Ice as a medium for RNA-catalysed RNA synthesis and evolution

James Stuart William Attwater

Medical Research Council Laboratory of Molecular Biology

> St. John's College May 2011

This dissertation is submitted for the degree of Doctor of Philosophy

Acknowledgements

I am greatly indebted to my supervisor, Dr. Philipp Holliger, for his support, ideas and enthusiasm. I also much appreciate the generous help and cooperation I received from Alan Coulson and Aniela Wochner. My gratitude extends to my colleagues Vitor Pinheiro, Chris Cozens, Alex Taylor, Pete Jones, Graeme Boocock and Claudia Baar who patiently furnished me with much useful advice, and Mike Gait, John Sutherland and David Loakes for their input on manuscripts. I also thank the MRC Laboratory of Molecular Biology and St. John's College, Cambridge, for financial support.

Declaration

This thesis does not exceed 60,000 words in length, and describes work carried out at the MRC Laboratory of Molecular Biology, Cambridge, between October 2007 and March 2011. All work described was carried out by myself, with the exception of the collaborations and contributions described below. This work has not been submitted whole or in part for any other degree.

> James Attwater May 2011

Chapter 2: Preliminary experiments testing R18 activity in ice were performed by Dr. Alan Coulson (MRC-LMB), who also generated PCR products encoding the R18 and R18i ribozymes, and a range of hybridisation templates including HybI. Dr. Aniela Wochner (MRC-LMB) demonstrated the synthesis of 5' fluorescein-labelled ribozyme for use in ribozyme degradation assays. Dr. Jennifer Ong (NEB) sourced the mineral water. Dr. Vitor Pinheiro (MRC-LMB) designed and executed the replicator simulations in (Attwater et al. 2010), and carried out MUSCLE alignment of sequenced extension products. Dr. Jernej Ule (MRC-LMB) facilitated high-throughput sequencing, and Dr. Jeremy Skepper (University of Cambridge) enabled and oversaw cryo-scanning electron microscopy.

Chapter 3: The CBT selection system as described herein was developed in collaboration with Dr. Aniela Wochner, based on earlier work by Dr. Alan Coulson. Much of the RPA screening protocol development was carried out by Dr. Aniela Wochner, based on an ELISA assay developed by Dr. Vitor Pinheiro. Dr. Richard Grenfell (MRC-LMB) and Dr. Maria Daly (MRC-LMB) provided FACS facilities and advice. Dr. Aniela Wochner performed the model selection in Figure 3.5A, all the selections and experiments outlined in Section 3.10, and the sequencing described in Figure 3.17B&C.

Ice as a medium for RNA-catalysed RNA synthesis and evolution

-James Attwater

A critical event in the origin of life is thought to have been the emergence of a molecule capable of self-replication and evolution. According to the RNA World hypothesis, this could have been an RNA polymerase ribozyme capable of generating copies of itself from simple nucleotide precursors. *In vitro* evolution experiments have provided modern examples of such ribozymes, such as the R18 RNA polymerase ribozyme, exhibiting basic levels of this crucial catalytic activity; R18's activity, however, falls far short of that required of an RNA replicase, leaving unanswered the question of whether RNA can catalyse its self-replication.

This thesis describes the development and use of a novel *in vitro* selection system, Compartmentalised Bead-Tagging (CBT), to isolate variants of the R18 ribozyme with improved sequence generality and extension capabilities. CBT evolution and engineering of polymerase ribozymes, together with RNA template evolution, allowed the synthesis of RNA molecules over 100 nucleotides long, as well as the RNA-catalysed transcription of a catalytic hammerhead ribozyme. This demonstrates the catalytic capabilities of ribozyme polymerases.

The R18 ribozyme was also exploited as an analogue of a primordial replicase, to determine replicase behaviour in different reaction environments. Substantial ribozyme polymerisation occurred at -7° C in the liquid eutectic phase of water-ice; increased ribozyme stability at these low temperatures allowed longer extension products to be generated than at ambient temperatures. The concentration effect of eutectic phase formation could also yield RNA synthesis from dilute solutions of substrates, and provide quasicellular compartmentalisation of ribozymes. These beneficial physicochemical features of ice make it a potential protocellular medium for the emergence of primordial replicases.

Ice also could serve as a medium for CBT, allowing the isolation of a polymerase ribozyme adapted to the low temperatures in the ice phase, demonstrating the primordial potential and modern feasibility of ribozyme evolution in ice.

Table of Contents

Ackn	OWLEDGEMENTS	I
DECL	ARATION	II
ICE AS	S A MEDIUM FOR RNA- CATALYSED RNA SYNTHESIS AND EVOLUTIO	N III
TABLE	E OF CONTENTS	IV
LIST	OF FIGURES	VII
LIST	of Tables	x
1 IN	NTRODUCTION	1
1.1	The RNA world	1
1.2	RNA REPLICATION	2
1.3	THE R18 RNA POLYMERASE RIBOZYME	4
1.4	R18 RIBOZYME ACTIVITY	7
2 R	IBOZYME POLYMERASE ACTIVITY IN ICE	9
2.1	INTRODUCTION	9
2.2	RIBOZYME ACTIVITY IN ICE	10
2.3	REACTION CONCENTRATION THROUGH ICE CRYSTAL GROWTH	15
2.4	RIBOZYME STABILITY	20
2.5	RIBOZYME FIDELITY IN ICE	22
2.6	COMPARTMENTALISATION IN ICE	25
2.7	DISCUSSION	33
3 D:	IRECTED EVOLUTION OF RIBOZYME RNA POLYMERASE ACTIVITY	35
3.1	INTRODUCTION	35
3.2	IMPROVED RIBOZYME POLYMERASE SELECTION SYSTEMS	35
3.3	BEAD-BASED SELECTION	37

3.4	COMPARTMENTALISED BEAD-TAGGING	38						
3.5	ROLLING CIRCLE AMPLIFICATION	41						
3.6	MODEL SELECTIONS	44						
3.7	EVOLUTION OF R18 MUTAGENISED VARIANTS	46						
3.8	ISOLATION OF IMPROVED VARIANTS							
3.9	ENGINEERING OF Z, A RIBOZYME WITH IMPROVED GENERALITY							
3.10) HARNESSING TEMPLATE RECOGNITION FOR RNA SYNTHESIS							
3.11	1 TEMPLATE EVOLUTION							
3.12	SS _{C19} -MEDIATED SYNTHESIS OF LONG RNAS	60						
3.13	FIDELITY OF SYNTHESIS OF LONG RNAS	63						
3.14	DISCUSSION	65						
4 RI	BOZYME EVOLUTION IN ICE	68						
4.1	INTRODUCTION	68						
4.2	SELECTION FOR RIBOZYME POLYMERASE ACTIVITY IN ICE	68						
4.3	EVOLUTION OF R18 MUTAGENISED VARIANTS IN ICE	71						
4.4	ISOLATION OF IMPROVED CLONES							
4.5	LOW TEMPERATURE ADAPTATION OF Y7							
4.6	LONG-RANGE RNA REPLICATION IN ICE	79						
4.7	TEMPLATE SELECTION IN ICE	82						
4.8	SYNTHESIS OF RIBOZYME SEQUENCE IN ICE	84						
4.9	DISCUSSION	85						
		06						
5 CC	DNCLUSIONS	80						
5.1	ICE AS A PROTOCELLULAR MEDIUM FOR RNA REPLICATION	86						
5.2	IN VITRO EVOLUTION OF RNA REPLICASES	87						
6 M	ATERIALS AND METHODS	00						
0 14	ATERIALS AND METHODS							
6.1	RIBOZYME POLYMERASE ASSAY	90						
	6.1.2 TEMPLATE DESIGN	90						
	6.1.3 RIBOZYME PREPARATION	91						
	6.1.4 Assay setup	91						
	6.1.5 RESOLUTION AND QUANTIFICATION OF PRIMER EXTENSION REACTIONS	92						
	6.1.6 DILUTED EXTENSION REACTIONS	94						
	6.1.7 POLYCLONAL ACTIVITY ASSAY	94						

6.2	SCANNING ELECTRON MICROSCOPY96					
6.3	RIBOZYME DEGRADATION ASSAY					
6.4	RIBOZYME DIFFUSION ASSAY					
6.5	GENERATION OF MUTAGENISED RIBOZYME LIBRARY					
6.6	COMPARTMENTALISED BEAD-TAGGING	98 98 98 98 99 99				
6.7	RECOMBINATION OF SELECTION POOLS USING STEP					
6.8	SCREENING OF RIBOZYME CLONES					
6.9	EXTENSION PRODUCT SEQUENCING6.9.1 HIGH-THROUGHPUT SEQUENCING (TABLE 2.1, FIGURE 2.9)6.9.2 LONG PRODUCT SEQUENCING (FIGURE 3.17, TABLE 4.2)					
6.10	TEMPLATE SELECTION 6.10.1 SELECTION AT 17°C 6.10.2 SELECTION IN ICE					
6.11	OLIGONUCLEOTIDE SEQUENCES					
6.12	ABBREVIATIONS					
Refer	ENCES	110				

List of Figures

FIGURE 1.1. STRUCTURE OF THE R18 RNA POLYMERASE RIBOZYME
FIGURE 2.1. R18 RETAINS ACTIVITY IN ICE11
FIGURE 2.2. TIME COURSES OF AVERAGE PRIMER EXTENSION AT 17°C AND IN ICE
FIGURE 2.3. ENHANCED EXTENSION THROUGH UPSTREAM HYBRIDISATION 14
FIGURE 2.4. SUPERCOOLING AND FREEZING DILUTED REACTIONS16
FIGURE 2.5. THE CONCENTRATION EFFECT OF ICE CRYSTAL GROWTH17
FIGURE 2.6. MAGNESIUM COUNTERION EFFECTS UPON EXTENSION
FIGURE 2.7. RIBOZYME DEGRADATION20
FIGURE 2.8. LONG EXTENSIONS IN ICE
FIGURE 2.10. ERROR SPECTRA OF R1824
FIGURE 2.11. PRINCIPLE OF THE IN-ICE DIFFUSION ASSAY
FIGURE 2.12. DIFFUSION RESTRICTION IN FROZEN DILUTE SOLUTIONS27
FIGURE 2.13. RIBOZYME DIFFUSIVITY IN ICE
FIGURE 2.14. 3D-ULTRASTRUCTURES OF EUTECTIC PHASES

FIGURE 3.1. SECONDARY STRUCTURE OF R18 LIGATED TO BEAD-BOUND
HAIRPINS
FIGURE 3.2. COMPARTMENTALISED BEAD-TAGGING
FIGURE 3.3. MINICIRCLE STRINGENCY43
FIGURE 3.4. SENSITIVITY OF ROLLING CIRCLE AMPLIFICATION
FIGURE 3.5. MODEL SELECTIONS45
FIGURE 3.6. SCREENING PRINCIPLE
FIGURE 3.7. ISOLATION OF THE IMPROVED CLONE C37
FIGURE 3.8. ENGINEERING OF Z50
FIGURE 3.9. PREDICTED SECONDARY STRUCTURES OF R18 AND Z51
FIGURE 3.10. GENERALITY OF PRIMER EXTENSION BY Z
FIGURE 3.11. TEMPLATE RECOGNITION ENHANCES EXTENSION
FIGURE 3.12. SECONDARY STRUCTURES OF C19-DERIVED RIBOZYMES55
FIGURE 3.13. GENERALITY AND HAMMERHEAD SYNTHESIS
FIGURE 3.14. SELECTION OF REPLICABLE TEMPLATES
FIGURE 3.15. SEQUENCE-TAG MEDIATED SYNTHESIS OF LONG RNAS61

FIGURE 3.16. SYNTHESIS OF LONG RNAS BY TC19Z62
FIGURE 3.17. ERROR SPECTRA OF RNA SYNTHESIS BY TC1963
FIGURE 4.1. MODEL SELECTION IN ICE
FIGURE 4.2. COMPARTMENTALISED BEAD-TAGGING IN ICE
FIGURE 4.3. ISOLATION AND ENGINEERING OF Y73
FIGURE 4.4. SECONDARY STRUCTURES OF RIBOZYMES EVOLVED AND ENGINEERED IN ICE
FIGURE 4.5. ACTIVITY OF Y AND R18 AT -7° C in ice and at 17° C75
FIGURE 4.6. PRIMER EXTENSION IN ICE BY RIBOZYMES WITH COMBINATIONS OF MUTATIONS
FIGURE 4.7. TIME COURSES OF PRIMER EXTENSION BY Y AND R1877
FIGURE 4.8. Y AND WILD TYPE ACTIVITY AT A RANGE OF TEMPERATURES78
FIGURE 4.9. SS _{C19} -MEDIATED RNA SYNTHESIS BY TC19Y80
FIGURE 4.10. SELECTION OF NOVEL TEMPLATES IN ICE
FIGURE 4.11. SYNTHESIS OF MINIZYME IN ICE

List of Tables

TABLE 2.1. RIBOZYME POLYMERASE FIDELTY. 23
TABLE 2.2. CALCULATION OF RIBOZYME DIFFUSIVITY
TABLE 3.1. PARAMETERS USED FOR SELECTION AT 17°C. 47
TABLE 4.1. PARAMETERS USED FOR SELECTION IN ICE. 71
TABLE 4.2. ACCURACY OF TC19Y-CATALYSED RNA POLYMERISATION. 81
TABLE 6.1. RELATIONSHIP BETWEEN POLYCLONAL ACTIVITY AND POOL COMPOSITION IN MODEL SELECTIONS
TABLE 6.2. Assaying polyclonal activity

1 Introduction

1.1 The RNA world

To explain the origins of life on Earth, the complexity of modern life must be reconciled with the presumed simplicity of prebiotic chemistry from which early life emerged. The power of evolution to generate new biological functions and improve existing ones could drive the increasing complexity and diversification of life from its simplest single-celled progenitors; the greatest conceptual obstacle to overcome is therefore the question of how an evolvable entity first emerged.

Existing life is founded upon two classes of informational biopolymers: proteins and nucleic acids. Proteins provide the bulk of catalytic function, carrying out replication of the DNA that encodes them. However, the co-emergence in a primordial environment of two mutually-dependent coded biopolymer systems, yielding life, is extremely improbable. A more parsimonious solution invokes the emergence of a single prebiotic polymer that was capable of both coding and catalysis (Woese 1967; Crick 1968; Orgel 1968). This would manifest both phenotype and genotype – the elementary properties of life that allow evolution – within a single entity.

This "RNA world" hypothesis (coined by (Gilbert 1986); Gesteland et al. 2006) postulates that RNA fulfilled both of these roles in early life. RNA, like DNA, can encode heritable genetic information in the sequence of nucleobases along its length, whereas the information in protein sequences cannot be recovered and copied. Crucially, however, the discovery of ribozymes (Kruger et al. 1982; Guerrier-Takada et al. 1983) demonstrated that some RNA sequences are capable of catalysis. Although a number of DNA sequences can act as catalysts (Joyce 2004; Fiammengo and Jaschke 2005), they have not exhibited the catalytic diversity of ribozymes; the 2' hydroxyl of RNA provides additional reactivity and facilitates folding into a range of diverse structures, allowing RNA to exhibit extraordinary versatility in forming specific binders and catalysts (Wilson and Szostak 1999; Ellington et al. 2009). Naturally-occurring ribozymes almost exclusively catalyse phosphoryl transfer reactions such as cleavage and ligation (Talini et al. 2009), but ribozymes exhibiting a range of activities, potentially supporting a primitive metabolism, have been

isolated from libraries of random sequence RNAs using directed evolution (Bartel and Unrau 1999).

Several features of modern biochemistry provide strong if circumstantial evidence that RNA did indeed precede DNA and proteins in early life. For example, in modern translation, RNA has both an informational role (as mRNA) and, significantly, a catalytic role: despite protein catalysis of most other cellular processes, peptide bond formation is thought to be catalysed by the RNA component of the ribosome (Nissen et al. 2000). Furthermore, coenzymes containing fragments of RNA such as NAD⁺, ATP, coenzyme A and S-adenosylmethionine are prevalent throughout modern metabolism, and could represent the cofactors or even vestiges of ribozymes that exploited these functional groups to enhance their catalytic range (Jadhav and Yarus 2002). Along with other features of modern biochemistry (such as the spliceosome and self-splicing introns (Toor et al. 2009; Valadkhan et al. 2009), riboswitches (Tucker and Breaker 2005), and RNAse P (Reiter et al. 2010)), these may represent "relics" from the RNA world (Jeffares et al. 1998), present in the breakthrough organism (Benner et al. 1989) that first developed coded protein synthesis to allow evolution of protein catalysts. The conservation of proteinaceous ribonucleotide reductases throughout all domains of life (Logan et al. 1999), along with the involvement of RNAs throughout modern translation but not DNA synthesis (Freeland et al. 1999), suggests the existence of a subsequent protein-RNA world before the emergence of DNA genomes. The stable, easily readable information storage provided by DNA yielded an organism whose descendent was the last common ancestor of all extant life.

1.2 RNA replication

How could this RNA world have emerged under the abiotic conditions on the early Earth? While many key processes remain obscure, including any potential roles of amino acids and lipids, significant progress has been made in mapping out a plausible path. A chemical synthesis of the pyrimidine nucleotides has been demonstrated under prebiotically plausible conditions (Powner et al. 2009), and selective crystallisation could act upon any enantiomeric excesses (Blackmond 2010) to yield enantiopure β -D-ribonucleotides. Nonenzymatic polymerisation of activated nucleotides into oligonucleotides (in both nontemplated and templated modes) has been performed (Monnard 2005; Ferris 2006; Schrum et al. 2009), although doing so using nucleotides activated in a prebiotically plausible manner has proven challenging (Verlander et al. 1973; Robertson and Joyce 2010). If (and at the moment, it's a big "if") such nonenzymatic sequence generation occurred with predominantly 3'-5' regiospecificity, with sequence generality and activity sufficient to outpace degradation, it would yield pools of random RNA sequences from which the first replicase ribozymes could have arisen. These first replicators would thus necessarily have been simple. The simplest possible evolvable entity would have been a heterotrophic RNA replicase ribozyme, capable of catalysing the synthesis of copies of itself from activated nucleotide precursors (Hager et al. 1996). Although cycles of self-replication and adaptation can emerge from cross-catalytic ligase ribozyme networks (Lincoln and Joyce 2009), specific substrate sequences are required that were likely absent under primordial conditions. A nucleotide polymerase ribozyme could have utilised activated monomers (and also, potentially, oligomers) to perform templated replication of RNA, analogous to that performed by proteinaceous polymerases. Such a replicase would have been central to both the emergence and evolution of ribo-organisms of an RNA world, facilitating RNA-based heredity and expression of 'RNA genes' (Muller 2006).

To support the RNA world hypothesis, RNA's capacity for self-replication must be demonstrated. However, the primordial replicase appears to have been lost. Given the widespread, almost complete replacement of RNA world ribozymes with protein enzymes, directed evolution of functional RNAs represents the only path to understanding the catalytic capabilities and behaviour of ancient ribozymes, by attempting to generate and study modern analogues (Ellington et al. 2009). *In vitro* selection applies the principles of Darwinian evolution – variation, selection, and propagation – at the molecular level, allowing the isolation of diverse, complex aptamers and catalysts from libraries of random sequences. During each round of selection, a series of biochemical treatments is applied to the library to ensure that recovery of each sequence is based upon its desired functional properties, enriching the selection pool in active sequences, and eventually allowing screening (after

sufficient rounds) to isolate improved sequences. This powerful technology has facilitated the development of a number of novel ribozyme activities (Joyce 2007).

The catalysis of RNA replication is a particularly complex process, involving sequence-independent interactions between ribozyme, template, primer and nucleotide. Although some naturally-occurring ribozymes can be engineered or evolved to catalyse limited 3'-5' bond formation (Doudna and Szostak 1989; Vicens and Cech 2009), the constellation of activities required for RNA replication – including templated nucleotide addition and template translocation – are only exhibited to a significant extent by one ribozyme: the R18 polymerase ribozyme, selected *in vitro* from random sequence RNAs (Johnston et al. 2001).

1.3 The R18 RNA polymerase ribozyme

R18 is an RNA template-dependent RNA polymerase ribozyme, capable of the successive addition of nucleotides to the 3' end of an RNA primer opposite an RNA template. The isolation of this sophisticated activity required a complex, stepwise series of directed evolution experiments. An initial groundbreaking in vitro selection study isolated ribozymes with RNA ligase activity from a pool of random-sequence RNAs (Bartel and Szostak 1993), by retrieving sequences that could ligate their 5' end to an oligonucleotide (allowing capture on an oligonucleotide affinity column and specific amplification by PCR). One family of sequences, the class I ligases, catalysed the formation of 3'-5' phosphodiester bonds between RNA oligonucleotides substrates (Ekland et al. 1995), analogous to the reaction performed by proteinaceous polymerases. b1-207, a class I ligase obtained by further evolutionary optimisation and engineering, could catalyse ligation with a $k_{cat} > 1/s$ (Ekland and Bartel 1995). An engineered variant could even catalyse limited polymerisation of nucleoside triphosphates (NTPs) to the 3' end of an RNA oligonucleotide at the ligation junction, opposite a hybridised template, with an average fidelity of 85% (Ekland and Bartel 1996).

To discover ribozymes able to polymerise NTPs opposite a separate RNA template, an additional random-sequence domain was added to the variably-mutated catalytic core of the class I ligase (Johnston et al. 2001). The resulting library was subjected to a different *in vitro* selection scheme, again based upon the principle of

4

ribozyme self-modification: ribozymes that could extend a covalently linked primer using 4-thioUTP could be purified after separation upon denaturing polyacrylamide gels poured with small amounts of N-acryloyl-aminophenylmercuric acetate (which selectively impedes migration of 4-thioU). After ten rounds of selection, 2 of 23 families of ribozymes were capable of extending primers in a template-dependent manner; a further eight rounds of selection were performed upon one isolate, followed by engineering to yield an optimised RNA polymerase ribozyme – R18.

R18 is a 189 nucleotide-long RNA molecule, consisting of two domains: the catalytic core derived from the class I ligase, and an additional 'processivity' domain allowing the core to polymerise nucleotides on a separate primer/template duplex (Figure 1.1A). The sequence of the catalytic core changed little during the polymerase selection, indicating a high degree of evolutionary optimisation (Johnston et al. 2001). Thus, although the crystal structure of the complete polymerase ribozyme has not been solved, the known crystal structure of a class I ligase variant provides a foundation for understanding polymerase structure and function (Shechner et al. 2009).

Extrapolating from this structure, a model can be built of the spatial relationship between primer/template duplex, incoming nucleotide, and polymerase core domain (Figure 1.1B). The polymerase catalytic core would consist of two coaxially-stacked domains, with the 5' single-stranded sequence making minorgroove interactions with the double-stranded part of the primer/template duplex (Figure 1.1C), positioning it relative to the active site. The ligation junction in the crystal structure corresponds to the site of nucleotide incorporation in the polymerase, where the 5' single-stranded sequence and bulged residues from the coaxial stacks form an active site. This is proposed to function like that of proteinaceous polymerases, with backbone phosphates providing potential coordination sites for two Mg²⁺ ions, to bind the triphosphate of the nucleotide and stabilise the pentacoordinate α -phosphorous transition state (Shechner et al. 2009).



Figure 1.1. Structure of the R18 RNA polymerase ribozyme.

(A) Secondary structure of R18, as deduced from analysis of covarying residues (Johnston et al. 2001) and mutation analysis (Wang et al. 2011). The catalytic and processivity domains (black) are depicted surrounding the primer A/template I duplex (orange/purple). Residues forming the catalytic centre are highlighted in pink. A short stem oligonucleotide (GGCACCA) completes the catalytic domain. (B) The crystal structure of the class I ligase ribozyme (Shechner et al. 2009), the catalytic core of R18, cropped and annotated to

highlight important features in the polymerase context. The structure depicts the bold/italicised ribozyme residues in (A), and a short stretch of primer/template duplex; an added nucleotide at the ligation junction/active site is highlighted in green. Residues are coloured as in (A); those ligase residues differing from the polymerase sequence are marked in tan. Crosslinking studies suggest that the processivity domain resides on top of the structure (Wang et al. 2011). **(C)** An A-minor triad at the 5[′] end of the ribozyme mediates sequence-general interactions between the ribozyme and primer/template duplex (Shechner et al. 2009).

However, significant features of the ligase structure would differ in the polymerase. A short 7-nucleotide 'stem' oligonucleotide completes the R18 ribozyme by partially hybridising to the linker sequence between the two domains, reconstituting a helix present in the ligase. Despite its prominent position in the ligase structure, though, omission of this stem oligonucleotide exerts only subtle effects upon primer extension by the polymerase (Zaher and Unrau 2007). Secondly, the helix in the ligase structure corresponding to the primer/template duplex is held in place as a contiguous part of the ligase molecule. The polymerase processivity domain must, therefore, compensate by correctly positioning the independent primer/template duplex. However, it is unknown how it achieves this, or whether it also interacts with the active site, or the single-stranded template downstream of the primer. Deletion and mutation analysis suggests that residues 129-163 comprise the functional region, a series of stem-loops positioned through interaction with the apical region of the core domain (residues 18-23) (Wang et al. 2011).

1.4 R18 ribozyme activity

R18 adds nucleotides to a primer in a template-dependent manner; the correct Watson-Crick base pair forming nucleotide is added opposite each template base with high fidelity – an average of 96.7% of the time, as judged by the relative efficiencies of incorporation and misincorporation (Johnston et al. 2001). However, it can add no more than 14 nucleotides over 24 hours at 17°C, before the ribozyme degrades due to the ligase core's preference for a high concentration of Mg²⁺ in the reaction buffer (0.2 M) (Glasner et al. 2002), which promotes hydrolysis of the ribozyme.

This weak activity compared to that of the ligase core arises partly from the ribozyme's low affinity for primer/template duplex ($K_m > 0.4 \text{ mM}$; (Lawrence and Bartel 2003)). The ribozyme possesses no covalent or hybridisation linkage to the primer/template duplex substrate, relying instead upon sequence-general interactions – for example, hydrogen bonding to 2' hydroxyls and adenine-mediated minor groove interactions (Muller and Bartel 2003; Shechner et al. 2009) – to keep together these two polyanionic complexes. This low affinity, forgoing sequence-specific interactions, is the price paid for the capacity to act upon a duplex in multiple registers and continue to extend primers, and for generality to allow extension upon any primer-template duplex. Nonetheless, it severely limits ribozyme processivity (Lawrence and Bartel 2003), and highlights a potential area for improvement of the ribozyme.

A critical point is that although these interactions can form with any primer/template duplex sequence, they form better with some sequences than others; processivity and duplex affinity vary depending upon duplex sequence (Lawrence and Bartel 2003), as does the effect of replacing duplex ribonucleotides with deoxyribonucleotides (Muller and Bartel 2003). Furthermore, the ribozyme adds different nucleotides with different efficiencies (Johnston et al. 2001). As a result, although the polymerase ribozyme is technically general (in that it is able to add at least one nucleotide to any primer/template duplex), it typically adds only a handful of nucleotides; extension by 14 nucleotides only occurs on a highly favourable template sequence (Lawrence and Bartel 2005).

For these reasons, R18's RNA polymerase activity is not sufficient to copy the variety of complex sequences encoding it, and falls far short of allowing contemporary RNA self-replication. Nonetheless, the nature of this activity is remarkable – it is a genuine glimpse of the activity of a true replicase. Further evolution experiments would be required to assess whether more capable ribozyme polymerases with improved replicase potential could be generated based on R18. Such experiments would allow the central tenet of the RNA world hypothesis – that RNA can catalyse its own replication – to be tested.

8

2 Ribozyme Polymerase Activity in Ice

2.1 Introduction

R18 can be viewed not just as a stepping-stone on the path to a modern RNA replicase, but also as a tool to study the behaviour of the earliest replicases. By examining the activity of R18 under different reaction conditions, inferences can be made regarding the behaviour of ancient ribozymes in primordial environments. The earliest polymerase ribozymes would necessarily have been simple molecules, perhaps lacking some of the functionalities of a fully-fledged autonomouslyreplicating ribo-organism. Reaction media that could enhance their function and compensate for their simplicity would represent likely environments for the emergence of self-replication.

One key feature of such protocellular environments would be the capacity to facilitate compartmentalisation of replicating ribozymes. Emergent replicases must have acted upon a separate encoding template molecule; however, aqueous solution would have been populated by a range of unrelated sequences competing for the replicase's activity. Unless replicases performed preferential polymerisation upon their encoding templates, any advantageous mutations arising would be dissipated through fruitless replication of unrelated sequences; likewise, detrimental mutations that emerged would accumulate in the absence of any mechanism to selectively disfavour their persistence. In short, evolution could not occur. However, if replicases and their encoding templates were together isolated from unrelated molecules through protocellular compartmentalisation, emergent beneficial phenotypes would act upon the encoding genotype alone, increasing their relative abundance. By providing a primitive form of kin selection, co-localisation in a protocellular environment allows Darwinian evolution to occur (Szostak et al. 2001), and protects replicators from emergent fast-replicating molecular parasites (Szabo et al. 2002).

Many structured environments have been proposed as protocellular media capable of providing compartmentalisation. Membraneous vesicles bear the closest resemblance to modern life, and are capable of growth and division coupled to replication of their contents (Szostak et al. 2001; Schrum et al. 2010), but their inability to tolerate high Mg²⁺ ion concentrations renders them unsuitable vessels for

ribozymes such as R18 that require high concentrations of divalent cations for activity (Monnard et al. 2002; Chen et al. 2005). Fatty acid micelles can support ribozyme activity: R18 and primers linked to cholesterol anchors congregate on micelles, promoting encounters and primer extension (Muller and Bartel 2008). However, such a linkage is of uncertain prebiotic plausibility, and ribozymes exhibited fast exchange rates between the micelles studied, limiting the compartmentalisation effect. Atmospheric aerosols (Dobson et al. 2000) and microchannels within carbonate rocks (Baaske et al. 2007; Budin et al. 2009) could serve as compartments that can also concentrate substrates through evaporation and thermophoresis respectively. These environments, though, risk exposing the ribozyme to transient increases in temperature; the 3' phosphodiester bond of RNA is susceptible to degradation through nucleophilic cleavage due to the presence of a vicinal 2' OH (Li and Breaker 1999), rendering ribozymes vulnerable to degradation at high temperatures (Pace 1991) – particularly in the presence of divalent cations required for activity. Replicase success is governed as much by ribozyme stability as by RNA synthesis rate. I thus investigated ribozyme activity in a cold environment water-ice.

2.2 Ribozyme activity in ice

When an aqueous solution of ions or other solutes is cooled below its freezing point, a biphasic system is formed, whereby solutes are excluded from the growing ice crystals and are concentrated in an interstitial liquid brine – the eutectic phase – depressing its freezing point to the incubation temperature and preventing further freezing. Ice crystal growth causes the progressive dehydration and concentration of solutes (Vajda 1999), accelerating many chemical reactions, notably the formation of RNA oligomers by nonenzymatic polymerisation of activated nucleotides (Monnard et al. 2003; Trinks et al. 2005; Monnard and Szostak 2008; Monnard and Ziock 2008). Thus, ice has the potential to provide, *in situ*, starting materials for prebiotic RNA evolution.

However, ice did not seem an auspicious medium for RNA replication. The R18 ribozyme exhibits slow polymerisation at ambient temperatures. Whilst a hairpin ribozyme variant has been reported to exhibit some vestigial ligation activity when frozen (Vlassov et al. 2004), this proceeded 2,400× slower than the best ribozyme ligases at 37°C (Vlassov et al. 2005). Low temperatures can negatively affect macromolecular catalysis in multiple ways impinging on both chemistry and catalyst, e.g. by impacting molecular motions and conformational transitions essential for catalysis and template translocation by polymerases (Tindall and Kunkel 1988). Indeed, freezing completely abolished the activity of the proteinaceous T7 RNA polymerase (Figure 2.1). However, the R18 RNA polymerase ribozyme retained substantial RNA polymerase activity in frozen reactions at -7° C (Figure 2.1). Further cooling (to -25° C) of the frozen reaction mixture towards the eutectic point abolished activity, indicating that a liquid brine phase is required for polymerase activity.

	T7 RNA polymerase		R18 ribozyme RNA polymerase				
T (°C):	37 1	–7 F 1	17	17 8	-7 F	-7 F -	-25 F
r (days).							

Figure 2.1. R18 retains activity in ice.

Denaturing PAGE of fluorescent primer extension reactions (primer A, template HybI) using a proteinaceous polymerase (T7 RNA polymerase) and the R18 ribozyme RNA polymerase at ambient temperatures (red) and in ice (blue). The lowest band represents the unextended primer in this gel, as well as in all subsequent primer extension gels.

Ribozyme polymerase activity in ice, although slower, persisted for much longer, even after eight days' incubation. Time-courses of average primer extension at 17°C and in ice illustrate the reduction in polymerisation rate after days at 17°C (Figure 2.2A). Estimates of early primer extension suggested 10.7× faster initial extension at 17°C than in ice at -7°C. However, some of this difference was caused by the higher proportion of primers remaining unextended in ice; this could be on account of solute sequestration upon freezing, or altered ribozyme/primer/template ternary complex formation at low temperatures. While reactions at ambient temperatures yielded only modest additional extension after two days, replication activity in ice continued for days and weeks, allowing it to "catch-up", i.e. to synthesise equally long extension products. After seven days, the average length of

primers extended by at least two nucleotides was similar in ice (8.2 nucleotides/primer) to at 17°C (9.1 nucleotides/primer) (Figure 2.2B). Thus, while proteinaceous RNA polymerases are inactivated by freezing, ribozyme-catalyzed RNA polymerisation can proceed within the eutectic phase of ice unexpectedly effectively.





(A) Primer extension reactions (R18, primer A, template HybI) were incubated at 17°C or at -7° C in water-ice for the indicated times; separation by denaturing PAGE allowed quantification of the average number of nucleotides added to each primer in the reaction (Section 6.1.5) (means ± s.d.; N = 3). Estimates of the initial rates of nucleotide addition in these reactions were derived from the first time points in each series (indicated) where primers had been extended by comparable levels. This suggested initial extension rates of 7.3 nucleotides/primer a day at 17°C and 0.68 nucleotides/primer a day at -7° C in water-ice. However, extension is not distributed evenly amongst the population of primers; even after a week, a steadily declining fraction (26% at 17°C, 57% in water-ice) remains unextended, suggesting that initiation of extension upon a primer/template duplex is slow, particularly in ice, using this system. (**B**) To lessen the influence of initiation rate on measured extension, the average length of primers extended by two or more nucleotides can be examined; such average lengths at -7° C in ice are closer to those at 17° C (means ± s.d.; N = 3).

These polymerisation assays were carried out upon a template, HybI, that yielded enhanced primer extension. HybI shares the template sequence of template I, upon which R18 exhibits the best primer extension. However, HybI also possesses a 3' sequence that allows it to hybridise to the 5' terminus of the ribozyme (Figure 2.3A). This template tethering increases the local concentration of the primer/template duplex near the ribozyme, compensating for its high K_m for primer/template duplex (Lawrence and Bartel 2003); when using low concentrations of primer, template, and ribozyme, enhanced extension was observed upon this template (Figure 2.3B).

This system represented a useful assay for polymerase activity, allowing significant ribozyme extension to be attained at the lower, more prebiotic RNA concentrations used in these experiments (0.25 μ M each of primer, template and ribozyme). Low temperatures exerted similar effects upon ribozyme-catalysed polymerisation on the untethered template I – long incubations in ice yielded substantial primer extension (Figure 2.3C).

Figure 2.3. Enhanced extension through upstream hybridisation.

(A) Secondary structure of R18
 interacting with the primer A
 (orange) / template HybI (blue)
 duplex.

(B) Denaturing PAGE of extensions of primer A by R18 upon untethered and tethered templates. The concentrations RNA of each component in the reactions were varied simultaneously, maintaining a 1:1:1 ratio; identical reaction volumes were, however, run on the gel. Quantification of the average extent of primer extension is displayed below each lane.

(C) Denaturing PAGE of extensions of primer A upon both untethered and tethered templates (0.25 μ M each RNA), incubated for different times in ice and at ambient temperatures.



2.3 Reaction concentration through ice crystal growth

Incubation of reactions at -7° C allowed ice crystals to persist, but was not sufficient to induce ice crystal formation; reactions cooled to -7° C remained supercooled as aqueous solutions. Frozen extension reactions required a short freeze at -25° C to induce ice crystal formation, before incubation at -7° C allowed a eutectic phase to thaw out. Comparing frozen and supercooled reactions at -7° C isolated the effect of ice upon extension from that of temperature; frozen reactions yielded notably more extension than supercooled reactions (Figure 2.4A). Repeated freeze/thaw cycles, or other methods of freezing – such as flash-freezing in dry ice/ethanol baths, or introduction of an ice crystal to a supercooled reaction – had similar effects. This suggested that the benefits of freezing stemmed from the final state of the eutectic phase, rather than the freezing treatment.

This increased activity could be caused by the concentration of solutes by ice crystal growth upon eutectic phase formation. I therefore determined the volume of the eutectic phase (V_E) in reactions at $-7^{\circ}C$ by assessing ice survival as a function of reaction concentration. Standard 40 μ l extension reactions were lyophilised, and the resulting salts resuspended in a range of reduced volumes (V_B) of water, yielding a series of more concentrated reactions. These were frozen at -25°C to induce ice crystal formation, and then incubated at -7° C for one week to ensure crystalline and eutectic phase equilibration. If V_R were smaller than V_E, the solute concentration would have been higher than in the original eutectic phase, depressing the freezing point of the solution below that of the original eutectic phase. Hence, upon shifting the frozen aliquot from -25° C to -7° C, the ice would melt. Conversely, if V_R were greater than V_E , then ice would remain upon transferral to $-7^{\circ}C$. The transition from ice survival to ice loss at -7° C occurred at V_R = 10 µl, defining the volume of the eutectic phase in reactions at -7°C as ca. 25% of the initial reaction volume. Thus, we estimate the solute concentrations in such eutectic phases to be 0.8 M MgCl₂, 200 mM Tris·HCl, 16 mM of each NTP and 1 μ M of each RNA, in equilibrium with water ice at -7°C. The ligase core continues to benefit from Mg²⁺ ion concentrations above 0.2 M (Glasner et al. 2002), potentially explaining some of the increased polymerase activity in ice.



Figure 2.4. Supercooling and freezing diluted reactions.

(A) Denaturing PAGE of extension of primer A upon template HybI by R18 at $-7^{\circ}C$ (7 d) in frozen or supercooled reactions. (B) Quantification of average extension (7 d) of primer A upon both tethered and untethered template by R18, observed after denaturing PAGE. Solute levels were varied by reaction dilution (1 = undiluted), which dramatically reduced extension in aqueous solution (both at $17^{\circ}C$ (red circles, N = 1) and supercooled at $-7^{\circ}C$ (purple circles, means \pm s.e.m.; N = 3)), but barely affected extension in ice (blue diamonds, N = 1) over this range. Extension activity on the untethered template is more severely affected by dilution presumably due to the reduction of tertiary ribozyme/primer-template complex formation.

Freezing the reactions made ribozyme polymerase activity remarkably robust to depletion of crucial solutes. High concentrations of ribonucleoside triphosphates (NTPs) and magnesium salts (MgCl₂) are both essential for ribozyme activity (Glasner et al. 2002); reducing them by dilution resulted in a sharp decrease in ribozyme activity in solution, both at ambient temperatures (17°C) and in super-cooled solutions (-7° C) (Figure 2.4B). In contrast, RNA polymerase ribozyme activity in ice remained unchanged, even after dilution of reactions by 50-fold, and persisted after dilutions of up to 200-fold (Figure 2.5A).



Figure 2.5. The concentration effect of ice crystal growth.

(A) Average extension in diluted reactions (7 d, primer A, template HybI, R18) at 17°C (red circles) or in ice at -7°C (blue diamonds) relative to undiluted reactions at these temperatures, as judged by denaturing PAGE (means \pm s.e.m.; N = 3). (B) Denaturing PAGE of extensions (6.25 nM each of primer A, template HybI, and R18) in mineral water (NOAH's California spring water, with a favourable Mg²⁺:Ca²⁺ ratio), supplemented only with 5 μ M of each NTP, in ice at -7°C and at 17°C (16 d).

Due to the equilibrium between the ice phase and the eutectic phase, ice crystal growth only ceases when a certain solute concentration has been reached in the eutectic phase (Vajda 1999). Thus, more dilute starting conditions do not lead to a more dilute eutectic phase but rather reduce its volume until the same equilibrium concentration is reached.

This effect seemed to be unable to fully compensate for the most severe dilutions tested – 100-fold and 200-fold. The reasons for this effect are unclear but could include absorption of solutes into growing ice crystals or solute interactions with the increased plastic area of the reaction vessels. Upon freezing 50-fold diluted starting reactions, solutes were concentrated a total of 200-fold, enabling near optimal RNA replication activity from micromolar nucleotide concentrations and from Mg²⁺ ion concentrations closer to those found in present-day freshwater sources. Indeed, ribozyme-catalysed RNA polymerisation could proceed in a frozen mineral water supplemented only with 5 μ M of each NTP (Figure 2.5B). The capacity of water-ice to concentrate scarce substrates and ions would have been particularly

beneficial to a primordial replicase, lessening the requirement for high substrate affinity, and allowing replicases to thrive in substrate poor environments.

Positive counterions like Mg²⁺ are critical cofactors of ribozyme structure and activity. Indeed, the cationic requirements of the class I RNA ligase ribozyme have been investigated in detail and its activity has been found to be strictly dependent on Mg²⁺, with all other cations (including Mn²⁺) strongly inhibitory (Glasner et al. 2002). However, in contrast to the well-studied functional roles of positive counterions like Mg²⁺, the influence of negative counterions had not previously been considered. I replaced the main negative counterion chloride (CI-) with a range of alternative counterions (SO₄²⁻, CH₃COO⁻, Br⁻, NO₃⁻), while maintaining equimolar levels of Mg²⁺, and studied their effect on ribozyme polymerase activity both in ice and at ambient temperature. The identity of the negative counterion exerted limited influence on ribozyme activity at 17°C, but affected ribozyme polymerase activity noticeably in ice (Figure 2.6A). While some counterions were inhibitory, replacement of CI^- with sulphate (SO_4^{2-}) substantially enhanced in-ice polymerase ribozyme activity, despite inhibiting it in solution. Intriguingly, ranking of counterion effects in ice (but not at ambient temperature or in supercooled solutions) followed the Hofmeister series (Pegram and Record 2007). The more chaotropic anions such as nitrate and bromide reduced in-ice extension, whereas sulphate, a kosmotropic anion, substantially enhanced in-ice polymerase ribozyme activity.

Hofmeister effects on macromolecular function are well-known and ubiquitous (Zhang and Cremer 2006). However, the restriction of the effects to in-ice polymerase activity argues against solely temperature-dependent interactions of the negative counterion with the ribozyme or its hydration shell, but rather suggests indirect effects on water structure and ice crystal growth. Notably, MgSO₄ is capable of structuring and slowing a large fraction of water molecules, particularly at high concentrations and low temperatures (Tielrooij et al. 2010), with unknown implications for ribozyme polymerase activity. Additionally, the eutectic point of MgSO₄ is -4° C (McCarthy et al. 2007), nearly 30°C higher than that of MgCl₂ (-33°C). Although attenuated by the presence of the other solutes in the reaction, the higher eutectic point and consequently much weaker depression of the freezing

temperature by equimolar amounts of MgSO₄ compared to MgCl₂ would result in more extensive freezing, affecting eutectic phase microstructure. Scanning electron microscopy (SEM) imaging of freeze-fractured eutectic ice phases (Figure 2.6B) indeed suggests a decreased eutectic phase volume in MgSO₄ ices, and consequently an increased concentration effect upon eutectic phase formation.



Figure 2.6. Magnesium counterion effects upon extension.

(A) Denaturing PAGE of primer extension (primer A, template HybI, R18) in reactions (8 d) where the CI counterion of magnesium was replaced with a range of other counterions, displayed in order of position in the Hofmeister series. (B) The eutectic phase in ices formed at -7° C from extension reactions with MgCl₂ or MgSO₄, imaged by freeze-fracture scanning electron microscopy. The raised veins of eutectic phase are on average visibly narrower in ice formed with equimolar amounts of MgSO₄.

2.4 Ribozyme stability

What was the cause of the persistence of ribozyme activity in ice? The dependence of the RNA polymerase ribozyme on high Mg^{2+} concentrations for optimal activity accelerated the hydrolysis of its RNA backbone, limiting its half-life to < 52 hours at 17°C. However, at -7°C in ice, its half-life was increased to > 16 days, a 7.4× increase in stability (Figure 2.7). As a result, after 15 days of incubation, less that 1% of ribozyme remained full-length at 17°C, but over half was intact in ice at -7°C. Ribozyme activity was observed to slow more abruptly that this, though (Figure 2.2), perhaps due to inhibition by accumulated degradation products.



Figure 2.7. Ribozyme degradation.

(A) Denaturing PAGE of 5' fluorescein-labelled RNA polymerase ribozyme incubated in extension buffer for 7 days. (B) Degradation as a function of incubation time at 17°C (red circles) or in ice at $-7^{\circ}C$ (blue diamonds) (means \pm s.d.; N = 3). Exponential decay functions were fitted to the data, suggesting decay constants of 0.319 (17°C; $R^2 = 0.9945$) and 0.043 ($-7^{\circ}C$; $R^2 = 0.973$) per day.

To explore to what extent this stability in ice could be harnessed, extension reactions were allowed to continue for longer. While ribozyme activity at ambient temperatures ceased after a few days (extending a primer by up to 16 nucleotides, Figure 2.8A), ribozyme polymerase activity in ice continued for weeks, yielding extension products of up to 23 nucleotides length after 25 days (Figure 2.8B). These long extensions in ice required low starting NTP concentrations (0.5-1 mM of each); due to the concentration effect of eutectic phase formation, standard concentrations (4 mM of each) were concentrated to inhibitory levels (16 mM of each, which attenuated the formation of longer products at 17°C) (Figure 2.8A). Replacing the Cl⁻ counterion with SO_4^{2-} allowed RNA products up to 32 nucleotides in length to be synthesised in ice (Figure 2.8C), corresponding to almost three turns of an RNA double helix. Thus, despite slowing down polymerisation, freezing can preserve ribozymes and prolong their activity, achieving RNA syntheses inaccessible at ambient temperatures.



Figure 2.8. Long extensions in ice.

Denaturing PAGE of extensions of primer A upon various templates by R18 under the indicated temperature/buffer conditions, for 25 days (red = aqueous, blue = frozen). (A) Lowering NTP concentrations facilitated synthesis of the longest products in ice (0.2 M $MgCl_2$, template HybI). (B) Extension upon a longer template (0.2 M $MgCl_2$, template HybI2). (C) In ice, extension proceeded better in 0.2 M $MgSO_4$; up to 32 nucleotides were added opposite a longer template (HybI41; the inset shows a high-sensitivity scan of the indicated area).

2.5 Ribozyme fidelity in ice

Replication must exceed a certain degree of accuracy – the error threshold – to avoid an 'error catastrophe' and dissipation of the genotype through generation of mutated competitors (Eigen 1971; Kun et al. 2005). Furthermore, a higher fidelity than this error threshold would likely be required at the origin of life: a replicase must synthesise active copies of itself faster than it degrades, and errorprone synthesis would render a substantial fraction of offspring inactive. Thus, polymerase fidelity is a critical attribute for a replicase ribozyme. Temperature influences the strength of molecular interactions and is known to affect the substrate discrimination and fidelity of proteinaceous polymerases (Tindall and Kunkel 1988). Indeed, the recognition specificity of the hairpin ribozyme was found to be notably relaxed at subzero temperatures (Vlassov et al. 2004).

To assess whether polymerisation in ice was achieved at the price of fidelity, ribozyme error rates were determined by high-throughput sequencing of primers extended by 12 nucleotides from the ambient temperature ($17^{\circ}C$, 1355 sequences) and in-ice ($-7^{\circ}C$, 2070 sequences) reactions in Figure 2.8B (Table 2.1). These extension products yielded a value for R18 substitution fidelity at ambient temperatures (97.1%) in good agreement with previous estimates (96.7%) obtained by comparison of the relative rates of incorporation and misincorporation of NTPs (Johnston et al. 2001). The total fidelity at $17^{\circ}C$, including deletions, was 96.2%, and a similar value of 93.4% was observed for ribozyme fidelity in ice (Table 2.1), the difference being mainly due to an elevated rate of deletion mutations in ice (2.4%) compared to at ambient temperatures (0.8%).

A Template base	C	G	U	А	Mean
Incorporation (%) 17°C	at				
G	99.32	0.26	6.54	0.37	
С	0.14	98.01	2.33	0.70	
А	0.08	0.00	88.99	0.97	
U	0.19	1.22	0.15	96.88	
Substitution rate	e 0.41	1.48	9.02	2.04	
Corrected substitutio	n rate 0.02	1.10	8.69	1.67	2.93
Deletion rate	0.28	0.52	1.99	1.08	
Corrected deletion	rate 0.15	0.39	1.87	0.95	0.84
Fidelity	99.84	98.52	89.44	97.38	96.21
B Template base:	С	G	U	A	Mean
Incorporation (%) -7°C in ice	at				
G	98.35	0.51	6.79	1.21	
С	0.32	96.40	2.73	1.02	
А	0.21	0.12	82.84	1.43	
U	0.17	2.49	0.44	94.88	
Substitution rate	e 0.70	3.12	9.96	3.66	
Corrected substitutio	n rate 0.31	2.75	9.66	3.30	4.07
Deletion rate	0.96	0.48	7.20	1.46	
Corrected deletion	rate 0.83	0.35	7.08	1.33	2.44
Fidelity	98.86	96.90	83.26	95.37	93.39

Table 2.1. Ribozyme polymerase fidelity.

Extension products generated by R18 were cloned and subjected to high-throughput sequencing, allowing the patterns of errors generated at $17^{\circ}C$ (A) and in ice at $-7^{\circ}C$ (B) to be compared. Under both conditions, G-U wobble pairing accounted for the majority of substitutions. Sequencing an oligonucleotide corresponding to a chemically-synthesised extension product (CompI) allowed estimation of the background error rate in the sequencing process (0.13% deletions, 0.38% substitutions) and determination of the underlying polymerase-derived errors (after correcting for background mutation and reversion rates). Insertion rates in the ribozyme-synthesised sequences were slightly lower than in the background, and so were assumed to be negligible. Average fidelities were calculated using geometric means of the fidelity opposite each base.

Furthermore, detailed inspection of the error spectrum and sequence dependency revealed closely matching mutation signatures and hotspots at both temperatures (Figure 2.9). While error rates were highly base- and position-dependent, and thus to some degree template-specific, similar patterns were observed at 17° C and -7° C, suggesting that the mechanisms governing ribozyme fidelity are intact at subzero temperatures. The sequences also show that the ribozyme is capable of continuing polymerisation after misincorporations and

deletions at both temperatures. Thus, eutectic ice phases would be able to support and enhance RNA replication without substantially compromising replication fidelity.



Figure 2.9. Error spectra of R18.

(A) Substitution spectrum at 17° C. (B) Substitution spectrum in ice at -7° C. (C) Positional error rates at 17° C and (D) in ice at -7° C (extension proceeded right—left; correct and unknown rates left blank; errors with ambiguous position evenly assigned).
2.6 Compartmentalisation in ice

Compartmentalisation and colocalisation of a replicase and its offspring is a prerequisite for Darwinian evolution (Szostak et al. 2001). Thus, RNA replication in ice does not per se imply a capacity to support RNA evolution. Elegant in silico experiments have previously shown that diffusion limitation, by keeping molecules progeny together, can provide functional colocalisation and their and compartmentalisation and support replicase evolution by ensuring kin selection through reciprocal altruism (Szabo et al. 2002). Diffusion restriction links phenotype and genotype and allows evolution, while simultaneously restricting the spread of parasitic RNAs. Ice is highly structured at the microscale, with the eutectic phase forming a complicated network of brine-filled spaces and channels interspersed between the ice crystals. Could such lattice structures restrict ribozyme diffusion and afford a form of quasicellular compartmentalisation?

I sought to detect this beneficial trait using a sensitive assay for ribozyme diffusion through the ice (Figure 2.10A). 1 μ m-diameter microbeads decorated with RNA primer/template duplexes were randomly dispersed within an ice column formed from (diluted) extension buffer, as a 'scoring grid'. A thin layer of undiluted extension buffer containing a high concentration of R18 was frozen atop this, and the column was incubated at -7° C in ice to allow diffusion to occur (Figure 2.10B). Ribozymes that encountered beads extended primers upon them using NTPs in the eutectic phase, and these extended primers were fluorescently detected using rolling circle amplification (see Section 3.5), allowing quantification of bead population fluorescence by fluorescence activated cell sorting (FACS). By assessing the fraction of fluorescent beads after incubation, the extent to which the ribozyme had diffused through the ice could be deduced. As incubations were allowed to continue for longer, more beads acquired fluorescence in FACS (Figure 2.10C).

The results showed a dramatic restriction of ribozyme diffusion through ice, with ribozyme requiring over a week to reach 70% of the beads (Figure 2.10C). Furthermore, ribozyme diffusion through ice columns formed from more dilute extension buffer was decreased further (Figure 2.11A, B), dropping sharply as a function of higher dilutions.



Figure 2.10. Principle of the in-ice diffusion assay.

(A) If ice microstructure limits the exposure of primer-coated 1μ m-diameter beads to diffusing ribozyme, only a subset of the beads would experience primer extension (red). (B) Experimental set-up. 1 µm-diameter beads labelled with primer were distributed randomly throughout a conical column of eutectic ice inside an envelope of preformed pure ice. A small volume of ribozyme-rich ice on top provided a pool of ribozyme molecules that were allowed to diffuse down through the eutectic phase of the underlying ice column. As the ribozyme passed the beads embedded throughout the eutectic phase of the ice column, it engaged and extended bead-tethered primer/template duplexes. This generated a spatial distribution of two distinct bead populations: one that had encountered ribozyme, and was coated with extended primers, and one that ribozyme had not reached, and on which primers remained unextended. This distribution thus reflected the extent of diffusion of functional ribozyme through the ice phase and could be determined by FACS by scoring the proportion of beads with extended primers. (C) FACS dot plots showing ribozyme diffusion through ice formed from undiluted MgCl₂ extension buffer, as a function of time. Positive beads (red, fluorescence > 30.5) emerged early in the incubation, and increased in number as the ribozyme front advanced through the ice phase.



Figure 2.11. Diffusion restriction in frozen dilute solutions.

(A) FACS profiles of beads from ices formed from undiluted and diluted MgCl₂ extension buffer (after 8 days' diffusion). Beads with many extended primers (red) gained fluorescence during processing. (B) The extent of ribozyme diffusion after 8 days (as measured by the proportion of fluorescent beads) through ices with different solute levels and magnesium counterions (means \pm s.e.m.; open diamonds = MgCl₂, N=4; filled diamonds = MgSO₄, N = 3). A background level of 0-3% positive beads was observed in pure water-ice columns, presumably from localised small-scale thawing of the top of the column upon ribozyme addition, and the displayed values are corrected for this. (C) FACS histograms showing fluorescence of beads after eight days in ice with the starting ribozyme evenly distributed through the column (red line = positive threshold fluorescence). The columns were formed from (diluted) MgCl₂ extension buffer, and similar results were seen using MgSO₄ extension buffer. In-ice primer extension was not significantly biased by ice column composition per se; thus, the percentage of positive beads after ribozyme diffused through the column (B) is a true measure of ribozyme diffusivity. (D) The structure of the eutectic phase in (diluted) MgCl₂ reactions, imaged by freeze-fracture SEM. The lighter, raised web of channels represents the eutectic phase.

Scanning electron microscopy (SEM) imaging of freeze-fractured eutectic ice phases revealed a potential physical basis for this effect, as striking changes in eutectic phase volume and topological structure were observed upon the progressive reduction of the starting solute concentration (Figure 2.11D). As adjacent ice crystals fused, the connectivity of the brine channel network was reduced, yielding a sparse, thin and increasingly fragmented network of brine veins. Ices in which the main negative counterion chloride (Cl⁻) was replaced by sulphate (SO₄²⁻) displayed an even narrower brine vein structure (Figure 2.6B); this more constricted ice microstructure may account for the enhanced RNA replicase activity (Figure 2.6A) and the further reduced ribozyme diffusion observed in sulphate ices (Figure 2.11B). Thus, solute identity as well as concentration can determine ice microstructure and alter ribozyme replication activity and diffusion.

These data were used to calculate the diffusivity of the ribozyme in different MgCl₂ ices (Table 2.2). To achieve this, the experiment was modelled as ribozyme diffusing down a uniform tube (Figure 2.12A), necessitating the knowledge of a ribozyme concentration at a specific distance from the ribozyme phase at the end of the incubation. Therefore, the minimum 'critical' concentration of ribozyme that yielded a positive signal was determined. In columns formed from undiluted MgCl₂ containing a uniform starting concentration (1 nM or higher) of ribozyme, beads gained a positive signal after eight days' incubation, indicating that in the diffusion experiments at most a 4 nM or higher ribozyme concentration was required in the eutectic phase to yield positive signal. As beads were distributed uniformly throughout the column (here modelled as a 2 cm deep cone), the corrected positive bead percentages ($B_{\%}$ – Figure 2.11B) could be converted to distances to the critical concentration (*x*, in cm):

$$x = 2 - \sqrt[3]{(8 - B_{\%}/12.5)}$$

These corresponded to the distances of the critical ribozyme concentration from the start (beyond which diffused ribozyme concentration is too low to yield positive signals). Means of these distances were used to solve the diffusion equation for the diffusivity of ribozyme within each ice (Table 2.2), using the model in Figure 2.12A.

Column composition (fold-dilution of extension buffer)	1	2.5	10	20	50
Critical ribozyme concentration (nM)	1	0.4	0.1	0.05	0.02
Depth of positive beads (cm) Standard error (cm)	0.69 0.024	0.48 0.070	0.34 0.079	0.19 0.055	0.038 0.014
Diffusivity (D) (cm²/s)	2.6×10⁻³	1.0×10⁻ ⁸	4.3×10⁻⁰	1.1×10-9	3.7×10 ⁻¹¹
$D = D_{aq} \frac{\varepsilon_{t} \delta}{\tau}$					
Diffusivity/D _{aq}	0.19	0.076	0.032	0.0084	0.00028
Porosity (ϵ_t)	0.25	0.1	0.025	0.0125	0.005
Tortuosity (τ)	1.3	1.3	0.8	1.5	18

Table 2.2. Calculation of ribozyme diffusivity.

The percentage of fluorescent beads observed after ribozyme diffusion through different $MgCl_2$ ices was converted to a diffusivity value, and interpreted as described in the text.

While the microstructure of the eutectic phase in itself impeded diffusive (and, crucially, convective) transport of macromolecules (compared to the solution phase), a striking further drop in diffusivity was observed in ice phases formed from more dilute starting mixtures. This diffusivity D, in a porous medium such as ice, is related to the diffusivity in liquid (D_{aq}) by three dimensionless parameters – porosity (ϵ_t), constrictivity (δ) and tortuosity (τ) (Boving and Grathwohl 2001) (Table 2.2). The diffusivity of lysozyme in aqueous solution is 1.11×10^{-6} cm²/s at 25°C (Brune and Kim 1993); as the ribozyme's diameter is approximately 3× greater (10 nm (Muller and Bartel 2008; Shechner et al. 2009) vs. 3.4 nm (Cardinaux et al. 2007)), I estimated its diffusivity in aqueous solution to be 3.8×10^{-7} cm²/s at 25°C, and 1.35×10^{-7} cm²/s at -7° C (D_{aq}). ϵ_t is the liquid fraction of the ice and was known from the concentration effect of eutectic phase formation, and could account for the majority of the difference between the predicted and measured diffusivities in ices formed from buffers diluted up to 20-fold. Thus, diffusion in ice is limited by the low volumes of the eutectic phase.

However, the diffusion coefficient in ice formed from the highest dilutions (50-fold) was much lower than predicted by changes in ε_t alone. δ is a measure of the influence of the channel walls on diffusion, and is only relevant when the size of the particle is comparable to the width of the channel (Ternan 1987). Eutectic phase

channels typically have diameters in the μ m range (Figure 2.13), and thus, while surface effects could not be ruled out, pore wall effects were unlikely to significantly impede the diffusion of the much smaller ribozyme molecules, so δ was set at unity. τ is the square of the ratio of the length of a path connecting 2 points and their distance apart, characterising the curvature of the eutectic phase. Paths about perfect hexagonal crystals would yield $\tau \approx 1.8$; the sharp increase in τ at higher dilutions (50-fold) suggested reduction in the connectivity of the eutectic phase – through channel closure during freezing. Such fragmentation of the eutectic phase would manifest itself as a steep increase in diffusive path length, and a drop in diffusivity (Figure 2.12B).



Figure 2.12. Ribozyme diffusivity in ice.

(A) Ribozyme diffusion was modelled as an amount Q $(1.9 \times 10^{-10} \text{ moles/cm}^2)$ of ribozyme starting at the closed end of a tube and diffusing as an even front through the ice at a rate governed by the diffusivity D. C(x,t) is the concentration of ribozyme (moles/cm³) after t seconds, x cm from the start; at the end of the incubation (194 h in our assays), beads at a distance x (derived from bead population fluorescence, calculated in Table 2.2) had been exposed to the critical concentration ($C_{x,t}$) of ribozyme (4 nM in the eutectic phase) sufficient to exhibit a positive FACS signal, allowing this equation to be solved for D. (B) The calculated in-ice diffusivity (D, orange diamonds) of ribozyme and corresponding tortuosity (τ , blue triangles) of the eutectic phases in ices formed from different dilutions of MgCl₂ extension buffer (Table 2.2).

These drastic reductions in ribozyme diffusivity could thus be fully rationalised by the decreasing volume and increasing fragmentation of the eutectic phase, as observed by SEM imaging (Figure 2.11D). The sharp increase in relative path length for ribozyme diffusion in the most dilute ices was symptomatic of a breakdown of the connectivity of the brine channel network as a function of starting solute concentration, noted previously as a sharp reduction in ice electrical conductivity (Grimm et al. 2008). SEM imaging of partially sublimated ices further illustrated the three-dimensional structure of the eutectic phase and its fragmentation in ices formed from more dilute solutions (Figure 2.13). Although the relationship between diffusion restriction and compartmentalisation is complex (Szabo et al. 2002), our *in silico* simulations demonstrated how a low rate of lattice fragmentation leads to the emergence of compartmentalised grid sectors, protecting replicases from parasitic molecules (Attwater et al. 2010). Additionally, another effect was observed that may contribute compartmentalisation upon freezing of dilute solutions: although diluted reactions had much lower concentrations of solutes prior to freezing, they contained similar levels of dissolved gas (or potentially higher levels, due to the decreased solubility of gases in the presence of salts), which were concentrated to a greater degree upon freezing. As a result, bubbles were common in ice made from 50-fold diluted extension buffer, often associated with discrete and self-contained torus-shaped inclusions of eutectic phase observed in SEM (Figure 2.13D) which could potentially serve as protocellular compartments.

Eutectic phase structure is dependent upon solute level, temperature conditions and freezing regime. Sea ice, for example, can contain distinct eutectic phase compartments, visible under a light microscope (Krembs et al. 2011). In some conditions, particularly when using dilute solutions, freezing can yield structures potentially capable of providing quasicellular compartmentalisation.





Ices formed from undiluted (A), 12-fold diluted (B) and 50-fold diluted (C, D) MgCl₂ extension reactions were flash-frozen in liquid N₂ to preserve the eutectic phase structure and freezefractured, as in Figure 2.6B & Figure 2.11D; however, the samples then underwent prolonged sublimation before gold deposition and scanning electron microscopy imaging. This preferential sublimation of the ice crystals exposed the three-dimensional ultrastructure of the eutectic phase, revealing the prominent sheets of a contiguous eutectic phase in the undiluted reaction (A). Dilution led to a decrease in the volume of the eutectic phase, manifested by fusion of adjacent ice crystals and shrinking of contiguous sheets of eutectic phase to a network of tubular channels along the vertices of crystals (B), and eventually to sparsely distributed clusters of filamentous tubes (C). In the most dilute eutectic phases (D), gas bubbles containing discrete eutectic tori were frequent, resulting in their deposition on the ice surface upon sublimation.

2.7 Discussion

Using the R18 RNA polymerase ribozyme as the best available modern day analogue of a primordial replicase, I have shown that ice has the capacity not only to support but to enhance accurate ribozyme-catalysed RNA replication through substrate and solute concentration and attenuation of replicase degradation. Furthermore, some eutectic phase microstructures could enable RNA evolution through quasicellular compartmentalisation within the eutectic phase. Ice may therefore have harboured replicating ribozymes in a "cold RNA world" (Bada et al. 1994; Vlassov et al. 2005; Monnard and Ziock 2008).

Prebiotic RNA self-replication within eutectic ice phases presupposes the existence of significant bodies of surface ice on the early earth. After the Earth's crust formed, heat flux from within the mantle would have faded; to maintain a temperate climate in the face of the faint early sun (~75% as bright as today) (Sleep et al. 2001), a greenhouse effect would be required. In the absence of biogenic methane, this would have been CO₂-dominated, and continental weathering and subduction would have eventually scoured this gas from the atmosphere (Zahnle 2006). Previous assumptions that the surface and ocean temperatures on the Hadean Earth were high, and therefore incompatible with surface ice, have recently been challenged by studies suggesting temperate climatic conditions for the Archean aeon, compatible with substantial polar and seasonal ice deposits (Wilde et al. 2001; Valley et al. 2002; Hren et al. 2009; Rosing et al. 2010).

Due to the concentrating effect of eutectic phase formation, ribozymecatalyzed RNA replication can proceed from starting mixtures containing as little as 4 mM Mg²⁺ and 20 μ M of each NTP. The strength of this concentration effect is inversely proportional to the total solute level (Vajda 1999; Monnard and Ziock 2008) and consequently dependent on the salinity of the environment; it would therefore have been best realised in a freshwater environment. It is notable that saline environments also inhibit other prebiotically relevant processes, such as vesicle formation, concentration of solutes by thermophoresis, and the non-enzymatic condensation of activated mononucleotides (Monnard et al. 2002; Baaske et al. 2007). Frozen dilute solutions could provide compartmentalisation of RNA replication, as well as relieving replicase dependence upon prebiotically implausible substrate concentrations, expanding the range of environments colonisable by replicases.

While protocellular compartmentalisation can take many forms, it is often considered that molecular self-replication could originate within membraneous protocellular vesicles in an ambient, aqueous environment (Mansy et al. 2008; Schrum et al. 2010). My results demonstrate that neither RNA replication nor compartmentalisation is necessarily confined to the solution phase, or indeed ambient temperatures, but that both are provided within the aqueous eutectic phase of water-ice at subzero temperatures. Ice had previously been shown to stabilise nucleotide components (Levy and Miller 1998) and catalyse both the *de novo* as well as templated synthesis of random RNA oligomers (Monnard et al. 2003; Trinks et al. 2005; Monnard and Szostak 2008; Monnard and Ziock 2008): processes that take advantage of some of the same features of ice that benefit replicases. Such an environment appears conducive to both replicase emergence and success. In contrast, although replicase ribozymes require compartmentalisation, protocells offer few benefits to nonenzymatically-replicating sequences; rather, product inhibition through hybridisation to concentrated complementary strands would disfavour sequences that reside inside vesicles - RNA sequences outside the protocell would compete effectively with compartmentalised ones, suffering less from inhibition by related molecules.

An argument can therefore be made that the first replicators emerged in ice, independent of any abiotic synthesis of membrane components. The properties of ice can promote steps from prebiotic oligomer synthesis to the emergence of RNA self-replication and Darwinian evolution. Emerging replicases sheltered in the ice would then have a better chance to adapt to a range of less favourable environments, evolving faster replication speeds to outpace degradation at ambient temperatures, and low divalent cation requirements to operate in membraneous protocells. Eventual colonisation of protocells would provide enhanced compartmentalisation and protection from parasitic and predatory ribo-organisms, yielding the ancestors of cellular life.

3 Directed Evolution of Ribozyme RNA Polymerase Activity

3.1 Introduction

Despite the improvements in ribozyme stability in ice that allow synthesis of longer extension products, the central tenet of the RNA world hypothesis – RNA's capacity to catalyse its self-replication – remains to be verified. While the R18 polymerase ribozyme shows all the basic activities needed for self-replication, i.e. templated nucleotide incorporation, 3´-5´ regiospecificity, template translocation etc., an activity substantially superior to that of R18 must be isolated to enable a ribozyme to synthesise a copy of itself. Directed evolution remains the most successful strategy for obtaining new and improved RNA catalysts (Wilson and Szostak 1999; Joyce 2007), as evidenced by the isolation of as complex a ribozyme as R18 from random sequence.

A functioning selection system requires molecules exhibiting the desired activity to be recovered more frequently than inactive molecules after each round of selection; the difference in recovery will determine how quickly the selection proceeds. However, the success of a selection will rest upon two other parameters: the recovery of active molecules relative to parasitic molecules (those that circumvent the selection criteria or exploit active molecules), and the relative abundance of active and parasitic molecules in the selection pools. The emergence of desired molecules must outpace the emergence of parasites to allow screening to isolate successful clones.

3.2 Improved ribozyme polymerase selection systems

Further evolution to obtain a replicase would thus be more likely to succeed starting from the pre-existing polymerase ribozyme; however, efforts to improve on R18's polymerase activity by directed evolution in the decade since its isolation have met with limited success. For example, refining the selection strategy that yielded R18 and reapplying it to earlier stages in that selection uncovered a number of families of ribozyme polymerases with diverse processivity domains, but none performed better than the family that yielded R18 (Lawrence and Bartel 2005).

Furthermore, limited correlation was observed between family activity and family abundance in the selection pools, and families of inactive ribozymes emerged and dominated later rounds of selection; these were presumed to disseminate through parasitic mechanisms, perhaps by serving as an efficient substrate for 4-thioU modification by active polymerases. This suggested that the selection system suffered from a fundamental weakness that prevents its continued application to ribozyme polymerase optimisation: any improved ribozyme faces competition from the abundant evolving parasitic molecules in the pool.

A further drawback of the original selection scheme for evolving an improved ribozyme concerns the criteria used for recovery of a ribozyme. 4-thioUTP incorporation allowed selection based upon addition of one or two nucleotides to a primer, but this is a capacity that R18 already exhibits. Direct selection for multiple incorporations or synthesis of a stretch of RNA would be necessary to allow the acquisition of properties such as improved processivity and sequence generality. Ideally, the ribozyme would also be selectable in an *in trans* catalytic context to promote activity as a true catalyst; although this property is not a key prerequisite for replicase activity, such a selection pressure might be expected to promote the emergence of processivity.

To satisfy these requirements, Zaher and Unrau designed a novel *in vitro* selection system to allow the isolation of superior ribozyme polymerases (Zaher and Unrau 2007). They exploited *in vitro* compartmentalisation in the aqueous compartments of water-in-oil emulsions to allow transcribed ribozyme to extend a primer attached to its parental DNA molecule in the same compartment, linking phenotype to genotype; this technology had previously been applied to the directed evolution of proteinaceous polymerases (Ghadessy et al. 2001). Extended primers were bound to biotinylated probe molecules to recover the attached genes. By basing gene recovery upon modification of the parental gene and not of the ribozyme itself, they were able to select for catalytic activity *in trans*; furthermore, probe binding requires synthesis of a stretch of RNA, with gene recovery more likely the further a ribozyme extends, allowing direct selection for processive polymerisation.

This selection protocol, however, led to takeover by a different class of molecular parasite: DNA molecules containing binding sites for the probe molecule.

Further rounds of selection based on the original 4-thioU incorporation scheme (to favour polymerase spread within the pool) rescued the selection, yielding a ribozyme (B6.61) with modestly improved RNA polymerase activity capable of extending a primer by up to 20 nucleotides.

However, this selection highlighted the risks associated with selection criteria that base gene recovery upon binding of a single molecule to a specific DNA or RNA sequence; the challenge of obtaining an improved ribozyme that synthesises such a sequence contrasts with the ease of emergence of probe-complementary sequences within the ribozyme gene. Selections based upon reverse transcription of ribozyme using a primer complementary to ribozyme-synthesised sequence would be expected to face similar challenges. Given the presumed infrequency of replicase activity amongst RNA molecules, a selection system was needed that reduces opportunities for parasite emergence yet still recovers genes based on synthesis of an RNA sequence.

3.3 Bead-based selection

Fluorescence activated cell sorting (FACS) of microbeads provided a potential mechanism to overcome these difficulties. If individual ribozyme genes were linked to such beads, the opportunity would arise to base recovery upon ribozyme synthesis of not just one, but multiple bead-linked RNA sequences. Any such sequence in a single parasitic DNA molecule would thus be eclipsed by multiple extension sequences generated by an improved ribozyme. Fluorescent detection of extended primers would then allow FACS of beads to recover only those bound to genes encoding active ribozymes. Micelles assembled from amphiphiles can be used to allow ribozymes and RNA primers to congregate (Muller and Bartel 2008), improving primer extension, but do not exhibit sufficient stability to link phenotype to genotype on a selection timescale.

Microbead-based FACS selection had been demonstrated for the recovery of an active ribozyme from a library of class I ligase variants (Levy et al. 2005). This selection harnessed *in vitro* compartmentalisation of beads using waterin-oil emulsions to maintain the genotype-phenotype linkage, ensuring that ribozymes transcribed from one gene modify *in trans* RNA primers attached only to that gene's bead. However, the ribozyme RNA polymerase exhibits optimal activity in buffers with a much higher Mg²⁺ concentration (0.2 M) than that tolerated by the proteinaceous T7 RNA polymerase used to transcribe the ribozyme. Thus, transcription of ribozyme and primer extension by ribozyme cannot both proceed at optimal levels in the same water-in-oil compartment. Previous studies had used a buffer composition compatible with both transcription and ribozyme-catalysed primer extension (Zaher and Unrau 2007), but I observed that both activities suffered significantly under these conditions; furthermore, the non-canonical RNA-dependent RNA polymerase activity of T7 upon the primer/template duplex was able to match or exceed that of the transcribed ribozyme, potentially resulting in a high level of background in any selection.

Both of these problems could be avoided if transcription and extension were performed in separate emulsions. Each stage could then benefit from optimal buffer composition, and primer/template duplexes would only be bound to beads after transcription was complete, avoiding extension by proteinaceous polymerase. However, this separation required a means of linking ribozyme to beads after transcription, to maintain the linkage of ribozyme phenotype to bead-bound genotype.

3.4 Compartmentalised bead-tagging

Simultaneous transcription of ribozyme and ligation to bead-bound hairpins in emulsion provided a suitable mechanism to ensure phenotype-genotype linkage. Ribozyme could be efficiently transcribed with a 5' terminal phosphate by including high concentrations of guanosine monophosphate (GMP) in the transcription reaction. This allowed ribozyme ligation to short RNA-DNA hairpins coating the bead via a short 5' terminal sequence (Figure 3.1).



* R18i: [5'-CAAAUCUGCCCUCAGAUCUUG-3']

Figure 3.1. Secondary structure of R18 ligated to bead-bound hairpins.

Beads (brown) were coated with RNA-DNA hairpins (5Hairpin) prior to transcription/ligation. The 3' sequence of these hairpins was RNA (green) to allow ligation to the eight-nucleotide fixed 5' ribozyme sequence of a ribozyme (black) upon the DNA part of the hairpin (blue) by T4 RNA Ligase 2 (NEB). The hairpin was linked to the bead via a disulphide bond-containing linker (red), and contained two 2'-O-methyl RNA residues (purple) to discourage hairpin extension by non-canonical T7 RNA polymerase activity. The asterisk indicates the position of the inactivating duplication sequence in the R18i variant used in model selections. Two 2'-O-methyl RNA residues were also present in the template strand of the ribozyme gene, to terminate transcription at the final base of the ribozyme; however, the majority of MSS T7 RNA polymerase molecules proceeded transcribing beyond them, generating a 3' run-through transcript (RTT) sequence. Stem2 was present in selections to complete the ribozyme structure.

This 'compartmentalised bead-tagging' (CBT) protocol (Figure 3.2) created stable clonal repertoires of bead-bound ribozymes. Quantification of bead-bound products by denaturing PAGE and SYBR Gold staining indicated that typically ~3,000 ribozyme molecules are ligated to each bead, derived from a single PCR

product. This was dependent on the high efficiency of the MegaShortScript kit (MSS) T7 RNA polymerase enzyme (Ambion), and allowed selection based upon the activity of thousands of ribozyme copies, not the stochastic activity of a single molecule. By increasing the extension available for detection, this allowed the application of more stringent selection criteria, aiding isolation of superior ribozymes.



Figure 3.2. Compartmentalised bead-tagging.

To begin a round of selection, a library of ribozyme genes was bound to streptavidin-coated microbeads at a density of up to one gene per bead (i). Ribozyme was transcribed and ligated to RNA hairpins on beads in the compartments of a water-in-oil emulsion (ii; inset shows light microscopy of selection emulsion, field diameter ~0.15 mm), generating clonal

bead-bound repertoires of ribozymes. After recovery from emulsion and decoration with primers (iii), the beads were resuspended in buffer with RNA template and NTPs, emulsified, and incubated at 17°C to allow ribozyme-catalyzed primer extension. After recovery of beads, extended primers were used to prime rolling circle amplification (RCA) (v); the resulting single-stranded DNA concatemers were converted to a double-stranded form and stained by Picogreen (vi), allowing separation of fluorescent beads using FACS (vii). The genes bound to such beads, encoding active ribozymes, were recovered using PCR (viii).

Ribozyme-coated beads were then further decorated with primers (~60,000 per bead), before emulsification in extension buffer with RNA template and NTPs. The hairpins are linked to the bead via a disulphide linker; introduction of DTT by mixing with a second emulsion containing DTT in the aqueous phase reduced this bond and released the ribozyme into the emulsion compartment, allowing selection based upon true *in trans* activity. This also eliminated the possibility that parasitic ribozyme variants emerge that confer a fluorescent signal to beads by displaying the target RNA sequence. Alternatively, when using libraries presenting a low risk of emergence of such parasites (e.g. mutagenised ribozyme could be left ligated to beads, increasing local concentration around primers and extension whilst still requiring some *in trans* interaction.

Incubation at 17°C allowed active ribozymes to 'tag' beads by extending bead-bound RNA primers; the emulsion was then broken and beads were recovered. Denaturing washes removed RNA template, exposing synthesised RNA sequences for fluorescent detection to allow bead separation by FACS.

3.5 Rolling circle amplification

The level of extension performed by ribozymes was too low to confer a fluorescence signal through incorporation of fluorescently labelled nucleotides or hybridisation to a fluorescent probe; some form of signal amplification was required.

This was achieved by using the extended primers to trigger rolling circle amplification (RCA; Section 6.6.3). DNA minicircles were annealed to beads, and captured by ribozyme-extended primers via a stretch sharing the same sequence as the RNA template. Any weakly bound minicircles were washed off before the addition of phi29 DNA polymerase; at 37°C, this enzyme performs RCA primed by the ribozyme-synthesised RNA primer capturing the minicircle, generating bead-linked concatemers of single-stranded DNA. To detect these enlarged nucleic acid tags, DNA probes were annealed to a repeated sequence in the single-stranded DNA. These probes were used to prime conversion of the concatemers to double-stranded DNA to allow specific staining with PicoGreen (Invitrogen). Fluorescently-labelled beads, bound to active ribozyme genes, could then be isolated by FACS, allowing generation of the output selection pool by PCR. Recovery of genetic information directly from DNA rather than via RNA using reverse transcription avoided introducing a selection bias towards sequences that are transcribed and reverse transcribed well. Furthermore, after prolonged incubation in high [Mg²⁺] extension buffer, bead bound ribozymes had suffered significant degradation.

The unextended primers upon beads linked to inactive ribozyme genes are too short to capture minicircles, and so cannot trigger RCA or gain fluorescent signal. Furthermore, this mechanism distinguished between template-dependent extension and untemplated nucleotide transferase activity: the template sequence must be replicated with reasonable accuracy to allow it to capture the minicircle, providing selection pressure not just for RNA polymerase activity but, to some degree, for RNA polymerisation fidelity.

At least ten hybridised nucleotides between primer and minicircle were required to allow minicircle persistence through washing until RCA; employing minicircles that bind to positions further along the extended primers allowed the application of more stringent selection pressure (Figure 3.3). Polymerases unable to extend primers beyond a certain length would not generate fluorescent signal – only polymerases capable of extending primers processively would yield signal when using stringent minicircles.



Figure 3.3. Minicircle stringency.

Unextended bead-bound RNA primers (black) cannot hybridise sufficiently to the recognition site of DNA minicircles (red) to trigger RCA; only RNA sequences synthesised by a template-dependent RNA polymerase (grey) retain minicircles past washing. A low-stringency minicircle (**A**) requires less ribozyme-catalysed primer extension to become captured than a high-stringency minicircle (**B**).

During FACS, measurement of particle size using forward- and sidescatter can distinguish distinct populations of events corresponding to single beads, pairs of beads and larger aggregates of multiple beads. As aggregates of beads exhibit intrinsically high fluorescence, selective gating of the single bead events during analysis and sorting was necessary to accurately compare the extensionderived fluorescence of beads. The more extended primers are bound to a bead, the higher the fluorescence signal exhibited by the bead after RCA (Figure 3.4); bead fluorescence was approximately proportional to the square root of the number of extended primers bound. Furthermore, beads bound to single extended primers could be clearly distinguished from untagged beads, demonstrating the extent of signal amplification possible using RCA; previous studies have also used RCA to visualise and identify single molecules (Lizardi et al. 1998). The DNA concatemers generated by RCA can even, in sufficient numbers, yield a size increase of the bead detectable by forward-scatter (Figure 2.10C).



Figure 3.4. Sensitivity of rolling circle amplification.

Histograms of fluorescence of single-bead gated FACS events from populations of beads after RCA (minicircle: DNAcirc+5). Bead populations were bound to varying densities of BioU10-Aext, corresponding to fully-extended RNA primers, before RCA.

Further tailoring of selection pressure could be achieved through manipulation of the FACS fluorescence threshold values used to gate beads for sorting and recovery. Raising the gate fluorescence value would promote sorting of beads carrying genes encoding ribozymes able to extend more primers. However, application of stringency should be balanced to promote the selection of ribozymes able to extend more primers further.

3.6 Model selections

In order to validate CBT, model selections were performed upon mixtures of R18 and R18i ribozyme genes (R18 genes bearing a 21-nucleotide insertion) (Figure 3.1). The insertion in R18i inactivates the ribozyme and allowed differentiation of R18 from R18i by gel electrophoresis of the amplification products from sorted beads. When beads bound to either R18 or R18i genes were mixed at a ratio of 1:10, the fluorescent beads recovered in FACS after a model selection yielded exclusively R18-sized amplification products, confirming a tight genotype-phenotype linkage in the CBT workflow. Recovery of active R18 ribozymes was possible from high starting dilutions in inactive ribozymes (up to 1:10⁵ R18:R18i), indicating that enrichment factors of up to ~70,000-fold are possible in a single round of CBT (Figure 3.5A).



Figure 3.5. Model selections.

(A) Agarose gel electrophoretic separation of amplification products before (–) and after (+) one round of CBT from R18 gene-bound beads mixed in various ratios with R18i gene-bound beads. (B) 10% PAGE (SYBR Gold-stained) of amplification products after one round of CBT (96 h, 17°C; +0 minicircle stringency) from 'libraries' of R18 genes mixed with R18i genes and bound to beads at different densities.

CBT was also able to generate high-purity pools (to 84%) of R18 genes in single rounds of selection (Figure 3.5B). These model selections, rather than using a mixture of beads bound to active and inactive ribozyme genes, were set up using beads bound to a mixture of active and inactive ribozyme genes; this situation more closely reflected the mixed DNA pools encountered in selections. As a result, output pools were not as pure as in Figure 3.5A, because stochastic binding of inactive ribozyme genes to beads carrying active ribozyme genes before selection resulted in their 'hitchhiking' through that round of selection. DNA cannot be uniformly distributed amongst beads before selection, so this will result in an inefficiency in each round of selection; however, as this effect is stochastic, parasites cannot adapt to exploit it.

Lowering gene densities on beads during binding resulted in higher-purity selection outputs, as the lower density of inactive ribozyme genes leads to a reduced rate of hitchhiking of inactive genes on beads with active genes. Given the composition of the starting pool in Figure 3.5B, stochastic binding before selection would be expected to limit the output purities obtainable to 53% (1 DNA molecule per bead) or 85% (0.2 DNA molecules per bead) R18, confirming that the remainder of the CBT protocol keeps phenotype and genotype tightly linked. Thus, in later rounds of selection, where high-purity output pools are desired for screening, lower gene densities should be used; but earlier rounds should maintain higher gene densities to

increase library size. Raising the gene density significantly above 1 DNA molecule per bead could allow the exploration of larger libraries at the cost of significant hitchhiking of background genes, but also risks reducing the activity of genes bound by overwhelming the ligated ribozyme binding capacity of beads.

3.7 Evolution of R18 mutagenised variants

Having established that CBT can effectively distinguish active ribozymes from inactive ones, I assessed whether it could be used to isolate genes encoding improved ribozymes from a pool of mutagenised R18 genes. The B6.61 ribozyme (Zaher and Unrau 2007) performed more extension than R18 upon some templates, but less on others; because R18 was better characterised, and to avoid the risk of starting from a more specialised ribozyme, R18 was chosen as the wild type. A starting library of variants was generated by mutagenic PCR of the R18 gene, supplementing the amplification reaction with 8-oxo-dGTP and dPTP (containing the pyrimidine analogue 3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one) (Section 6.5). Each member of the resulting pool exhibited a 4.4% mutation rate on average per position, mainly comprising transition mutations.

This starting library used a low mutation rate to allow the isolation of individual mutations or groups of mutations that enhance ribozyme activity, without inactivating such ribozymes through a high mutational load. The library had an initial size (as measured by the total number of single-bead gated events examined for fluorescence by FACS in the first round of CBT) of 4.8×10^7 individuals. The primary limitation upon the size was the number of beads sorted during FACS, similar to that sorted in the previous ribozyme selection using FACS (Levy et al. 2005). This library was small by the standards of other *in vitro* evolution experiments; indeed, the selection leading to the isolation of the polymerase ribozyme used a library comprising over 10^{15} molecules (Johnston et al. 2001). CBT partially compensated for this by examining the activity of thousands of copies of each ribozyme variant; this rendered it sensitive to polymerase activities, allowing ribozymes that are capable of performing long extensions to be recovered even if individually they only perform such extensions rarely. Selection pressures could be applied that would otherwise allow many improved ribozymes to be only stochastically recovered.

Improved variants of a ribozyme would be expected to occur more frequently than novel ribozyme activities, but nevertheless the selection (Table 3.1) was structured to maximise the sequence space explored: the larger the library, the higher the likely activity of the best variant. The first three rounds of selection represented 'genetic drift' of the population, using long incubation times and a low-stringency minicircle requiring wild type or better activity for recovery, to deplete the pool of detrimental mutations. Ribozyme genes in the resulting pools were recombined using a staggered extension process (StEP) (Zhao and Zha 2006) optimised for short genes (Section 6.7), selected, and recombined again to generate new combinations of neutral and beneficial mutations.

Round	Starting gene density /bead	Primer	Template	Extension time (h)	DNA minicircle stringency	Library size	Gene recovery	Post- recovery	Pool polyclonal activity vs. wt (Starting pool: 6%)
1	1	BioU10-A	Ι	116	0	4.8×10 ⁷	2.85%		22%
2	0.2	BioU10-A	Ι	114	0	1×10 ⁷	25.4%		56%
3	0.2	BioU10-A	Ι	100	0	2.2×10 ⁷	17.2%	StEP	47% (71% pre-StEP)
4	0.4	BioU10-A	Ι	64	0	3.8×10 ⁷	0.78% High, 4.18% Low	StEP	35%
5	0.4	BioU10-A	II	77	0	2.2×10 ⁷	0.638% High, 2.46% Low	Gel purification	80%
6	0.2	BioU10-A	Ι	51	-3	1×10 ⁷	0.41% High, 1.83% Low		114%
7	0.2	BioU10-A	Ι	51	-3	1.2×10 ⁷	0.186% High, 0.865% Low		120%
8	0.1	BioU10-A	Ι	29	-3	4×10 ⁶	0.16% High, 0.789% Low	Gel purification	144% (High) 123% (Low)
S	creening	BioU10-A	Ι	107	P2 (+5) P3 (+3)	44 clones from each of the High and Low pools			

Table 3.1. Parameters used for selection at 17°C.

Minicircle stringency denotes the overlap (in nucleotides) with the unextended primer. 'Library size' was calculated as the total number of sorted single beads × the gene density per bead. 'Gene recovery' represents the number of positive sorted beads relative to the library size, as, regardless of the gene density, beads with the highest fluorescence likely carried a gene. In later, 'stringent' rounds, high-fluorescence beads were sorted into two gates: the brightest into a 'High' gate and the rest into a 'Low' gate (beads in each gate are given as % of total bead count), which were amplified separately and then combined at a 1:1 ratio. This recombined pool was then subjected to four additional rounds of increasingly stringent selection, recovering two gates of fluorescent beads per round and further amplifying the top one (to enrich genes encoding the most active ribozymes within the pool). Selection progress was monitored using a polyclonal primer extension assay (Section 6.1.7); after eight rounds of CBT, polyclonal pool activity had risen to exceed wild type levels (Table 3.1), and individual genes were cloned for analysis.

3.8 Isolation of improved variants

To identify candidate ribozyme genes for detailed PAGE analysis of primer extension activities, a plate-based assay was developed to screen large numbers of clones, employing conditions resembling those during the selection (ribozyme polymerase plate assay, RPA, Figure 3.6, Section 6.8).



Figure 3.6. Screening principle.

primer extension was in separate solution reactions by each ribozyme clone (i); the primer was then bound to the wells of streptavidin-coated plates, template removed *(ii)*, and а horseradish peroxidase-linked DNA probe was hybridised to the extension allowing colorimetric of wells containing extended primers (iv) generated by active ribozymes. Eluting the bound probe from the plate, and re-probing with a more stringent probe sequence that binds further downstream on the extension product (v) allows the ribozyme of clones synthesis of longer extension products (vi).

Screening 44 clones from each of the Round 8 'High' and 'Low' pools (Figure 3.7A) highlighted a number of clones that were then assayed for primer extension activity using PAGE. Many exhibited quantitatively higher extension than the wild type, but in a similar pattern; two, however – C35 and C37 – performed substantially more extension upon both the selection primer/template duplex (BioFITCU10A/I) (Figure 3.7B) and an unrelated duplex with a different sequence (B/IV), with extension by C37 superior to extension by C35.



Figure 3.7. Isolation of the improved clone C37.

(A) Scatter plot of ribozyme polymerase plate assay (RPA) signals of clones from Round 8 'High' & 'Low' recovery PCR pools, using stringent (P3) and less stringent (P2) probes. (B) Further analysis of clone C37 using denaturing PAGE of primer extension upon the selection and screening primer/template duplex sequence (BioFITCU10A/I) (17°C, 24 h). (C) The secondary structure of ribozyme C37; RTT = run-through transcript. Mutations relative to wild type are highlighted in red.

3.9 Engineering of Z, a ribozyme with improved generality

C37 comprised five mutations relative to wild type (A2 Δ , C60U, G93A, G95A, A159C) (Figure 3.7C). The ribozyme C35 was closely related to C37, missing the C60U and G95A mutations, but C37 outperformed it; thus, further engineering focused on C37 (Figure 3.8), yielding the ribozyme Z (Figure 3.9).



Figure 3.8. Engineering of Z.

Denaturing PAGE of primer extensions (17°C, 40 h) upon two distinct primer/template duplexes was used to judge the influence of selected mutations and sequence elements on ribozyme polymerase activity. All ribozymes were examined with the 5'-GGACAACC-sequence ('R10', from the R10 variant (Johnston et al. 2001), used as a ligation tag in the selection) present; the grey boxed lanes represent the activity of wild type in the selection context (although these ribozymes lacked ligated 5' hairpin; ribozymes transcribed with this hairpin encoded in the gene retained activity on primer B/template IV, but lost some on primer A/template I). Of the five mutations present in C37 (red), the removal of one (A2 Δ) did not affect activity, yielding the ribozyme Z; the other four all contributed to maximal activity (lanes 1-4). Lanes 5 & 6 show the influence of the G133U mutation isolated in a separate clone. Extension by R18 on the selection primer/template (A/I) was inhibited by the

run-through transcript (RTT), while C37 (selected in its presence) and Z (derived from C37) were not; Z was actually adapted to its presence, showing modestly reduced activity upon removal of this sequence. Due to the arbitrary nature of the RTT sequence, it seems probable that any sequences that emerge during further evolution could compensate for its absence; indeed, replacing the RTT with a lengthened terminal hairpin mostly attenuated the reduction in Z's activity.



Figure 3.9. Predicted secondary structures of R18 and Z.

Z possesses four mutations (in red) relative to R18; it also comprises a 3' run-through transcript (RTT) and a 5' ligation tag sequence (GGACAACC) absent in R18, but shows no dependence upon and reduced potential base-pairing with the stem oligonucleotide.

Z exhibited improved activity compared to R18 on all primer-template duplexes tested (Figure 3.10), extending more primers upon the classical primer A/template I duplex, and generating longer extension products opposite other templates. The four mutations relative to wild type confer a more general polymerase activity upon Z, better tolerating the synthesis of different sequences.

The C60U mutation incrementally enhanced RNA polymerase activity; it was the only mutation selected in the catalytic core, and changed a G-C base-pair to a weaker G-U wobble-pair in a central stem (with unknown consequences for local structure or dynamics). The catalytic core has generally proved resistant to mutation in selection experiments, although this mutation is present in the core of a different family of ribozyme polymerases (Lawrence and Bartel 2005).



Figure 3.10. Generality of primer extension by Z.

Denaturing PAGE of polymerisation reactions (17°C, 40 h) using R18 and Z upon a range of primer/template duplexes (some chosen to approximately equally represent each nucleotide and each dinucleotide pair).

The G93A and G95A mutations in the linker region disrupt interaction with (and render ribozyme activity independent of) the stem oligonucleotide. Mutations in the stem-pairing region were selected independently several times in separate selections, suggesting that stem presence, though beneficial for the wild type on some templates, is ultimately limiting to the ribozyme's polymerisation potential. G93A and G95A may merely be necessary to tolerate a single-stranded stem-pairing region in the absence of stem, or they could promote new interactions with primer-template duplex or the ribozyme processivity domain.

A159C is located in the processivity domain (residues 98-187) as part of a large asymmetric interior loop, later suggested to form a four-base helix (Wang et al. 2011). A159C allows the formation of a new G133:C159 base pair, augmenting this four base-pair stem formed by bulge segments A129-U132 (5'-ACCU) and A160-U163 (5'-AGGU). Indeed, G133U, a separate beneficial mutation that was isolated in the selection, would stabilise the stem in an identical way, by promoting formation of an U133:A159 base pair (Figure 3.8, lane 5). These two mutations, while beneficial individually (with A159C superior to G133U), negate one another's effects when combined (Figure 3.8, lane 6). This strongly suggests that the selected trait is indeed the formation of a new base pair between positions 133 and 159, presumably to enhance this structure in the processivity domain. The effects of the introduction of

other mutations into Z upon polymerase activity at 17°C were studied using primer extension assays on two different primer/template duplexes (see Figure 4.6 for analogous examples). However, activity was not further improved either by the introduction of mutations present in other isolated clones (or the A168G mutation from B6.61 (Zaher and Unrau 2007)), or by strengthening of the 129-133 helix (through lengthening the helix by a base pair or replacing A-U base pairs with C-G base pairs).

Examination of the overall secondary structure of the Z ribozyme reveals several instances of unpaired adenines in non-helical regions; indeed, over half (18/35) of non-helix, non-tetraloop residues in the processivity domain were adenines. Unpaired adenines may participate in tertiary interactions with other structural elements – for example, in the ribosome, where they are the base most frequently involved in such interactions (Nissen et al. 2001) – and may be crucial for primer/template duplex interaction in R18, where A-minor patches interact with the minor groove of separate RNA helices in a largely sequence independent manner (Shechner et al. 2009). Unpaired adenines may be key both to stabilising ribozyme tertiary structure and to imparting sequence generality to polymerase ribozymes, and may mediate such interactions in the polymerase domain as well as the ligase core (Wang et al. 2011).

How did the selection lead to the emergence of generality? Template I was used throughout the selection, with the exception of one round where template II was used – which differs from template I only in its downstream sequences, and was employed to discourage the emergence of some specific ribozyme-template interactions. Yet, the largest improvements in extension are seen using different primers and templates. One factor could have been the context of the selection duplex: a 5' U10 linker is used to space the primer out from the bead surface (and is critical for maximising on-bead extension). This linker, along with the RTT sequence, results in much lower primer extension on template I in the selection context (Figure 3.7B) by the wild type than it is capable of. Perhaps this inefficiency resulted in selection to overcome these inhibitory obstacles, yielding a ribozyme capable of coping with a wider range of duplexes and sequences, and obscuring the polymerase preference for the classical primer/template duplex.

3.10 Harnessing template recognition for RNA synthesis

A selection using a similar CBT protocol was carried out upon a library of R18 variants with an additional 5['] 48-nucleotide stretch of random sequence (Wochner et al. 2011), to promote the emergence of a third ribozyme domain that interacts in a non-sequence-specific manner with the upstream duplex, in analogy to the 'thumb' domain possessed by many proteinaceous polymerases. Subsequently, the structure of the class I ligase indicated that the wild type already interacts with all the upstream duplex provided in the selection; however, after three rounds of selection using template I, a clone was isolated (C19, Figure 3.12A) capable of improved extension upon the selection primer/template duplex (Figure 3.11A).

This enhancement was specific to this primer/template duplex, and appeared to be mediated by the interaction between a short single-stranded stretch in the evolved domain of the ribozyme (ss_{C19}) and a complementary sequence at the 5' end of the template downstream of the primer. Indeed, mutation of this template sequence led to a reduction in primer extension, which could be partially restored by compensating mutations in ss_{C19} (Figure 3.11B). This 'template recognition' likely allows hybridisation of C19 upon the template, compensating for the wild type's high K_M for primer/template duplex.



Figure 3.11. Template recognition enhances extension.

Denaturing PAGE of primer extension reactions (17°C, 24 h). (A) Extension of primer A by the R18 and C19 ribozymes on the selection template I. (B) Extension or primer A by C19

upon templates based on template I with successive point mutations (red) in the ss_{C19} binding site ('template mutation'), and with compensating mutations (to reconstitute a 6 nucleotide hybridisation site) in C19 ('template mutation + ss_{C19} mutation'). ss_{C19} and the respective binding site are depicted in green. **(C)** Extension of primer A on template I-3 by the engineered ribozyme tC19 compared to by the R18 and C19 ribozymes.

There is potentially a further element of positioning in this interaction; C19 seemed structurally adapted to the extension of primers near to the template binding site, as this ribozyme exhibited little improved extension upon a primer/template duplex containing a longer template (I-3) with a recognition site more distant from the primer. However, truncation of the 5[′] domain of C19 yielded an engineered variant (tC19, Figure 3.12B) capable of extending 13% of primer A molecules by up to 26 nucleotides opposite such a template (Figure 3.11C), unlocking the benefits of downstream hybridisation upon longer templates.



Figure 3.12. Secondary structures of C19-derived ribozymes.

(A) C19; evolved 5´48-nucleotide domain in orange. C19 also has a G93A mutation (orange) that renders its activity stem-independent (ref). (B) The truncated C19 ribozyme tC19; engineered residues in green. (C) tC19Z; mutations introduced from Z in red.

tC19 allowed the synthesis of some long RNAs, but its polymerase activity, like that of its parent R18, remained template dependent. While capable of long primer extension on favourable templates, synthesis upon a majority of RNA template sequences was limited. To improve generality, Z's core mutations were combined with the 5'-extension of tC19 to yield the hybrid ribozyme tC19Z (Figure 3.12C), exhibiting the improved sequence generality and extension capabilities from each ribozyme. Though its activity was still not independent of the template sequence, tC19Z outperformed all its parent ribozymes (R18, tC19, Z), synthesising longer extension products on a range of different primer/template sequences (Figure 3.13A).



Figure 3.13. Generality and hammerhead synthesis.

(A) Denaturing PAGE of polymerisation by R18, Z, tC19 and tC19Z on different primer/template duplexes containing a downstream ss_{C19} binding site (17°C, 24 h). (B)

Secondary structure of hammerhead endonuclease minizyme with ribozyme-synthesised segment (green) and substrate (red). Essential catalytic residues are boxed. **(C)** Denaturing PAGE of extensions of primer A on the minizyme template MzTemp by ribozymes R18, Z, tC19 and tC19Z (left panel; 17° C, 24 h). tC19Z-synthesised extension products long enough to form a symmetrical minizyme with the substrate (+24/≥+27, red boxes) were purified and tested for endonuclease activity (right panel, denaturing PAGE of cleavage reactions; S, substrate (MzSub); CP, cleavage product). Substrate (red) is specifically cleaved by the control chemically-synthesised minizyme (+, Mz) and by tC19Z-synthesised minizymes (+24/≥+27, green), but not in their absence (–); see (Wochner et al. 2011).

The tC19Z ribozyme exhibits RNA synthesis opposite a range of arbitrary template sequences. This raised the question of whether this ribozyme could complete the synthesis of an RNA sequence with a phenotype. An RNA template was designed that encodes an RNA sequence possessing a catalytic activity: a hammerhead nuclease ribozyme. To facilitate the synthesis of sufficient amounts of full-length ribozyme for characterisation, a minimal version of the hammerhead endonuclease designed for therapeutic applications was chosen (McCall et al. 2000) (Figure 3.13B). In contrast to R18, the tC19Z RNA polymerase ribozyme could synthesise full-length hammerhead minizymes, harnessing an ss_{C19} binding site on the 5' end of the template to polymerise a \geq 24 nucleotide stretch representing all the catalytic residues and a stem of the minizyme (Figure 3.13C, left panel). Purified extension products corresponding to both the 24-nucleotide and \geq 27-nucleotide bands exhibited catalytic activity and performed sequence-specific cleavage of a cognate substrate RNA (Figure 3.13C, right panel).

The ribozyme-catalysed synthesis of a functional ribozyme corresponds to a key aspect of ribo-organism life cycles – the RNA-catalysed transcription of RNA genes. The ss_{C19} sequence tag, which emerged rapidly in the selection, represents a straightforward, beneficial mechanism for enhancing ribozyme polymerisation opposite upstream RNA sequences constituting the template 'genes' of a riboorganism. Specific interactions between a replicase and cognate templates via a recognition tag that promotes polymerisation could allow a ribo-organism to comprise a number of separate RNA molecules – a replicase, additional ribozymes performing other functions, and their corresponding template genes.

3.11 Template evolution

To fulfil such a role for a replicase, a recognition tag must be capable of enhancing polymerisation when separated from the primer by many more template residues, which themselves must not interfere with tag-based replication. To assess whether the ss_{C19} sequence can promote replication when located further downstream, a template selection scheme was developed to obtain longer replicable template sequences. A library of 5×10^{13} RNA template molecules was prepared based on template I-3, with a region of template between the primer and ss_{C19} binding site replaced by a 36-nucleotide random-sequence stretch (Figure 3.14A). Biotinylated RNA primers were extended on these templates by tC19; fully-extended primers were then purified (Figure 3.14B), and RT-PCR was used to recover the sequences of templates that could be transcribed by the ribozyme (Figure 3.14C). This selection would be expected to favour template sequences that are easily replicable by ribozyme polymerases, mimicking evolutionary processes in the RNA world - where replication templates as well as replicases themselves would have been under selective pressure to co-evolve towards maximum replication efficiency. The absence of selection for function in this system, however, would allow the isolation of sequences that directly reflect ribozyme template sequence preferences.

Template sequences cloned from the first round of selection could typically be extended only by a few nucleotides; however, many appeared to exhibit less secondary structure than random sequences of the same length, suggesting some degree of selection. One possible explanation was that templates remained bound to primers during bead washing after extension (promoted by limited ribozyme-catalysed extension upon them), and their recovery during gel purification would have been promoted by their resulting migration nearer to desired extension products. Exposure to proteinaceous polymerases during RT-PCR could then have favoured completion of primer extension upon them and sequence recovery by RT-PCR.

To promote recovery of templates fully transcribed by tC19, a second round of template selection was performed upon the first round templates; these were transcribed in the presence of guanosine monophosphate. After extension and initial template removal, XRN-1 exoribonuclease was used to degrade remaining bound templates, before extended primers were 3' blocked through addition of dideoxyGTP by terminal transferase, to prevent any later protein-catalysed extension. Over half of the template clones isolated in this round of selection appeared to be derived from a single template, I-5 (Figure 3.14D), upon which tC19 could extend 1.5% of primers by \geq 47 nucleotides.



Figure 3.14. Selection of replicable templates.

(A) RNA primers (BioFITCU10-A, BioU10-A) were extended by the tC19 ribozyme (17°C, 84 h) on a library of template sequences (ss_{C19} binding site in red). (B) Extended primers were bound to beads, to allow removal of template strands through repeated heating and washing in urea buffer. Extended primers were then stripped from the beads and resolved by denaturing PAGE, allowing selective purification and recovery of extension products of ~50 nucleotides length. (C) RT-PCR was then used to recover the DNA sequences of templates that had been replicated by the tC19 ribozyme. The presence of a primer-specific mutation confirmed that these sequences were derived from extended primers (and not from contaminating template). The round 1 template pool was subjected to a second round of selection ($17^{\circ}C$, 92 h) with additional safeguards against template persistence, which yielded template *I*-5 **(D)**, upon which extension by tC19 could proceed beyond 50 nucleotides.

Template I-5 resembles template I-3, but with two additional repeats of a central 11-nucleotide repeat resembling the sequence of the classical template I. Although this sequence likely derived from recombination amongst low levels of contaminating template I-3-encoding DNA after the first round, its isolation demonstrates the potential of the template selection scheme to identify replicable templates.

3.12 ss_{C19}-mediated synthesis of long RNAs

A series of templates was generated based on template I-3 with increasing numbers of central 11-nucleotide repeats. Upon these templates tC19 could synthesise up to 95 nucleotides of RNA, fully extending primers up to near the ss_{C19} binding site on the template, demonstrating that the benefits of downstream recognition tags can apply even to replication of long templates (Figure 3.15). Indeed, although no fully extended product was observed when primer extension by over 100 nucleotides was probed, substantial polymerisation enhancement was nevertheless observed on template I-10.

Yields of fully extended products are limited by a number of factors, including ribozyme dissociation and ribozyme and product degradation. Together these result in, on average, 7% of extension products terminated at each template position by the end of the incubation, as judged by the quantified proportion of primer extended beyond each position in Figure 3.15. Combined with the ~40-60% of primers that remain unextended, this yielded only 0.035% of products fully extended by \geq 91 nucleotides opposite template I-9. The terminated rate varied within each extension in a template sequence-dependent manner. However, the average terminated rate was broadly similar (between 6.8% and 7.5% per position) for extensions on templates I-4 to I-9, where template length differed by over 50 nucleotides – indicating that the effectiveness of the recognition tag is relatively independent of the distance of the ss_{C19} binding site from the primer over this range.


Figure 3.15. Sequence-tag mediated synthesis of long RNAs.

Denaturing PAGE of extension of BioFITC-A on the engineered template series I-n by the tC19 and R18 ribozymes (17°C, 7 d). 'n' indicates the number of repeats of the central 11-nucleotide sequence between the primer and ss_{C19} binding sites. The schematic depicts primer extension by tC19 on TI-n. The inset shows a high-resolution scan of the boxed area.

tC19Z was also capable of significant RNA synthesis upon this series of templates, but seemed to pause more often along the sequence, hampering its ability to synthesise the longest products (Figure 3.16); polymerisation of up to 74 nucleotides was detectable. This limitation represents, perhaps, the price of generality: tC19Z exhibited attenuated extension capabilities when the template consists of repeats based on the template sequence most easily transcribed by R18.



Figure 3.16. Synthesis of long RNAs by tC19Z.

Denaturing PAGE of extension of BioFITC-A on the engineered template series I-n by the tC19Z and R18 ribozymes (17°C, 7 d). 'n' indicates the number of repeats of the central 11-nucleotide sequence between the primer and s_{C19} binding sites.

3.13 Fidelity of synthesis of long RNAs

To confirm the template-dependent synthesis of long RNAs, fully extended products from Figure 3.15 and Figure 3.16 were excised, cloned and sequenced (Figure 3.17). While comparisons are imperfect due to differences between the sequences sampled for the different ribozymes, tC19 (97.3% fidelity) seemed to polymerise RNA more accurately than R18 (95.7% fidelity).

Z contains the C60U mutation in the catalytic core as well as several other mutations that could potentially affect its fidelity. The effect of these mutations upon the fidelity of Z could be assessed by comparison of the error spectra of RNA synthesised by tC19 and tC19Z; tC19Z generated very few errors (99.1% fidelity), particularly when synthesising minizyme catalytic subunit, where only two $G \rightarrow A$ transitions occurred in the 999 positions sequenced.



Figure 3.17. Error spectra of RNA synthesis by tC19.

Error rates observed in RNA molecules synthesised at 17°C by (A) R18 (Table 2.1), (B) tC19 (comprising 46 fully extended products on template *I*-6, 3 upon *I*-7, 1 upon *I*-8, 8 upon *I*-9 and 1 upon *I*-10) and (C) tC19Z (comprising 21 fully extended products on template *I*-6 and 37 upon MzTemp). The error rate represents the average percentage of each error type occurring per position for each nucleobase. The fidelity of RNA polymerisation by each ribozyme can be estimated by generating a geometric average of the observed fidelities opposite each base.

These evolved ribozymes are clearly capable of synthesising long RNAs with great accuracy. But how well can sequencing fully extended primers represent the true accuracy of polymerisation carried out by these ribozymes? Errors made during polymerisation that stalled the ribozyme would not be detected upon sequencing of fully-extended products; given that such products represent only a small fraction of all synthesis, sequencing them alone risks overlooking a substantial fraction of mistakes. Indeed, lighter bands with altered mobility that run between two intermediate bands are often visible in ladders of extension products, and likely represent misincorporation or insertion/deletion of nucleotides. However, sequencing of intermediate bands does not reveal a substantial increase in mutation frequency at the final position (Figure 2.9D), indicating that most intermediate bands result from ribozyme stalling or degradation. Furthermore, the area of bands excised from gels prior to sequencing is large enough to encompass extension products with altered mobility due to errors. Indeed, fully-extended products contain errors distributed throughout their sequences, demonstrating that ribozyme polymerisation can continue to some degree after making an error, and thus that we would expect any tendency to generate errors during synthesis to be represented in fully-extended products.

It could be argued, from the point of view of the error threshold (Eigen 1971), that the accuracy of full-length extension is the most important fidelity criterion, in analogy to the behaviour of nonenzymatic templated polymerisation (Rajamani et al. 2010): truncation of sequences after error incorporation would grant the master sequence a substantial selective advantage, promoting the maintenance of genetic information through cycles of replication. Regardless, taking the above factors into account, sequencing of fully extended products likely gives a good picture of the relative fidelities of the ribozymes. How well these values correspond to the true fidelities depends upon the quantitative parameters of polymerase behaviour upon error generation. The error rate of R18 suggested by such experiments, however, closely resembles that obtained by assessing the relative efficiencies of nucleotide incorporation and misincorporation (Johnston et al. 2001). Improved measures of fidelity would be provided by sequencing all primers in a reaction; this would be facilitated by the characterisation of ribozymes able to substantially extended

the majority of primers in a reaction. A final caveat is that the error rate, like the polymerisation rate, is highly sequence-dependent, and a full picture of ribozyme polymerase fidelity would require assessment of accuracy in different sequence contexts.

3.14 Discussion

The resistance of the R18 ribozyme to further improvement by directed evolution had raised the possibility that the lineage represented a 'dead-end' on the road to a replicase (Joyce 2007). However, application of CBT to libraries of lightlymutated ribozymes demonstrates that the sequence space surrounding the wild type is in fact richly populated with improved variants. By selecting directly for processive extension and limiting the opportunities for parasite emergence, repeated rounds of selection and recombination can generate and enrich ribozymes with notably improved activity.

Z's additional sequence generality extracts the ribozyme lineage from the local activity maximum of R18 extension upon primer A/template I. However, its activity still exhibits substantial sequence dependence, which must be overcome to some degree to grant it the capacity to replicate its own sequence. Generality may depend upon the emergence of further non-sequence-specific interactions between the ribozyme (or any novel domains thereof) and the duplex, perhaps allowing current interactions to be relaxed (along with existing sequence preferences). To this end, selections should employ lengthier primer/template duplexes (both upstream and downstream of the primer) to provide elements for the ribozyme to interact with. Further evolution using CBT should also rotate between a number of distinct primer/template duplex sequences for each round of selection, both to encourage the emergence of generality, and reduce the risk of specialists such as C19 dominating the selection pool.

The C19 ribozyme-template interaction emerged rapidly during selection, likely because the template sequence was maintained during selection. Although at first glance this appeared to be detrimental to the emergence of generality, downstream hybridisation actually enhanced the transcription of a range of upstream sequences (Figure 3.13), and on some templates can exert beneficial effects over long intervening sequences (Figure 3.15).

The simplicity of such a beneficial interaction suggests that it could have been exploited by the very first replicases. These are thought to have emerged from pools of random RNA sequences generated by nonenzymatic polymerisation of suitably activated nucleotides (Robertson and Joyce 2010). The minimal complexity of a ribozyme capable of template-directed nucleotide polymerisation is not known, although they are thought to have been extremely rare. A process of templatedependent nonenzymatic polymerisation on single-stranded RNAs, coupled to physicochemical processes facilitating strand separation, could result in a bias towards longer sequences (if such a process occurs faster than nontemplated polymerisation) (Manapat et al. 2010); sequence information can be replicated accurately nonenzymatically, aided by a process of postmismatch stalling (Rajamani et al. 2010). Such a process would ensure that nascent replicases were generated together with their complementary template sequences to copy. Critically, in this "prelife" setting (Nowak and Ohtsuki 2008), sequences with good templating properties would be replicated faster than others competing for nucleotides, enriching the pool of sequences in those that are easily replicable – even before the emergence of enzymatic catalysis. As this benefit would likely apply to both nonenzymatic and enzymatic replication, the chance that any nascent polymerasetemplate pair could act as a replicator may be significantly increased.

In such a scenario, given the complementarity of polymerase and template, many positions along the ribozyme could serve as replication tags. A 3' tag recognition site on the template upstream of the primer (e.g. Figure 2.3) would only enhance polymerisation at initiation, due to the lengthening of the intervening nascent duplex. However, the polymerisation enhancement resulting from a downstream tag recognition site – as used by tC19 - can (so long as template folding allows) continue to promote primer extension; the closer this site to the end of the molecule, the more of the template benefits during replication. It is also possible for template folding to play a positive role, facilitating polymerisation by localising the 3' end of the primer near the tag recognition site where the polymerase binds; indeed, random sequence RNAs are predicted to tend to fold such that the ends of

the RNA are near to one another (Yoffe et al. 2011). Both the gene and the polymerase require a tag recognition site at their 5' ends to benefit from tagpromoted replication when serving as templates, and a complementary sequence at their 3' ends to encode the converse recognition site and (in the polymerase) act as a tag. Given that short sequences are sufficient to act as tags, such combinations would arise frequently amongst primordial random-sequence RNA molecules; all that is then required is the correct structural context to ensure their use, orienting target molecules for replication whilst avoiding intramolecular tag-tag site hybridisation. Tag sequence duplication in the replicase could allow its migration and use at any other position in the replicator, such as near the 5' end (as in tC19).

This sequence-specific interaction of the ribozyme with the 5'-end of the template is reminiscent of the recognition of target mRNAs by the prokaryotic ribosome through the Shine-Dalgarno sequence (Shine and Dalgarno 1975). Such recognition might have been particularly advantageous in a prebiotic setting, where an RNA polymerase ribozyme would not have evolved in isolation but in the presence of a large number of unrelated RNA oligomers. Related single-stranded RNA molecules encoding a polymerase and its template that share simple recognition tags could selectively cooperatively replicate, promoting a primitive form of kin selection and enabling evolution even in the absence of compartmentalisation within protocellular entities. For example, an RNA molecule that modestly enhances the rate of nonenzymatic nucleotide polymerisation, yet which does not dissociate completely from its complement (remaining hybridised by a tag sequence), would selectively promote the replication of its complement even in the presence of other RNA molecules with different sequences. Relatively simple polymerisation enhancements may then be sufficient to promote the emergence of more capable and specific replicases. Furthermore, although 'parasitic' molecules possessing the tag but no function would easily emerge and slow replicase reproduction, tag sequence drift could allow some degree of parasite evasion. Such replicases may represent precursors to more complex ribo-organisms than are linked to their template in series, and replicate intramolecularly (Pace and Marsh 1985). Ultimately, however, compartmentalisation would yield more genetically homogeneous and efficient ribo-organisms, better protected from the emergence of evolved parasites.

4 Ribozyme Evolution in Ice

4.1 Introduction

Ice, as a reaction environment, can provide a number of important benefits to ribozyme replication. Firstly, low temperatures stabilise ribozymes, prolonging activity and allowing the synthesis of longer RNA molecules. Ribozymes can also benefit from the concentration of solutes within the supercooled eutectic phase by ice crystal growth; this can significantly boost ribozyme activity in substrate-poor environments. Finally, freezing some solutions generates a fragmented eutectic phase that could provide cell-like compartmentalisation to ribozyme replicases, facilitating Darwinian evolution through kin selection. Although these properties of ice were described using the R18 ribozyme, they would be expected to benefit most ribozyme polymerases. R18 is active in ice at a temperature (-7°C) almost 30°C below the temperature at which it was evolved (22°C) (Johnston et al. 2001), and thus may not represent the full potential of RNA catalysis in this environment. This raises the question: can ribozyme polymerases be isolated that do not just tolerate freezing conditions, but actually take advantage of them? To explore this, ribozymes must be evolved in ice, with the eutectic phase hosting the selective step. Adaptation of R18 to low temperatures represented an accessible initial goal.

4.2 Selection for ribozyme polymerase activity in ice

In vitro selection of nucleic acids has traditionally been carried out at ambient temperatures, close to those at which ribozymes are active in nature. To explore evolution under the frozen conditions at which ancestral ribozymes may have operated, a selection system for improved RNA polymerase activity must satisfy certain criteria. Firstly, R18 exhibits a very low affinity for primer/template duplexes at ambient temperatures, and primer extension occurs even less frequently at subzero temperatures (Figure 2.2). Thus, the selection system must be capable of sensitively detecting the extension signature of ribozymes active in ice. Secondly, the longer incubations necessary to achieve significant synthesis of extension products by iceactive ribozymes necessitate maintenance of a tight genotype-phenotype linkage during the incubation. CBT fulfilled both of these criteria: RCA provides powerful signal amplification of ribozyme activity (Figure 3.4), and combined with the sampling of activity of thousands of ribozyme copies per gene, renders CBT sensitive to weak ribozyme polymerase activities. Furthermore, bead linkage of genes, ribozymes and extension products facilitates a tight genotype-phenotype coupling, allowing high-purity pools of active ribozyme genes to be obtained after model selections in ice (Figure 4.1).



Figure 4.1. Model selection in ice.

PCR amplification products after one round of CBT, with the selective step at –7°C in ice (132 h, +7 minicircle stringency; 10% PAGE, SYBR Gold-stained). The starting library comprised R18 genes mixed with genes encoding the inactivated R18 variant with a 21-nucleotide insertion (R18i), and was bound to beads at different densities prior to selection. Sorted pool activity is described by the percentage of R18 genes in the recovery PCR pool. Once again, lower gene densities on beads resulted in higher-purity selection outputs.

The CBT protocol for selection in ice (Figure 4.2) closely resembles the standard CBT protocol. However, after transcription/ligation and primer decoration, beads were not emulsified but were encased in the ice phase by freezing in fourfold-diluted extension buffer and incubated at -7° C. The concentration effect of eutectic phase formation counteracted any effect upon eutectic phase composition of this dilution, which served to better space out the beads within the ice. Ribozymes remained ligated to the beads during the incubation, limiting any ice-active ribozymes to extending primers on their own bead. Beads were recovered from the ice by thawing with EDTA, and processed as before.



Figure 4.2. Compartmentalised bead-tagging in ice.

To begin a round of selection, a library of ribozyme genes was bound to streptavidin-coated microbeads at a density of up to one gene per bead (i). Ribozyme was transcribed and ligated to RNA hairpins on beads in the compartments of a water-in-oil emulsion (ii; inset shows light microscopy of selection emulsion, field diameter ~0.15 mm), generating clonal bead-bound repertoires of ribozymes. After recovery from emulsion and decoration with substrate primers (iii), the beads were frozen (at -25° C) in diluted ribozyme extension buffer with RNA template, and incubated at -7° C to allow eutectic phase formation (iv; inset shows cryo-SEM of selection ice, field diameter 0.15 mm). After thawing and recovery of beads, extended primers were used to prime rolling circle amplification (RCA) (v); the resulting single-stranded DNA concatemers were converted to a double-stranded form and stained by Picogreen (vi), allowing separation of fluorescent beads using FACS (vii). The genes bound to such beads, encoding ice-active ribozymes, were then recovered using PCR (viii).

4.3 Evolution of R18 mutagenised variants in ice

To begin to explore the potential for ribozyme evolution in ice, CBT ice selection was applied to the library of mutagenised R18 ribozyme genes (Section 6.5), in parallel with the selection at 17°C. The low level of mutagenesis (4.4% per position) allowed a search for small sets of mutations sufficient to adapt the ribozyme to activity in ice. The polyclonal starting library exhibited 11% of wild type activity in ice; after eight rounds of CBT selection, the in-ice RNA polymerase activity of the polyclonal selection output (Section 6.1.7) had increased to up to 84% of wild type activity (Table 4.1).

Round	Starting gene density /bead	Primer	Template	Extension time (h)	Minicircle stringency	Library size	Gene recovery	Post- recovery	Pool polyclonal activity vs. wt (Starting pool: 11%)
1	1	BioU10-A	Ι	186	+7	5.2×10 ⁷	2.72%		15%
2	0.2	BioU10-A	Ι	550	+7	1.5×10 ⁷	35.35%		26%
3	0.2	BioU10-A	Ι	172	+7	2.8×10 ⁷	17.83%	StEP	26% (47% pre-StEP)
4	0.4	BioU10-A	Ι	325	0	4.7×10 ⁷	1.06% High, 3.74% Low	StEP	24%
5	0.4	BioU10-A	II	256	0	2.3×10 ⁷	0.548% High, 1.93% Low	Gel purification	36%
6	0.2	BioU10-A	Ι	115	0	1.5×10 ⁷	0.35% High, 1.67% Low		56%
7	0.2	BioU10-A	Ι	124	0	1.1×10 ⁷	0.235% High, 0.995% Low		69%
8	0.1	BioU10-A	Ι	62	0	4×10 ⁶	0.171% High, 0.819% Low	Gel purification	53% (High) 84% (Low)
Screening		BioU10-A	Ι	334	P2 (+5) P3 (+3)	44 clones from each of the High and Low pools			

Table 4.1. Parameters used for selection in ice.

Minicircle stringency denotes the overlap (in nucleotides) with the unextended primer. 'Library size' was calculated as the total number of sorted single beads × the gene density per bead. 'Gene recovery' represents the number of positive sorted beads relative to the library size, as, regardless of the gene density, beads with the highest fluorescence likely carried a gene. In later, 'stringent' rounds, high-fluorescence beads were sorted into two gates: the brightest into a 'High' gate and the rest into a 'Low' gate (beads in each gate are given as % of total bead count), which were amplified separately and then combined at a 1:1 ratio.

Initial rounds of selection, as before, focused on low-stringency 'drift' to deplete the pool of detrimental mutations, followed by StEP-mediated recombination and subsequent selection with increasing stringency, through reduction of incubation time down to ~ 2 days in the final round to select for faster ribozymes. This had the detrimental effect of necessitating selection based on relatively low FACS signals; the corresponding sorting gates were set at much lower fluorescence values than in the 17°C selection. Polyclonal analysis indicates that, as a result, the 'High' gate in the final round contained a large proportion of 'noise' - the low rate of MoFlo events that are gated independent of extension as high-fluorescence single bead events due to noise in optics and fluorescence detection, carrying with them whichever unselected genes happen to be bound - lowering polyclonal pool activity. The 'Low' gate, corresponding to a much narrower fluorescence range, and thus much less vulnerable to sorting noise, continued to increase in activity. Although highstringency selection exposes the most active mutants, more FACS background events and molecules are picked up in lower, wider fluorescence gates; future selections must avoid 'sailing too close to the wind', and maintain sufficient extension-derived fluorescent signal to allow the use of high-fluorescent gates that catch less noise.

4.4 Isolation of improved clones

The enriched Round 8 pools were screened for individual genes encoding ribozymes with improved in-ice RNA polymerase activity using an RPA protocol modified for screening for activity in ice (Section 6.8). This screening highlighted a number of clones of interest; gel-based assays of primer extension in ice were performed using twelve of these ribozymes, many of which exhibited quantitatively higher extension than the wild type but in a similar pattern. Two clones, however, with distinctly improved activity were identified: C8 and C30 (Figure 4.3). Both originated from the Round 8 'Low' pool.



Figure 4.3. Isolation and engineering of Y.

(A) Scatter plot of ribozyme polymerase plate assay (RPA) signals of clones from Round 8 'High' & 'Low' recovery PCR pools, using stringent (P3) and less stringent (P2) probes. (B) Further analysis of clones C30 and C8 using denaturing PAGE of primer extension by these ribozymes and the selection wt upon the selection and screening primer/template duplex sequence (BioFITCU10A/I) and (C) an unrelated primer/template duplex (B/IV) ($-7^{\circ}C$ ice, 0.5 mM each NTP, 162 h). (D) The secondary structures of ribozymes C30 and C8; RTT = run-through transcript. Mutations relative to wild type are highlighted in colour. (E) Engineering of Y from clone C8. Denaturing PAGE of primer extension ($-7^{\circ}C$ ice, 0.5 mM each NTP, 162 h) upon two unrelated primer/template duplexes was used to judge the effects of addition/removal of mutations and elements, including a mutation (C60U) from Z. Each of the three mutations from C8 was important for the Y phenotype (ribozymes 1-3).

C30 gave an improved RPA signal with the lower-stringency probe, while C8 gave higher signals with the more stringent probe as well; this was reflected in the pattern of primer extension performed by these ribozymes upon the selection/screening primer/template duplex (Figure 4.3B). C30, however, performed more extension upon the unrelated primer B/template IV duplex (Figure 4.3C); of its eight mutations relative to wild type, two (in the stem-pairing region) were shared with Z. Indeed, the pattern of extension performed by C30 resembled to some extent that of Z; however, it transpired that Z retained much of its activity in ice, outperforming C30 (Figure 4.6). Thus, further engineering focused on C8, yielding the ribozyme Y (Figure 4.3E, Figure 4.4B).



Figure 4.4. Secondary structures of ribozymes evolved and engineered in ice. (A) The wild type R18 ribozyme construct; residues not mutagenised in the starting library are shown in grey. (B) The evolved ribozyme Y, with mutations relative to wild type shown in blue. (C) tC19Y, with the 5' engineered recognition tag (s_{C19}) shown in orange.

Y exhibited enhanced RNA polymerase activity in ice, particularly upon the template sequence present in the selection (Figure 4.5); it could efficiently extend such a primer-template duplex by up to 17 nucleotides. Y possesses four mutations relative to wild type: U72G, G93A and C97A (from clone C8), and C60U (derived from *Z*). C60U acts to marginally increase the activity of the ribozyme; further combinations of mutations from improved clones generally failed to yield more active ribozymes (examples are given in Figure 4.6). The improved extension by Y at low temperatures stemmed from the mutations isolated in the selection, which function together to provide a cold-active phenotype (Figure 4.3). The selection system was able to isolate several sets of mutually dependent mutations. Some mutations and sequences (e.g. C60U, ss_{C19}) can operate in different contexts, but many are incompatible, likely through effects upon either structure or folding.



Figure 4.5. Activity of Y and R18 at -7° C in ice and at 17° C.

Denaturing PAGE of primer extensions by wt and Y ribozymes upon three unrelated primer/template duplexes (left panels: $-7^{\circ}C$ in ice, 0.5 mM each NTP, 9 d; right panels: $17^{\circ}C$, 4 mM each NTP, 41.5 h).



Figure 4.6. Primer extension in ice by ribozymes with combinations of mutations. Denaturing 10% PAGE (stained by SYBR Gold) of primer extensions ($-7^{\circ}C$ ice, 8 d, 0.5 mM each NTP) by ribozymes containing combinations of mutations from different clones (from C37 in red, C8 in blue, C30 in green, from C37 & C30 in orange, from all three in brown). The best ribozymes derived from C37, C8 and C30 are Z (3^{rd} lane), Y (7^{th} lane) and C30+A159C (13^{th} lane) respectively. Extension was carried out by unpurified ribozymes upon two primer/template duplexes with 22 bases of additional upstream duplex (32A/32A, based upon A/I, and 32B/32B, based upon B/IV), assessing ribozyme tolerance of upstream sequence.

4.5 Low temperature adaptation of Y

The mutations in Y also improve ribozyme activity at 17° C (Figure 4.5). However, the stability of the ribozyme at low temperatures allows its polymerisation to continue for longer, and extension by Y at -7° C in ice overtakes that at 17° C after two days' incubation (Figure 4.7).



Figure 4.7. Time courses of primer extension by Y and R18.

Denaturing PAGE of extensions of primer A upon template I by wild type and Y ribozymes, incubated at $17^{\circ}C$ (upper panel) and at $-7^{\circ}C$ in ice (lower panel) for the indicated times. Furthermore, in aqueous solution, Y adds $36\times$ more nucleotides per primer than the wild type ribozyme upon the selection template at -7° C, but only $4\times$ more at 17° C, indicating a degree of cold-adaptation of Y (Figure 4.8); this ribozyme demonstrates impressive polymerase activity at a range of low temperatures.



Figure 4.8. Y and wild type activity at a range of temperatures.

(A) Denaturing PAGE of extension of primer A upon template I by wild type and Y ribozymes, at -7° C in ice (1 mM each NTP) and at a range of temperatures in aqueous solution (4 mM each NTP) (162 h).

(B) Quantification of nucleotides added per primer (means ± s.e.m.; N = 3)

4.6 Long-range RNA replication in ice

The benefits provided by the mutations in Y also extend to other modes of RNA polymerisation. Replacing the 5' terminus of Y with the ss_{C19} sequence tag allows the resulting ribozyme (tC19Y) to hybridise to the RNA template downstream of the primer (Figure 4.4C). tC19Y generates more long extension products than the tC19 ribozyme (lacking the C60U, U72G and C97A mutations from Y), and is able to synthesise RNA sequences over 100 nucleotides long at 17°C (Figure 4.9A, left panel). Time-courses of primer extension upon such templates demonstrate that tC19Y-synthesised extension products 63 nucleotides long are apparent after only 16 hours, and that synthesis at 17°C is essentially finished within 4 days (Figure 4.9B); indeed, allowing reactions to proceed for 25 days at 17°C yields little benefit, as ribozyme, template and product degradation take hold (Figure 4.9C).

The ss_{C19}-template interaction was evolved at 17°C, and tC19 exhibited only a limited capability to synthesise long RNA sequences in ice, even when extension in ice was enhanced by the replacement of the MgCl₂ in the extension buffer with MgSO₄. However, the ice-selected mutations allow tC19Y to extend primers by up to 85 nucleotides in frozen reactions (Figure 4.9A, right panel), improving the prospects that template recognition and hybridisation could also be exploited by ribozymes replicating in ice. In contrast to reactions at 17°C, very little synthesis to +63 nucleotides is observed after seven days' incubation in ice; accumulation of the longest products requires weeks of incubation (Figure 4.9D). Adapting the recognition interaction to low temperatures may yield improved synthesis of long RNA strands in ice.



Figure 4.9. ss_{C19}-mediated RNA synthesis by tC19Y.

Denaturing PAGE of extension of primer BioFITC-A by ribozymes able to hybridise downstream upon the I-n series of templates (where n is the number of central 11-nucleotide repeats). $17^{\circ}C$: 0.2 M MgCl₂. $-7^{\circ}C$ Ice: 0.2 M MgSO₄. (A) Extension by tC19 and tC19Y at $17^{\circ}C$ (7 d) and in ice at $-7^{\circ}C$ (42 d). (B) Time-courses of primer extension by tC19 and tC19Y at $17^{\circ}C$. (C) Extension by tC19Y at $17^{\circ}C$ (25 d); little improvement is seen over 7 day extensions (A). (D) Extension by tC19Y at $-7^{\circ}C$ in ice; the longest products accumulate later in the incubation.

The extension products synthesised by tC19Y were sequenced to assess the impact of the mutations in Y upon RNA polymerisation fidelity. Sequencing the primers extended to 63, 95 and 106 nucleotides by tC19Y at 17°C suggests an aggregate fidelity of 98.3% (Table 4.2) compared to that exhibited by tC19 (97.3%) and wild type (96.2%). The sequences of primers extended by tC19Y in ice to 63 nucleotides showed more errors, consistent with the more diffuse product bands observed after electrophoresis; nonetheless, a high polymerisation fidelity of 94.8% was maintained (Table 4.2), similar to that of the wild type in ice (93.4%).

Correct base:		G	С	Α	U
Total positions:		1722	697	730	685
Errors at 17°C :	G	Ι	1	9	1
	С	0	Ι	0	0
	А	1	0	Ι	0
	U	1	4	0	_
	Deletion	2	0	29	2
	Insertion	0	0	0	0
Positional fidelity (%):		99.8	99.3	94.8	99.6
Overall fidelity (%):		98.3			

С	orrect base:	G	С	Α	U
Total positions:		531	224	228	209
Errors at -7°C in ice :	G	Ι	0	10	1
	С	5	Ι	1	2
	А	3	0	Ι	4
	U	1	2	0	_
	Deletion	3	0	15	3
	Insertion	1	0	0	2
Positional fidelity (%):		97.6	99.1	88.6	94.3
Overall fidelity (%):		94.8			

Table 4.2. Accuracy of tC19Y-catalysed RNA polymerisation.

Fidelities are estimated by collating errors in sequenced extension products. 17°C fidelities were judged from 17 fully extended products on template *I*-6, 17 on *I*-9, and 11 on *I*-10; –7°C ice fidelities were judged from 19 fully extended products on template *I*-6 (Figure 4.9A). The overall fidelity represents a geometric mean of the positional fidelities of incorporation opposite each base.

4.7 Template selection in ice

The I-n series of templates was transcribed relatively poorly in ice compared to at 17° C; could synthesis of different sequences be enhanced by downstream hybridisation in ice? Template selection at 17° C (Section 3.11) was halted after template I-5 came to dominate the sequence pool, but a parallel template selection performed in ice suffered no such convergence. During four rounds of selection for templates that could be transcribed by tC19 in ice at -7° C, the pool of template sequences maintained diversity whilst yielding steadily increased extension. Many of the sequences investigated after the fourth round could, like template I-5, be fully transcribed by ~+50 nucleotides, whilst showing stronger earlier extension (Figure 4.10A).



Figure 4.10. Selection of novel templates in ice.

(A) Denaturing PAGE of extension of primer BioFITCU10-A by gel-purified tC19 (-7° C in ice, 17.5 d, 0.25 μ M each RNA, 0.2 M MgSO₄, 1 mM each NTP) upon gel-purified RNA templates: the starting template library (Start), the template libraries during successive

rounds of template selection in ice (Rounds 1-4, see Section 6.10.2), template I-5, and five templates (A50-E50) cloned from the Round 4 template pool. Several of these sequences yield densities near to the size of fully-extended products on template I-5, as do the later rounds of template selection. **(B)** Sequences of template I-5 and the five screened templates; each template also possessed a 3' A10 sequence. **(C)** Base composition of the central 36-nucleotide random stretch in the templates during evolution, estimated by sequencing cloned templates (Round 1: 7 templates sequenced, Round 2: 18, Round 3: 21, Round 4: 13). This stretch is presumed to exhibit equal representation of the four nucleobases before evolution, but biases mount over the course of the selection.

These novel templates do not consist of repeat sequences (Figure 4.10B); notably, they all exhibit very low predicted (by mfold) (Zuker 2003) secondary structure, similar to or less than template I-5 – further evidence for selection based upon ribozyme-catalysed extension. However, sequences sampled from each pool illustrated the emergence of a compositional bias in selected sequences: in the variable stretch, templates became very C-rich and G-poor (Figure 4.10C). This could represent the template sequence preferences of the ribozyme. C can form Watson-Crick base pairs with three hydrogen bonds, providing more nucleotide binding energy; G can too, but its wobble-pairing potential would predispose the template to formation of more inhibitory secondary structure. U is favoured over A in the template, reflecting a preference for purine incorporation (perhaps arising from increased nucleobase stacking areas).

It should be noted that the selected templates were also transcribed efficiently by T7 RNA polymerase; as recovery was also proportional to abundance in the template library, a significant selection pressure for transcription by protein existed that could have biased template sequences. Regardless, the greatest relative fitness evidently originated from acting as a good template for a ribozyme; this selective advantage was also able to overcome the introduction of errors by tC19, avoiding sequence degeneration and yielding diverse sampled sequences that could be fully transcribed by the ribozyme.

4.8 Synthesis of ribozyme sequence in ice

As well as transcribing arbitrary template sequences, tC19Y can be used to synthesise, in ice, RNA molecules possessing a phenotype. The Y mutations enhanced primer extension upon an RNA template encoding the catalytic subunit of a hammerhead minizyme; tC19Y converted over one-third of primer to full-length (+23 nucleotides) minizyme at 17°C, and one-fifth at -7°C in ice (Figure 4.11). tC19Z is also able to synthesise full-length minizyme in ice – indicating that the benefits of mutations obtained at ambient temperatures can potentially translate to low temperatures.



Figure 4.11. Synthesis of minizyme in ice.

Denaturing PAGE of extension of primer A upon gel-purified template MzTemp encoding the minizyme catalytic subunit. $17^{\circ}C$: 0.2 M MgCl₂, 4 mM each NTP, 7 d. $-7^{\circ}C$ Ice: 0.2 M MgSO₄, 1 mM each NTP, 29 d.

4.9 Discussion

The isolation of these clones demonstrates that ribozyme evolution can be carried out at subzero temperatures in the eutectic phase, a precondition for the potential persistence of primordial ice-bound ribo-organisms in a 'cold RNA world'. Primer extension catalysed by Y on some templates in ice can match and exceed that at 17°C. The demonstration of substantial ribozyme-catalysed polymerisation in ice after limited mutagenesis of a ribozyme bodes well for the further adaptation of the polymerase ribozyme to eutectic phase activity, as well as the isolation of novel ribozyme activities in water-ice. The use of CBT to select ribozymes in ice also offers the opportunity to evolve novel RNA domains at low temperatures, and to explore how such domains could contribute to ribozyme catalysis. Some RNA structures become accessible at low temperatures (Sun et al. 2007), and formation of other structures may be possible using shorter sequences than at 17°C.

Ice microstructure could be sufficient to provide the quasicellular compartmentalisation necessary to ensure kin selection and evolution amongst populations of ribozyme replicases. By avoiding the need to adapt any candidate replicase to activity in another compartmentalising medium such as membraneous protocells (which would require replicase activity in < 4 mM Mg²⁺) (Chen et al. 2005), the capacity to select for ribozyme activity directly within this potentially protocellular medium brings the onset of autonomous molecular replication a step closer.

5 Conclusions

5.1 Ice as a protocellular medium for RNA replication

Using the R18 RNA polymerase ribozyme as the closest modern analogue of a primordial replicase, I have shown that freezing could have promoted ribozyme replication by increasing RNA half-life and concentrating substrates and ions critical for polymerisation. Some aqueous solutions, when frozen, also yield eutectic phase microstructures potentially capable of compartmentalising replicases, providing a phenotype-genotype linkage necessary for replicase evolution.

Polar and seasonal water-ice deposits were likely abundant on the early earth, and these results suggest that they would have been readily colonised by ribo-organisms. It is also notable that water-ice is far more abundant than liquid water on planetary bodies. In ice, significant ribozyme activity is possible, such as that exhibited by the Y RNA polymerase; other ribozymes (such as the hammerhead minizyme (not shown) and the hairpin ribozyme) are also active in ice. It seems likely that further characterisation and evolution of ribozymes in ice would yield an array of cold-active ribozymes able to support diverse ribo-organism functions in this environment.

Could ice have promoted the initial emergence of replicating RNA? Speculation concerning which environment would have facilitated the emergence of life can be based upon three factors: how widespread upon the early Earth an environment was, the expected abundance and complexity of nonenzymatically-synthesised sequences therein, and the minimal complexity required of a replicating RNA in that environment. Environments that promote ribozyme polymerase activity allow replication by simpler, shorter ribozymes that are more likely to occur in abiotic random sequence pools. A simple set of environmental conditions – dilute solutions at sub-zero temperatures – provides both compartmentalisation, and concentration and preservation of molecular species. Although extensive exploration of RNA assembly and behaviour in ice is called for, its properties suggest that freezing could both enhance the variety of available sequences and reduce the ribozyme complexity necessary for a replicase, helping to bridge the gap between prebiotic chemistry and biology.

5.2 In vitro evolution of RNA replicases

Development and application of CBT, a novel *in vitro* selection system for ribozyme RNA polymerase activity, allowed the evolution of improved polymerase ribozymes, both at 17° C and in ice. These evolution experiments also highlighted an important replicase-template interaction: polymerisation-enhancing recognition tags that could allow rudimentary ribozyme evolution to occur even in the absence of a compartmentalising medium. The evolved ribozymes – tC19 at 17° C, and Y at -7° C in ice – are capable of performing substantial RNA polymerisation, demonstrating the synthetic power of RNA catalysts.

CBT represents a validated system for further evolution of ribozyme polymerases. It could be readily adapted to provide additional selection pressures absent in these initial selections: rotating primer/template duplexes between rounds would enrich pools in sequence-general polymerases, and use of structured RNA templates in selections would test whether ribozymes can transcribe RNA opposite folded templates.

Such capabilities would greatly expand the range of transcribable sequences; the new, unstructured template sequences isolated by template selection illustrate the present ribozymes' extremely low tolerance for template secondary structures. Furthermore, improved generality could facilitate polymerisation of some difficult sequences and increase extension upon a range of primer/template duplexes.

The improved ribozymes can synthesise some functional RNA sequences such as a hammerhead minizyme. However, from the point of view of the RNA world hypothesis, it will be crucial to demonstrate replication of RNA sequences – requiring synthesis of both coding and noncoding strands. The template selection scheme generates sequences that have one transcribable strand, but does not predispose the second to replication; the scheme could, however, be adapted to isolate sequences that could be replicated.

Although CBT provides some selection for polymerisation fidelity, minicircle binding to extended primers may still tolerate errors. A ribozyme compartmentalised self-replication (CSR) selection system – where, in its simplest format, genetic information is recovered from ribozyme-synthesised ribozyme

sequence – would provide a much more stringent selection for accuracy. Due to the high thresholds of polymerase activity and sequence generality that would be required for ribozyme recovery from CSR (or even short-patch CSR), CBT is needed to first obtain capable candidate ribozymes, filling the void between selections for addition of a single base and selections for replication of a full sequence.

CSR would couple selection for template readability with maintenance and improvement of ribozyme function, mimicking the critical evolutionary pressures experienced by the first replicases. Directed evolution of this form represents the most effective tool for uncovering complete ribozyme replicase activity, to allow validation of the central assumptions of the RNA world hypothesis, and reconstruction of elements of self-replicating molecular systems in the laboratory.

6 Materials and Methods

6.1 Ribozyme polymerase assay

6.1.1 Principle

The behaviour of ribozyme polymerases had been characterised under a limited number of conditions. I developed a fluorescence-based primer extension assay to facilitate the convenient exploration of polymerase activity in a wider range of conditions. The ribozyme-catalysed extension of 5' fluorescein-labelled RNA primers upon RNA templates using NTPs could be visualised by denaturing PAGE, allowing the reliable, sensitive detection of polymerase activity. The fluorescein group is stable over long incubations and repeated purifications, and avoided the requirement for 5' radiolabelling of primers. The presence of a bulky dye group at the 5' end of the primer risked interfering with ribozyme primer/template recognition; indeed, extension of a primer 5' labelled with Cy5 was noticeably inhibited. However, extension proceeded well using fluorescein-labelled primers, comparable to published extensions using radiolabelled primers; this may be due to a different dye size and linker length.

6.1.2 Template design

Templates used were longer than those used in previous studies - although more extension occurs on shorter templates (Johnston et al. 2001), the ribozyme must be able to cope with longer ones to self-replicate. A condition for efficient RNA synthesis was that the templates exhibited a low tendency to fold up into secondary structures, as these were found to obstruct primer binding and/or extension by the ribozyme polymerase. Although RNA secondary structure prediction is not perfect at low temperatures, mfold (Zuker 2003) was used when designing templates, to avoid highly structured sequences. Of the unstructured templates tested, R18 performed the most extension upon template I, indicating that this sequence is relatively easy to copy. Other template sequences used to assay sequence generality (Figure 3.10) were based on less favourable template sequences used previously (Lawrence and Bartel 2005), but lengthened and varied to achieve an approximately equal representation of all four bases and all sixteen dinucleotide sequence pairs. Both primer and template sequence influenced extension; within the small sample of sequences examined herein there was no obvious correlation between sequence and extension, although the inclusion of some template dinucleotide motifs (5'-3' AU or AA) frequently impeded ribozyme-catalysed polymerisation.

6.1.3 Ribozyme preparation

Ribozymes were transcribed from a PCR product containing an upstream T7 promoter and 'clamp' sequence (5'-GATCGAGATCTCGATCCCGCGAAATTAATACGACTC ACTATA-) using a MegaShortScript high-yield transcription kit (Ambion) optimised for the synthesis of short RNAs. For most reactions, ribozyme was purified using a Qiagen RNeasy Mini kit. Two R18 variants were used: the wild type as in the original paper (Johnston et al. 2001) (with a 5'-GG- sequence and equimolar amounts of the stem 5'-GGCACCA (stem), Figure 3.9), and a engineered variant (with a 5'-GGACAACC- sequence present in the R10 variant (Johnston et al. 2001) and a stem 3' blocked with a dideoxy residue to prevent its extension by the ribozyme (5'-GGCACCddC, stem2), Figure 4.4A) that exhibited modestly increased extension upon primer A/template I and could hybridise to template HybI. The former variant was used as the wild type throughout Chapter 3 to achieve comparison with the literature wild type, and the latter throughout Chapters 2 and 4 to assay maximal wild type activity. Stems were only included where indicated in secondary structures or figures; omitting the stem yielded shorter, stronger extension upon primer A/template I, but had negligible effects on extension upon primer A/template HybI (except in ice when using low NTP concentrations, where it enhanced synthesis of the longest products - resulting in its (ultimately unnecessary) inclusion in the mutagenised library selections).

6.1.4 Assay setup

To set up standard extension reactions, 10 pmol each of primer, template and ribozyme (+ stem) were annealed together in a small volume of water (1.3 μ l in Chapter 2, 2 μ l in Chapters 3 and 4). Annealing was carried out at either 50°C for 5 minutes or 80°C for two minutes, followed by incubation at 17°C for 10 minutes and storage on ice; both treatments yielded maximum activity, as did slowly ramping down the temperature (0.1°C/s) from 80°C to 17°C. Annealing at temperatures lower than 50°C reduced activity; heating not only ensured primer/template duplex formation, but also refolding of the ribozyme after purification-induced denaturation.

Annealing reactions in a large volume of water (before adding concentrated extension buffer) yielded similar extension at 17° C, but $4-6\times$ less at -7° C. Annealing in a large volume of RNeasy column eluate (which usually comprises 1% of the final reaction) did support full extension at -7° C, implying that solute levels during renaturation after RNeasy purification affected ribozyme activity in ice. Regardless, activity in ice was not dependent upon this, as unpurified ribozyme exhibited full activity in ice (Section 6.1.7).

Chilled extension buffer was then added to a final volume of 40 μ l (Chapter 2, f.c. 0.25 μ M each RNA) or 20 μ l (Chapters 3 and 4, f.c. 0.5 μ M each RNA), f.c. 0.2 M MgCl₂ (pre-treated with Bio-Rad Chelex 100 resin to remove heavy metal ions, though omitting this step did not affect extension, suggesting a negligible contribution to RNA degradation from heavy metal impurities), 50 mM Tris·HCl pH 8.3 and 4 mM of each NTP (Li⁺ salts, though extension was unaffected using the Na⁺ salts), unless otherwise indicated. To explore the influence on extension of Mg²⁺ counterions other than Cl- (SO₄²⁻, Br⁻, NO₃⁻, CH₃CO₂⁻, Figure 2.6A), the MgCl₂ in the extension buffer was replaced with equimolar concentrations of the relevant Mg²⁺ salt.

Reactions were then quickly placed at the appropriate temperature; in-ice primer extensions were first frozen at -25° C for 10 minutes, then transferred to a Techne RB-5 refrigerated bath, maintained at -7° C by a Techne Tempette TE-8D thermostat. Notable features of the buffer environment differed between reactions at 17° C and in the fourfold-concentrated eutectic phase at -7° C:

17°C: 0.17 M free Mg²⁺, 50 mM Tris @ pH 8.53, 64 mM Li⁺,

2.1 mM MgNTP², 13.9 mM Mg₂NTP.

Ice (-7°C): 0.674 M free Mg²⁺, 200 mM Tris @ pH 9.23, 256 mM Li⁺,

2.3 mM MgNTP², 61.7 mM Mg₂NTP.

Mg²⁺-NTP complex formation was calculated based on (Thomen et al. 2008). The concentrations of components of the buffer in frozen reactions were varied to explore whether any were at a surplus relative to other components, but only reducing NTP concentrations influenced extension (Figure 2.8).

Protein polymerase extension assays (Figure 2.1, left panel) were set up as above, but ribozyme was omitted and T7 RNA Polymerase (NEB) was added after buffer at the incubation temperature (f.c. $0.25 \,\mu$ M, in 1× T7 RNA Polymerase buffer (NEB) supplemented with 14 mM MgCl₂ and 4 mM of each NTP).

6.1.5 Resolution and quantification of primer extension reactions

Reactions were stopped by adding 0.5 volumes of 0.5 M EDTA. Extension products (typically 1-1.5 pmols/lane) were resolved by denaturing Polyacrylamide Gel Electrophoresis (PAGE), allowing analysis of fluorescence using a Typhoon Trio scanner. Primer extension was quantified using ImageQuant TL, as described previously (Muller and Bartel 2008). Each gel image lane was divided up into bands corresponding to unextended primer and each extension product; an intensity value for each band was given by subtraction of the bulk of background signal (using a baseline drawn between intensity minima in the

lane). For the quantification of weak extensions (Figure 2.5A), further background correction was then carried out by subtracting the intensity values of an empty lane. These corrected intensity values were used to calculate the average extension per primer in a lane (E) as follows:



where x_n is the intensity of the band corresponding to *n* base additions. Gel image brightness and contrast were adjusted to illustrate extension product banding patterns.

Due to the stability of the RNA:RNA duplexes synthesised by the ribozyme, complementary strands (longer than about 16 nucleotides) often reannealed after heating (94°C 5 min) in 6 M urea when cooled prior to gel loading, hindering resolution of banding patterns. Thus, during sample heating, a 10-20× excess of unlabelled competing oligonucleotide complementary to the template was present to sequester it after denaturation and allow primers to be resolved. RNA competing oligonucleotides were the most effective; even 100× excesses of DNA competing oligonucleotides, as used in Figure 4.11 (MzComp), were unable to fully segregate primer and template strands for longer syntheses.

To avoid the requirement for a competing oligonucleotide for every template tested, some extensions were performed using a biotinylated, fluorescein-labelled primer (BioFITC-A, BioFITCU10-A). After these extensions were stopped, primers in the reaction were bound to 10 μ l of MyOne Streptavidin C1 Dynabeads (Invitrogen), washed to remove excess template, then heated (60°C, 3 min) in 9.3 M UTET (9.3 M urea, 10 mM Tris·HCl, 1 mM EDTA, 0.1% Tween-20) to dissociate template from bead-bound primers, and washed in 9.3 M UTET to remove as much template as possible. This process was repeated to promote the unbound state in a binding equilibrium by minimising template concentration. The beads were then heated in 95% formamide/10 mM EDTA, and the supernatant was resolved by denaturing PAGE. A similar protocol was used to resolve extension on primer 32A/template 32A and primer 32B/template 32B. This protocol was harsh enough to lead to some primer loss from beads (~75%) during washing, but was still unable to completely separate some of the longest primers from their templates (as evidenced by the slowlymigrating smear in the template I-6 lane in Figure 3.15). This may be a downside of the unstructured templates used in this study that may not be able to adopt alternative intramolecular structures that can compete with extension product binding.

6.1.6 Diluted extension reactions

To set up diluted extension reactions, 10 pmols each of ribozyme construct, template and primer were annealed as above. Extension buffer was diluted with water and chilled on ice prior to addition; for an *n*-fold dilution, $(n-1) \times 40 \ \mu$ l of water were added to the extension buffer. After incubation, reactions were stopped with 20 μ l 0.5 M EDTA; aliquots containing 1.5 pmols of primer were taken from each, made up to equal volumes with water, dried, and analyzed by PAGE as above.

6.1.7 Polyclonal activity assay

To estimate the effect of a mutagenesis protocol or a round of selection upon a population of ribozymes, the DNA pool can be transcribed and the activity of the resulting polyclonal ribozyme pool assayed. Although this method is sensitive to the transcription efficiency of each ribozyme, and combinations of ribozymes do not necessarily yield their average extension, it provides a rapid measure of the progress of a selection.

Examining the output pools of the model selections (Figure 3.5, Figure 4.1), where wild type composition can be accurately judged due to the size difference between R18 and R18i genes, shows that polyclonal activity assays reflect the expected activity well at 17° C, but tend to underrepresent activity at -7° C (Table 6.1). This could arise from an altered persistence in ice of any inhibitory effect of annealing different RNA molecules together.

Solaction conditions:	Assay for	Wild type composition of ribozyme pool (%):				
	wild type:	Starting pool	1 PCR/bead	0.2 PCR/bead		
17°C	DNA	13	46	80		
170	Activity	14	41	68		
7°0 loo	DNA	13	47	83		
	Activity	5	25	64		

Table 6.1. Relationship between polyclonal activity and pool composition in modelselections.

The fraction of genes encoding wild type ribozyme was calculated by PAGE for the starting pool of PCR molecules used in the model selections (Figure 3.5, Figure 4.1) and in the outputs of model selections (at $17^{\circ}C$ and at $-7^{\circ}C$ in ice, using 1 or 0.2 PCR molecules per bead) (DNA). The activity of the corresponding polyclonal ribozyme pools relative to wild

type was determined by quantifying primer extension performed by RNeasy-purified polyclonal ribozyme under the respective selection conditions (17°C 22 h, 4 mM each NTP, or -7°C 108 h, 0.5 mM each NTP; 0.25 μ M ribozyme, stem2, primer A, template HybI).

Annealing is required to observe full activity when using RNeasy-purified ribozyme, particularly in ice, but this issue can be bypassed by setting up reactions using unpurified ribozyme: unexpectedly, adding ribozyme directly from transcription reactions (after agarose gel-quantification, using an RNeasy-purified standard) to reaction components without annealing yields full ribozyme activity, even for reactions in ice, and improves the activity of ribozyme mixtures relative to wild type (Table 6.2).

Reaction	Ribozyme pool extension						
preparation:		Starting pool	Round 1 output	Round 2 output	Wild type		
RNeasy-purified	nt/primer	0.05	0.14	0.42	1.79		
ribozyme, annealed	% wild type	2.8	7.6	23	100		
Unpurified	nt/primer	0.08	0.32	0.96	2.58		
annealed	% wild type	3.3	12	37	100		

Table 6.2. Assaying polyclonal activity.

The activity of polyclonal ribozyme pools produced from the early stages of the in-ice selection (Table 4.1), measured by quantifying average primer extension ($-7^{\circ}C$ 224 h, 0.5 mM each NTP; 0.25 μ M ribozyme, stem2, primer A, template HybI) as nucleotides (nt) added/primer, performed either by RNeasy-purified ribozyme annealed with primer/template, or by unpurified transcribed ribozyme without an anneal. The activity obtained relative to a similarly-purified wild type ribozyme is indicated.

Polyclonal activity of selection pools was measured in a format close to the selection conditions: ribozymes were transcribed with the 3' RTT, agarose gel-quantified, and 10 pmols of ribozyme in 0.4 μ l diluted transcription buffer were added to 10 pmols each of stem2, template I and primer BioFITCU10-A in 0.8 μ l water, before addition of chilled extension buffer (17°C, 20 μ l: 0.2 M MgCl₂, 50 mM Tris·HCl pH 8.3, 4 mM each NTP; -7°C Ice, 80 μ l: 50 mM MgCl₂, 12.5 mM Tris·HCl pH 8.3, 125 μ M each NTP) and incubation for 166 h under the respective conditions. After samples resolution by denaturing PAGE,

average primer extension was quantified as described, and compared to wild type extension to estimate relative activity.

The activity of unpurified ribozyme is also exploited during ribozyme screening (Section 6.8), allowing assessment of the full ribozyme activity or each clone without resorting to widespread ribozyme purification. This behaviour indicates that the ribozyme folds properly during transcription, in the presence of Mg²⁺. Correct cotranscriptional folding is a desirable trait for ribozyme replicases that reproduce through strand displacement, and future selections should seek to maintain it by avoiding denaturation between transcription and extension steps.

6.2 Scanning Electron Microscopy

Samples with the same composition as (diluted) extension/selection reactions were loaded into the ends of 1 mm diameter copper tubes, leaving a protruding drop of solution. After flash-freezing over liquid N₂, samples were transferred to a Reichert AFS (Leica) and incubated for 60-100 minutes under a N₂ atmosphere at -7° C, to allow eutectic phase formation. At the end of the incubation, to preserve eutectic phase structure, samples were transferred directly to liquid N₂ using chilled tweezers, where they were mounted on a transfer shuttle and moved to a PP2000 Cryo-SEM preparation chamber (Quorum) at -130° C. The sample was freeze-fractured at the neck of the tube, and the temperature was raised (-90°C for 10 minutes for Figure 2.6B, Figure 2.11D and Figure 4.2, -80°C for 30 minutes for Figure 2.13A, -85°C for 15 minutes for Figure 2.13B, C, D) to allow sublimation of water from ice crystals, leaving the salts of the eutectic phase protruding. Longer sublimations at higher temperatures lead to more water loss, exposing more extensive eutectic phase structures. Upon cooling and sputter-coating with ~30 nm of gold, images were acquired using an FEI XL30 FEG SEM.

6.3 Ribozyme degradation assay

5'-fluorescein-labelled R18 was transcribed using an Ambion MegaShortScript Kit, in the presence of 3.75 mM each of ATP, UTP, and CTP, 2 mM of GTP, and 2 mM of 5'-fluorescein-labelled AG RNA dinucleotide starter (Dharmacon), before purification using an RNeasy Mini Kit. 3 pmols each of fluorescently labeled ribozyme, template I and primer Au (lacking a fluorophore) were annealed together in 9 μ l of water. 2.4 μ l of 1 M MgCl₂ and 0.6 μ l of 1 M Tris·HCl (pH 8.3) were chilled on ice and added to the RNAs, before incubation at different temperatures and resolution by denaturing PAGE. The ratio of full-
length ribozyme to shorter fragments was determined, after comparison to starting samples to correct for the presence of premature transcription termination products.

6.4 Ribozyme diffusion assay

The experiments were carried out within a preformed water ice sheath to avoid formation of a liquid interface with the polypropylene tube walls, through which ribozyme could diffuse. These sheaths were formed by placing a TipOne 1-200 μ l Graduated Filter Tip (Starlab) in a 0.5 ml microcentrifuge tube. The top open end of the tip was sealed with water, allowing the tip to be surrounded with 440 μ l of water before freezing, creating a conical void in the ice. To prepare scoring beads, $0.5 \,\mu$ l of MyOne Streptavidin C1 Dynabeads (Invitrogen) were bound to 0.5 pmols each of 5'-biotinylated RNA primer BioU10-A, and 3'-biotinylated DNA oligonucleotide 3Hyb complementary to the 5' end of the ribozyme (to increase ribozyme local concentration near the bead surface, improving extension). These beads were resuspended in 195 μ l of extension buffer (with 200 mM MgCl₂ (or, if indicated, MgSO₄), 50 mM Tris·HCl (pH 8.3), 4 mM of each NTP and 0.5 μ M of template I), or 195 μ l of a dilution thereof, chilled at -7° C, and transferred to an ice sheath for incubation at -7° C overnight (to ensure complete eutectic phase equilibration). 30 pmols of ribozyme were annealed (50°C 5 min, 17°C 10 min) in 1.8 μ l of water, made up to 5 μ l in undiluted extension buffer, and added to the top surface of the ice column at -7° C to allow diffusion to begin.

Incubations were stopped by thawing with Tween-20 to 0.1% and an excess of EDTA. On-bead primer extension was detected using RCA (Section 6.6.3) (+7 minicircle stringency), and bead populations were analyzed on a FACSCalibur instrument (BD Biosciences) with a FL-1 detector voltage of 450V. Typically, about 20,000 FACS events corresponding to single beads (as gated by event forward and side scatter) were collected per sample. Flow cytometry data were analyzed using FlowJo software (TreeStar).

6.5 Generation of mutagenised ribozyme library

Mutagenic PCR was used to generate a library of R18 variants; an R18 gene was amplified (6 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 120 s) using primer 7 and a ten-fold excess of primer 8, in the presence of 400 μ M each of dATP, dGTP, dCTP, dUTP, 8-oxo-2'-deoxyguanosine-5'-triphosphate (TriLink) and 2'-deoxy-P-nucleoside-5'-triphosphate (TriLink) (Zaccolo and Gherardi 1999; Petrie and Joyce 2010). The resulting products were bound to beads and washed in 0.1 M NaOH to deplete the wild-type sequences, and amplified using primers 1 and 2 to generate ribozyme genes for selection

encoding a variety of wild type ribozyme variants with positions 1-173 mutagenised at an average rate of 4.4%, mostly transition mutations with wide positional variations in mutation rates.

6.6 Compartmentalised bead-tagging

6.6.1 Transcription/ligation

Biotinylated ribozyme genes were bound to streptavidin-coated paramagnetic microbeads (MyOne Streptavidin C1 Dynabeads, Invitrogen) in BWBT (0.2 M NaCl, 10 mM Tris HCl, pH 7.4, 1 mM EDTA, 0.1% Tween-20) at an average density of one ribozyme gene per bead (or less). Gene concentration in a stock in a lo-bind tube (Eppendorf) was quantified beforehand by comparison to Low Molecular Weight marker (NEB) by native PAGE, SYBR Gold (Invitrogen) staining, and gel densitometry. ~6×10⁴ 5Hairpin molecules were then bound per bead. To coat each bead with a clonal ribozyme population derived from the bound gene, ribozymes were transcribed in a 5'-monophosphorylated form and ligated via their 5'-termini to the bead-bound hairpin oligonucleotides in emulsion; beads were resuspended in 150 µl transcription/ligation mix (80 mM HEPES, pH 7.6, 22 mM MgCl₂, 1 mM spermidine, 3.75 mM ATP, 3.75 mM UTP, 3.75 mM CTP, 2 mM GTP, 10 mM GMP, 0.8 U/µI RNasin (Promega), 0.25 U/µI T4 RNA Ligase 2 (NEB), 0.48 µg/µI BSA (NEB), 5% MegaShortScript enzyme mix (Ambion), 1 U/ μ I T7 RNA Polymerase (NEB), 4.6 U/mI Yeast Inorganic Pyrophosphatase (NEB)), added to 600 μ l oil mix (7% (w/v) ABIL WE09, 20% (v/v) mineral oil and 73% (v/v) Tegosoft DEC (Diehl et al. 2006)) and emulsified as previously described (Diehl et al. 2006). After incubation for 16 h at 37°C in emulsion, the beads were extracted (Diehl et al. 2006) and washed with BWBT and UTET (8 M urea, 10 mM Tris HCl, pH 7.4, 1 mM EDTA, 0.1% Tween-20) to remove non-ligated ribozyme. Efficiency of transcription/ligation was quantified by heating a proportion of the beads in 10 μ l 95% formamide/10 mM EDTA for 4 min at 94°C, followed by gel electrophoresis of the supernatant, SYBR Gold staining, and gel densitometry (yielding ~3,000 ribozymes/bead).

6.6.2 Extension

5'-biotinylated RNA BioU10-A primers (~ 1.2×10^{5} /bead) were bound to the beads in BWBT, then the beads were washed and cooled in water (23°C 5 min, 17°C 10 min). Beads for selections at 17°C were resuspended in 150 µl chilled extension buffer (0.2 M MgCl₂, 50 mM Tris·HCl, pH 8.3, 4 mM each NTP, 0.5 µM RNA template, 0.5 µM stem2), emulsified as above in 600 µl oil mix, and incubated at 17°C, before recovery from the emulsion as before. Beads for selections in ice were resuspended in 600 μ l final volume chilled extension buffer (50 mM MgCl₂, 12.5 mM Tris·HCl, pH 8.3, 125 μ M each NTP, 0.125 μ M RNA template, 0.125 μ M stem2) and frozen at –25°C (10 min) before transfer and incubation at –7°C to allow eutectic phase formation. Beads were recovered by thawing with 75 μ l 0.5 M EDTA pH 7.5 and Tween-20 (to 0.1%).

After washing with BWBT, beads were blocked with Biocytin (1.33 mM in 187.5 μ I BWBT, 15 min) to reduce migration of extended primers during subsequent heating. Beads were heated in UTET for 3 min at 60°C to remove bound template, and washed in TBT (10 mM Tris·HCI, pH 7.4, 0.1 μ g/ μ I BSA, 0.1% Tween-20). Beads thus recovered were processed further in batches of 2×10⁷ beads.

6.6.3 Rolling circle amplification

A single-stranded DNA minicircle containing a sequence complementary to the extension product was annealed (50 nM DNA minicircle, 0.1 μ g/ μ l BSA, Phi29 buffer (NEB), 2 nM primer 1 (to restore double-stranded PCR product); 60°C for 1 s, 0.1°C/s to 17°C, 17°C for 20 min), and after removal of the supernatant, beads were given an additional wash in chilled buffer (0.1 μ g/ μ l BSA, 1× Phi29 buffer, 1% glycerol, 0.01% Tween-20) on ice to remove residual and weakly-bound DNA minicircle, before RCA (rolling circle amplification, as used previously to detect surface-bound molecules (Lizardi et al. 1998; Li, M. et al. 2006)) was performed (0.3 mM each dNTP, 0.1 µg/µl BSA, 0.2 U/µl Phi29 polymerase (NEB), Phi29 buffer, 15°C for 1 min, 0.1°C/s to 37°C, 37°C for 20 min). As RCA finished, the DNA oligonucleotide RCAprobe was added (to $1.8 \,\mu$ M) along with Picogreen (Invitrogen) (to 0.0025×). This probe was annealed to the RCA product (60°C for 1 s, 0.1°C/s to 37°C) to prime the conversion (by fresh Phi29 polymerase) of the RCA product to a double-stranded form specifically bound by Picogreen. The beads were then diluted in three volumes of PBST (phosphate buffered saline, 0.1% Tween-20) and filtered (CellTrics 30 μ m, Partec) prior to FACS. 'Positive' beads carrying the most fluorescent RCA products, and thus the most active ribozymes genes, were sorted using Purify 1 settings on a MoFlo High Speed sorter (Beckman Coulter). Genes from these positive beads were recovered via PCR (GoTag HotStart, Promega) using primers 1 and 2, purified (Qiaguick PCR purification, Qiagen), and gel-quantified before further rounds of selection.

6.6.4 DNA minicircle design for RCA

Linear single-stranded DNA molecules were designed to form a 'dumbbell' structure with five nucleotides at each end of the molecule hybridised together to a central ten-nucleotide stretch at 16°C (Wochner et al. 2011). These were 5' phosphorylated (Polynucleotide Kinase, NEB) and annealed in water (60°C for 1 s, 0.1°C/s to 16°C, 16°C for 10 min) to allow dumbbell formation, then incubated at 16°C to allow circularisation through ligation with T4 DNA ligase (NEB), and gel-purified.

6.7 Recombination of selection pools using StEP

StEP shuffling (Zhao et al. 1998) of libraries was achieved using hot-start PCR (f.c.: 1 μ M each primer 1 and 2, 0.1 M tetramethyl ammonium chloride, 1× Taq buffer, 50 μ M each dNTP (to balance slow extension with sufficient product formation), 0.1 U/ μ l SuperTaq (HT Biotechnology, added at 94°C), 0.3 pmol PCR product in 20 μ l; 120 cycles of 94°C for 30 s, 65°C for 1 s, in a thin-walled microtube), and genes were further amplified by six cycles of standard PCR. The resulting library contained 10% parent genes and 90% recombination products (with an estimated 2.9% crossing-over chance per backbone position and 0.3-1% mutation rate, as determined by sequencing and analysing the output of StEP performed on a mix of a wild type gene and a gene with 15 mutations). Gel-purification was necessary after recombination, as short PCR products began to emerge.

6.8 Screening of ribozyme clones

Polyclonal DNA pools were ligated into a pGEM-T Easy Vector (Promega), and used to transform NEB 10-beta competent E. coli cells (NEB). Ribozyme clones were screened and ranked by scoring primer extension in a ribozyme polymerase plate assay (RPA). Following colony PCR with primers 1 and 2 (0.4 μ M each primer, 0.25 mM each dNTP, 0.1 U/µI SuperTaq), 10 µI of the PCR reaction was transferred to Strep Thermo-Fast 96 plates (Thermo Scientific) to immobilise ribozyme genes. Wells were washed with PBS, 10 µl transcription mix was added (3.75 mM each NTP, 80 mM HEPES, pH 7.6, 22 mM MgCl₂, 1 mM spermidine, 0.24 μ g/ μ l BSA, 0.4 U/ μ l RNasin, 5 U/ μ l T7 RNA polymerase (NEB)) and plates were incubated for 16 h at 37°C. 1 μ l of each transcription reaction (5– 10 pmol of ribozyme) was transferred to 9 μ l of primer extension reaction mix (f.c.: 0.2 μ M BioU10-A, 0.2 μ M template I, 1 μ M stem2, 0.2 M MgCl₂, 50 mM Tris HCl pH 8.3, and 4 mM (17°C) or 0.5 mM (-7°C lce) of each NTP) and incubated at 17°C for 105 h, or frozen on dry ice (5 mins) and incubated at -7° C for 321 h. Reactions were stopped by adding 5 μ l of 0.5 M EDTA and were transferred into 100 µl BWBT in StreptaWell 96 well plates (Roche) to allow primers to bind. Between subsequent steps, wells were washed three times with PBS. RNA templates were removed with 200 μ I 9.3 M UTET (heated to 60°C for 3 min), wells were

blocked with $300 \ \mu$ l 1× Rotiblock (Roth), and incubated with $100 \ \mu$ l 50 nM 5' biotinylated probe, pre-bound to equimolar amounts of Neutravidin-HRP (Pierce/Thermo Scientific) in PBS/1× Rotiblock. Binding was detected colorimetrically by addition of $100 \ \mu$ l TMB substrate (Thermo Scientific), the reaction was stopped with $100 \ \mu$ l 1 M H₂SO₄, and product was analyzed using a SpectraMax 340 microplate reader (Molecular Devices). Plates could be washed with 9.3 M UTET and PBS, and re-probed with a less stringent probe to corroborate signal and estimate extension patterns.

6.9 Extension product sequencing

6.9.1 High-throughput sequencing (Table 2.1, Figure 2.9)

Extension reaction products were resolved using denaturing PAGE, and the bands corresponding to the desired extension products for sequencing (+12 nucleotides) were excised from the gel (from the top of the preceding band to the bottom of the subsequent band – to ensure that any products whose errors affected their gel mobility were sequenced). Precipitated RNAs were resuspended in 15 μ l H₂O, half of which was used for polyadenylation (1 mM ATP, 0.5 U/ μ l *E. coli* Poly(A) Polymerase (PAP) (NEB) in 1× PAP buffer (NEB)). Reactions were stopped after 60 s at 37°C by addition of EDTA to a final concentration of 10 mM. Reactions were reverse transcribed and PCR-amplified using Primer 3 (or a close variant) and Primer 4 (SuperScript One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen)), facilitating subsequent addition by PCR of DNA tags allowing Illumina sequencing.

To prevent contamination, we had used a competing oligonucleotide to resolve the products during gel purification (CompISeq) with a blocked 3' end (dideoxyC), which prevents poly-A tailing. To allow exclusion of sequences arising from potential degradation products of this oligonucleotide, it is modified to U instead of a G at the position corresponding to the last base of the primer. The few sequences picked up derived from this competing oligonucleotide were discarded. Sequences were aligned using MUSCLE (Edgar 2004) into groups of identical sequences to facilitate manual counting of errors. Although some sequences corresponded to shorter (or longer) extension products than +12 nucleotides, data from all sequences (originating from *bona fide* extended primers) were used to calculate error rates. Due to the poly-A tails, deletions or A substitutions could not be unambiguously identified at the 3' ends of the sequences and these positions were excluded when calculating rates of these errors. To determine the background levels of mutation that arise from the sequencing process, a chemically synthesised RNA corresponding to fulllength extended primer (CompI) was also sequenced, revealing low but significant levels of error. The 649 sequences thus obtained exhibited positional error rates of 0.13% for deletions, 0.38% for substitutions, and 0.2% for insertions. The total substitution and deletion rates were corrected for this (Table 2.1) to obtain the true ribozyme error rates.

6.9.2 Long product sequencing (Figure 3.17, Table 4.2)

5'-biotinylated primer extension products of desired lengths were excised from the gel and extracted. Precipitated RNAs were resuspended in 15 μ I H₂O, half of which was used for polyadenylation (1 mM ATP, 0.5 U/ μ I *E. coli* Poly(A) Polymerase (PAP) (NEB) in 1× PAP buffer (NEB)). Reactions were stopped after 60 s at 37°C by addition of EDTA to a final concentration of 10 mM. The reaction was bound to beads and washed with BWBT to remove polyadenylation side products, preventing them from competing with genuine products during RT-PCR. Bead-bound RNAs were reverse transcribed, PCR-amplified using Primer 3 and Primer 4 (SuperScript One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen)) and ligated into pGEM-T (Promega). Plasmids were used to transform NEB 10-beta competent *E. coli* cells (NEB) (Figure 3.17) or SURE 2 Supercompetent cells (Agilent Technologies) (Table 4.2) that were less able to recombine transformed sequences, better preserving repeat sequences.

6.10 Template Selection

6.10.1 Selection at 17°C

An RNA template library was transcribed from a DNA library (generated through fill-in of T7C19 and TempLib oligonucleotides) and gel-purified. 90 pmol of this library and 83.3 pmol of BioU10-A primer (with 6.67 pmol of fluorescent BioFITCU10-A primer as a marker) were annealed together (50°C for 5 min, 17°C for 10 min) in 200 μ l water. 90 pmol of tC19 ribozyme were annealed separately in 6 μ l water. These were combined in chilled extension buffer (f.c.: 200 mM MgCl₂, 50 mM Tris•HCl, pH 8.3, 4 mM each NTP, 0.25 μ M each RNA in 360 μ l) and incubated at 17°C for ~90 h, then were stopped with 150 μ l of 0.5 M EDTA. To thoroughly deplete template sequences, primers in the reaction were bound to microbeads (MyOne Streptavidin C1 Dynabeads, Invitrogen), washed in TBT, heated to 60°C in UTET for 3 min to remove bound template, washed in TBT, heated in UTET again and TBT washed once more, before resuspension in 95% formamide, 10 mM EDTA, and heating (95°C for 5 min) to release primer from the beads. Primers in the supernatant were separated by denaturing PAGE, and the gel zone was excised corresponding to where a

primer extended by 50 nucleotides would be expected to run. Eluate from the gel fragments was ethanol precipitated with glycogen as a carrier and the pellet was resuspended in 20 μ M TempRec primer, which was annealed (72°C for 3 min) to the fixed sequence at the end of extension products to prime reverse transcription of the extension product (f.c.: 10 U/µI SMART-Scribe RT (Clontech), 1× first-strand buffer (Clontech), 2 mM DTT, 1 mM of each dNTP, 10 μ M TempRec primer; 37°C for 5 min, 42°C for 10 min. This enzyme is able to add untemplated cytidine residues at the 3' end of cDNA to lengthen the fixed sequence available for downstream primer binding). An equal volume of BWBT containing microbeads was added to the product to re-bind biotinylated extended primers (and associated cDNA) and allow any reverse transcription side-products to be washed away. The cDNA bound to beads was then amplified by PCR (with $2 \mu M$ TempRec and PriRec primers and 1% Triton X-100 (Sigma)), and full-length PCR products were agarose gel-purified, re-amplified, and cloned to obtain individual sequences. These primers amplified outside of a primer-specific mutation that demonstrated that recovered sequences were derived from extended primers, and as this mutation induced a wobble-pair in the primer/template duplex, it was omitted from subsequent rounds. The library was re-amplified by U10D and T7C19 primers, and this product was transcribed to generate a second-round library of templates with 5' monophosphates; extension upon these was carried out as above, but prior to gel purification, in-between the UTET washes, the 5' monophosphate served to mark the template for degradation (0.01 U/ μ I XRN-1 (NEB), 1× NEB buffer 3, 0.1% Tween-20, 37°C for 30 min), and after a TBT wash primers were 3'-blocked (0.1 U/ μ l Terminal Transferase (NEB), 0.2 mM dideoxy-GTP, 1× Terminal Transferase buffer (NEB), 0.25 mM CoCl₂, 37°C for 40 min) to prevent any later extension by proteinaceous polymerases.

6.10.2 Selection in ice

The first two rounds of the template selection in ice were carried out as at 17° C, but extension reactions contained only 1 mM of each NTP, and reactions were frozen at -25° C before incubation at -7° C as normal (Round 1: 423 h incubation; Round 2: 452 h incubation). The output pool from the second round of selection was subjected to mutagenic PCR to introduce further variation: 5 cycles of 94°C for 30 s, 45°C for 30 s, 72°C for 120 s, followed by 15 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 120 s, using 0.5 μ M each of primers U10D and T7C19, in the presence of 400 μ M each of dATP, dGTP, dCTP, dUTP, 8-oxo-2'-deoxyguanosine-5'-triphosphate (TriLink) and 2'-deoxy-P-nucleoside-5'-triphosphate (TriLink), followed by recovery PCR with only standard dNTPs. The resulting RNA template library was used in two half-sized selection extensions in ice, one using 0.2 M MgCl₂, and

one using 0.2 M MgSO₄ (742 h incubation at -7° C). As no clear differences were observed in the outputs of these two rounds, they were combined and used in a final half-sized selection extension (0.2 M MgSO₄, 140 h, -7° C). Purification and recovery for rounds 3 and 4 used a similar protocol to round 2, without mutagenic PCR.

6.11 Oligonucleotide sequences

DNA is depicted in grey, RNA in black. On ribozyme templates, ss_{C19} binding sites are highlighted in red; RNA primer binding sites are underlined, and random sequences are in lowercase. RNAs generated by *in vitro* transcription (IVT) were RNeasy purified; all other RNAs, and those indicated, were gel purified. Ribozymes were transcribed from a PCR product generated by Primer 1 and Primer 2 (with 3' RTT) or Primer 9 (without 3' RTT); engineered mutant ribozyme genes were generated by PCR using primers encoding mutations. Some short RNAs were transcribed from double-stranded DNA generated by fill-in of the indicated DNA oligonucleotides. All double-stranded DNAs were purified by Qiaquick (Qiagen) before transcription. Free amino groups in 'hairpin' and 'BioFITCU10-A', and 'BioFITC-A' were biotinylated according to manufacturer's instructions with EZ-Link Sulfo-NHS-SS-Biotin (Thermo Scientific) and EZ-Link NHS-Biotin (Thermo Scientific), respectively.

Application	Name	Source	Sequence (5´→3´)
PCR primers	Primer 1	Sigma	GATCGAGATCTCGATCCCGCGAAATTAATACGACTCAC TATAGGACAACC
	Primer 2	Sigma	Biotin-Biotin-GGTAAGCCTTTTTTTTTGCGGCCGC- 2′OMeG-2′OMeG-AGCCGAAGC
	Primer 3	Sigma	ACACGACGCTCTTCCGATCTCACTGCCAACC
	Primer 4	Sigma	CCTTATTAGCGTTTGCCATTTTTTTTTTTTTTTTTTTTT
	Primer 5	Sigma	GATCGAGATCTCGATCCCGCGAAATTAATACGACTCAC TATA
	Primer 6	Sigma	GATCGAGATCTCGATCCCGCGAAATTAATACGACTCAC TATAGTCAATGA
	Primer 7	Sigma	Biotin-biotin-GATCGAGATCTCGATCCCGCGAAATTAATA CGACTCACTATAGGACAACC
	Primer 8	Sigma	GGAGCCGAAGCTCC
	Primer 9	Sigma	GGAGCCGAAGCTCCGGGGATTATGAC
CBT	5Hairpin Gel purified	IDT	GGTTGT-2´OMeC-2´OMeC-AGATCTT(C6-NH ₂ -T)TT GAUC U
	DNAcirc+7 [template I]	Sigma	GTTACCTTTCAATGAATCCACGCTTCGCACGGTTGGTG TAACGACTTTTCGGATTTCTAGGATCTCCAAGTATGTTC TAAGTC
	DNAcirc+5 [template II]	Sigma	GTTACTTTTGCCTCCCTTCGCACGGTTCTTTGTAACGAT CTTTCTGGATTTCTAGGATCTCGTCCCTATAGTGAGTCG GTTCTAGATC
	DNAcirc0 [template I]	Sigma	GTTACTTTTCAATGAATCCACGCTTCGCATGTAACGACT TTTCGGATTTCTAGGATCTCCAAGTATGTTCTAAGTC
	DNAcirc0 [template II]	Sigma	GTTACTTCTATCTCCCTTCGCATTCTTTGTAACGATCTTT CTGGATTTCTAGGATCTCGTCCCTATAGTGAGTCGGTT CTAGATC
	DNAcirc-3 [template I]	Sigma	GTTACTTTTCAATGAATCCACGCTTCTGTAACGACTTTT CGGATTTCTAGGATCTCCAAGTATGTTCTAAGTC
	RCAprobe	Sigma	GGATTTCTAGGATCTC
Beads	3Hyb Gel purified	In house	GGTTGTCCCATTG-C6-biotin

RPA	P2	Sigma	Biotin-TCCCTTCGCACGGTT
Probes	P3	Sigma	Biotin-TGAATCCACGCTTCGCACGG
Stem	stem	Dharmacon	GGCACCA
	stem2	Dharmacon	GGCACC-dideoxyC
RNA	А	Dharmacon	Fluorescein-CUGCCAACCG
primers	Au	Dharmacon	CUGCCAACCG
	BioU10-A	IDT	Biotin-UUUUUUUUUUUCUGCCAACCG
	BioU10- Aext	Dharmacon	Biotin-UUUUUUUUUUUCUGCCAACCGUGCGAAGGGAG
	BioFITC-A	IDT	Fluorescein-(C6-NH2-dT)-CUGCCAACCG
	BioFITCU1 0-A	Dharmacon	Fluorescein-(C6-NH ₂ -dT)-UUUUUUUUUUUUCUGCCAACCG
	В	Dharmacon	Fluorescein-GAAUCAAGGG
	С	Dharmacon	Fluorescein-GAUAGGUAG
	32A	IDT	Biotin-GAUUAAGUGCUAUUCAGGACGUCUGCCAACCG
	32B	IDT	Biotin-AUACCUGUUCGCCAGCGUUACUGAAUCAAGGG
Ribozvme	I	Dharmacon	CAAUGAAUCCACGCUUCGCACGGUUGGCAGAACA
templates	Hvbī	Dharmacon	CAAUGAAUCCACGCUUCGCACGGUUGGCAGAACAGG
			UUGUCC
	HybI22	Dharmacon	GUCAAUGAAUCCACGCUUCGCA <u>CGGUUGGCAG</u> AACA GGUUGUCC
	HybI41	Dharmacon	GUCAAUGACACGCUUCGCACACGCUUCGCACACGCUU
			CGCACGGUUGGCAGAACAGGUUGUCC
	I-1	Sigma, IVT	Forward:Primer 5
			Reverse:TTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTC
			ATTGACTATAGTGAGTCGTATTAATTTC
			Transcript:GUCAAUGACACGCUUCGCACGGUUGGCAGA
			AAAAAAAA
	I-2	Sigma, IVT	Forward:Primer 5
			Reverse:TTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG
			CGAAGCGTGTCATTGACTATAGTGAGTCGTATTAATTTC
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA
			<u>CGGUUGGCAG</u> AAAAAAAAA
	I-3	Sigma, IVT	Forward:Primer 5
			Reverse:TTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG
			CGAAGCGTGTGCGAAGCGTGTCATTGACTATAGTGAGT
			CGTATTAATTTC
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA
		-	CACGCUUCGCA <u>CGGUUGGCAG</u> AAAAAAAAA
	I-4	Sigma, IVT	Forward:Primer 5
			Reverse:TTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG
			CGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTCATTG
			ACTATAGTGAGTCGTATTAATTTC
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA
			CACGCUUCGCACACGCUUCGCA <u>CGGUUGGCAG</u> AAAAA
	1-5	Sigma, IVI	Forward:Primer 5
			Reverse: IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
	1-6	Sigma IV/T	Eorward: Drimer 5
	1-0	Sigina, ivi	
			GCGTGTGCGAAGCGTGTCATTGACTATAGTGAGTCGTA

			ΤΤΛΛΤΤΤΟ
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA CACGCUUCGCACACGCUUCGCACACGCUUCGCACACG
	1-7	Sigma, IV I	Forward:Primer 5 Reverse:TTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG CGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAA GCGTGTGCGAAGCGTGTGCGAAGCGTGTCATTGACTAT AGTGAGTCGTATTAATTTC Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA CACGCUUCGCACACGCUUCGCACACGCUUCGCACACG
			CUUCGCACACGCUUCGCA <u>CGGUUGGCAG</u> AAAAAAAAA
			A
	1-8	Sigma IVT	Forward Primer 5
	1-0	olgina, iv i	Reverse:TTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG CGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAA GCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAAGCGT GTCATTGACTATAGTGAGTCGTATTAATTTC Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA CACGCUUCGCACACGCUUCGCACACGCUUCGCACACG
1			CUUCGCACACGCUUCGCACACGCUUCGCACGGUUGG
	1.0	Sigmo IVT	<u>Orward</u>
	1-9	Sigma, IVI	Forward:Primer 5 Reverse:TTTTTTTTTTTCTGCCAACCGTGCGAAGCGTGTG
			GIGCGAAGCGIGICATIGACIAIAGIGAGICGIAIIAA
			TTTC
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA
			CACGCUUCGCACACGCUUCGCACACGCUUCGCACACG
			CUUCGCACACGCUUCGCACACGCUUCGCACACGCUU
			CGCACGGUUGGCAGAAAAAAAAA
	1.10	Sigma IV/T	Forward Drimor 5
	1-10	Sigilia, IVI	
			CGAAGCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAA
			GCGTGTGCGAAGCGTGTGCGAAGCGTGTGCGAAGCGT
			GTGCGAAGCGTGTGCGAAGCGTGTCATTGACTATAGTG
			AGTCGTATTAATTTC
			Transcript:GUCAAUGACACGCUUCGCACACGCUUCGCA
			CACGCUUCGCACACGCUUCGCACACGCUUCGCACACG
			CUUCGCACACGCUUCGCACACGCUUCGCACACGCUU
			CGCACACGCUUCGCACGGUUGGCAGAAAAAAAAAA
	П	Dharmacon	
	ш	Sigma IV/T	Forward Primer 5
	111	Sigina, IV I	
			Transcript:GUUCCGAAUUGACCUAGGGUAG <u>CGGUUGG</u>
			<u>CAG</u> AAAAAAAAA
	IIIC19	Sigma, IVT	Forward:Primer 5
			Reverse:TTTTTTTTTTTCTGCCAACCGCTACCCTAGGTCA
			ATTCGGAAGTCATTGTCTATAGTGAGTCGTATTAATTTC
			Transcript:GACAAUGACUUCCGAAUUGACCUAGGGUAG
			CGGUUGGCAGAAAAAAAAA
	IV	Sigma IV/T	Forward Primer 5
	1 V	Sigina, IV I	Reverse TTTTTTTTGAATCAACCCCCACCTCCAATC
			TTEATTETETATACTEACTECTATTAATTE

			UCAAAAAAAAA
	IVC19	Sigma IVT	Forward Primer 5
		olgina, ivi	Reverse TTTTTTTTTGAATCAAGGGCCGAGGTCCAATC
			TGTTTAATGTCATTGTCTATAGTGAGTCGTATTAATTTC
	V	Sigma IVT	Earward: Primer 5
	v	Sigilia, IV I	
	T 1/4		
	1 1 1	Sigma, IVI	Forward:Primer 5
			Reverse:
			Transcript:GUCUUUAGAACCCACGGCGUAG <u>CUACCUAU</u>
	TVIC19	Sigma, IVI	Forward:Primer 5
			Reverse:TTTTTTTTTGATAGGTAGCTACGCCGTGGGTT
			CTAAAGTGTCATTGTCTATAGTGAGTCGTATTAATTTC
			Transcript:GACAAUGACACUUUAGAACCCACGGCGUAG
			<u>CUACCUAUC</u> AAAAAAAAAA
	32A	Sigma, IVT	Forward:Primer 5
			Reverse:GATTAAGTGCTATTCAGGACGTCTGCCAACCG
			TGCGAAGCGTGGATTCATTGACTATAGTGAGTCGTATT
			AATTTC
			Transcript:GUCAAUGAAUCCACGCUUCGCACGGUUGGC
			AGACGUCCUGAAUAGCACUUAAUC
	32B	Sigma, IVT	Forward:Primer 5
			Reverse:ATACCTGTTCGCCAGCGTTACTGAATCAAGGG
			CCGAGGTCCAATCTGTTTAAGCTATAGTGAGTCGTATTA
			ATTTC
			Transcript:GCUUAAACAGAUUGGACCUCGG <u>CCCUUGAU</u>
			UCAGUAACGCUGGCGAACAGGUAU
	MzTemp	Sigma, IVT	Forward:Primer 6
	Gel purified		Reverse:CTGCCAACCGCTGATGAGCGAAAGTGTGGAGT
	,		GTGTGTGAGTGTCATTGACTATAGTGAGTCG
			Transcript:GUCAAUGACACUCACACACACUCCACACUU
			UCGCUCAUCAGCGGUUGGCAG
Competing	CompI	Dharmacon	CUGCCAACCGUGCGAAGCGUGGAUUCAUUG
oligonucle-	CompISea	Dharmacon	GGACAACCUGUUCUGCCAACCUUGCGAAGCGUGGAU
otides			U-dideoxvC
Minizvme	MzSub	IDT	Cv5-CACUCCACACUCCGGUUGGCAG
, -	Mz	IDT	CUGCCAACCGCUGAUGAGCGAAAGUGUGGAGUG
	MzComp	Sigma	CTGCCAACCGCTGATGAGCGAAAGTGTGGAGTGTGTG
	mzeemp	olgina	TGAGTGTCATTGACTATAGTGAGTCG
Template	T7C19	Sigma	GATCGAGATCTCGATCCCGCGAAATTAATACGACTCAC
library		Sigina	TATAGTCAATGACACGCTTCGCACACGCTTC
synthesis	Templ ib	Sigma	
		Sigina	nnGAAGCGTGTGCGAAGCGTG
	Template	IVT	
	library		
Tomplata	TompDoo	Sigma	
soloction	DriDoo	Sigma	
Selection		Sigilia	
1	עטרט ן	Sigma	TITITITICIGUCAACCG

6.12 Abbreviations

4-thioU	.4-thiouracil
4-thioUTP	.4-thiouridine triphosphate
8-oxo-dGTP	.8-oxo-deoxyguanosine triphosphates
ATP	Adenosine triphosphates
CBT	.Compartmentalised bead-tagging
cDNA	.Complementary DNA
CSR	.Compartmentalised self-replication
CTP	.Cytidine triphosphate
d	Days
dideoxyGTP	.Dideoxyguanosine triphosphate
dPTP	.6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-
	[4,5-c][1,2]oxazin-7-one triphosphate
DTT	Dithiothreitol
EDTA	.Ethylenediaminetetraacetic acid
f.c	.Final concentration
FACS	.Fluorescence-activated cell sorting
GMP	.Guanosine monophosphate
h	.Hours
IDT	Integrated DNA Technologies
IVT	. In vitro transcription
LMB	Laboratory of Molecular Biology
min	.Minutes
MRC	Medical Research Council
mRNA	messenger ribonucleic acid
MSS	.MegaShortScript
NAD ⁺	Nicotinamide adenine dinucleotide
NEB	New England Biolabs
nt	Nucleotides
NTP	.Nucleoside triphosphate
PAGE	.Polyacrylamide gel electrophoresis
PCR	.Polymerase chain reaction
RCA	.Rolling circle amplification
RNA	.Ribonucleic acid
RNAse P	.Ribonuclease P
RPA	.Ribozyme polymerase plate assay
RT-PCR	Reverse transcription polymerase chain reaction
RTT	.Run-through transcript
S	Seconds
s.d	.Standard deviation
s.e.m	.Standard error of the mean
SEM	.Scanning electron microscopy
StEP	.Staggered extension process
UTP	Uridine triphosphate
wt	.Wild type

References

Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. and Holliger, P. (2010). "Ice as a protocellular medium for RNA replication." <u>Nat Commun</u> **1**(6): doi:10 1038/ncomms1076.

Baaske, P., Weinert, F. M., Duhr, S., Lemke, K. H., Russell, M. J. and Braun, D. (2007). "Extreme accumulation of nucleotides in simulated hydrothermal pore systems." <u>Proc Natl Acad Sci U S A</u> **104**(22): 9346-51.

Bada, J. L., Bigham, C. and Miller, S. L. (1994). "Impact melting of frozen oceans on the early Earth: implications for the origin of life." <u>Proc Natl Acad Sci U S A</u> **91**: 1248-50.

Bartel, D. P. and Szostak, J. W. (1993). "Isolation of new ribozymes from a large pool of random sequences [see comment]." <u>Science</u> **261**(5127): 1411-8.

Bartel, D. P. and Unrau, P. J. (1999). "Constructing an RNA world." <u>Trends Cell Biol</u> **9**(12): M9-M13.

Benner, S. A., Ellington, A. D. and Tauer, A. (1989). "Modern metabolism as a palimpsest of the RNA world." <u>Proc Natl Acad Sci U S A</u> **86**(18): 7054-8.

Blackmond, D. G. (2010). "The origin of biological homochirality." <u>Cold Spring Harb</u> <u>Perspect Biol</u> **2**(5): a002147.

Boving, T. B. and Grathwohl, P. (2001). "Tracer diffusion coefficients in sedimentary rocks: correlation to porosity and hydraulic conductivity." <u>J Contam Hydrol</u> **53**(1-2): 85-100.

Brune, D. and Kim, S. (1993). "Predicting protein diffusion coefficients." <u>Proc Natl</u> <u>Acad Sci U S A</u> **90**(9): 3835-9.

Budin, I., Bruckner, R. J. and Szostak, J. W. (2009). "Formation of protocell-like vesicles in a thermal diffusion column." <u>J Am Chem Soc</u> **131**(28): 9628-9.

Cardinaux, F., Stradner, A., Schurtenberger, P., Sciortino, F. and Zaccarrelli, E. (2007). "Modeling equilibrium clusters in lysozyme solutions." <u>Europhysics Letters</u> **77**: 48004-p1-p5.

Chen, I. A., Salehi-Ashtiani, K. and Szostak, J. W. (2005). "RNA catalysis in model protocell vesicles." J Am Chem Soc **127**(38): 13213-9.

Crick, F. H. (1968). "The origin of the genetic code." <u>J Mol Biol</u> **38**(3): 367-79.

Diehl, F., Li, M., He, Y., Kinzler, K. W., Vogelstein, B. and Dressman, D. (2006). "BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions." <u>Nat</u> <u>Methods</u> **3**(7): 551-9. Dobson, C. M., Ellison, G. B., Tuck, A. F. and Vaida, V. (2000). "Atmospheric aerosols as prebiotic chemical reactors." <u>Proc Natl Acad Sci U S A</u> **97**(22): 11864-8.

Doudna, J. A. and Szostak, J. W. (1989). "RNA-catalysed synthesis of complementary-strand RNA." <u>Nature</u> **339**(6225): 519-22.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." <u>Nucleic Acids Res</u> **32**(5): 1792-7.

Eigen, M. (1971). "Selforganization of matter and the evolution of biological macromolecules." <u>Naturwissenschaften</u> **58**(10): 465-523.

Ekland, E. H. and Bartel, D. P. (1995). "The secondary structure and sequence optimization of an RNA ligase ribozyme." <u>Nucleic Acids Res</u> **23**(16): 3231-8.

Ekland, E. H. and Bartel, D. P. (1996). "RNA-catalysed RNA polymerization using nucleoside triphosphates." <u>Nature</u> **383**(6596): 192.

Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995). "Structurally complex and highly active RNA ligases derived from random RNA sequences." <u>Science</u> **269**(5222): 364-70.

Ellington, A. D., Chen, X., Robertson, M. and Syrett, A. (2009). "Evolutionary origins and directed evolution of RNA." Int J Biochem Cell Biol **41**(2): 254-65.

Ferris, J. P. (2006). "Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life." <u>Philos Trans R Soc Lond B Biol Sci</u> **361**(1474): 1777-86; discussion 1786.

Fiammengo, R. and Jaschke, A. (2005). "Nucleic acid enzymes." <u>Curr Opin</u> <u>Biotechnol</u> **16**(6): 614-21.

Freeland, S. J., Knight, R. D. and Landweber, L. F. (1999). "Do proteins predate DNA?" <u>Science</u> **286**(5440): 690-2.

Gesteland, R. F., Cech, T. R. and Atkins, J. F. (2006). <u>The RNA World 3rd ed.</u> Cold Spring Harbor Laboratory Press, N.Y.

Ghadessy, F. J., Ong, J. L. and Holliger, P. (2001). "Directed evolution of polymerase function by compartmentalized self-replication." <u>Proc Natl Acad Sci U S A</u> **98**(8): 4552-7.

Gilbert, W. (1986). "Origin of life: The RNA world." <u>Nature</u> **319**(6055): 618-618.

Glasner, M. E., Bergman, N. H. and Bartel, D. P. (2002). "Metal ion requirements for structure and catalysis of an RNA ligase ribozyme." <u>Biochemistry</u> **41**(25): 8103-12.

Grimm, R. E., Stillman, D. E., Dec, S. F. and Bullock, M. A. (2008). "Low-frequency electrical properties of polycrystalline saline ice and salt hydrates." <u>J Phys Chem B</u> **112**(48): 15382-90.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." <u>Cell</u> **35**(3 Pt 2): 849-57.

Hager, A. J., Pollard, J. D. and Szostak, J. W. (1996). "Ribozymes: aiming at RNA replication and protein synthesis." <u>Chem Biol</u> **3**(9): 717-25.

Hren, M. T., Tice, M. M. and Chamberlain, C. P. (2009). "Oxygen and hydrogen isotope evidence for a temperate climate 3.42 billion years ago." <u>Nature</u> **462**(7270): 205-8.

Jadhav, V. R. and Yarus, M. (2002). "Coenzymes as coribozymes." <u>Biochimie</u> **84**(9): 877-88.

Jeffares, D. C., Poole, A. M. and Penny, D. (1998). "Relics from the RNA world." J Mol Evol **46**(1): 18-36.

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. and Bartel, D. P. (2001). "RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension." <u>Science</u> **292**(5520): 1319-25.

Joyce, G. F. (2004). "Directed evolution of nucleic acid enzymes." <u>Annu Rev</u> <u>Biochem</u> **73**: 791-836.

Joyce, G. F. (2007). "Forty years of in vitro evolution." <u>Angew Chem Int Ed Engl</u> **46**(34): 6420-36.

Joyce, G. F. (2007). "Structural biology. A glimpse of biology's first enzyme." <u>Science</u> **315**(5818): 1507-8.

Krembs, C., Eicken, H. and Deming, J. W. (2011). "Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic." <u>Proc Natl Acad Sci U S A</u> **108**(9): 3653-8.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. and Cech, T. R. (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." <u>Cell</u> **31**(1): 147-57.

Kun, A., Santos, M. and Szathmary, E. (2005). "Real ribozymes suggest a relaxed error threshold." <u>Nat Genet</u> **37**(9): 1008-11.

Lawrence, M. S. and Bartel, D. P. (2003). "Processivity of ribozyme-catalyzed RNA polymerization." <u>Biochemistry</u> **42**(29): 8748-55.

Lawrence, M. S. and Bartel, D. P. (2005). "New ligase-derived RNA polymerase ribozymes." <u>Rna</u> **11**(8): 1173-80.

Levy, M., Griswold, K. E. and Ellington, A. D. (2005). "Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization." <u>Rna</u> **11**(10): 1555-62.

Levy, M. and Miller, S. L. (1998). "The stability of the RNA bases: implications for the origin of life." <u>Proc Natl Acad Sci U S A</u> **95**(14): 7933-8.

Li, M., Diehl, F., Dressman, D., Vogelstein, B. and Kinzler, K. W. (2006). "BEAMing up for detection and quantification of rare sequence variants." <u>Nat Methods</u> **3**(2): 95-7.

Li, Y. and Breaker, R. R. (1999). "Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2,Äò-Hydroxyl Group." <u>Journal of the American Chemical Society</u> **121**(23): 5364-5372.

Lincoln, T. A. and Joyce, G. F. (2009). "Self-sustained replication of an RNA enzyme." <u>Science</u> **323**(5918): 1229-32.

Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C. and Ward, D. C. (1998). "Mutation detection and single-molecule counting using isothermal rolling-circle amplification." <u>Nat Genet</u> **19**(3): 225-32.

Logan, D. T., Andersson, J., Sjoberg, B. M. and Nordlund, P. (1999). "A glycyl radical site in the crystal structure of a class III ribonucleotide reductase." <u>Science</u> **283**(5407): 1499-504.

Manapat, M. L., Chen, I. A. and Nowak, M. A. (2010). "The basic reproductive ratio of life." <u>J Theor Biol</u> **263**(3): 317-27.

Mansy, S. S., Schrum, J. P., Krishnamurthy, M., Tobe, S., Treco, D. A. and Szostak, J. W. (2008). "Template-directed synthesis of a genetic polymer in a model protocell." <u>Nature</u> **454**(7200): 122-5.

McCall, M. J., Hendry, P., Mir, A. A., Conaty, J., Brown, G. and Lockett, T. J. (2000). "Small, efficient hammerhead ribozymes." <u>Mol Biotechnol</u> **14**(1): 5-17.

McCarthy, C., Cooper, R. F., Kirby, S. H., Rieck, K. D. and Stern, L. A. (2007). "Solidification and microstructures of binary ice-I/hydrate eutectic aggregates." <u>American Mineralogist</u> **92**: 1550-60.

Monnard, P. A. (2005). "Catalysis in abiotic structured media: an approach to selective synthesis of biopolymers." <u>Cell Mol Life Sci</u> **62**(5): 520-34.

Monnard, P. A., Apel, C. L., Kanavarioti, A. and Deamer, D. W. (2002). "Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium." <u>Astrobiology</u> **2**(2): 139-52.

Monnard, P. A., Kanavarioti, A. and Deamer, D. W. (2003). "Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases." J Am Chem Soc **125**(45): 13734-40.

Monnard, P. A. and Szostak, J. W. (2008). "Metal-ion catalyzed polymerization in the eutectic phase in water-ice: a possible approach to template-directed RNA polymerization." <u>J Inorg Biochem</u> **102**(5-6): 1104-11.

Monnard, P. A. and Ziock, H. (2008). "Eutectic phase in water-ice: a self-assembled environment conducive to metal-catalyzed non-enzymatic RNA polymerization." <u>Chem Biodivers</u> **5**(8): 1521-39.

Muller, U. F. (2006). "Re-creating an RNA world." Cell Mol Life Sci 63(11): 1278-93.

Muller, U. F. and Bartel, D. P. (2003). "Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization." <u>Chem Biol</u> **10**(9): 799-806.

Muller, U. F. and Bartel, D. P. (2008). "Improved polymerase ribozyme efficiency on hydrophobic assemblies." <u>Rna</u> **14**(3): 552-62.

Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T. A. (2000). "The structural basis of ribosome activity in peptide bond synthesis." <u>Science</u> **289**(5481): 920-30.

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. and Steitz, T. A. (2001). "RNA tertiary interactions in the large ribosomal subunit: the A-minor motif." <u>Proc Natl Acad</u> <u>Sci U S A</u> **98**(9): 4899-903.

Nowak, M. A. and Ohtsuki, H. (2008). "Prevolutionary dynamics and the origin of evolution." <u>Proc Natl Acad Sci U S A</u> **105**(39): 14924-7.

Orgel, L. E. (1968). "Evolution of the genetic apparatus." J Mol Biol **38**(3): 381-93.

Pace, N. R. (1991). "Origin of life--facing up to the physical setting." <u>Cell</u> **65**(4): 531-3.

Pace, N. R. and Marsh, T. L. (1985). "RNA catalysis and the origin of life." <u>Orig Life</u> <u>Evol Biosph</u> **16**(2): 97-116.

Pegram, L. M. and Record, M. T., Jr. (2007). "Hofmeister salt effects on surface tension arise from partitioning of anions and cations between bulk water and the airwater interface." <u>J Phys Chem B</u> **111**(19): 5411-7.

Petrie, K. L. and Joyce, G. F. (2010). "Deep sequencing analysis of mutations resulting from the incorporation of dNTP analogs." <u>Nucleic Acids Res</u>.

Powner, M. W., Gerland, B. and Sutherland, J. D. (2009). "Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions." <u>Nature</u> **459**(7244): 239-42.

Rajamani, S., Ichida, J. K., Antal, T., Treco, D. A., Leu, K., Nowak, M. A., Szostak, J. W. and Chen, I. A. (2010). "Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication." <u>J Am Chem Soc</u> **132**(16): 5880-5.

Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T. and Mondragon, A. (2010). "Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA." <u>Nature</u> **468**(7325): 784-9.

Robertson, M. P. and Joyce, G. F. (2010). "The Origins of the RNA World." <u>Cold</u> <u>Spring Harb Perspect Biol</u>.

Rosing, M. T., Bird, D. K., Sleep, N. H. and Bjerrum, C. J. (2010). "No climate paradox under the faint early Sun." <u>Nature</u> **464**(7289): 744-7.

Schrum, J. P., Ricardo, A., Krishnamurthy, M., Blain, J. C. and Szostak, J. W. (2009). "Efficient and rapid template-directed nucleic acid copying using 2'-amino-2',3'dideoxyribonucleoside-5'-phosphorimidazolide monomers." <u>J Am Chem Soc</u> **131**(40): 14560-70.

Schrum, J. P., Zhu, T. F. and Szostak, J. W. (2010). "The origins of cellular life." <u>Cold</u> <u>Spring Harb Perspect Biol</u> **2**(9): a002212.

Shechner, D. M., Grant, R. A., Bagby, S. C., Koldobskaya, Y., Piccirilli, J. A. and Bartel, D. P. (2009). "Crystal structure of the catalytic core of an RNA-polymerase ribozyme." <u>Science</u> **326**(5957): 1271-5.

Shine, J. and Dalgarno, L. (1975). "Determinant of cistron specificity in bacterial ribosomes." <u>Nature</u> **254**(5495): 34-8.

Sleep, N. H., Zahnle, K. and Neuhoff, P. S. (2001). "Initiation of clement surface conditions on the earliest Earth." <u>Proc Natl Acad Sci U S A</u> **98**(7): 3666-72.

Sun, X., Li, J. M. and Wartell, R. M. (2007). "Conversion of stable RNA hairpin to a metastable dimer in frozen solution." <u>Rna</u> **13**(12): 2277-86.

Szabo, P., Scheuring, I., Czaran, T. and Szathmary, E. (2002). "In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity." <u>Nature</u> **420**(6913): 340-3.

Szostak, J. W., Bartel, D. P. and Luisi, P. L. (2001). "Synthesizing life." <u>Nature</u> **409**(6818): 387-90.

Talini, G., Gallori, E. and Maurel, M. C. (2009). "Natural and unnatural ribozymes: back to the primordial RNA world." <u>Res Microbiol</u> **160**(7): 457-65.

Ternan, M. (1987). "The Diffusion of Liquids in Pores." <u>The Canadian Journal of</u> <u>Chemical Engineering</u> **65**: 244-249.

Thomen, P., Lopez, P. J., Bockelmann, U., Guillerez, J., Dreyfus, M. and Heslot, F. (2008). "T7 RNA polymerase studied by force measurements varying cofactor concentration." <u>Biophys J</u> **95**(5): 2423-33.

Tielrooij, K. J., Garcia-Araez, N., Bonn, M. and Bakker, H. J. (2010). "Cooperativity in ion hydration." <u>Science</u> **328**(5981): 1006-9.

Tindall, K. R. and Kunkel, T. A. (1988). "Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase." <u>Biochemistry</u> **27**(16): 6008-13.

Toor, N., Keating, K. S. and Pyle, A. M. (2009). "Structural insights into RNA splicing." <u>Curr Opin Struct Biol</u> **19**(3): 260-6.

Trinks, H., Schroder, W. and Biebricher, C. K. (2005). "Ice and the origin of life." <u>Orig</u> <u>Life Evol Biosph</u> **35**(5): 429-45.

Tucker, B. J. and Breaker, R. R. (2005). "Riboswitches as versatile gene control elements." <u>Curr Opin Struct Biol</u> **15**(3): 342-8.

Vajda, T. (1999). "Cryo-bioorganic chemistry: molecular interactions at low temperature." <u>Cell Mol Life Sci</u> **56**(5-6): 398-414.

Valadkhan, S., Mohammadi, A., Jaladat, Y. and Geisler, S. (2009). "Protein-free small nuclear RNAs catalyze a two-step splicing reaction." <u>Proc Natl Acad Sci U S A</u> **106**(29): 11901-6.

Valley, J. W., Peck, W. H., King, E. M. and Wilde, S. A. (2002). "A cool early Earth." <u>Geology</u> **30**: 351-354.

Verlander, M. S., Lohrmann, R. and Orgel, L. E. (1973). "Catalysts for the self-polymerization of adenosine cyclic 2', 3'-phosphate." J Mol Evol **2**(4): 303-16.

Vicens, Q. and Cech, T. R. (2009). "A natural ribozyme with 3',5' RNA ligase activity." <u>Nat Chem Biol</u> **5**(2): 97-9.

Vlassov, A. V., Johnston, B. H., Landweber, L. F. and Kazakov, S. A. (2004). "Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world." <u>Nucleic Acids Res</u> **32**(9): 2966-74.

Vlassov, A. V., Kazakov, S. A., Johnston, B. H. and Landweber, L. F. (2005). "The RNA world on ice: a new scenario for the emergence of RNA information." <u>J Mol Evol</u> **61**(2): 264-73.

Wang, Q. S., Cheng, L. K. and Unrau, P. J. (2011). "Characterization of the B6.61 polymerase ribozyme accessory domain." <u>Rna</u> **17**(3): 469-77.

Wilde, S. A., Valley, J. W., Peck, W. H. and Graham, C. M. (2001). "Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago." <u>Nature</u> **409**(6817): 175-8.

Wilson, D. S. and Szostak, J. W. (1999). "In vitro selection of functional nucleic acids." <u>Annu Rev Biochem</u> **68**: 611-47.

Wochner, A., Attwater, J., Coulson, A. and Holliger, P. (2011). "Ribozyme-catalyzed transcription of an active ribozyme." <u>Science</u> **332**(6026): 209-12.

Woese, C. R. (1967). <u>The Genetic Code: The Molecular Basis for Genetic</u> <u>Expression</u>. Harper and Row, New York.

Yoffe, A. M., Prinsen, P., Gelbart, W. M. and Ben-Shaul, A. (2011). "The ends of a large RNA molecule are necessarily close." <u>Nucleic Acids Res</u> **39**(1): 292-9.

Zaccolo, M. and Gherardi, E. (1999). "The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase." <u>J Mol</u> <u>Biol</u> **285**(2): 775-83.

Zaher, H. S. and Unrau, P. J. (2007). "Selection of an improved RNA polymerase ribozyme with superior extension and fidelity." <u>Rna</u> **13**(7): 1017-26.

Zahnle, K. J. (2006). "Earth's earliest atmosphere." Elements 2(4): 217-222.

Zhang, Y. and Cremer, P. S. (2006). "Interactions between macromolecules and ions: The Hofmeister series." <u>Curr Opin Chem Biol</u> **10**(6): 658-63.

Zhao, H., Giver, L., Shao, Z., Affholter, J. A. and Arnold, F. H. (1998). "Molecular evolution by staggered extension process (StEP) in vitro recombination." <u>Nat</u> <u>Biotechnol</u> **16**(3): 258-61.

Zhao, H. and Zha, W. (2006). "In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP)." <u>Nat Protoc</u> **1**(4): 1865-71.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." <u>Nucleic Acids Res</u> **31**(13): 3406-15.