and with certain PT complexes the rate of extension beyond a helical turn was nearly two orders of magnitude faster than the WT ribozyme, albeit typical rate enhancements were only 2-4 fold superior (Figure 4-8, 4-9

and 4-10). At least part of this superiority could be attributed to the ability of B6.61 to correctly incorporate Watson-Crick pairs on templates where the WT made coding errors (Figure 4-10). The combination of enhanced extension kinetics and improved fidelity resulted in a ribozyme able to polymerize at least 20-nt of sequence (Figure 4-9)– nearly two complete helical turns, setting a new extension record for an artificial polymerase.

Perhaps the most interesting aspect of the selection reported here is its relevance to the construction of complex RNA-based systems. This IVC polymerase selection attends to two key points that are thought to be pertinent to early life namely replication and compartmentalization (Bartel & Unrau, 1999). Although the water-in-oil droplets used here are prebiotically irrelevant, they can serve as a starting point for future endeavours where both catalysis and compartmentalization are required. In particular, the IVC technique can be used to construct either RNA pathways or systems requiring more than one RNA strand that would be extremely difficult or impossible to achieve using conventional techniques. Round 6 from the branch B selection contained ~80% inactive genomes, which implies that the catalytic power of active ribozyme polymerases such as B6.61 are sufficient to carry forward a mixed population of DNA genomes through the selection process. If these inactive genomes could be replaced with a high diversity pool and placed under selective pressure in the presence of B6.61 so as to improve its polymerization capability, it might be possible to isolate a truly processive multi-subunit polymerase ribozyme.

All modern protein DNA polymerases, in addition to being complex multi-subunit machineries, depend on other enhancing factors to reach their full processive potential. For instance, the core DNA polymerase III of *E. coli* polymerizes DNA with a rate of 20

nucleotides s^{-1} and a processivity only of tens of nucleotides. By interacting with the β-clamp protein, its rate of polymerization is improved by a factor of ~40 with a processivity exceeding 50,000 nucleotides (Benkovic et al., 2001). B6.61 has a respectable polymerization rate with incorporation rates reaching 0.6 nucleotides min^{-1} (Figure 4-8), still far less than the 100 min^{-1} k_{cat} observed for phosphodiester bond formation by the class I ligase core (Bergman et al., 2000). As a result, it is reasonable to assume that selection for a B6.61 RNA cofactor could dramatically improve the polymerization rate of the B6.61 complex. Furthermore, as the polymerase currently exhibits a low affinity for the PT complex, it is conceivable that the association of B6.61 with this RNA cofactor may dramatically improve the processivity of this ribozyme polymerase to a point that auto-replication might seriously be considered.

CHAPTER 5:
T7 RNA polymerase mediates fast promoter independent extension of unstable nucleic acid complexes

This chapter is largely based on the manuscript; Zaher HS, Unrau PJ. 2004. "T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes". *Biochemistry* **43**:7873-7880.

5.1 Introduction:

T7 RNA polymerase is a DNA-dependent RNA polymerase consisting of a single polypeptide chain of 883 amino acids (Dunn & Studier, 1983) that requires no additional transcription factors; making it ideal for the *in vitro* preparation of RNA from synthetic or non-synthetic DNA templates containing the promoter sequence (Milligan et al., 1987; Krupp, 1988). After binding with nanomolar affinity to its promoter sequence (Oakley et al., 1979; Martin & Coleman, 1987), T7 RNAP forms a stable elongation complex capable of sustained, high speed processive transcription (Cheetham et al., 1999; Cheetham & Steitz, 1999; Yin & Steitz, 2002). The transition from promoter binding to elongation is initiated by the slower synthesis of a short RNA fragment (10-12 nt long) that induces a substantial rearrangement of the N terminal region of the polymerase (Cech, ; Cheetham et al., 1999; Cheetham & Steitz, 1999; Yin & Steitz, 2002). This process is imperfect and short abortive and 5′ inhomogeneous transcripts are often created by the enzyme failing or incorrectly transitioning from an initiation to elongation complex (Milligan & Uhlenbeck, 1989; Pleiss et al., 1998; Tuschl et al., 1998).

Given the highly specificity of the T7 promoter it is surprising that RNA synthesis can still occur in the absence of a T7 promoter sequence. RNA sequences that can self-

replicate (RNA X) in the presence of T7 RNAP have been reported by Konarska and Sharp (Konarska & Sharp, 1989, 1990). It has also been shown that the extension of transcribed RNA can occur using RNA as a template (Cazenave & Uhlenbeck, 1994). The synthesis of these RNAs, as inferred from their sequence, appears to involve a complicated iterative process of extension and template remodelling (Cazenave & Uhlenbeck, 1994). This activity appears to be quite unlike the highly processive elongation that occurs during regular transcription, and is poorly characterized.

In this report, we have studied the transient aspects of T7 RNAP promoter independent polymerization by limiting the processive ability of the enzyme. By choosing appropriate oligonucleotide sequences and incubating them with individual nucleotide triphosphates, priming regions within an oligonucleotide can be targeted and their ability to sustain nucleotide incorporation evaluated. By limiting incorporation to a single nucleotide type we were able characterize the minimal substrate requirements for transient nucleotide incorporation and determine the basic kinetics of this process.

5.2 Materials and methods:

5.2.1 T7 RNA polymerase.

T7 RNAP was purified from *E. coli* strain BL21 carrying the His-tagged plasmid pT7-911Q (Ichetovkin et al., 1997). The enzyme was purified as previously described (Smith et al., 1988) using immobilized nickel ion affinity chromatography (NTA resin, Qiagen). Protein concentration was determined by optical density at 280 nm using an extinction coefficient of $1.4 \times 10^5\ M^{-1}cm^{-1}$ (King et al., 1986). The activity of our enzyme preparation compared favourably to commercial T7 RNAP (Roche) in a normal

transcription assay and the enzyme was found to have a specific activity of ~4.3×10^7 U/μmole (1 unit of enzyme can incorporate 1 nmole of CTP into acid-precipitable RNA products in 60 minutes at 37°C at pH 8.0 (Davanloo et al., 1984).

5.2.2 Oligonucleotide synthesis.

RNA oligonucleotides were purchased from Dharmacon and deprotected according to the company's protocol. RNA was dried using a SpeedVac, and resuspended in water. DNA oligonucleotides were synthesized on an ABI 392 DNA synthesizer using Expedite chemistry.

5.2.3 5′-oligonucleotide labelling.

Oligonucleotides at ~500 nM, were incubated in T4 polynucleotide kinase buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6) supplemented with 74 kBq/μl of [γ-^{32}P]-ATP (111 TBq/mmol, NEN) and 0.5 U/μl of T4 polynucleotide kinase (Invitrogen) at 37°C for 30 minutes. The resulting radiolabelled oligonucleotides were purified using 15% preparative PAGE and excised from the gel. Oligonucleotides shorter than 10 nt were eluted in 1 mM EDTA for at least 12 hours. The eluate was passed over an equilibrated (50 mM triethylammonium acetate, pH 7.6) C18 SPICE cartridge (Analtech). The cartridge was then washed with the same buffer, and the bound nucleic acids eluted with 2.5 ml of acetonitrile. The recovered oligonucleotides were dried on a SpeedVac overnight. Oligonucleotides longer than 10 nt were eluted from gel fragments in 300 mM NaCl for at least 12 hours and then ethanol precipitated.

5.2.4 Extension reactions.

The standard assay was carried out in 40 mM Tris-HCl (pH 7.9), 2.5 mM spermidine, 26 mM $MgCl_2$, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 4 mM ATP, 4 mM CTP and 2 mM UTP. Standard assays using single NTPs were performed at 4 mM, with oligonucleotides at 10 nM unless otherwise stated. Final enzyme concentration was 0.1 μM. Incubations were carried out at 37°C and stopped by the addition of an equal volume of 2 × formamide loading dye (95% formamide, 5 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue). Reactions were resolved on 20% polyacrylamide sequencing gels unless otherwise specified.

5.2.5 Kinetic studies of promoter-dependent initiation.

The method used was similar to that of Martin and Coleman (Martin & Coleman, 1987). The T7 promoter complex was formed by annealing PT7 (5′-CTT TAA TAC GAC TCA CTA TAG G) and CT7 (5′-AGT CCT ATA GTG AGT CGT ATT AAA G), each at 4 μM, in TE (pH 7.8) at 94°C for 2 minutes before cooling to room temperature.

5.2.6 dsDNA inhibition of extension.

A short template sequence 5′-AAA GGG AAC C (SC) complementary to the 3′ end of TG_2L_4 (see Table 5-1 for sequence) and a fully complementary sequence 5′-AAA GGG AAC CAA AAA AAA AA (LC) were titrated from ~1 nM to 1 μM in the presence of 10 nM 5′-labelled TG_2L_4 in CTP or UTP extension buffer and annealed as before. Reactions were started by addition of enzyme and stopped with loading dye and a 10 fold excess of unlabelled TG_2L_4.

5.2.7 Analysis.

The fraction reacted for 5′ labelled extension products was calculated by dividing the intensity of each band by the total intensity in that lane. Total CMP incorporation was obtained by multiplying the fraction reacted at the n^{th} extension by n, the number of extensions. These weighted extension values were summed and reflect the total CMP incorporation per oligonucleotide. For initiation data (labelled with [γ-^{32}P]-GTP), the fraction of total radiolabel incorporation from all products was added and multiplied by the concentration of GTP to give the molarity of initiated strands. K_m and k_{app} values were determined using the Michaelis-Menten equation and fit using GraphPad Prism V4.00.

5.3 Results:

A short RNA oligonucleotide, 5′-r(CUG CCA A), was observed to be specifically extended by GMP when incubated in the presence of T7 RNAP and GTP (Figure 5-1). A deoxynucleotide having the sequence 5′-CTG CCA A, was also only extended by GMP (extension was ~1.5 times faster than the RNA sequence). The extension of these and other oligonucleotides was not dependent on enzyme preparation, as commercial T7 RNAP (Roche) had nearly identical transcriptional and extension properties to our enzyme. Both enzyme preparations exhibited a slow exonuclease activity of 0.001-0.01 min^{-1} per oligonucleotide at 0.1 μM enzyme concentration (Figure 5-1) (Sastry & Ross, 1997). The appearance of a faint second band in the GTP lane for both the RNA and DNA sequences (Figure 5-1) suggested to us the possibility that these constructs form extendable loop structures primed by a single bp between the second and most terminal

residue in the oligonucleotide and using the most 5′ residue as a template. This arrangement might, if the template could slip relative to the extension product, allow the incorporation of a second GMP residue by priming from the rG:rU (rG:T) wobble pair created by the first nucleotide incorporation.

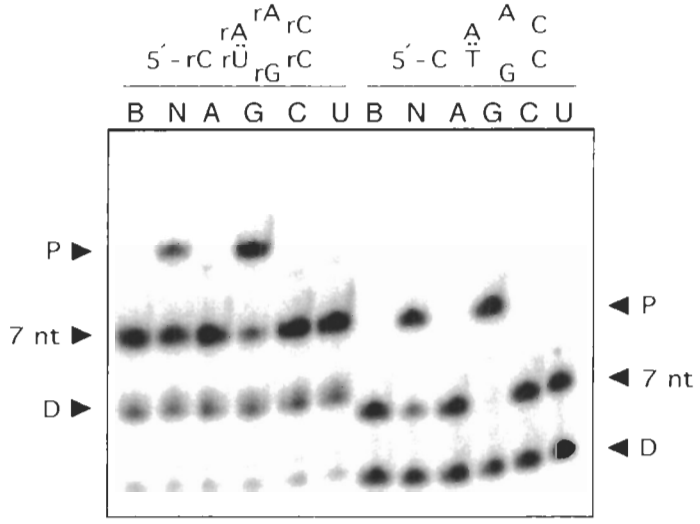


Figure 5-1: Extension reactions of short RNA and DNA oligonucleotides having analogous sequences. Sequence and proposed transient extension structures are indicated. Extension by T7 RNAP in the presence of: no NTPs (B), all 4 NTPs (N), ATP (A), GTP (G), CTP (C), or UTP (U). Extension is only observed in the presence of GTP (indicated by P). A weak second extension suggests the possible formation of a rG:rU or (rG:T) wobble bp. A weak nuclease activity intrinsic to T7 RNAP creates a 6 nt degradation product (indicated by D) after incubation in standard conditions for 45 minutes.

5.3.1 A single arbitrary bp can initiate polymerization.

To test this hypothesis, a series of mutants derived from the DNA sequence 5′-CTG CCA A were synthesized and screened using our single nucleotide triphosphate T7 RNAP extension assay. The DNA sequence 5′-GGC ACC C (named: G_2L_4, with L standing for the hypothetical loop size between G(2) and C(7)), was found to specifically incorporate CMP and no other nucleotide. However, in contrast to the sequence 5′-CTG CCA A, numerous CMP incorporations were observed (Figure 5-2). Single and double mutations of G_2L_4 were synthesized that either preserved or destroyed this hypothetical

G(2):C(7) bp. Mutations that preserved base-pairing were found to be active in our single nucleotide primer-extension assay, while mutations that disrupted base-pairing had no detectable extension activity (Figure 5-2a). Multiple extensions were not observed when the initial bp was changed to A(2):T(7), presumably because freshly incorporated C(8) cannot hybridize with the residue A(2).

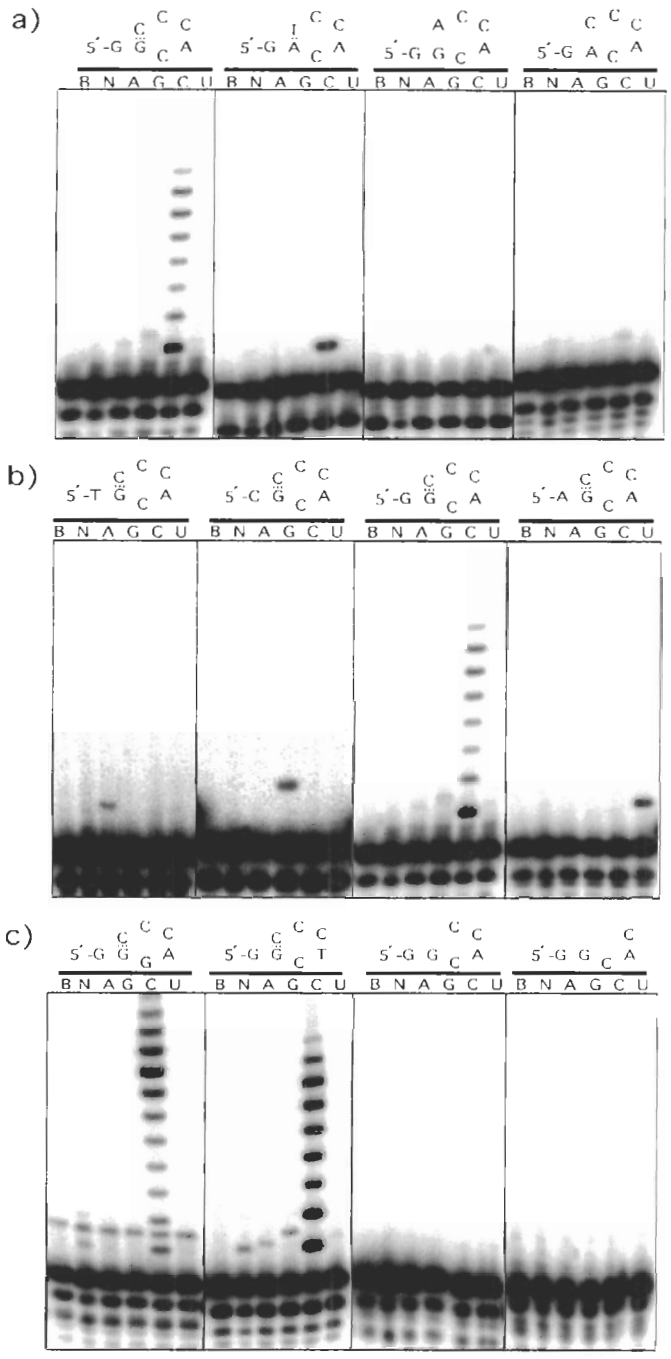


Figure 5-2: Extension of sequence variants of 5′-GGCACCC. Sequences and proposed transient extension structures are shown. **a)** Mutations of the transient base-pair G(2):C(7). Extension is only observed when the bp is preserved. **b)** Mutations of the coding nucleotide N(1) from left to right: T, C, G and A. Extension is only observed in the presence of the correct complementary NTP. Multiple extensions are observed only for the 5′-GGCACCC construct in the presence of CTP. **c)** Mutations of the loop sequence can result in increased extension rates, while point deletions eliminated extension. All panels show extension by T7 RNAP in the presence of: no NTPs (B), all 4 NTPs (N), ATP (A), GTP (G), CTP (C), or UTP (U). Oligonucleotides were incubated using standard conditions for 45 minutes.

5.3.2 The incorporated nucleotide was encoded by the template.

All 7 nt constructs appeared to use the terminal 5′ residue of the oligonucleotide to code for nucleotide extension. Constructs that varied the proposed coding nucleotide from G(1) in G_2L_4 to A, C or T, while preserving the proposed G(2):C(7) bp, all showed extension by the correct complementary nucleotide monomer when single nucleotide triphosphates were fed to the enzyme (Figure 5-2b). Extension in the presence of all four NTPs often took place at a reduced rate or was not observed. The exact dependence of this effect with respect to template sequence was not fully explored.

5.3.3 Unimolecular and bimolecular extension depends on template choice and concentration.

To investigate the nature of the primer-template complex further, the kinetics of G_2L_4 extension were measured. The initial rate of nucleotide incorporation per enzyme climbed rapidly with substrate concentration, plateauing at a k_{app} of 2.0 min^{-1}. A Hill coefficient very close to one was observed (Figure 5-3a, Coefficient = 0.96 ± 0.10, Table 5-1). Consistent with the apparent unimolecularity of the G_2L_4 extension, no activity was observed for truncated forms of G_2L_4 where the loop closing the proposed G(2):C(7) bp interaction was decreased from 4 nt to 3 nt or 2 nt (Figure 5-2c). Mutations that preserved loop size resulted in similar or even enhanced activity. In contrast, a 5 nt DNA construct designed to anneal with a copy of itself and encode for CMP incorporation (5′-GGG CC) resulted in completely different kinetics. This construct displayed a sigmoidal rate dependence with concentration, resulting in a Hill coefficient very close to two and a k_{app} of 48 min^{-1} (Figure 5-3b, Coefficient = 1.99 ± 0.10, Table 5-1).

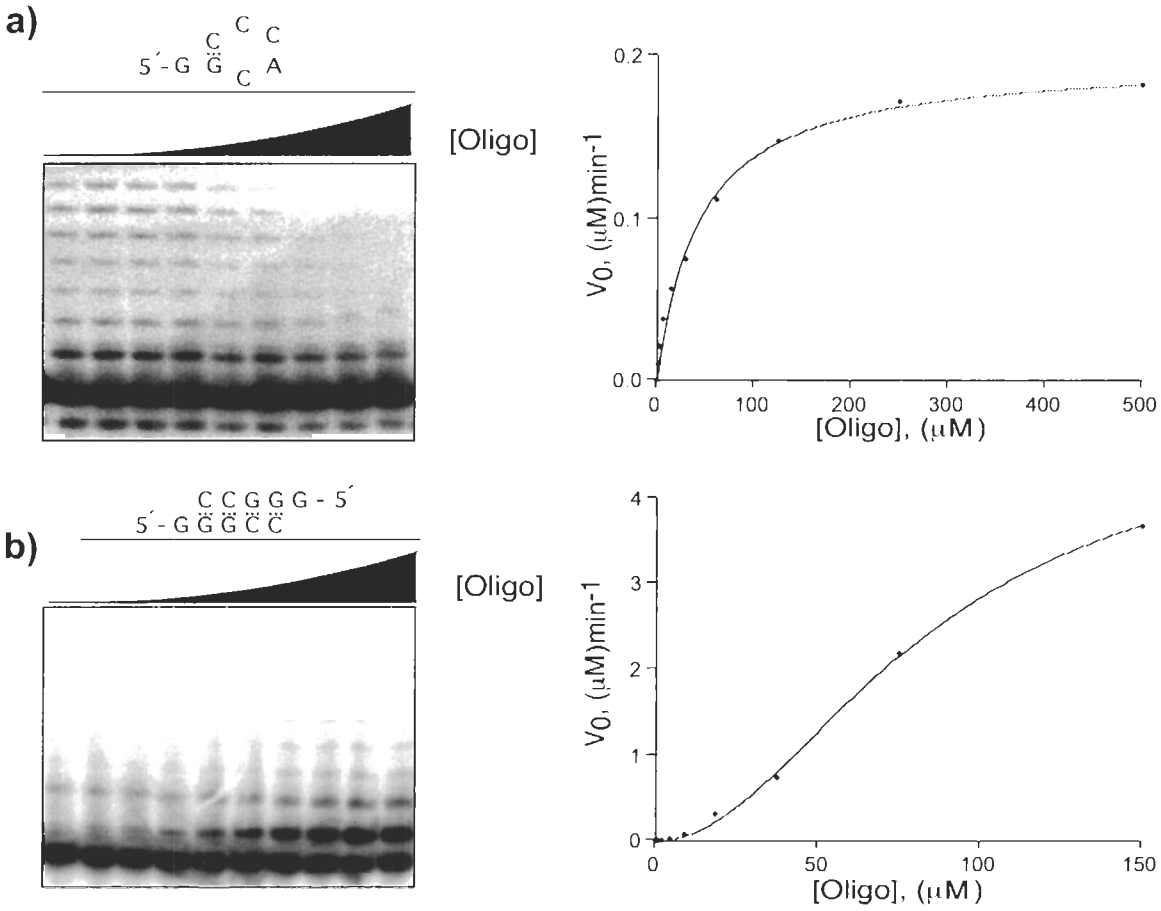


Figure 5-3: Kinetics of CTP extension for a unimolecular and bimolecular type sequence. **a)** The concentration of 5′-GGCACCC was titrated from ~2 μM to 500 μM, while keeping the total amount of radioactivity within each lane fixed. A Hill coefficient of 0.96 ± 0.10 fits well to the data. **b)** A self complimentary oligo 5′-GGGCC was also titrated from 0.3 μM to 150 μM. In this case sigmoidal kinetics were observed with a Hill coefficient of 1.99 ± 0.10 fitting the data.

5.3.4 Extension kinetics, K_m and k_{app} determination for short oligonucleotide substrates.

To explore the extension properties of substrates containing coding nucleotides embedded within longer homogenous sequence, AG_2L_4, AG_3L_4 and TG_2L_4 (see Table 5-1 for sequences) were synthesized and compared to G_2L_4. The effects of length, base-pairing and sequence composition on K_m and k_{app} for CMP incorporation were examined. Time-courses were carried out for each oligonucleotide using the standard extension assay (Figure 5-4). All of the oligonucleotides exhibited an initial rapid extension for 8-9

nt followed by a much slower but sustained extension. Total CMP incorporation (using 10 nM substrate and 0.1 μM enzyme) was linear with respect to time early in the time course. G_2L_4, AG_2L_4, AG_3L_4 and TG_2L_4 had nucleotide incorporation rates per oligonucleotide of 0.007 min^{-1}, 0.055 min^{-1}, 1.5 min^{-1} and 1.7 min^{-1} for up to 4 h, 30 min, 10 min and 5 min, respectively.

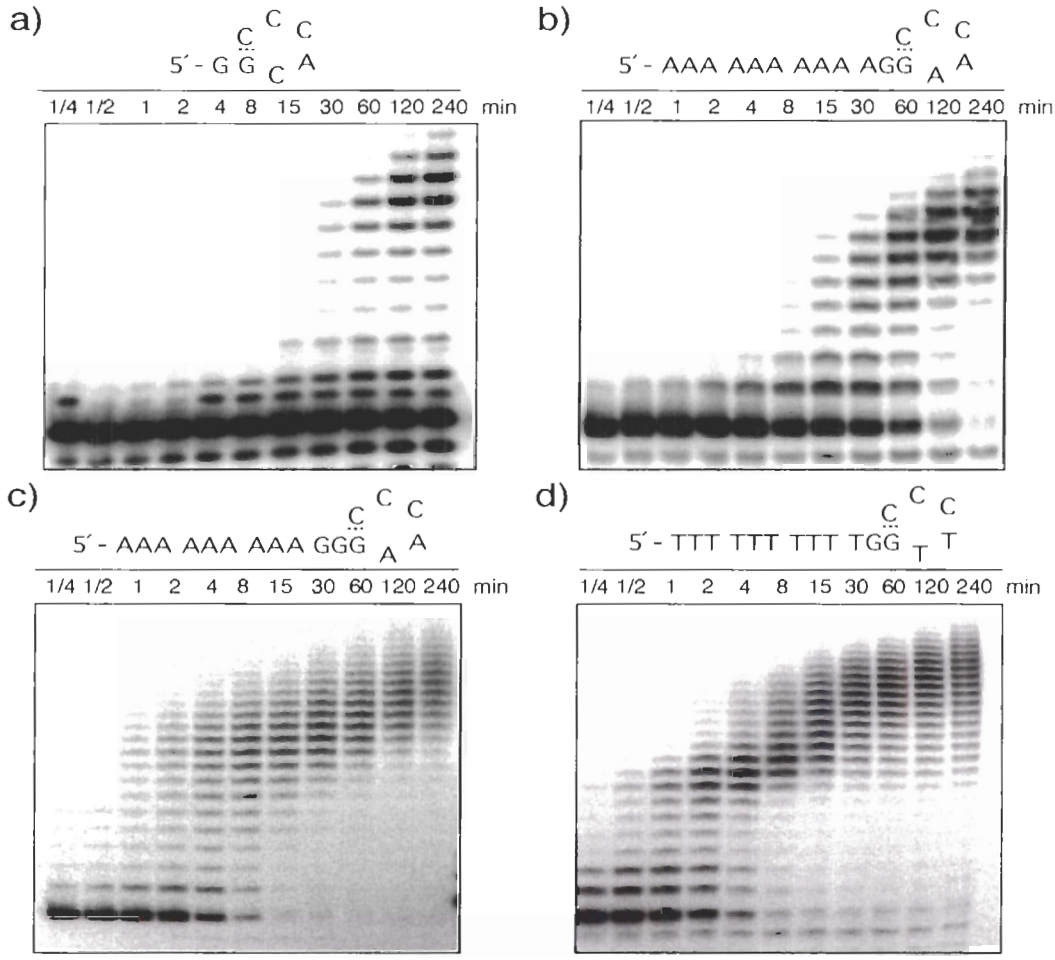


Figure 5-4: Time-courses of CTP extension reactions of a) SG_2L_4, **b)** AL_4G_2, **c)** AG_3L_4 and **d)** TG_2L_4. These sequences, which have extension rates spanning two orders of magnitude (Table 5-1), all slow their extension rate after incorporating 8-9 nucleotides. All constructs were incubated using standard conditions.

For each oligonucleotide, initial velocities were measured at template concentrations spanning 1 μM to 500 μM while keeping enzyme concentration fixed at 0.1 μM. Initial apparent velocities were determined using incubation times that were

short compared to the time required for trace amounts of oligonucleotide to behave in a non-linear fashion (G_2L_4 for 30 minutes, AG_2L_4 for 10 minutes and both AG_3L_4 and TG_2L_4 for 2 minutes). The best fit to the Michaelis-Menten equation was obtained for each construct. The k_{app} values were found to be in the range of 2 to 240 min^{-1} while the K_m values were in the 10 μM range (Table 5-1). A three-parameter fit was used to determine Hill coefficients. In all cases the observed Hill coefficients were close to one, consistent with a unimolecular extension process.

Table 5-1: Kinetic parameters determined for oligonucleotide extension.

Substrate	Sequence	K_m (μM)	k_{app} (min^{-1})	Hill Coefficient
Promoter -complex		0.016 ± 0.002	500 ± 22[a]	1.27 ± 0.13
G_2L_4	GGCACCC	43 ± 4	2.0 ± 0.1	0.96 ± 0.10
AG_2L_4	AAAAAAAAAAGGAACCC	18 ± 5	16 ± 2	1.19 ± 0.26
AG_3L_4	AAAAAAAAAGGGAACCC	41 ± 6	240 ± 10	1.18 ± 0.15
TG_2L_4	TTTTTTTTTGGTTCCC	1.8 ± 0.2	14 ± 1	0.79 ± 0.09
	GGGCC	6700 ± 2400[b]	48 ± 2	1.99 ± 0.10

[a] Initiation.
[b] units of $(\mu M)^2$.

These values were compared to the rate of initiation of the enzyme in our reaction conditions using an assay developed by Martin and Coleman (Martin & Coleman, 1987; Gardner et al., 1997). We observed a K_m of 16 nM and a k_{app} of 500 min^{-1} for a T7 promoter construct capable of run off transcription (Table 5-1). While binding affinity was comparable, our catalytic rate of 500 min^{-1} was ten times higher than observed by Martin and Coleman, likely due to our use of saturating NTP conditions and slightly different buffer conditions (Gardner et al., 1997). Extension experiments using G_2L_4 (10 nM concentration) were used to measure the binding affinity of CTP in our single nucleotide extension assay. CTP was titrated from 0.16 mM to 20 mM. A K_m(CTP) of

0.65 mM was obtained after fitting the data to Michaelis-Menten kinetics. This value is comparable to the K_m observed for GTP during initiation (Gardner et al., 1997).

5.3.5 Sequence and pairing ability weakly affects the bounds of extension.

Having established that constructs G_2L_4, AG_2L_4, AG_3L_4 and TG_2L_4 all had dramatically different extension rates as a function of sequence and pairing ability, we explored the decrease in extension rate observed after 8-9 nt of extension in all of these constructs (Figure 5-4). Seven variants of AG_2L_4 (AG_2L_l series), each having the same length and overall nucleotide composition were synthesized. The initial loop size of these variants was systematically varied from 2 to 14 nt by moving the GG sequence element in the 5′ direction along the poly(A) leader sequence. These constructs were incubated for 45 min in the presence of CTP using our standard assay. An initial loop length of 2 nt was found to be inactive (as would be expected for a unimolecular process). The remaining constructs all showed good initial extension followed by a sharp decrease in extension rate when the total loop size (initial DNA + RNA extension products) reached ~11-12 nt (Figure 5-5a). The effect of increasing the intrinsic pairing ability from one to two bp was studied by synthesizing variants of AG_3L_4 (AG_3L_l series). Even though members of the AG_3L_l series were dramatically faster than those of the AG_2L_l series (Table 5-1, Figure 5-4) after 45 minutes of incubation the relative loop sizes of the AG_3L_l series were again approximately the same with constructs having loops ~14 nt long (Figure 5-5b).

The only other nucleotide incorporated into the AG_2L_4 substrate by T7 RNAP was UMP. The enzyme quickly added more than 30 UMP nucleotides to the AG_2L_4 substrate

after only 45 minutes of incubation at a rate approximately three times as fast as CMP incorporation. When given all four NTPs, AG_2L_4 appeared to be preferentially extended by UMP (as judged by gel mobility shift patterns) at a rate only slightly slower than that observed using UTP alone.

The phosphorylation state of oligonucleotide substrates does not affect extension: Since the measurement of K_m required the use of large amounts of oligonucleotide substrate and only a trace amount of this substrate was phosphorylated we were concerned, particularly for small substrates such as G_2L_4, that the presence or absence of a 5′ phosphate was important for activity. We therefore performed experiments with unlabelled G_2L_4 and AG_3L_4 at 10 μM oligonucleotide concentration. Reactions were incubated for 15 and 30 min (AG_3L_4) or 1, 2 and 4 hours (G_2L_4) in our standard CTP extension buffer. Reactions were stopped by heat inactivating the enzyme at 65°C for five minutes, diluting 1000 fold and then radiolabelling using polynucleotide kinase and a six fold excess of [γ -^{32}P]-ATP. The extension pattern observed was very similar to our standard pre-reaction labelling strategy, indicating that the phosphorylation state of the oligonucleotide was of minor or no importance.

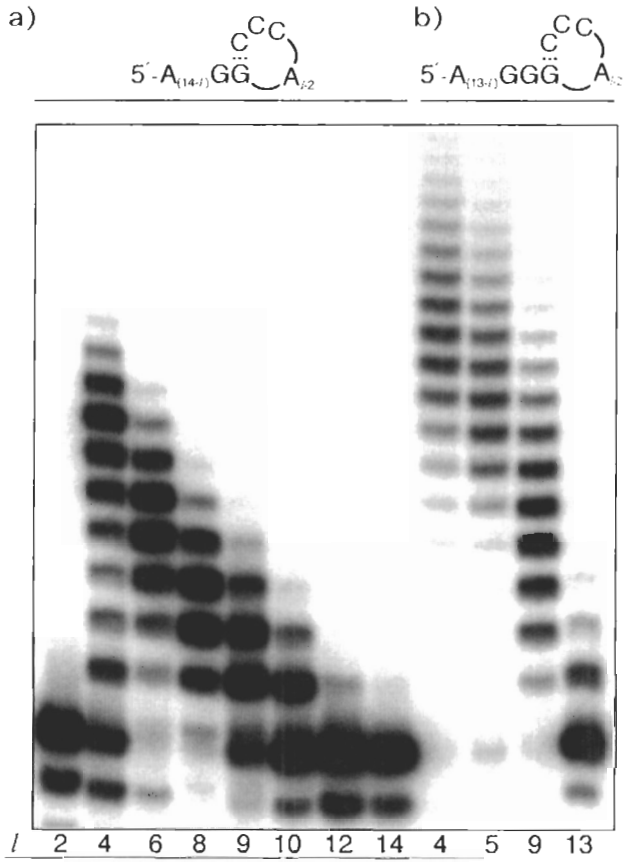


Figure 5-5: Maximum effective extension depends on loop size and transient base-pairing ability. Standard CTP extension reactions of: **a)** Substrates that can only form one base-pair, with initial loop sizes (*l*) ranging from 2 to 14 nt. **b)** Substrates that can potentially form two base-pairs, with initial loop sizes ranging from 4 to 13 nt. As the initial loop size increases, the extension rate slows notably (see also Figure 5-4), but appears independent of initial loop size. Substrates were incubated for 45 minutes using standard conditions.

5.3.6 Competitive inhibition found using inactive substrates.

The ability of the inactive substrate AG_2L_2 (5′-AAA AAA AAA AAA GGC CC, Figure 5-5a) to compete with the active construct AG_2L_4 was determined by adding it at 0 μM, 25 μM, 50 μM and 100 μM concentrations to a series of AG_2L_4 extension reactions. For each reaction AG_2L_4 was titrated from 8 μM to 500 μM (Figure 5-6a) in the presence of CTP. AG_2L_2 was found capable of inhibiting AG_2L_4 extension, and a Lineweaver-Burk plot (Figure 5-6b) suggested that this inhibition was competitive. Apparent K_m values were obtained for each inhibitor concentration and plotted against AG_2L_2

concentrations (Figure 5-6c) to yield a K_i of 45 μM. This result is similar to the K_m values observed for other oligonucleotides (Table 5-1).

5.3.7 Stable primer-template DNA complexes inhibit extension.

The 17 nt long sequence TG_2L_4, efficiently incorporated CMP and AMP. TG_2L_4 was also found to poorly incorporate UMP, at a rate at least 50 times slower than the observed rate of CMP incorporation using our single NTP extension assay. This unanticipated incorporation, possibly resulting from coding using a G:U wobble pair (see also Figure 5-1), was not observed in AG_2L_4, which however was ~30 fold slower than TG_2L_4. TG_2L_4 was incubated with either a short (SC) or a long (LC) complementary oligonucleotide, each having the potential to code for up to 3 UMP incorporations. The shorter sequence can form a 7 bp interaction with TG_2L_4 with an expected T_m of 24°C, while the longer sequence forms 17 bp of interaction and has a theoretical T_m of 43°C (Figure 5-7). In the presence of the shorter oligonucleotide, intra-moleeular CMP incorporation was not noticeably slowed even at high concentrations of SC, while UMP incorporation was observed to increase only marginally from intrinsic background levels (Figure 5-7). In contrast, the presence of stoichiometric or higher amounts of LC completely inhibited both CMP and UMP incorporation.

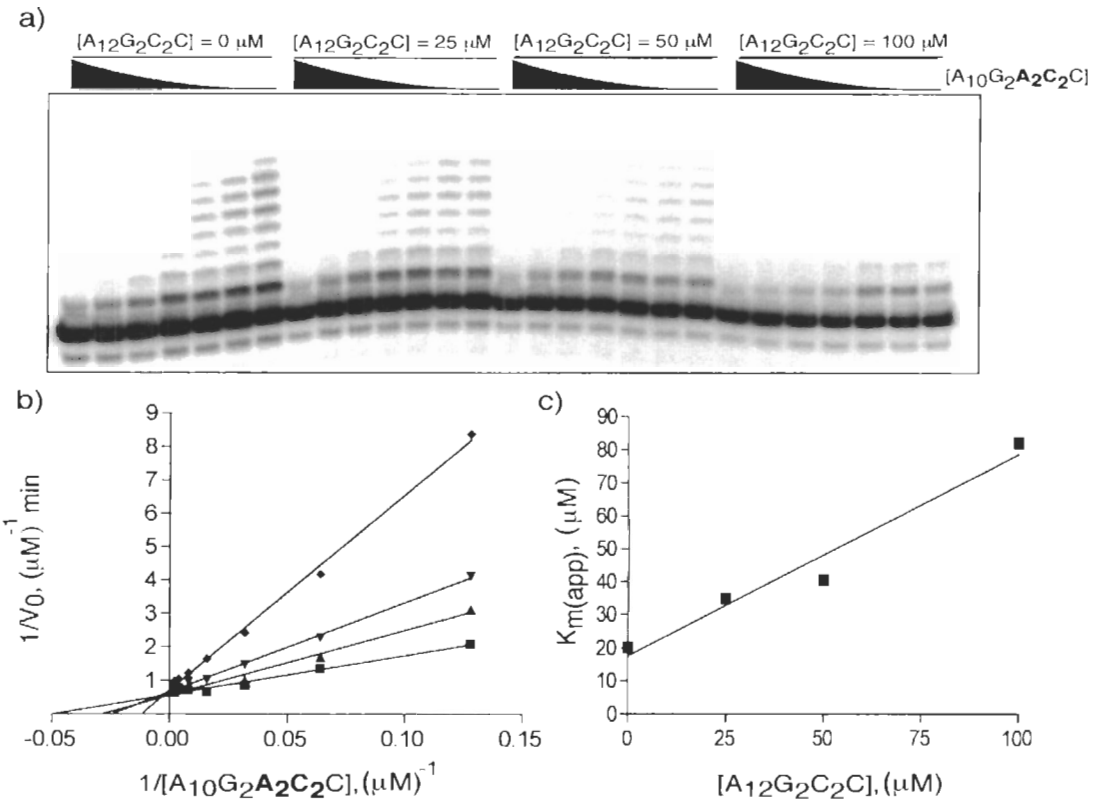


Figure 5-6: The ability of inactive AG_2L_2 to inhibit AL_4G_2 extension. a) CTP extension reactions varying AL_4G_2 concentrations from 8 µM to 500 µM. The inactive sequence AG_2L_2 was titrated from 0 µM to 100 µM. **b)** A Lineweaver-Burk plot of the inverse initial velocities against inverse AL_4G_2 concentrations. **c)** A plot of $K_{m(app)}$ against AG_2L_2 concentration, resulting in a K_i of 45 µM.

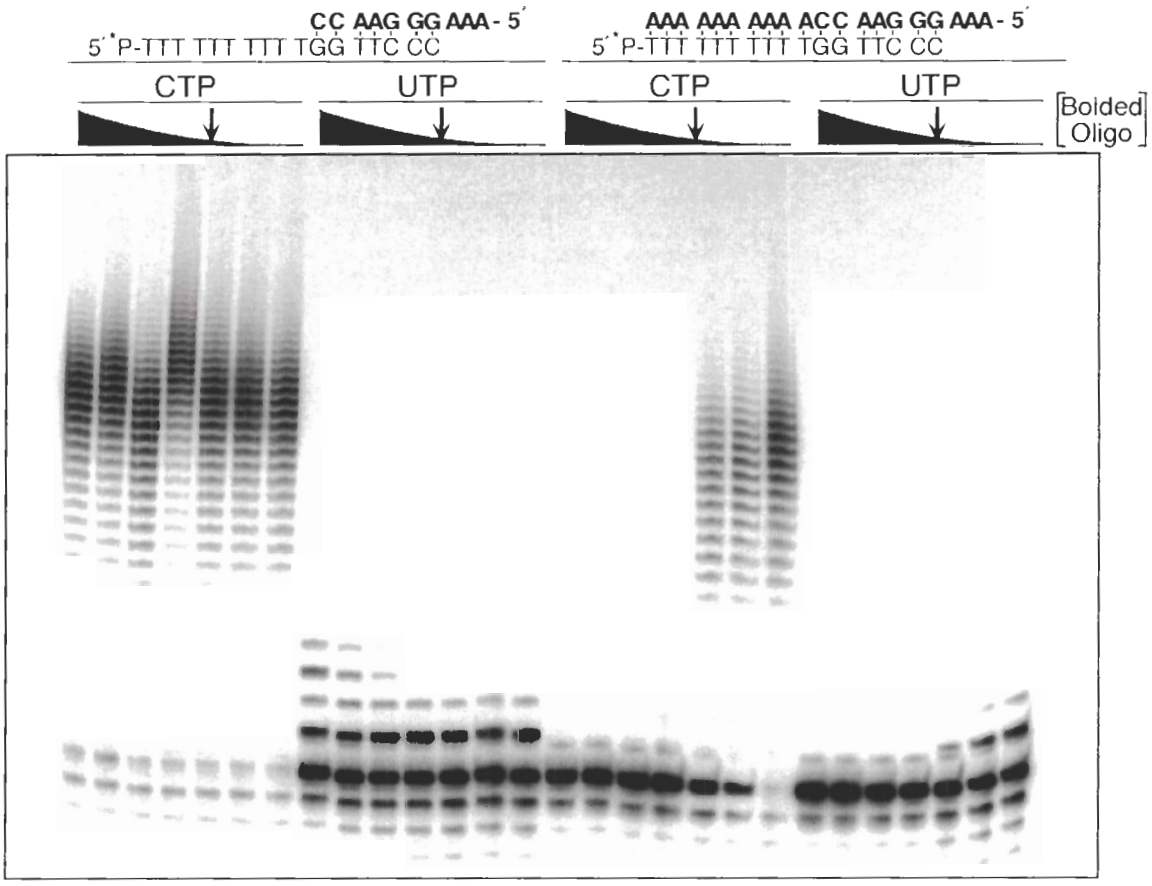


Figure 5-7: dsDNA inhibition of unimolecular extension. 5′-labelled TG_2L_4 (bottom oligonucleotide) was held at 10 nM, while SC or LC (top oligonucleotides in bold, left and right panel respectively) were titrated from 1 μM to 0 μM (1, 1/4, 1/16, 1/64, 1/256, 1/1024, and 0 μM respectively) in standard CTP or UTP extension reactions. Arrows indicate the point at which the hybridizing oligonucleotides and TG_2L_4 are at approximately equimolar concentration.

5.4 Discussion:

T7 RNAP initiates transcription of RNA from a single nucleotide triphosphate after first tightly binding its DNA promoter sequence. Prior to the formation of an elongation complex, the enzyme forms a small (~ 4 bp) bubble of DNA between the bound T7 RNAP promoter and the active site of the enzyme (Cheetham & Steitz, 1999). The polymerase synthesizes a short 10-12 nt long RNA fragment in a template dependent fashion that induces a conformational rearrangement of the protein (Yin & Steitz, 2002). This rearrangement liberates the T7 RNAP promoter sequence and creates an RNA exit

tunnel that stabilizes the larger (~7 bp) replication bubble required for efficient elongation (Cheetham & Steitz, 1999). In the absence of a promoter, we have found that short oligonucleotides are surprisingly good substrates for T7 RNAP. The observed extension rates compare well to the rate of initiation in the presence of T7 promoter sequence (Table 5-1). This, combined with the substrates respectable binding affinities (2-43 μM), indicates that the active site of the enzyme retains considerable catalytic ability in the complete absence of a T7 promoter sequence.

A notable feature of promoter independent extension is that extension products are predictable when suitable incubation conditions are used. The extension of 7 nt long DNA substrates was coded by the most 5′ residue in the sequence. Systematically varying this nucleotide resulted in extension by the complementary nucleotide in every case tested (Figure 5-2b). This extension appeared to rely only on the formation of a transient primer-template complex consisting of a single Watson-Crick base-pair (or less efficiently with a wobble pair, see Figures 5-1 and 5-7) that can form between the 3′ terminus of the oligonucleotide and the nucleotide immediately adjacent and 3′ to the coding nucleotide. Consistent with this hypothesis, mutants that disrupted pairing were inactive while a double mutant that preserved pairing was found to be active (Figure 5-2a).

This unimolecular model of extension, where active constructs must form unstable loops by folding back their 3′ ends so as to make an interaction within their own sequence, is supported by a number of other lines of evidence. The Hill coefficients for constructs predicted to be extended in a unimolecular fashion were always observed to be close to one (Table 5-1, Figure 5-3a). This effect was observed for a range of substrate

lengths and sequence compositions. Further, constructs having Hill coefficients of one tolerated the insertion of sequence into their loop region in all cases tested, but became inactive whenever the loop region became smaller than 4 nt (Figure 5-2c, 5a). This loss of activity, would be expected owing to the relatively high thermodynamic instability of such short loop structures (Woese et al., 1990). In contrast, a self-complementary oligonucleotide (5′-GGG CC), which could plausibly be extended in a bimolecular and not a unimolecular reaction, was observed to have a Hill coefficient of two; consistent with a bimolecular extension process (Table 5-1, Figure 5-3b).

For the first time we have isolated the transient extension abilities of T7 RNAP away from its processive polymerization properties. Extension was controlled by both the choice of construct sequence and of the nucleotide triphosphate in the extension reaction. For example, construct AG_2L_4, which in the presence of CTP can use G(11) as a template and G(12):C(17) to prime a single extension, must be reorganized in order to incorporate a second CMP (after translocation A(10) can not serve as a template for CMP incorporation) and was therefore used to study the ability of this complex to dissociate from or slip within the enzyme active site. In contrast substituting UTP for CTP allows both fast processive extension as well as transient extension. Indeed, given that over 30 rapid UMP incorporations were observed when using AG_2L_4 as a substrate, a complicated and irresolvable combination of processive and transient extension steps must have occurred given that the template sequence contains only 10 sequential dA residues.

Another important feature of promoter independent extension is that stable double-stranded constructs do not appear to be efficient substrates. The only

characterized bimolecular substrate 5′-GGG CC, had 4 bp of stabilization and was extended efficiently albeit only at high oligonucleotide concentrations. The construct TG_2L_4, when deliberately complexed with a short template sequence to form a relatively unstable bimolecular complex coding for up to three UMP additions (SC) was a poor substrate for UMP addition (Figure 5-7). At the same time the fast unimolecular incorporation of CMP intrinsic to TG_2L_4 was not significantly inhibited. The same construct hybridized to a longer, more stable template (LC, 17 bp of hybridization) failed to incorporate either UMP or CMP once stochiometric or higher amounts of template were added to the reaction. Taken together this data strongly suggests that the enzyme cannot deal efficiently with long stable dsDNA in the absence of a promoter sequence, but must initially recognize and extend substrates that spend a substantial fraction of their time in an unstructured, single-stranded form.

A model consistent with our data is that unimolecular substrates are first bound to the enzyme and then fold transiently into an extendable complex that allows the enzyme to incorporate a nucleotide. If after incorporation the construct is unable to support further addition, either because the end of the template has been reached (G_2L_4 type constructs), or the correct cognate NTP was not supplied (AG_2L_4 type constructs), the enzyme-substrate complex can explore two options: either the construct disassociates from the enzyme and rebinds in a later step, or it slips within the bound enzyme-substrate complex allowing the incorporation of a second nucleotide. The ability of the inactive substrate AG_2L_2 to act as a competitive inhibitor for AG_2L_4, with a K_i similar to the binding affinity of AG_2L_4, indicates that substrate release does occur during extension and favours a transient extension mechanism that includes both a dissociative and a

slippage component. The interplay between these two effects appears likely to be complicated (observe the huge difference in initial rate between AG_2L_4 and AG_2L_6 as implied by Figure 5-5a) and was not dissected further.

This model of promoter independent extension is consistent with X-ray structures that capture T7 RNAP complexes during the initiation of polymerization. Yin and Steitz have shown that prior to the formation of an elongation complex, the enzyme forms a small four nucleotide bubble of single-stranded DNA between the bound T7 RNAP promoter and the template found in the active site of the enzyme. The coding strand of this bubble is bound to an active site pocket during initiation and appears to accumulate or be 'scrunched' during early RNA synthesis (Cheetham et al., 1999; Cheetham & Steitz, 1999; Yin & Steitz, 2002; Kukarin et al., 2003). Constructs lacking the T7 promoter sequence might plausibly bind into this catalytic pocket and be able to loop back their 3′ ends and in this way self prime extension. The finite size of this pocket (6-9 nt), which has been estimated by crystallography (Cheetham & Steitz, 1999), may also explain the marked decrease in rate observed for a variety of unimolecular constructs after loop sizes of ~11-14 nt are reached and is consistent with the high initial velocity observed for constructs with small initial loop sizes of 4 to 8 nt (Figure 5-4, 5). Entropically, larger loop sizes may not be favoured within the binding pocket and unlooped structures may be preferentially stabilized, slowing the rate of extension.

The ability of T7 RNAP to extend short unimolecular substrates using only a single bp of priming may help to explain the evolution of viroid-like nucleic acid sequences capable of self-replication in the presence of the enzyme (Konarska & Sharp, 1989, 1990; Biebricher & Luce, 1996). These short RNA sequences are substantially self

complimentary and contain numerous repeats that can form a variety of highly symmetrical secondary structures. Our work, although mainly performed with DNA oligonucleotides, suggests that efficient self-replicating sequences should not form loop regions and double-stranded elements near the site of nucleotide synthesis much larger than 14 nt and 10 bp respectively (see Figure 5-5 and 7). It is interesting to note that both RNA X and RNA Y appear to respect these constraints (Konarska & Sharp, 1990) as do sequences generated using a completely template free replicative system (Biebricher & Luce, 1996). The transient nature of priming and the fact it requires as little as one bp to initiate may also help to explain the significant sequence diversity observed among known self-replicating RNA sequences and the sensitivity of their emergence to incubation conditions. The implication of DNA dependent RNA polymerases in the replication of PSTV viroids and the hepatitis delta virus together with the considerable symmetry evidence in PSTV viroid sequences (Symons) suggests that the promoter independent properties of T7 RNAP may also prove useful in understanding the emergence of viral RNA sequences.

CHAPTER 6: Conclusion

6.1 Two distinct capping ribozymes have highly convergent catalytic properties:

We have isolated a new general capping ribozyme, called 6.17, that shares extensive catalytic properties with the previously isolated capping ribozyme called Iso6. Both ribozymes accelerate the formation of 5′-5′ RNA cap structures when incubated with a variety of unblocked phosphate containing nucleophiles. In addition, both ribozymes promote cap exchange and hydrolysis reactions. The most striking similarities are the detailed kinetics that both ribozymes exhibit: 6.17 and Iso6 display similar affinities towards nucleotide triphosphates and diphosphates, while discriminating against nucleotide monophosphates. However, the maximal chemical rate of capping remains fixed, independent of the nucleotide's identity. Both ribozymes also display similar pH dependence. This extensive list of similarities is unusual for any pair of indendently selected ribozymes and is made more remarkable as both ribozymes have different secondary structures and exhibit markedly different metal ion requirements; Iso6 requires Ca^{2+}, while 6.17 prefers Mg^{2+} for capping.

Together, the catalytic similarities shared by these two ribozymes suggest that the chemistry of capping imposes a fairly rigid restriction on how newly emergent ribozymes mediate the reaction. We suggest that, like protein enzymes that catalyze similar reactions, capping is substantially different from previously characterized ribozyme reactions and most likely involves a covalent intermediate. If true, then the highly symmetric reaction of capping constrains catalysis in a way that has not been encountered by previously isolated ribozymes.

As both Iso6 and 6.17 exhibit general substrate requirements, we believe that future evolution of these capping ribozymes using selection conditions where substrate discrimination is required will allow the isolation of RNA capping ribozymes that utilize a specific nucleotide substrate. The additional sequence information required to specify this substrate discrimination would provide important clues as to the evolutionary potential of RNA.

6.2 6.17 capping ribozyme retains stereochemistry during cap exchange:

Our work using thio-substituted substrates demonstrated that the stereochemistry of the 6.17 capping ribozyme proceeds with an overall retention of stereo-configuration. This demonstration is significant as all previous ribozymes where stereochemistry has been measured, inversion is observed. Inversion of stereochemistry is normally expected in a catalytic reaction involving a single chemical step. The retention of stereochemistry observed for our capping ribozyme implies a more unusual two-step reaction. This interpretation is reinforced by the kinetics of ribozyme mediated capping. We found a rate-limiting step that does not depend on substrate choice. This would be expected for a two-step reaction that utilizes a covalent intermediate.

The finding that the 6.17 ribozyme utilizes a retaining mechanism could be explained by the chemistry of capping. Capping involves a highly symmetrical reaction where the initial nucleofuge and final nucleophile are nearly identical. Both are highly charged phosphates making a single-step mechanism difficult. A two-step mechanism, although more complicated, is much better suited to deal with the highly charged groups

involved in capping; as shown by protein capping enzymes, which also utilize this strategy.

The isolation and subsequent identification of the covalent intermediate formed by the capping ribozyme is not an easy task given that its steady state concentration is expected to be low (section 3.3.1). Consequently, efforts should be aimed at increasing its concentration as a key first step in identifying the catalytic residue responsible for its formation. Formally this could be done by either inhibiting the second step, that is the attack of the nucleophile, or by accelerating the first step, the formation of the intermediate itself. The latter is more difficult as it requires some prior knowledge of the nature of the intermediate, but could still be achieved by altering the pyrophosphate leaving group. It is possible that changing the linkage between the α and β phosphate of the ribozyme into an imido or thio-ester could increase the reaction rate of the first step to a point where the intermediate could be observed. Inhibition of the second step would require the addition of nucleotides that have modifications at their terminal phosphate, in order to decrease their nucleophilicity but at the same time still allow them to bind to the ribozyme's substrate binding pocket, inhibiting the attack of water.

Alternatively, a systematic investigation of residues contribution to catalysis using classical interference experiments or nucleotide analogue interference mapping could potentially identify catalytically critical nucleotides. Subsequent detailed investigation of these residues may well identify the nucleotide responsible for the formation of the intermediate.

6.3 Accurate and processive polymerization by an improved RNA polymerase ribozyme:

We selected and characterized a superior ribozyme polymerase. This ribozyme, called B6.61, is faster and elongates primer-templates better than its parent, the Round-18 polymerase, on all primer template complexes tested. In particular, one primer-template was extended by at least 20-nt – a record for a ribozyme polymerase.

The improvements in extension rate and overall elongation were correlated with improved fidelity. This improved fidelity may be achieved by increasing the stabilization of correctly paired incoming nucleotides with respect to their Watson-Crick partner found on the template strand. This stabilization would not only improve fidelity but would also increase extension rate by raising the local concentration of the correct nucleotide in the enzyme active site prior to phosphodiester bond formation. This hypothesis is consistent with the range of data discussed in chapter 4.

The improvements in B6.61 result from a total of 6 nucleotide modifications. These modifications are all outside of the catalytic core of the ribozyme, which is formed from the Class I ligase motif. The modifications replace interactions that were previously provided by the P2 RNA element, that was added in *trans* to the Round-18 ribozyme to form the P2 stem of the Class I ligase. This replacement is demonstrated by the finding that the B6.61 polymerase extension pattern reverts nearly exactly to that of the Round-18 polymerase upon addition of the P2 RNA.

An additional feature of this work was the selection itself. We used 3 L of emulsion to select for polymerization activity using a total pool diversity of ~9 X 10^{14}. While emulsions have been used previously to select for RNA catalysts, scaling has never approached that required to sample a library of the sort used in conventional *in vitro*

selections (i.e. ~10^{15}). Moreover the IVC technique can be used to construct either RNA pathways or systems requiring more than one RNA strand that would be extremely difficult or impossible to achieve using conventional techniques. In particular, the IVC methodology described in chapter 4 can be easily modified to isolate a helper RNA that may convert B6.61 into a truly processive multi-subunit polymerase ribozyme.

6.4 Promoter-independent extension by T7 RNA polymerase:

We characterized, for the first time, the surprisingly fast extension kinetics of short nucleic acid substrates lacking a promoter. Oligonucleotides able to form transient unimolecular looped structures closed by as little as one base-pair proved to be viable substrates. These short constructs were recognized by the polymerase with micromolar binding affinities and were utilized with rates that approach that of the enzyme's rate of initiation (catalytic rates of up to 250 min^{-1} were observed). We propose that these constructs are able to interact with the catalytic binding pocket of the enzyme that is normally used to 'scrunch' the DNA template during initiation and in this way stabilize an otherwise improbable priming event. Such transient behavior is likely to play an important and nontrivial role in the emergence of viroid like RNA sequences that exploit promoter independent features of this polymerase.

In particular, replication of RNA X (see section 1.8) by T7RNAP is likely to provide a simplified model for the study of the replication of virusoids by the host DNA-dependent RNA polymerases. The ability of these RNA sequences to hijack enzymes that normally accept double-stranded DNA substrates is poorly understood. RNA X contains an impressive level of symmetry and self-complimentary. These features enable

the molecule to adopt three separate structures; a barbell-like structure, a hairpin and a cloverleaf. By slightly altering the sequence of RNA X, these three structures could be targeted to unravel the mechanism of RNA X replication. These findings are expected to shed light into the role of RNA structure in mediating its own replication, which in turn can be used in the ultimate design of an RNA replicase ribozyme that bypasses the difficult task of strand-displacemnt.

APPENDIX:
Nucleic acid library construction using synthetic DNA constructs

This appendix is largely based on the manuscript; Zaher HS, Unrau PJ. 2005. "Nucleic acid library construction using synthetic DNA constructs". *Methods Mol. Biol.* **288**:359-378.

This appendix outlines seven synthetic and molecular biology techniques that allow the controlled synthesis of nucleic acid libraries. Specifically: 1) The high diversity chemical synthesis of point mutations. 2) The high diversity chemical synthesis of point deletions. 3) The split-bead approach for constructing point mutation or deletion libraries with limited sequence diversity. 4) Pool deprotection, gel purification, and quality control techniques. 5) Large-scale PCR amplification for the generation of high diversity double-stranded DNA libraries. 6) Type II restriction enzyme digestion techniques for the construction of long sequence libraries containing minimal fixed sequence. 7) Extension techniques for the rapid synthesis of long low diversity oligonucleotide sequences.

A.1 Introduction:

The diversity or the number of distinct sequences in a nucleic acid library can span an enormous range. High diversity libraries containing 10^{12}-10^{16} different sequences have been used to screen for nucleic acid aptamers, ribozymes and functional proteins (Ellington & Szostak, 1990; Tuerk & Gold, 1990; Bartel & Szostak, 1993; Roberts & Szostak, 1997; Keefe & Szostak, 2001). Pools with lower diversity, in the range of 10^{8}-10^{10}, are commonly used in phage display libraries (Scott & Smith, 1990;

Barbas et al., 2001). Finally, pools with very low or no sequence variation are often highly desirable, as in the recent *de novo* synthesis of a 7.5 kilobase viral genome (Cello et al., 2002). This chapter discusses techniques that allow the synthesis of nucleic acid libraries spanning this range of diversity.

When planning the construction of a nucleic acid pool it is important to consider how different synthesis strategies can affect the overall diversity of the pool. Incomplete sampling may result if the desired diversity cannot be achieved by a given method. This limitation may be unavoidable (for example, a 50-nt long random sequence pool can only be completely represented using 1.3×10^{30} different sequences in other words two million moles of nucleic acid!), but should be carefully considered before building a library.

A typical 0.2 μmole solid phase DNA synthesis cartridge contains control poured glass (CPG, Figure AI-1) beads chemically derivatized to allow the simultaneous synthesis of up to 10^{17} different DNA oligonucleotides. This large number of reactive sites makes possible the construction of high diversity pools containing extensive point mutations or deletions using chemical synthesis techniques (section A.3.1 and A.3.2). The synthesis of pools containing point mutations or deletions in limited areas of the pool is often useful and can easily be performed without changing the chemistry on a DNA synthesizer using the split-bead strategy (section A.3.3). During the split-bead synthesis, fractions of the CPG are removed and subject to independent DNA synthesis. Remixing these fractions and continuing synthesis creates sequence diversity that is limited ultimately by the number of beads (10^6-10^7 for a 0.2 μmole scale column) available in the synthesis.

Understanding the limitations of DNA phosphoramidite chemistry (Eckstein, 1991) is also critical for constructing optimal nucleic acid pools. The yield for long oligonucleotides can suffer from steric problems resulting from the pores of the CPG resin filling during synthesis. Therefore to maximize yield 1000 Å CPG should be used for oligos longer than about 50-nt and 2000 Å CPG for oligos 110-nt or longer. Smaller pore sizes can greatly inhibit final yield and should be avoided. Phosphoramidite coupling efficiency (typically 98-99%) also limits oligonucleotide length and yield. Chemical capping of the small amount of uncoupled oligonucleotide produced in each synthesis cycle is not completely efficient and consequently low levels of point deletions (typically ~0.2% per residue) are also introduced during the synthesis. Together, incomplete coupling and capping constrain the design of nucleic acid libraries by limiting both sequence diversity and oligonucleotide length.

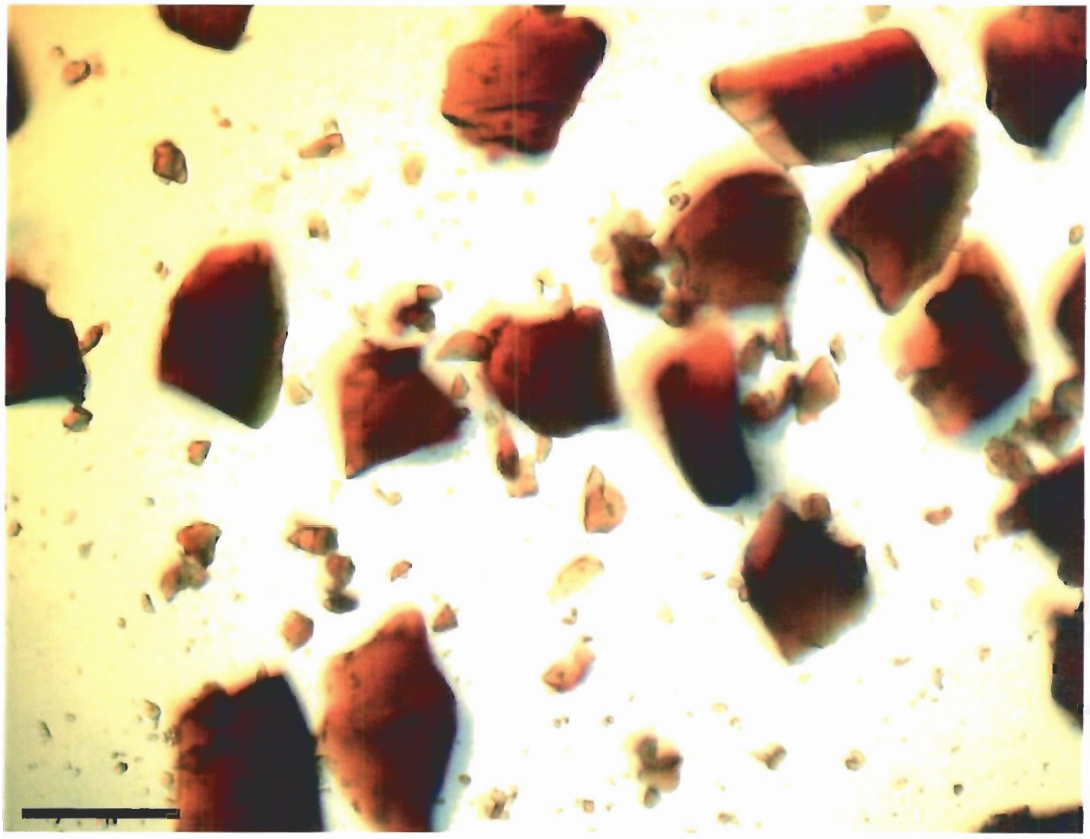


Figure A-1: 1000 Å CPG (dA) beads under the microscope. Scale bar is 20 μm.

After the chemical synthesis of the single-stranded DNA required to construct a pool, it must be deprotected, purified and the quality of the synthesis determined (section A.3.4). Long sequence high diversity pools can be constructed by first PCR amplifying (section A.3.5) short pool segments (100- to 140-nt long) that contain restriction endonuclease sites and ligating the digested fragments together using T4 DNA ligase (section A.3.6). Finally, a simple method allows the synthesis of long low diversity pools by extending partially overlapping sequences in a PCR-type reaction (section A.3.7).

A.2 Materials:

Procedures in section A.3.1, A.3.2 and A.3.3 require access to a DNA synthesizer. The synthesis of point deletions (section A.3.3) requires a programmable DNA synthesizer with a spare reagent port.

A.2.1 Point mutation.

- 0.5 g septum sealed bottles of N^6–Benzoyl-dA, N^4–Benzoyl-dC, N^2–Isobutyryl-dG, and T phosphoramidites (5'-dimethoxytrityl nucleoside 3'-(β-cyanoethyl)) (Applied Biosystems, CA. Cat 400330 to 400333).
- Acetonitrile diluent (<50 ppm H_2O), (Applied Biosystems, Cat GEN902005).
- DNA synthesis columns 1000 Å or 2000 Å CPG (Glen research, VA. Cat 20-2142-42).
- Dry inert gas source (Argon).
- Syringes with 21 gauge needles.
- Silanized glass DNA phosphoramidite synthesis bottles and rubber septums.

A.2.2 Point deletion.

- 4,4'-Dimethoxytrityl chloride (DMT-Cl) (Fluka, MO, part No. 38827).
- Anhydrous pyridine and dichloromethane.
- Deblock (10% trichloroacetic acid in dichloromethane, Applied Biosystems, Part No. 400236).

A.2.3 Split-bead.

- Glass plate, razor blades, vacuum line, and empty DNA synthesis columns.

A.2.4 Pool deprotection, purification and quality control.

- Saturated aqueous ammonia and butanol.
- 55°C heating block.
- 1.5 ml screw-capped Eppendorff tubes (Sarstedt, Germany).
- 2 × formamide loading dye (95% formamide, 5 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue).
- Polyacrylamide gel electrophoresis reagents.
- TA cloning kit (Invitrogen, CA) or other cloning kits.

A.2.5 PCR.

- 10 × PCR buffer stock: 100 mM Tris-HCl, 500 mM KCl, 15 mM $MgCl_2$, 0.1% gelatine, pH 8.3.
- 10 × dNTPs stock: 2 mM each of dATP, dCTP, dGTP, dTTP each at pH 7.5. (Amersham Pharmacia, NJ).
- *Taq* Enzyme (note 1).
- Thermocycler (PCR machine).
- Agarose gel electrophoresis reagents and equipment.
- Three temperature-controlled water baths at: 94°C, 45°C and 72°C.
- 15 ml polypropylene Falcon tubes (Becton Dickinson, NJ).
- EDTA, anhydrous ethanol, equilibrated phenol and chloroform.

- Resuspension solution: 50 mM Tris-HCl, 300 mM NaCl, pH 7.6
- 10 × TE buffer: 100 mM Tris-HCl, 10 mM EDTA, pH 7.6
- Preparative centrifuge and large centrifuge bottles.

A.2.6 inking pool segments using type II restriction enzymes.

- PCR reagents.
- *Ear* I or *Bbs* I Type II Restriction enzymes (NEB, MA).
- T4 DNA Ligase (Invitrogen, CA) and 20 mM ATP.

A.2.7 Synthetic strand extension.

- PCR reagents.

A.3 Methods:

The methods described below outline: 1) The chemical synthesis of point mutations. 2) The chemical synthesis of point deletions. 3) The split-bead approach for constructing point mutation/deletion libraries with limited sequence diversity. 4) Pool deprotection, gel purification, and quality control. 5) Large-scale PCR for the generation of high diversity libraries. 6) Type II restriction enzyme digestion techniques for the construction of long sequence libraries containing minimal fixed sequence. 7) Extension techniques for the rapid synthesis of long low diversity oligonucleotides.

A.3.1 Point mutations.

The art of making points mutations with a DNA synthesizer depends on the controlled mixing of phosphoramidite stocks in ratios that precisely control the mutagenic frequency of the nucleotide being coupled (Bartel & Szostak, 1993; Unrau & Bartel, 1998). The most common form of mutagenesis introduces mutations in an unbiased manner by mixing 'N' mix (an 'equiactive' stock containing the four phosphoramidites at concentrations such that each phosphoramidite is equally likely to couple) with a particular phosphoramidite. By mixing equiactive phosphoramidite with 'N' mix in different ratios it is possible to create a continuous spectrum of mutagenic frequencies from zero (coupling only a pure phosphoramidite) to completely random (coupling using only 'N' mix).

After designing the pool make a table summarizing the number of couplings required to synthesize the pool broken down by mutagenic phosphoramidite type.

- Calculate the total volume of each mutagenic phosphoramidite required to perform the synthesis (Check with the operator of your DNA synthesizer to determine phosphoramidite consumption per coupling at the scale you will be performing the synthesis). Allow for losses that will occur during reagent line priming on the DNA synthesizer.
- For each type of mutagenic phosphoramidite calculate the amount of equiactive and 'N' mix required and enter it into the table started in part 1. A phosphoramidite stock that produces mutations X% of the time is generated by mixing 'N' mix and the appropriate equiactive stock in the ratio 4X/(300-4X) to one (valid up to a mutational frequency of 75%, which corresponds to pure 'N'

mix). Where possible increase the amount of equiactive and 'N' to facilitate measurement with a syringe (note 2).

- Calculate the total amount of 'N' mix required for the synthesis. As the 'N' mix will be generated by mixing equal amounts of the four equiactive phosphoramidite stocks together again increase the total volume to facilitate accurate measurement of each equiactive stock.
- Calculate the total amount of each of the four equiactive stocks that are required for whole synthesis. Each stock will be used to generate 'N' mix (step 4) and mutagenic phosphoramidite mixes (step 3). Again increase volumes to facilitate measurement (note 3).
- Prepare clean and dry phosphoramidite bottles suitable for your DNA synthesizer that will contain all the mutagenic phosphoramidite solutions required for the synthesis. Include one extra bottle to hold the random 'N' mix required to generate the mutagenic phosphoramidite solutions and that can be mounted on the DNA synthesizer if pure 'N' type couplings are required during the synthesis. Wash previously used phosphoramidite bottles (which should be silanized) carefully with acetonitrile, followed by hot water. Rinse with ethanol. Dry the glassware in a drying oven (50°C overnight, ideally in vacuum) and apply rubber septums.
- Generation of equiactive phosphoramidite stocks: fill 10 ml syringes (having 21 gauge needles) with dry argon (this is easily achieved using an argon tank and regulator connected to a length of plastic tubing, insert the needle into the end of the tubing and fill the syringe with argon) and exchange this gas with anhydrous

acetonitrile from a septum sealed bottle. Carefully adjust the volume in the syringes to the following empirically determined volumes: dA 7.0 ml, dC 7.3 ml, dG 8.5 ml and T 10 ml. Inject through the septum of sealed 0.5 g phosphoramidite (cyanoethyl protected) bottles by slowly exchanging the acetonitrile in the syringe with the inert gas in the phosphoramidite bottle. Leave the syringe needle in the septum of the bottle. After letting the phosphoramidite dissolve for at least 5 minutes (making sure to dissolve the powder on the septum of the bottle) exchange the fluid in the bottle with any residual acetonitrile in the syringe to ensure a uniform concentration of equiactive phosphoramidite in each phosphoramidite bottle. If a 10 ml syringe is not required in further steps discard the syringe at this point (note 4).

- If the total volume of a particular equiactive phosphoramidite stock is larger than the volumes specified in step 7, resuspend the appropriate number of phosphoramidite bottles. To ensure uniformity mix all of the resuspended phosphoramidite together in a single septum sealed container before proceeding.
- Generation of 'N' mix: in the septum sealed anhydrous container generated in step 6, carefully mix equal volumes of the four equiactive phosphoramidite stocks together to make the desired amount of 'N' mix. Use a convenient gradation on an appropriately sized syringe to ensure equal volumes of the four equiactive phosphoramidites are added.
- Mutagenic phosphoramidite mixes: according to the ratios calculated in step 3, mix 'N' mix and appropriate equiactive phosphoramidite into the a labelled septum covered bottle being careful not to cross contaminate the syringes.

- Follow the appropriate machine specific procedures to mount phosphoramidite bottles on your DNA synthesizer. General concerns to be respected are: a) Reagents should be kept as anhydrous as possible. b) Any previous reagent should be thoroughly rinsed away by appropriate wash steps before mounting the mutagenic phosphoramidite mix on the synthesizer. Phosphoramidite solutions not mounted on the machine should be stored under argon in septum covered bottles.
- Synthesize the pool respecting the bottle numbering scheme used by your DNA synthesizer. If the number of mutagenic phosphoramidite solutions required exceeds the number of phosphoramidite ports available on your machine (remembering that four positions are required for regular unmutagenized DNA synthesis), carefully plan the minimum number of bottle changes required to perform the synthesis and change phosphoramidite bottles during the synthesis. This will require synthesizing the pool in segments. If the pool requires mutagenesis of the 3′ most residue use either a universal support (Glen Research) or mix CPG beads from unused columns using the methods outlined in section A.3.3 to generate the desired level of mutation.
- Follow the pool deprotection, gel purification, and quality control steps outlined in section A.3.4.

A.3.2 Point deletions.

This method is completely compatible with directed chemical mutagenesis (section A.3.1) and is simply implemented and optimized on a DNA synthesizer having a

spare reagent port. The protocol is an extension of a method originally developed by Treiber and Williamson (Treiber & Williamson, 1995) and allows the synthesis of high diversity libraries containing controlled point deletions with variable deletion frequencies (Chapple et al., 2003). While optimized using an 8909 Expedite DNA synthesizer (Applied Biosystems) with software controller, this method should be easily adaptable to other machines (section A.3.2.3).

- Make up 100 mM DMT-Cl (M.W. 338.8) in 95% dichloromethane and 5% pyridine. Weigh out DMT-Cl on an analytical balance into clean dry glassware that can be mounted on the spare reagent port of your DNA synthesizer. Add solvent using a clean, dry graduated cylinder under a blanket of argon gas. Mount the dissolved solution promptly on the DNA synthesizer after first washing (with acetonitrile) and drying the spare reagent delivery lines.
- A deletion coupling can be performed at any point during a synthesis by using the following modified coupling cycle (The Expedite 8909 synthesizer used for this application delivered a ~16 µl pulse of reagent in 0.36 sec. The column volume was ~70 µl, see note 5).
- Deblock the column, using the standard protocol of 10 rapid pulses of 3% trichloroacetic acid in dichloromethane (deblock) followed by 50 pulses of deblock over 49 sec. Wash with 40 pulses of acetonitrile.
- Deliver 100 mM DMT-Cl to the deblocked column (10 rapid pulses from the spare reagent port, followed by 16 pulses spread over either a 2- or 4-minutes coupling interval). Immediately wash the column (40 pulses of acetonitrile).

- Deliver deblock (7 rapid pulses) followed immediately by an extensive wash of the column (40 pulses of acetonitrile). (note 6).
- Standard coupling, capping and oxidation steps followed as usual.
- The two-minute reblocking step gave products that when cloned and sequenced (section A.3.4), were found to have an average deletion frequency of 17%. The four-minute deletion protocol yielded an average deletion frequency of 25%. Deletion frequencies were observed to be well balanced, having less than a two-fold bias with respect to sequence (Chapple et al., 2003).
- Overall deletion frequencies can most easily be estimated after the synthesis of a library by cloning and sequencing members of the pool population (section A.3.4). Deletion rates can be measured directly by synthesizing four pentanucleotide homopolymeric DNA sequences, where the third coupling cycle uses the deletion protocol being optimized. Deprotected oligonucleotides are butanol precipitated (section A.3.4) and radiolabelled with an excess of [^{32}P]-γ-ATP using polynucleotide kinase. Deletion products are resolved on a 20% polyacrylamide sequencing gel and quantified by phosphorimager.

A.3.3 Split-bead synthesis.

This procedure is used to make pools of limited sequence diversity containing point mutations, deletions or insertions. Common uses include the synthesis of degenerate codon libraries and the synthesis of sequences containing limited diversity in relatively localized regions of sequence. While the relatively small number of CPG

beads in a DNA synthesis column limits maximum sequence diversity, this disadvantage is offset by the relative ease with which simple pools can be constructed.

- Pause DNA synthesis immediately before the residue requiring modification, but after the DNA synthesizer has completed washing the column with acetonitrile in preparation for the next coupling cycle (note 7).
- Take the column off the machine and connect it to a vacuum line using appropriate connectors (usually a LUER syringe type fitting) and remove any residual acetonitrile resulting from the final column wash by applying vacuum for several minutes.
- While still applying vacuum, ensure all the CPG beads in the column are pulled against the frit on the vacuum side of the column by tapping the column gently. Carefully dismantle the column while still attached to the vacuum line.
- Detach the open column from the vacuum line and spread the dry CPG beads on a clean glass plate (note 8).
- Divide the beads using a clean razor blade into fractions of the desired size. The size of each fraction depends on the ratio of the desired modifications within the final pool. Manual division can be reasonably accurate and is much easier than weighing out bead fractions using a micro-balance (which in our hands incurs significant losses due to static electricity).
- *Point deletions or insertions* are constructed by placing the bead fraction destined to have the longer sequence length (i.e. the wild type sequence for a deletion or the mutated sequence for an insertion) into a column and coupling the desired sequence. Afterward the column is opened as in step 3 and the leftover beads on

the glass plate are sucked back into the column using the vacuum line (Figure AI-2). Synthesis is resumed until the next modification is required. It is important to realize when entering sequence data into a DNA synthesizer, that most synthesizer software assumes that the first nucleotide in the synthesis is attached to the CPG column thus, all oligonucleotide sequences used part way through a synthesis must include an extra residue at their 3′ ends in order to yield the correct final sequence.

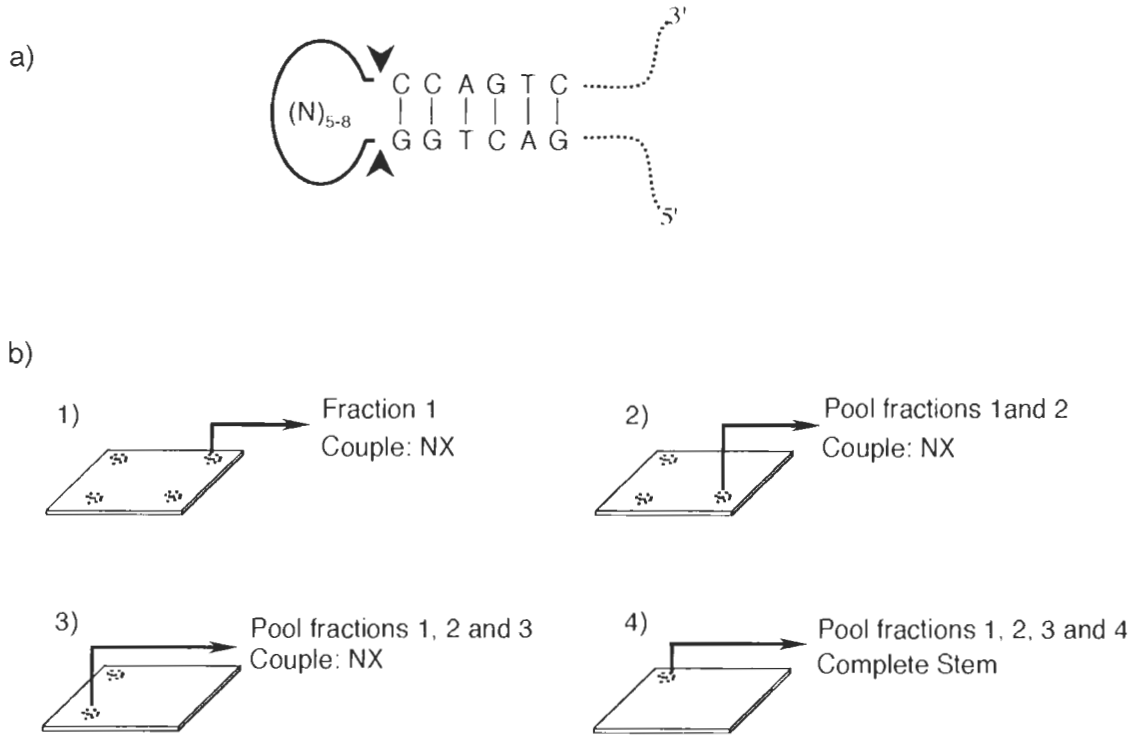


Figure A-2: a) Split-bead synthesis of a 6-bp stem loop possessing a variable 5- to 8-nt long random sequence loop delimited by the arrows. **b)** After synthesizing the first constant part of the construct and five of the eight random nucleotides using the sequence 5′-NNN NNC CAG TC… (where N is a random position synthesized by coupling 'N' mix equiactive phosphoramidite solution). The column is opened and the beads divided into four fractions. Fractions 1 to 3 are sequentially added back to the column and the sequence 5′-NX (X is an arbitrary nucleotide representing the synthesis currently on the column) is coupled after each addition. Finally the 4th fraction is added and the remainder of the sequence, 5′-…GAC TGG X is synthesized.

- ***Point mutation:*** proceed up to step 5 as described above. Place each fraction requiring a distinct coupling type into a clearly labelled DNA synthesis column. Couple the desired sequence onto each fraction using the DNA synthesizer (Figure AI-3). Random or degenerate sequences can easily be synthesized using mutagenized phosphoramidite mixes as described in section A.3.1. Once the mutation types for each fraction have been synthesized, open all of the columns and mix their beads together on a clean glass plate. Leave one column attached to the vacuum line as in step 3 and vacuum all of the beads into it. Resume synthesis until the next modification is encountered (note 9).

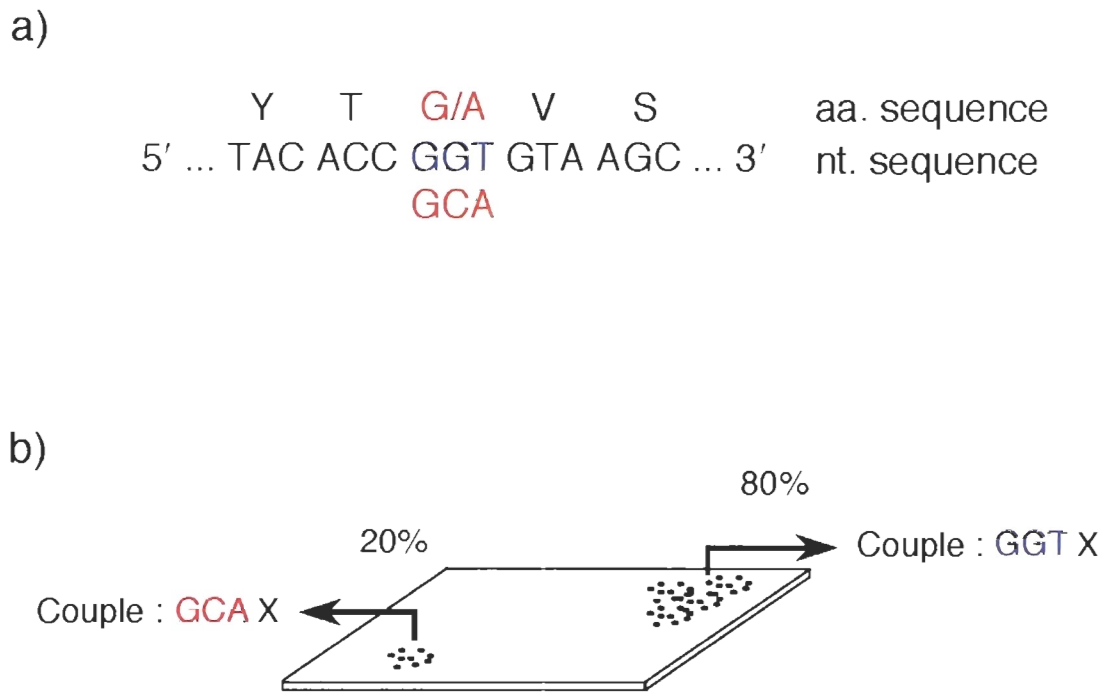


Figure A-3: Split-bead procedure for making a degenerate amino acid codon, having glycine 80% of the time and alanine the remainder. **a)** The first unmodified region, 5′-GTA AGC…, is synthesized normally. **b)** The column is opened and beads are divided into two piles in the ratio of 4:1. The larger fraction is coupled with a glycine codon, 5′-GGT X, while the smaller is coupled in a second column with an alanine codon, 5′-GCA X. The beads are then pooled together in a single column and the remainder of the sequence, 5′-…TAC ACC X, is synthesized (X is an arbitrary nucleotide representing the synthesis currently on the column).

A.3.4 Pool deprotection, purification and quality control.

After completing DNA synthesis, the resulting single-stranded DNA population must be deprotected, purified and quality accessed.

- Deprotect the synthesis by transferring the CPG beads from the synthesis column into a 1.5 ml screw-capped Eppendorff tube. Fill nearly to the top with saturated aqueous ammonia and leave at 55°C for at least 12 hours. Ensure that the tube is well sealed (This is easily done by smell). Transfer the ammonia solution into a 7.5 fold excess of butanol after first cooling it to prevent bumping from the hot ammonia gas. Vortex well and centrifuge at 12000 × g for 10 minutes. Decant the supernatant and resuspend the dried pellet in 1 ml of TE. Store the resulting crude single-stranded DNA at –20°C.
- The pool should be purified using preparative polyacrylamide gels in order to remove failed synthesis products and to confirm the general quality of the pool. This step is particularly important for long pools, as the capped material that results during a normal synthesis can inhibit the efficiency of PCR. Mix an aliquot of resuspended pool (step 1) with an equal volume of 2 × formamide loading dye. Load samples onto a preparative polyacrylamide gel of the appropriate percentage. Gel loading should not exceed 1 nmole of DNA per square millimetre of gel well surface area.
- Visualization by UV shadowing: wrap the gel on both sides with saran wrap and place a TLC plate containing 254 nm absorbing fluor under the resulting sandwich. Shine a handheld 254 nm UV light from above to produce shadows corresponding to the location of synthesis products (note 10). A uniform band of

the expected pool size with few or no discernable bands under this product should be observed. A long synthesis will produce a noticeable smear below the desired product resulting from the capping of incompletely coupled oligonucleotide sequences. The point deletion protocol (section A.3.2), if used extensively during the synthesis, will produce a noticeable broadening of the product band.

- Excise the single-stranded DNA using a clean razor blade and place the gel fragments into a five fold (by volume) excess of 300 mM NaCl. Elute on a rotator overnight at room temperature. Precipitate the nucleic acid out of the eluate by adding 2.5 volumes of ethanol and placing the resulting mixture at –20°C for at least one hour. Centrifuge at 15,000 × g for 30 minutes at 4°C. Resuspend the resulting pellet in TE and estimate concentration by OD (note 11).
- Using a small fraction of the purified DNA pool, generate double-stranded material by PCR (section A.3.5.1). Clone (using a TA cloning kit) and sequence at least 10 individual pool isolates (or more depending on the level of mutagenesis used to create the pool). Confirm that the desired pool properties are represented in these sequences before proceeding.

A.3.5 Large-scale pool amplification.

The synthesis of a high diversity double-stranded DNA pool can require hundreds of millilitres of PCR volume (Bartel & Szostak, 1993; Unrau & Bartel, 1998). Large volumes are required because the maximum concentration of single-stranded synthetic DNA template that can be effectively amplified in a PCR reaction is only in the 25-75 nM range. Moreover, synthetic single-stranded DNA does not extend as efficiently as

enzymatically produced DNA, further reducing the effective concentration of template in the PCR reaction. Taken together, these two effects imply that a pool containing 10^{15} different double-stranded sequences may require up to 200 mls of PCR volume. Reagent quality is critical when performing reactions on such scale. Pool fragments and PCR primers should be carefully designed, synthesized (sections A.3.1 though A.3.3) and gel purified (section A.3.4) before use.

A.3.5.1 Pilot PCR.

It is essential that analytical scale PCR reactions be performed before attempting amplification on a large-scale. Pilots varying template, primer, magnesium and enzyme concentrations should all be performed in order to minimize the number of PCR cycles and guarantee reagent quality. An obvious but important consideration is that pilot PCR reactions must use the same reagents to be employed in the large-scale PCR reaction.

- Program a thermocycler (PCR machine) for the following cycles. 94°C: 45 seconds (denaturing step), 50°C: 105 seconds (annealing step), 72°C: 115 seconds (extension step).
- Using the buffers and reagents prepared for the large-scale PCR reaction screen an array of conditions varying primer concentration (0.5 to 1.5 μM) and template concentration (10 to 75 nM). Add *Taq* (0.05-0.1 U/μl) enzyme to each reaction once they have reached 94°C.
- Collect aliquots from each cycle at the end of each 72°C extension step. Run aliquots on an agarose gel to see by which cycle PCR product forms (note 12).

PCR conditions that produce optimal PCR product in as few cycles as possible should be considered for large-scale PCR.

A.3.5.2 Large-scale PCR.

- Prepare three water baths set at 95°C, 72°C and 45°C (note 13).
- Mix the reagents (with the exception of *Taq* polymerase) in the ratios determined during the optimization process (3.5.1) in a sterile plastic container.
- Aliquot 10 ml fractions of the reaction mix into 15 ml polypropylene Falcon tubes and place in the 95°C water bath. Add aliquots of *Taq* to each tube after letting the fractions warm up for 10 minutes. Make sure to cap each tube tightly. Practice this step a few times using water to confirm that the caps seal well at high temperature (note 14).
- Carry out the following thermal cycles: 95°C: 4 minutes, 45°C: 5 minutes, 72°C: 7 minutes. Times have been lengthened from the pilot PCR to allow for the increased thermal mass of the large-scale reaction tubes. Invert tubes once every minute to ensure even temperatures are reached throughout the reaction volume. If different cycle temperatures are desired, modify the recommended cycle temperatures in the appropriate manner and monitor temperature changes throughout the PCR cycle using a low thermal mass thermometer (note 15).
- Carry out the number of cycles determined by the pilot PCR. After finishing the amplification, place the large-scale reaction at 4°C and check the cycle aliquots on an agarose gel. The large scale amplification should compare very well with the pilot scale reactions performed previously. If not perform another cycle of

PCR amplification may be required. Put aside an aliquot of the large scale PCR for comparison with the precipitated pool generated later (step 8).

- If amplification is satisfactory, pool the individual amplification aliquots together in an appropriately sized high speed centrifuge bottle. Chelate free magnesium by the addition of one molar equivalent of EDTA and bring the NaCl concentration to 300 mM. After mixing well add 2.5 volumes of ethanol, balancing centrifuge bottles pairwise. Shake well and place the bottles at -20°C, ideally overnight. Spin at 12,000 × g for 30 min at 4°C. DNA will be pelleted down the outside wall of the centrifuge bottle and a solid white pellet should be observed at the bottom of the centrifuge bottle. While still chilled, decant the supernatant and remove with a pipette any excess ethanol solution. Without drying the pellet, dissolve the DNA using resuspension solution making sure to wash the sides of the container. Pipette out the resuspended DNA into a clean container and repeat the resuspension process a second time to ensure full recovery. Minimize the recovery volume; concentration of the pool by at least ten fold should easily be attainable (i.e. A total of 10 ml of resuspension solution should be used to dissolve a pellet resulting from a 100 ml PCR). (note 16).
- Perform a phenol-chloroform extraction to remove the large amounts of *Taq* enzyme co-precipitated by the first precipitation. Add an equal volume of equilibrated phenol to the pool and shake vigorously. Centrifuge at 12,000 × g for 20 minutes and carefully remove the upper aqueous layer. The phenol/aqueous interface is likely to contain visible amounts of protein (note 17). Add an equal volume of chloroform to the extracted material, shake vigorously and centrifuge

(note 18). Add 2.5 volumes of ethanol and leave overnight at -20°C. The solution should become flocculent upon addition of ethanol. Centrifuge at 12,000 × g to pellet the DNA and resuspend in TE supplemented with 50 mM NaCl.

- Carefully dilute a small amount of the final pool by the ratio of the initial large scale PCR volume to the final resuspension volume. Compare with an aliquot taken from the initial large-scale PCR (step 5) on an agarose gel; recovery should be nearly quantitative.

A.3.6 Linking pool segments using type II restriction enzymes.

This technique can be used to make long high diversity dsDNA libraries that cannot be manufactured in one piece using conventional DNA synthesis. The technique utilizes the ability of some type II restriction enzymes to cut outside of their recognition sequence. The 'sticky' ends resulting from digestion with enzymes such as *Ear* I and *Bbs* I make possible the specific ligation of two high diversity pool segments through a short region of defined sequence the length of the 'sticky' ends. This approach extends a technique developed to construct long high diversity libraries using enzymes such as *Ban* I and *Sty* I that leave the restriction recognition sequence within the final pool (Figure AI-4a), (Bartel & Szostak, 1993; Unrau & Bartel, 1998).

- Design the pool in a modular fashion, taking into account that DNA synthesis yield decreases exponentially with sequence length (100 to 140-nt segments are practical). Each module of the pool should be linked to its neighbour through a 3-bp (*Ear* I) or 4-bp (*Bbs* I) nonpalindromic spacer. This spacer region allows the

specific ligation of adjacent pool fragments having complementary 'sticky' ends generated by restriction digestion of the individual fragments.

- The 'sticky' end of each fragment is generated by appending to the end of each pool module a primer binding sequence containing 10-nt of arbitrary defined sequence (for efficient PCR and restriction digestion), the appropriate restriction enzyme recognition sequence, and the desired spacer sequence. The following sequence elements are required within the primer binding sequence: 5'-$(X)_{10}$CTCTTCXNNN (*Ear* I) or 5'-$(X)_{10}$GAAGACXXNNNN (*Bbs* I). The nucleotides indicated by a 'X' are arbitrary but defined sequence elements, while the nucleotides defined by 'N' specify the spacer sequence required for ligation. Each primer binding sequence requires a partner that is used to amplify the adjacent pool fragment. In order to generate sticky ends, the partner sequence must be synthesized with the reverse complement of the original 'N' sequence (Figure A-4b).

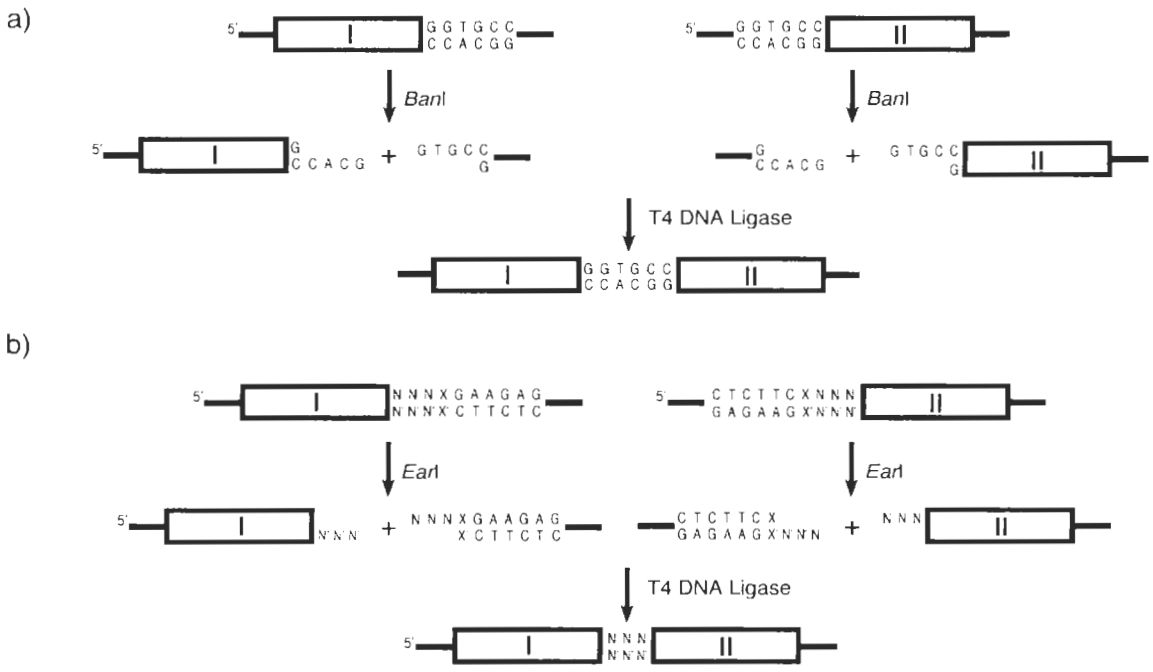


Figure A-4: Type II restriction digestion followed by ligation to generate long high diversity pools using *Ban*I or *Ear*I. Solid lines represent defined sequence elements required to anneal primers during PCR. Boxes represents regions of arbitrary sequence. **a)** Two PCR amplified DNA segments containing a *Ban*I site are digested to completion. The enzyme is heat inactivated (which denatures the cut primer sequences allowing them to reanneal with the PCR primers used in the PCR amplification) and the pool fragments are precipitated after phenol-chloroform extraction. The resulting segments are ligated together using T4 DNA ligase. The restriction site for *Ban*I is still found within the pool sequence. **b)** PCR amplified DNA segments containing the *Ear*I restriction site are digested to completion. T4 DNA ligase and ATP are then added to the reaction, allowing the ligation between pool segments having complementary 3-nt long sticky ends (sequence NNN and N′N′N′ respectively). The final pool construct cannot be cleaved by *Ear*I as it lacks a restriction site. Undesirable religation of cut primers onto pool segments is minimized by the presence of functional restriction enzyme during the ligation.

- Synthesize the segments and PCR primers required for amplification. Gel purify the resulting oligonucleotides (section A.3.4) and use large-scale PCR (section A.3.5) to make double-stranded pool segments.
- Independently digest small aliquots of the appropriate upstream and downstream fragments at high concentration (~1 μM) in the appropriate reaction buffer. NEB buffer #2 (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1mM DTT, pH 7.9) works well for both *Ear* I and *Bbs* I. When the segments are completely digested, mix the two fragments together, add 0.05 Weiss U/μl T4 DNA Ligase, and ATP

to a final concentration of 1 mM. Incubate samples overnight at 16°C. Systematically vary the relative fragment concentrations to maximize ligation yield (note 19).

- Once total ligation yield is acceptable scale up the digestion/ligation reaction by the desired amount. Phenol-chloroform extract and precipitate the pool as described in section A.3.5.7, resuspend the ligated pool in TE supplemented with 50 mM NaCl, and store at -20°C.

A.3.7 Synthetic oligonucleotide strand extension

This robust procedure can easily be used to construct long DNA strands from small partially overlapping synthetic DNA oligonucleotides. If mutations and deletions are particularly undesirable, the synthetic strands must be carefully gel purified before use.

- Design a sequence, which should be no longer than ~450 nucleotides in length. We have found that carefully gel purified DNA synthesis products contain deletions at a frequency of about 0.2%. This implies that a construct 450 nucleotides long is likely to contain a deletion mutation 60% of the time. Screening five or six isolates by cloning and sequencing should be sufficient to identify a perfect construct.
- Break the oligo into segments that overlap their neighbours by at least 16 nucleotides of sequence, such that no one segment is longer than 140 nucleotides. The two exterior sequence should be synthesized as though they were long PCR primers (5′ primer sense and 3′ primer antisense). Inner fragments should be

synthesized with overlapping 3′ ends allowing each oligo to prime the other for extension. Take care that no sequence can form extensive secondary structure either with itself or with other oligonucleotides except in the designed hybridization regions. An important consideration is the well known ability of *Taq* polymerase to add untemplated dA residues (Clark, 1988). The inner fragments therefore should be designed such that any untemplated incorporation of dA is compatible with the outer flanking sequence. Figure AI-5 summarizes these basic design principles.

- Synthesize the desired oligonucleotide segments using 2000 Å CPG if their length is 110 residues or more. After deprotection, carefully gel purify a small amount (section A.3.4) of each oligo (note 20).
- Add the two innermost oligonucleotides in equimolar amounts, at ~50 nM concentration each and perform one cycle of PCR (section A.3.5.1) adding enzyme after incubating the mixture at 94°C for one minute.
- Check extension on a polyacrylamide or agarose gel. For longer oligonucleotides, only a small amount of double-stranded material would be expected due to synthetic lesions in the starting DNA.
- Dilute the extended product 1000 fold into fresh PCR mix containing the external primer like sequences at 0.5 μmole concentration. Perform at least 10 PCR amplification cycles (section A.3.5.1) stopping when a PCR product of the desired length emerges (as determined by agarose gel).

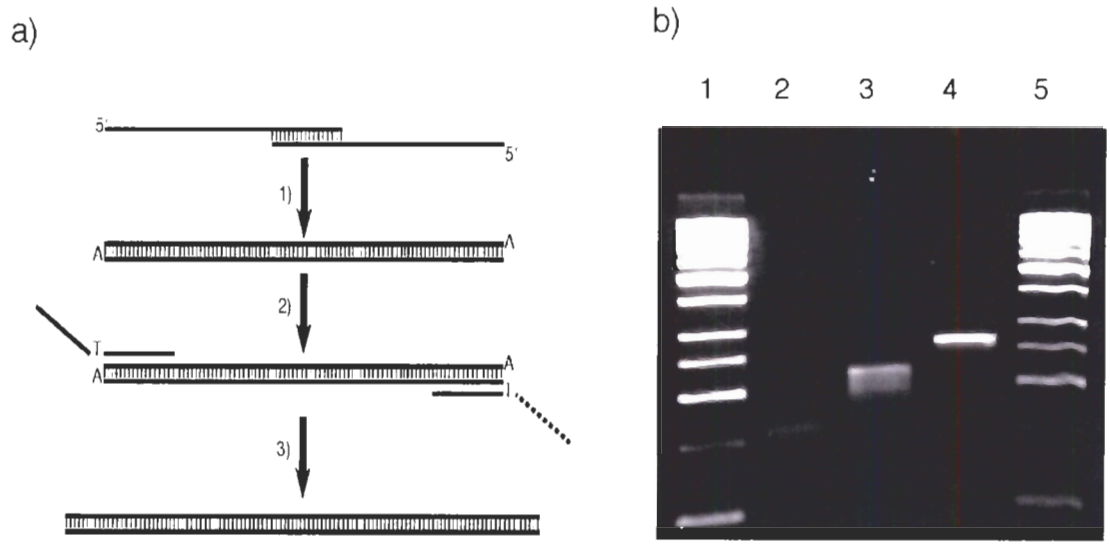


Figure A-5: Synthetic strand extension procedure. **a)** Schematic representation of the overall procedure. 1) Two long oligonucleotides with complementary sequence at their 3' ends are extended using *Taq* (possibly incorporating untemplated dA residue at their ends). 2) The DNA is diluted 1000 times and long flanking primers added. 3) PCR amplification is carried out to introduce the flanking sequence. **b)** An agarose gel showing the extension of two 118-nt oligos to make a 220-bp product, followed by extension with a single flanking PCR primer to make a 268-bp product. Lane 1 and 5, 50-bp ladder. Lane 2, oligonucleotides before extension. Lane 3, extended products, notice the smear of incomplete extension products beneath a definite band. Lane 4, PCR product with 5' flanking primer 65-nt long, increasing the final sequence length to 268-bp.

- Depending on application, cloning and subsequent sequencing may be required to isolate a construct free of mutation.

A.4 Notes:

1. Large-scale PCR consumes a considerable quantity of this enzyme. Over-expression plasmids for *Taq* are available from a number of laboratories for research purposes.
2. For example, 7 ml of a 10% mutagenized dA phosphoramidite stock, which couples a dA residue 90% of the time and equal percentages of dC, dG, and T otherwise, could be produced by mixing 6.07 ml of equiactive dA with 0.93 ml of

'N' mix. Increasing these amounts to 6.5 and 1 ml respectively is much easier to measure and will allow for minor losses that can occur during measurement.

3. Phosphoramidite reagents are cheap enough that making twice the desired volume of each phosphoramidite is often feasible, making it possible to synthesize the pool twice if required.

4. The different acetonitrile volumes used to make each equiactive mix reflects the differences in molecular weight and reactivity of each phosphoramidite and assumes that packing and reagent quality are uniform. We have found that the phosphoramidites supplied by ABI give reproducible results within a batch, but it is highly recommended that pools be cloned and sequenced after synthesis to confirm mutagenic frequencies (section A.3.4). Liquid volumes are most accurately determined by drawing a slight excess of acetonitrile into a syringe, tapping the syringe with needle pointed upwards to remove bubbles and then slowly setting the plunger to the correct volume. Work should be performed in a fume-hood.

5. Using this protocol on another DNA synthesizer will be sensitive to the length of time reagents in particular DMT-Cl and deblock are on the DNA synthesis column.

6. The reactivity of dG phosphoramidite with respect to deblocking and reblocking with DMT-Cl is about six fold higher than the other three protected phosphoramidites. This factor would lead to significant deletion sequence bias if DMT-Cl reblocking and deblocking are used independently.

7. Pausing DNA synthesis with trityl 'on' rather than 'off' is preferable as it allows observation of trityl release after the split-bead procedure. Neither choice will effect the ultimate quality of the synthesis.

8. CPG beads are easily scattered by static, which can be generated by latex gloves.

9. Empty DNA synthesis columns can be purchased from companies such as Glen Research. Alternatively, a used column can be washed thoroughly with ammonium hydroxide to remove traces of the previous synthesis and rinsed with ethanol. Dry thoroughly before use.

10. Make sure eyes and skin are well shielded from UV light. Minimize exposure of the nucleic acid to UV light.

11. There are other elution procedures, notably electroelution, that can give higher recovery of nucleic acid fragments. The soaking procedure outlined here is technically simple and allows the routine recovery of 50-60% of pool material. Ethanol precipitation will not work well for oligonucleotides shorter than about 30-nt. In which case recovery by C18 cartridge can be efficient.

12. Samples taken each cycle can be used to determine if an exponential amplification is taking place by loading 1/2 the amount of the n+1 cycle adjacent to the n cycle on an agarose gel.

13. Caution: take steps to prevent burns from the nearly boiling water in the baths. At high temperature settings, a water bath is likely to change temperature quickly upon removal of its cover.

14. Do not use polystyrene tubes as they will melt in the 95°C water bath. For large volumes of PCR, the construction of a tube holder greatly facilitates the ease with which the large-scale PCR is accomplished and can help prevent scalding.
15. Our experience shows little difference in yield when comparing pilot scale amplification with larger scale conditions. This is likely due in part to the short length of the synthetic strands being amplified.
16. When working with random pools, make sure that the DNA is always resuspended in a buffered solution containing monovalent cations to prevent denaturation. Once denatured (by resuspension in water or by application of heat) a high diversity pool will not spontaneously renature.
17. Re-extracting the aqueous/phenol region with resuspension buffer can increase final pool yield.
18. Avoid the use of polystyrene or polycarbonate plasticware during chloroform extraction as they will dissolve.
19. The restriction enzymes are deliberately kept active in the ligation reaction to recleave any cut ends that are inappropriately ligated back onto pool segments. Ligated segments will not be recognized, as correctly ligated fragments lack a restriction site. Low ligation yields can result if during the overnight ligation the restriction enzymes become inactive. This can be prevented by the addition of more enzyme. Alternatively, DNA digests can be heated to 60°C for 20 minutes to denature the short cleaved DNA fragments and encourage them to reanneal with the excess PCR primer present in the reaction before addition of ligase and ATP.

20. Increasing coupling and capping times during DNA synthesis can help to ensure high quality synthesis. Radiolabelling the oligonucleotides using polynucleotide kinase and ^{32}P γ-ATP greatly facilitates the ease in which analytical amounts of DNA can be gel purified in high purity.

Bibliography

Agresti JJ, Kelly BT, Jaschke A, Griffiths AD. 2005. Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using *in vitro* compartmentalization. *Proc. Natl. Acad. Sci. USA 102*:16170-16175.

Anastasi C, Crowe MA, Sutherland JD. 2007. Two-step potentially prebiotic synthesis of alpha-d-cytidine-5'-phosphate from d-glyceraldehyde-3-phosphate. *J. Am. Chem. Soc. 129*:24-25.

Apirion D. 1980. Genetic-mapping and some characterization of the *RNPA*49 mutation of *Escherichia-coli* that affects the RNA-processing enzyme ribonuclease-P. *genetics 94*:291-299.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science 289*:905-920.

Barbas CF, III., Scott JK, Silverman G, Burton DR. 2001. *Phage Display: A Laboratory Manual*. Plainview, NY: Cold Spring Harbor Laboratory Press.

Bartel DP, Doudna JA, Usman N, Szostak JW. 1991. Template-directed primer extension catalyzed by the *Tetrahymena* ribozyme. *Mol. Cell Biol. 11*:3390-3394.

Bartel DP, Szostak JW. 1993. Isolation of New Ribozymes From a Large Pool of Random Sequences. *Science 261*:1411-1418.

Bartel DP, Unrau PJ. 1999. Constructing an RNA world. *Trends Cell Biol. 9*:M9-M13.

Been MD, Cech TR. 1988. RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme. *Science 239*:1412-1416.

Benkovic SJ, Valentine AM, Salinas F. 2001. Replisome-mediated DNA replication. *Annu. Rev. Biochem. 70*:181-208.

Benner SA, Carrigan MA, Ricardo A, F. F. 2006. Setting the Stage: The History, Chemistry, and Geobiology behind RNA. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 1-21.

Bergman NH, Johnston WK, Bartel DP. 2000. Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry 39*:3115-3123.

Biebricher CK, Luce R. 1996. Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase. *Embo J. 15*:3458-3465.

Breslow R. 1959. On the Mechanism of the Formose Reaction. *Tetrahedron Letters*:22-26.

Buskirk AR, Kehayova PD, Landrigan A, Liu DR. 2003. *In vivo* evolution of an RNA-based transcriptional activator. *Chem. Biol. 10*:533-540.

Butlerow A. 1861. Formation synthetique d'une substabce sucree. *Compt. Rend. Acad. Sci. 53*:145-147.

Buzayan JM, Hampel A, Bruening G. 1986. Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res. 14*:9729-9743.

Cadwell RC, Joyce GF. 1994. Mutagenic PCR. *PCR-Methods and Applications 3*:S136-S140.

Carothers JM, Oestreich SC, Davis JH, Szostak JW. 2004. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc. 126*:5130-5137.

Cassano AG, Anderson VE, Harris ME. 2004. Analysis of solvent nucleophile isotope effects: Evidence for concerted mechanisms and nucleophilic activation by metal coordination in nonenzymatic and ribozyme-catalyzed phosphodiester hydrolysis. *Biochemistry 43*:10547-10559.

Cazenave C, Uhlenbeck OC. 1994. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA 91*:6972-6976.

Cech TR, and Golden B. L. 1999. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp. 708.

Cech TR, Zaug AJ, Grabowski PJ. 1981. *In vitro* splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell 27*:487-496.

Cello J, Paul AV, Wimmer E. 2002. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science 297*:1016-1018.

Chamberlain JR, Lee Y, Lane WS, Engelke DR. 1998. Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev. 12*:1678-1690.

Chapple KE, Bartel DP, Unrau PJ. 2003. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA 9*:1208-1220.

Chaput JC, Ichida JK, Szostak JW. 2003. DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc. 125*:856-857.

Chaput JC, Szostak JW. 2003. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc. 125*:9274-9275.

Cheetham GM, Jeruzalmi D, Steitz TA. 1999. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature 399*:80-83.

Cheetham GM, Steitz TA. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science 286*:2305-2309.

Chowrira BM, Berzal-Herranz A, Burke JM. 1993. Novel RNA polymerization reaction catalyzed by a group I ribozyme. *Embo J. 12*:3599-3605.

Clark JM. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res. 16*:9677-9686.

Cohen HM, Tawfik DS, Griffiths AD. 2004. Altering the sequence specificity of HaeIII methyltransferase by directed evolution using *in vitro* compartmentalization. *Protein Eng. Des. Sel. 17*:3-11.

Conn MM, Prudent JR, Schultz PG. 1996. Porphyrin metalation catalyzed by a small RNA molecule. *J. Am. Chem. Soc. 118*:7012-7013.

Crick FHC. 1968. Origin of genetic code. *J. Mol. Biol. 38*:367.

Curtis EA, Bartel DP. 2001. The hammerhead cleavage reaction in monovalent cations. *RNA 7*:546-552.

Dahm SC, Derrick WB, Uhlenbeck OC. 1993. Evidence for the role of solvated metal hydroxide in the hammerhead cleavage mechanism. *Biochemistry 32*:13040-13045.

Daros JA, Marcos JF, Hernandez C, Flores R. 1994. Replication of avocado sunblotch viroid - evidence for a symmetrical pathway with 2 rolling circles and hammerhead ribozyme processing. *Proc. Natl. Acad. Sci. USA 91*:12813-12817.

Das SR, Piccirilli JA. 2005. General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol. 1*:45-52.

Davanloo P, Rosenberg AH, Dunn JJ, Studier FW. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA 81*:2035-2039.

Decker P, Schweer H, Pohlmann R. 1982. Bioids .10. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas-chromatography gas-chromatography mass-spectrometry of normal-butoxime trifluoroacetates on ov-225. *J. Chrom. 244*:281-291.

Doherty AJ, Suh SW. 2000. Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res. 28*:4051-4058.

Doherty EA, Doudna JA. 2000. Ribozyme structures and mechanisms. *Annu. Rev. Biochem. 69*:597-615.

Doudna JA, Cech TR. 2002. The chemical repertoire of natural ribozymes. *Nature 418*:222-228.

Doudna JA, Lorsch JR. 2005. Ribozyme catalysis: not different, just worse. *Nat. Struct. Mol. Biol. 12*:395-402.

Doudna JA, Szostak JW. 1989. RNA-catalysed synthesis of complementary-strand RNA. *Nature 339*:519-522.

Doudna JA, Usman N, Szostak JW. 1993. Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry 32*:2111-2115.

Dunn JJ, Studier FW. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol. 166*:477-535.

Eckstein F. 1985. Nucleoside phosphorothioates. *Annu. Rev. Biochem. 54*:367-402.

Eckstein F. 1991. *Oligonucleotides and analogues: a practical approach*. Oxford: IRL Press.

Ekland EH, Bartel DP. 1995. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res. 23*:3231-3238.

Ekland EH, Bartel DP. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature 382*:373-376.

Ekland EH, Szostak JW, Bartel DP. 1995. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science 269*:364-370.

Ellington AD, Szostak JW. 1990. *In-vitro* selection of RNA molecules that bind specific ligands. *Nature 346*:818-822.

Eriksson M, Christensen L, Schmidt J, Haaima G, Orgel L, Nielsen PE. 1998. Sequence dependent N-terminal rearrangement and degradation of peptide nucleic acid (PNA) in aqueous solution. *New J. Chem. 22*:1055-1059.

Eschenmoser A. 1999. Chemical etiology of nucleic acid structure. *Science 284*:2118-2124.

Fedor MJ. 2002. The role of metal ions in RNA catalysis. *Curr. Opin. Struct. Biol. 12*:289-295.

Fedor MJ, Williamson JR. 2005. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol. 6*:399-412.

Ferber MJ, Maher LJ, 3rd. 1998. Combinatorial selection of a small RNA that induces amplification of IncFII plasmids in *Escherichia coli*. *J. Mol. Biol. 279*:565-576.

Ferre-D'Amare AR, Zhou KH, Doudna JA. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature 395*:567-574.

Ferris JP. 2006. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Phil. Tran. Roy. Soc. B-Biol. Sciences 361*:1777-1786.

Ferris JP, Orgel LE. 1965. Aminomalononitrile and 4-amino-5-cyanoimidazole in hydrogen cyanide polymerization and adenine synthesis. *J. Am. Chem. Soc. 87*:4976-4977.

Ferris JP, Orgel LE. 1966a. Studies on prebiotic synthesis. I. Aminomalononitrile and 4-amino-5-cyanoimidazole. *J. Am. Chem. Soc. 88*:3829-3831.

Ferris JP, Orgel LE. 1966b. An unusual photochemical rearrangement in synthesis of adenine from hydrogen cyanide. *J. Am. Chem. Soc. 88*:1074-&.

Ferris JP, Sanchez RA, Orgel LE. 1968. Studies in prebiotic synthesis. 3. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol. 33*:693-704.

Fiammengo R, Jaschke A. 2005. Nucleic acid enzymes. *Curr. Opin. Biotechnol. 16*:614-621.

Flores R, Delgado S, Gas ME, Carbonell A, Molina D, Gago S, De la Pena M. 2004. Viroids: the minimal non-coding RNAs with autonomous replication. *FEBS Lett. 567*:42-48.

Flores R, Hernandez C, de Alba AEM, Daros JA, Di Serio F. 2005. Viroids and viroid-host interactions. *Annu. Rev. Phytopathol. 43*:117-139.

Forster AC, Symons RH. 1987a. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell 49*:211-220.

Forster AC, Symons RH. 1987b. Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. *Cell 50*:9-16.

Fresco LD, Buratowski S. 1994. Active site of the mRNA-capping enzyme guanylyltransferase from Saccharomyces cerevisiae: similarity to the nucleotidyl attachment motif of DNA and RNA ligases. *Proc. Natl. Acad. Sci. USA 91*:6624-6628.

Frey PA, Richard JP, Ho HT, Brody RS, Sammons RD, Sheu KF. 1982. Stereochemistry of selected phosphotransferases and nucleotidyltransferases. *Methods Enzymol. 87*:213-235.

Fuller WD, Sanchez RA, Orgel LE. 1972. Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. *J. Mol. Biol. 67*:25-33.

Gardner LP, Mookhtiar KA, Coleman JE. 1997. Initiation, elongation, and processivity of carboxyl-terminal mutants of T7 RNA polymerase. *Biochemistry 36*:2908-2918.

Gesteland RF, Cech TR, Atkins JF. 2006. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Gilbert W. 1986. The RNA world. *Nature 319*:618.

Gold L, Polisky B, Uhlenbeck O, Yarus M. 1995. Diversity of oligonucleotide functions. *Annu. Rev. Biochem. 64*:763-797.

Gordon PM, Sontheimer EJ, Piccirilli JA. 2000. Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: Extending the parallels between the spliceosome and group II introns. *RNA 6*:199-205.

Griffiths AD, Potter BV, Eperon IC. 1987. Stereospecificity of nucleases towards phosphorothioate-substituted RNA: stereochemistry of transcription by T7 RNA polymerase. *Nucleic Acids Res. 15*:4145-4162.

Griffiths AD, Tawfik DS. 2003. Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization. *Embo J. 22*:24-35.

Griffiths AD, Tawfik DS. 2006. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol. 24*:395-402.

Gross HJ, Domdey H, Lossow C, Jank P, Raba M, Alberty H, Sanger HL. 1978. Nucleotide-sequence and secondary structure of potato spindle tuber viroid. *Nature 273*:203-208.

Grudzien E, Stepinski J, Jankowska-Anyszka M, Stolarski R, Darzynkiewicz E, Rhoads RE. 2004. Novel cap analogs for *in vitro* synthesis of mRNAs with high translational efficiency. *RNA 10*:1479-1487.

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*:849-857.

Hampel A, Cowan JA. 1997. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol. 4*:513-517.

Han J, Burke JM. 2005. Model for general acid-base catalysis by the hammerhead ribozyme: pH-activity relationships of G8 and G12 variants at the putative active site. *Biochemistry 44*:7864-7870.

Hartmann E, Hartmann RK. 2003. The enigma of ribonuclease P evolution. *Trends Genet. 19*:561-569.

Hertweck M, Mueller MW. 2001. Mapping divalent metal ion binding sites in a group II intron by Mn^{2+}- and Zn^{2+}-induced site-specific RNA cleavage. *Eur. J. Biochem. 268*:4610-4620.

Ho CK, Wang LK, Lima CD, Shuman S. 2004. Structure and mechanism of RNA ligase. *Structure (Camb) 12*:327-339.

Holley RW. 1965. Structure of an alanine transfer ribonucleic acid. *JAMA 194*:868-871.

Horton TE, DeRose VJ. 2000. Cobalt hexammine inhibition of the hammerhead ribozyme. *Biochemistry 39*:11408-11416.

Huang F, Yarus M. 1997a. 5'-RNA self-capping from guanosine diphosphate. *Biochemistry 36*:6557-6563.

Huang F, Yarus M. 1997b. A calcium-metalloribozyme with autodecapping and pyrophosphatase activities. *Biochemistry 36*:14107-14119.

Huang F, Yarus M. 1997c. Versatile 5' phosphoryl coupling of small and large molecules to an RNA. *Proc. Natl. Acad. Sci. USA 94*:8965-8969.

Huang F, Yarus M. 1998. Kinetics at a multifunctional RNA active site. *J. Mol. Biol. 284*:255-267.

Huang WH, Ferris JP. 2003. Synthesis of 35-40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Comm.*:1458-1459.

Ichetovkin IE, Abramochkin G, Shrader TE. 1997. Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J. Biol. Chem. 272*:33009-33014.

Ichida JK, Zou K, Horhota A, Yu B, McLaughlin LW, Szostak JW. 2005. An *in vitro* selection system for TNA. *J. Am. Chem. Soc. 127*:2802-2803.

Igloi GL. 1988. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry 27*:3842-3849.

Inoue T, Orgel LE. 1981. Substituent control of the poly(c)-directed oligomerization of guanosine 5'-phosphoroimidazolide. *J. Am. Chem. Soc. 103*:7666-7667.

Jencks WP. 1963. Mechanism of enzyme action. *Annu. Rev. Biochem. 32*:639-.

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. 2001. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science 292*:1319-1325.

Joyce GF. 1987. Nonenzymatic template-directed synthesis of informational macromolecules. *Cold Spring Harbor Symp. Quant. Biol. 52*:41-51.

Joyce GF. 2002. The antiquity of RNA-based evolution. *Nature 418*:214-221.

Joyce GF. 2004. Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem. 73*:791-836.

Joyce GF, Visser GM, Vanboeckel CAA, Vanboom JH, Orgel LE, Vanwestrenen J. 1984. Chiral selection in poly(c)-directed synthesis of oligo(g). *Nature 310*:602-604.

Kazantsev AV, Pace NR. 2006. Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol. 4*:729-740.

Keefe AD, Szostak JW. 2001. Functional proteins from a random-sequence library. *Nature 410*:715-718.

Kempeneers V, Vastmans K, Rozenski J, Herdewijn P. 2003. Recognition of threosyl nucleotides by DNA and RNA polymerases. *Nucleic Acids Res. 31*:6221-6226.

Kikovska E, Svard SG, Kirsebom LA. 2007. From the Cover: Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proc. Natl. Acad. Sci. USA 104*:2062-2067.

King GC, Martin CT, Pham TT, Coleman JE. 1986. Transcription by T7 RNA polymerase is not zinc-dependent and is abolished on amidomethylation of cysteine-347. *Biochemistry 25*:36-40.

Konarska MM, Sharp PA. 1989. Replication of RNA by the DNA-dependent RNA polymerase of phage T7. *Cell 57*:423-431.

Konarska MM, Sharp PA. 1990. Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase. *Cell 63*:609-618.

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena. *Cell 31*:147-157.

Krupp G. 1988. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene 72*:75-89.

Kukarin A, Rong M, McAllister WT. 2003. Exposure of T7 RNA polymerase to the isolated binding region of the promoter allows transcription from a single-stranded template. *J. Biol. Chem. 278*:2419-2424.

Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J. 1988. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol. 62*:4439-4444.

Lau MW, Cadieux KE, Unrau PJ. 2004. Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc. 126*:15686-15693.

Lawrence MS, Bartel DP. 2003. Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry 42*:8748-8755.

Lawrence MS, Bartel DP. 2005. New ligase-derived RNA polymerase ribozymes. *RNA 11*:1173-1180.

Lazcano A, Miller SL. 1996. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell 85*:793-798.

Levy M, Griswold KE, Ellington AD. 2005. Direct selection of trans-acting ligase ribozymes by *in vitro* compartmentalization. *RNA 11*:1555-1562.

Li Y, Liu Y, Breaker RR. 2000. Capping DNA with DNA. *Biochemistry 39*:3106-3114.

Li YF, Sen D. 1996. A catalytic DNA for porphyrin metallation. *Nat. Struct. Biol. 3*:743-747.

Lohrmann R. 1977. Formation of nucleoside 5'-phosphoramidates under potentially prebiological conditions. *J. Mol. Evol. 10*:137-154.

Lohrmann R, Orgel LE. 1968. Prebiotic synthesis: phosphorylation in aqueous solution. *Science 161*:64-66.

Lohrmann R, Orgel LE. 1971. Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science 171*:490-494.

Lohrmann R, Orgel LE. 1978. Preferential formation of (2'-5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron 34*:853-855.

Lohrmann R, Orgel LE. 1980. Efficient catalysis of polycytidylic acid directed oligoguanylate formation by Pb^{2+}. *J. Mol. Biol. 142*:555-567.

Lohse PA, Szostak JW. 1996. Ribozyme-catalysed amino-acid transfer reactions. *Nature 381*:442-444.

Lorsch JR, Szostak JW. 1994. *In-vitro* evolution of new ribozymes with polynucleotide kinase-activity. *Nature 371*:31-36.

Lorsch JR, Szostak JW. 1996. Chance and necessity in the selection of nucleic acid catalysts. *Accounts Chem. Res. 29*:103-110.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature 437*:376-380.

Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell 126*:309-320.

Martin CT, Coleman JE. 1987. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry 26*:2690-2696.

McConnell TS, Cech TR, Herschlag D. 1993. Guanosine binding to the *Tetrahymena* ribozyme - thermodynamic coupling with oligonucleotide binding. *Proc. Natl. Acad. Sci. USA 90*:8362-8366.

McGinness KE, Joyce GF. 2002. RNA-catalyzed RNA ligation on an external RNA template. *Chem. Biol. 9*:297-307.

McGinness KE, Joyce GF. 2003. In search of an RNA replicase ribozyme. *Chem. Biol. 10*:5-14.

McGinness KE, Wright MC, Joyce GF. 2002. Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol. 9*:585-596.

McKay DB. 1996. Structure and function of the hammerhead ribozyme: An unfinished story. *RNA 2*:395-403.

McSwiggen JA, Cech TR. 1989. Stereochemistry of RNA cleavage by the *Tetrahymena* ribozyme and evidence that the chemical step is not rate-limiting. *Science 244*:679-683.

Michel F, Ferat JL. 1995. Structure and activities of group-II introns. *Annu. Rev. Biochem. 64*:435-461.

Miller OJ, Bernath K, Agresti JJ, Amitai G, Kelly BT, Mastrobattista E, Taly V, Magdassi S, Tawfik DS, Griffiths AD. 2006. Directed evolution by *in vitro* compartmentalization. *Nat. Methods 3*:561-570.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res. 15*:8783-8798.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol. 180*:51-62.

Miyakawa S, Cleaves HJ, Miller SL. 2002a. The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph. 32*:195-208.

Miyakawa S, Cleaves HJ, Miller SL. 2002b. The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Orig. Life Evol. Biosph. 32*:209-218.

Miyakawa S, Murasawa K, Kobayashi K, Sawaoka AB. 2000. Abiotic synthesis of guanine with high-temperature plasma. *Orig. Life Evol. Biosph. 30*:557-566.

Mizuno T, Weiss AH. 1974. Synthesis and utilization of formose sugars. In: Tipson RW, Horton DH, eds. *Advances in Carbohydrate Chemistry and Biochemistry*. New York London: Academic Press. pp 173-227.

Monnard PA, Kanavarioti A, Deamer DW. 2003. Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J. Am. Chem. Soc. 125*:13734-13740.

Mueller D, Pitsch S, Kittaka A, Wagner E, Wintner CE, Eschenmoser A. 1990. Chemistry of alpha aminonitriles. Aldomerization of glycolaldehyde phosphate to racemic hexose 2,4,6triphosphates and (in presence of formaldehyde) racemic pentose 2,4-diphosphates: rac -Allose 2, 4, 6-triphosphate and racemic ribose 2,4-diphosphate are the main reaction products. *Helvetica Chimica Acta 73*:1410–1468.

Muhlbach HP, Sanger HL. 1979. Viroid replication is inhibited by alpha-amanitin. *Nature 278*:185-188.

Muller UF. 2006. Re-creating an RNA world. *Cell Mol. Life Sci. 63*:1278-1293.

Muth GW, Ortoleva-Donnelly L, Strobel SA. 2000. A single adenosine with a neutral pK(a) in the ribosomal peptidyl transferase center. *Science 289*:947-950.

Nakano S, Chadalavada DM, Bevilacqua PC. 2000. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science 287*:1493-1497.

Narlikar GJ, Herschlag D. 1997. Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem. 66*:19-59.

Nelson KE, Levy M, Miller SL. 2000. Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. USA 97*:3868-3871.

Nielsen PE, Egholm M, Berg RH, Buchardt O. 1991. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science 254*:1497-1500.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science 289*:920-930.

Noeske J, Richter C, Grundl MA, Nasiri HR, Schwalbe H, Wohnert J. 2005. An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc. Natl. Acad. Sci. USA 102*:1372-1377.

Noller HF, Hoffarth V, Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science 256*:1416-1419.

O'Rear JL, Wang SL, Feig AL, Beigelman L, Uhlenbeck OC, Herschlag D. 2001. Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations. *RNA 7*:537-545.

Oakley JL, Strothkamp RE, Sarris AH, Coleman JE. 1979. T7 RNA polymerase: promoter structure and polymerase binding. *Biochemistry 18*:528-537.

Orgel LE. 1968. Evolution of genetic apparatus. *J. Mol. Biol. 38*:381-.

Orgel LE. 2003. Some consequences of the RNA world hypothesis. *Orig. Life Evol. Biosph. 33*:211-218.

Orgel LE. 2004a. Prebiotic adenine revisited: eutectics and photochemistry. *Orig. Life Evol. Biosph. 34*:361-369.

Orgel LE. 2004b. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol. 39*:99-123.

Oro J. 1961. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. *Nature 191*:1193-1194.

Oro J, Kimball AP. 1961. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys. 94*:217-227.

Oro J, Kimball AP. 1962. Synthesis of purines under possible primitive earth conditions. II. Purine intermediates from hydrogen cyanide. *Arch. Biochem. Biophys. 96*:293-313.

Ortin J, Parra F. 2006. Structure and function of RNA replication. *Annu. Rev. Microbiol. 60*:305-326.

Padgett RA, Podar M, Boulanger SC, Perlman PS. 1994. The stereochemical course of group II intron self-splicing. *Science 266*:1685-1688.

Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. 1986. A self-splicing RNA excises an intron lariat. *Cell 44*:213-223.

Peracchi A, Beigelman L, Scott EC, Uhlenbeck OC, Herschlag D. 1997. Involvement of a specific metal ion in the transition of the hammerhead ribozyme to its catalytic conformation. *J. Biol. Chem. 272*:26822-26826.

Pleiss JA, Derrick ML, Uhlenbeck OC. 1998. T7 RNA polymerase produces 5' end heterogeneity during *in vitro* transcription from certain templates. *RNA 4*:1313-1317.

Pley HW, Flaherty KM, McKay DB. 1994. 3-dimensional structure of a hammerhead ribozyme. *Nature 372*:68-74.

Pollack SJ, Jacobs JW, Schultz PG. 1986. Selective chemical catalysis by an antibody. *Science 234*:1570-1573.

Prabahar KJ, Ferris JP. 1997. Adenine derivatives as phosphate-activating groups for the regioselective formation of 3',5'-linked oligoadenylates on montmorillonite: Possible phosphate-activating groups for the prebiotic synthesis of RNA. *J. Am. Chem. Soc. 119*:4330-4337.

Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G. 1986. Autolytic processing of dimeric plant-virus satellite RNA. *Science 231*:1577-1580.

Prudent JR, Uno T, Schultz PG. 1994. Expanding the scope of RNA catalysis. *Science 264*:1924-1927.

Pyle AM. 1993. Ribozymes - a distinct class of metalloenzymes. *Science 261*:709-714.

Pyle AM, Green JB. 1994. Building a kinetic framework for group II intron ribozyme activity: Quantitation of interdomain binding and reaction rate. *Biochemistry 33*:2716-2725.

Rajagopal J, Doudna JA, Szostak JW. 1989. Stereochemical course of catalysis by the *Tetrahymena* ribozyme. *Science 244*:692-694.

Ramirez F, Marecek JF, Szamosi J. 1980. Magnesium and calcium ion effects on hydrolysis rates of adenosine 5'-triphosphate. *J. Org. Chem. 45*:4748-4752.

Reimann R, Zubay G. 1999. Nucleoside phosphorylation: A feasible step in the prebiotic pathway to RNA. *Orig. Life Evol. Biosph. 29*:229-247.

Ricardo A, Carrigan MA, Olcott AN, Benner SA. 2004. Borate minerals stabilize ribose. *Science 303*:196.

Roberts RW, Szostak JW. 1997. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA 94*:12297-12302.

Rodnina MV, Beringer M, Wintermeyer W. 2006. Mechanism of peptide bond formation on the ribosome. *Q. Rev. Biophys. 39*:203-225.

Rupert PB, Massey AP, Sigurdsson ST, Ferre-D'Amare AR. 2002. Transition state stabilization by a catalytic RNA. *Science 298*:1421-1424.

Saladino R, Crestini C, Costanzo G, Negri R, Di Mauro E. 2001. A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3H)-pyrimidinone from formamide: implications for the origin of life. *Bioorg. Med. Chem. 9*:1249-1253.

Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW. 2006. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science 313*:1788-1792.

Salehi-Ashtiani K, Szostak JW. 2001. *In vitro* evolution suggests multiple origins for the hammerhead ribozyme. *Nature 414*:82-84.

Sanchez RA, Ferris JP, Orgel LE. 1967. Studies in prebiotic synthesis. II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol. 30*:223-253.

Sanchez RA, Ferris JP, Orgel LE. 1968. Studies in prebiotic synthesis. IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Biol. 38*:121-128.

Sanchez RA, Orgel LE. 1970. Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol. 47*:531-543.

Sastry SS, Ross BM. 1997. Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes. *J. Biol. Chem. 272*:8644-8652.

Saville BJ, Collins RA. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell 61*:685-696.

Sawai H, Orgel LE. 1975. Letter: Oligonucleotide synthesis catalyzed by the Zn^{2+} ion. *J. Am. Chem. Soc. 97*:3532-3533.

Schedl P, Primak P. 1973. Mutants of *Escherichia-coli* thermosensitive for synthesis of transfer-RNA. *Proc. Natl. Acad. Sci. USA 70*:2091-2095.

Schmeing TM, Huang KS, Strobel SA, Steitz TA. 2005. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature 438*:520-524.

Schoning KU, Scholz P, Guntha S, Wu X, Krishnamurthy R, Eschenmoser A. 2000. Chemical etiology of nucleic acid structure: The alpha-threofuranosyl-(3 '-> 2 ') oligonucleotide system. *Science 290*:1347-1351.

Scott EC, Uhlenbeck OC. 1999. A re-investigation of the thio effect at the hammerhead cleavage site. *Nucleic Acids Res. 27*:479-484.

Scott JK, Smith GP. 1990. Searching for Peptide Ligands with an Epitope Library. *Science 249*:386-390.

Seelig B, Jaschke A. 1999. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol. 6*:167-176.

Shan S, Kravchuk AV, Piccirilli JA, Herschlag D. 2001. Defining the catalytic metal ion interactions in the *Tetrahymena* ribozyme reaction. *Biochemistry 40*:5161-5171.

Shuman S, Lima CD. 2004. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr. Opin. Struct. Biol. 14*:757-764.

Sigel RKO, Vaidya A, Pyle AM. 2000. Metal ion binding sites in a group II intron core. *Nat. Struct. Biol. 7*:1111-1116.

Sleeper HL, Lohrmann R, Orgel LE. 1979. Template-directed synthesis of oligoadenylates catalyzed by Pb^{2+} ions. *J. Mol. Evol. 13*:203-214.

Sleeper HL, Orgel LE. 1979. Catalysis of nucleotide polymerization by compounds of divalent lead. *J. Mol. Evol. 12*:357-364.

Slim G, Gait MJ. 1991. Configurationally defined phosphorothioate-containing oligoribonucleotides in the study of the mechanism of cleavage of hammerhead ribozymes. *Nucleic Acids Res. 19*:1183-1188.

Smith D, Pace NR. 1993. Multiple magnesium-ions in the ribonuclease-P reaction-mechanism. *Biochemistry 32*:5273-5281.

Smith MC, Furman TC, Ingolia TD, Pidgeon C. 1988. Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem. 263*:7211-7215.

Soukup GA, Maher JJ, 3rd. 1998. Selection and characterization of RNAs that relieve transcriptional interference in *Escherichia coli*. *Nucleic Acids Res. 26*:2715-2722.

Stahley MR, Strobel SA. 2005. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science 309*:1587-1590.

Symons RH. 1997. Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res. 25*:2683-2689.

Symons, R. H. 2001. In: Soll, D., Nishimura, S., and Moore, P. B., eds. *RNA*. Elsevier Science, Oxford, U.K.

Szostak JW. 2003. Functional information: Molecular messages. *Nature 423*:689.

Szostak JW, Bartel DP, Luisi PL. 2001. Synthesizing life. *Nature 409*:387-390.

Tarasow TM, Tarasow SL, Eaton BE. 1997. RNA-catalysed carbon-carbon bond formation. *Nature 389*:54-57.

Tawfik DS, Griffiths AD. 1998. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol. 16*:652-656.

Teixeira A, Tahiri-Alaoui A, West S, Thomas B, Ramadass A, Martianov I, Dye M, James W, Proudfoot NJ, Akoulitchev A. 2004. Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature 432*:526-530.

Tramontano A, Janda KD, Lerner RA. 1986. Catalytic antibodies. *Science 234*:1566-1570.

Treiber DK, Williamson JR. 1995. A simple method for preparing pools of synthetic oligonucleotides with random point deletions. *Nucleic Acids Res. 23*:3603-3604.

Tsukiji S, Pattnaik SB, Suga H. 2004. Reduction of an aldehyde by a NADH/Zn^{2+} - dependent redox active ribozyme. *J. Am. Chem. Soc. 126*:5044-5045.

Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage-T4 DNA-polymerase. *Science 249*:505-510.

Tuschl T, Sharp PA, Bartel DP. 1998. Selection *in vitro* of novel ribozymes from a partially randomized U2 and U6 snRNA library. *Embo J. 17*:2637-2650.

Uhlenbeck OC. 1987. A small catalytic oligoribonucleotide. *Nature 328*:596-600.

Unrau PJ, Bartel DP. 1998. RNA-catalysed nucleotide synthesis. *Nature 395*:260-263.

Unrau PJ, Bartel DP. 2003. An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc. Natl. Acad. Sci. USA 100*:15393-15397.

Wadhwa R, Yaguchi T, Kaur K, Suyama E, Kawasaki H, Taira K, Kaul SC. 2004. Use of a randomized hybrid ribozyme library for identification of genes involved in muscle differentiation. *J. Biol. Chem. 279*:51622-51629.

Wang QS, Unrau PJ. 2002. Purification of histidine-tagged T4 RNA ligase from *E. coli*. *Biotechniques 33*:1256-1260.

Warnecke JM, Furste JP, Hardt WD, Erdmann VA, Hartmann RK. 1996. Ribonuclease P (RNase P) RNA is converted to a Cd^{2+}-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase p cleavage site. *Proc. Natl. Acad. Sci. USA 93*:8924-8928.

Watson JD, Crick FH. 1953a. Genetical implications of the structure of deoxyribonucleic acid. *Nature 171*:964-967.

Watson JD, Crick FH. 1953b. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature 171*:737-738.

Waugh DS, Pace NR. 1990. Complementation of an RNAse-P RNA (*RNPB*) gene deletion in *Escherichia-coli* by homologous genes from distantly related eubacteria. *J. Bacteriol. 172*:6316-6322.

Weinger JS, Parnell KM, Dorner S, Green R, Strobel SA. 2004. Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nat. Struct. Mol. Biol. 11*:1101-1106.

Wettich A, Biebricher CK. 2001. RNA species that replicate with DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry 40*:3308-3315.

White HB, 3rd. 1976. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol. 7*:101-104.

Wilson DS, Szostak JW. 1999. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem. 68*:611-647.

Winkler WC, Breaker RR. 2005. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol. 59*:487-517.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature 428*:281-286.

Woese CR. 1967. The Genetic Code. New York: Harper and Row. pp 179-195.

Woese CR, Winker S, Gutell RR. 1990. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Natl. Acad. Sci. USA 87*:8467-8471.

Yin YW, Steitz TA. 2002. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science 298*:1387-1395.

Youngman EM, Brunelle JL, Kochaniak AB, Green R. 2004. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell 117*:589-599.

Zaher HS, Unrau PJ. 2004. T7 RNA polymerase mediates fast promoter-independent extension of unstable nucleic acid complexes. *Biochemistry 43*:7873-7880.

Zaher HS, Unrau PJ. 2005. Nucleic acid library construction using synthetic DNA constructs. *Methods Mol. Biol. 288*:359-378.

Zaher HS, Unrau PJ. 2006. A general RNA-capping ribozyme retains stereochemistry during cap exchange. *J. Am. Chem. Soc. 128*:1384-900.

Zaher HS, Watkins RA, Unrau PJ. 2006. Two independently selected capping ribozymes share similar substrate requirements. *RNA 12*:1949-1958.

Zaug AJ, Cech TR. 1986. The *Tetrahymena* intervening sequence ribonucleic acid enzyme is a phosphotransferase and an acid phosphatase. *Biochemistry 25*:4478-4482.

Zimmerman JM, Maher LJ, 3rd. 2002. *In vivo* selection of spectinomycin-binding RNAs. *Nucleic Acids Res. 30*:5425-5435.

Zubay G. 1998. Studies on the lead-catalyzed synthesis of aldopentoses. *Orig. Life Evol. Biosph. 28*:13-26.

Zuker M, Mathews, D.H. and Turner, D.H. 1999. Algorithms and Thermodynamics for RNA Secondary Structure Prediction. In: Barciszewski J, Clark BFC, eds. *A Practical Guide in RNA Biochemistry and Biotechnology*: NATO ASI Series, Kluwer Academic Publishers.