[RNA Polymerase Ribozymes]

Michael S. Lawrence

B.A., Biochemistry and Linguistics & Cognitive Science
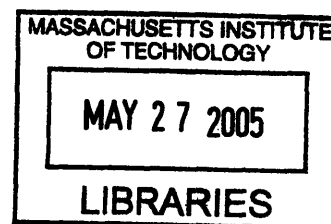Brandeis University, 1998

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2005

© 2005 Michael S. Lawrence. All Rights Reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part.

Signature of Author:

Michael S. Lawrence
Department of Biology
(May 20, 2005)

Certified by:

David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by:

Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

RNA Polymerase Ribozymes

by

Michael S. Lawrence

Submitted to the Department of Biology
on April 29, 2005
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Biology

ABSTRACT

The RNA World is a hypothetical ancient evolutionary era during which RNA was both genome and catalyst. During that time, RNA was the only kind of enzyme yet in existence, and one of its chief duties was the replication of RNA. This scenario presupposes that among all possible RNA sequences, there exist RNA replicase ribozymes, capable of synthesizing RNA using the information in an RNA template. The goal of the present work is to provide experimental evidence in support of this conjecture, by isolating such ribozymes in the laboratory. We created a large pool of RNA molecules each containing a previously isolated RNA ligase ribozyme and a large stretch of random RNA. Applying *in vitro* evolution to select for molecules that could extend a tethered RNA primer using nucleoside triphosphates, we isolated nine distinct classes of polymerase ribozymes. Two of these rudimentary polymerases were further evolved to the point that they each could add 14 nucleotides to an untethered primer-template. One of them was subjected to a detailed further characterization. The polymerization it catalyzes was shown to be accurate, with an average fidelity of nearly 97%. It was shown to be general, with primer-templates of all sequences and lengths being accepted as substrates. Finally, it was shown to be partially processive, with the polymerase achieving processivity as high as 90% in a few instances. The polymerase is currently limited by its low affinity for the primer-template. Future work will focus on improving primer-template binding, in order to produce a polymerase that can synthesize longer RNA.

Thesis Supervisor: David P. Bartel
Title: Professor of Biology

Dedication

To my poncho Steve.

Acknowledgements

First and foremost I thank my advisor Dave Bartel, whose unwavering support and kindness toward me set the tone for a very enjoyable and fulfilling graduate program. Dave's formidable insight and expertise, always generously shared when I turned to him for help, were an invaluable resource. Back when he still occasionally donned the lab gloves, he taught me how to run a nine-bottle part-doped, part-random DNA pool synthesis without squirting or puncturing myself with the amidite-filled syringes. Later, when I was struggling with the paper-writing process, he taught me to set manageable intermediate goals, probably the single most valuable skill I learned in grad school.

Next I thank all the former and present members of the Bartel Lab with whom I have had the good fortune to share my time there, especially:

Wendy Johnston, who graciously welcomed me as a collaborator on the polymerase project. Wendy was a primary teacher for me when I was new in the lab, and she taught me how to do *in vitro* selection, as well as most of the other techniques that formed the basis of this thesis work.

Peter Unrau, who impressed upon me the value of including extra controls in every experiment—because the less you have to assume, the better. He also taught me to write down my conclusions in my lab notebook, instead of (as he too once had done) letting the raw data speak for itself, uncontaminated by vulgar things like interpretations and ideas. Peter taught me to use the Phosphorimager, which became my moody inseparable lab sidekick for many years, and he helped me in my early stages of thinking about a kinetic model for ribozyme processivity. He also sparked my interest in astronomy by letting me see the rings of Saturn, for the first time with my own eyes, through his homemade telescope.

Uli Müller, whose enthusiasm for new ideas and methodologies is a tremendous positive influence to be around. He taught me to use the dynamic light-scattering machine, even if just for the sake of learning something new, because in the end, that's the best possible reason. Uli also provided a continually renewed source of news and entertainment, the "Wall of Other Interests", perfectly stationed within the lab to occupy those hour-long minutes when a Phosphorimager plate is being scanned.

And finally, Ed Curtis, whose careful intellect and tremendous personal integrity I deeply admire.

Table of Contents

New RNA Polymerase Ribozymes

Processivity of Ribozyme-Catalyzed RNA Polymerization

Chapter Two has been published previously as:
M.S. Lawrence and D.P. Bartel, "Processivity of Ribozyme-Catalyzed
RNA Polymerization." Biochemistry 42, 8748–8755
© 2003 American Chemical Society.

Secondary Structure and Sequence Optimization
of a Second RNA Polymerase Ribozyme

A New Tool for Polymerase Ribozyme Evolution:
Selecting for Addition of Many Nucleotides

RNA-Catalyzed RNA Polymerization:
Accurate and General RNA-Templated Primer Extension

The Appendix has been published previously as:
W.K. Johnston, P.J. Unrau, M.S. Lawrence, M.E. Glasner, and D.P. Bartel,
"RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated
Primer Extension." Science 292, 1319–1325. © 2001 American Association
for the Advancement of Science.

INTRODUCTION

One thing everyone can agree on is the origin of life. The conclusion is inescapable, because we see it all around us: our gardens, our pets, our friends, our children, ourselves. We are united by a universal consensus that at some point in the history of the Earth, living things appeared on its surface. Furthermore, we share a natural curiosity about where they came from. Seen on the scale of our own lifetimes, the question is a trivial one: everyone came from their parents. Every plant came from a seed, every chicken from an egg. The question gets harder when we escape our familiar timescales and take a longer view, asking: Where did the first parents come from? With this question, we begin to reach the limits of our consensus, branching out instead with a flowering diversity of hypotheses. Of these, the one that has proved the most compelling and inspiring to the greatest number of people throughout history is that our first ancestors were created by an intelligent and loving God. It may be fair to say that this explanation remains predominant.

A second hypothesis has succeeded in attracting the devotion of a passionate minority who view themselves as the product of billions of years of mindless, indifferent, random chemical confusion, a wispy chaotic flicker dissipating on the margins of the overwhelming blast of power from the sun's detonation.

Despite the enormous divergence of their implications, there are ways of partially reconciling these two views. Admittedly some do assert detailed knowledge of God's operating protocols, but a more common attitude acknowledges the humble limits of the human intellect: if God chose to accomplish his goals by designing and executing the Big Bang, thereby setting into motion the processes of physical and Darwinian evolution, then let none find fault. If on the other hand, there actually was no plan or architect, and the process unfolding around us was indeed indeliberate, then what disgrace is there in imagining a purposeful spirit to the natural laws of the universe? Scientists find such personifying metaphor irresistible.[1]

Why study the origin of life? For those certain that God created us in the straightforward way a toymaker might fashion a doll, no further enquiry is justified. But for those who admit the possibility of a subtler scheme, whereby his complex ideas unfolded organically and fuguelike from simple seeds—as well as for all those certain he played no part—there is much to be learned about the process that brought us forth. It is the first chapter in our history book, essential for a complete account of the human experience, and we are only just starting to write it. In the words of one contemporary researcher:

> *It is not difficult to argue that research on the beginning of life represents one of the last bastions of classical science, defined by the significance of its central goal, its breadth of scope, and a ratio of hypothesis to fact approaching infinity.*[2]

Where to begin

Many unanswered questions confront the student of life's origins: What was the early earth like?[3] How long did it take for life to emerge?[4] Were the first living things similar to modern-day organisms, or very different?[5] In the face of these nearly consummate mysteries, it can be useful to start with what we do know.

All modern cells have several things in common: first, a cell membrane, to keep the "self" inside and the environment outside; second, a set of highly sophisticated protein machines for all manner of household chores, such as controlling what gets into and out of the cell, catalyzing the breakdown of food into energy, synthesizing the cell's unique building materials, maintaining and

repairing the structures that hold the cell together, and sensing and transducing environmental cues. These enzymes, critical for every part of the cell's business, are not assembled randomly; instead, they are built from precise blueprints encoded in DNA. The third feature common to all modern cells, DNA dictates how and when to synthesize every protein, as well as orchestrating many other processes we are still only just beginning to understand.

When a cell divides, it must partition all its components in such a way that both daughter cells have what they need to survive. This includes passing on a complete set of instructions to each of them, and thus the cell needs a way to copy its DNA. This duplication is performed by protein enzymes, with dozens of them working together in huge complexes to faithfully copy the genetic material. The task is facilitated by the highly regular and self-complementary structure of double-stranded DNA, which is built from only two components: the A–T base-pair and the G–C base-pair. The shapes of these two base-pairs are essentially identical, and they are also symmetric, with the result that when they are stacked together and joined up to make double-stranded DNA, the repeating pattern makes a simple helical structure, largely irrespective of the specific order and orientation of the successive base-pairs.[6] When the time comes to replicate the DNA, the two strands can simply be pulled apart, exposing unpaired bases. At each exposed base, only one of the four available nucleotide monomers can form a proper match and reconstitute a base-pair, and this natural affinity for the complementary base is fostered and amplified by the attentive supervision of a replication enzyme. When the enzyme sees that the unpaired template base has transiently attracted a new pairing partner from the cellular soup, it solemnizes the union by permanently coupling the new base to the growing strand of new DNA. Then it shifts one position down the DNA strand to the next unpaired base, where it waits for the next nucleotide to arrive and pair. By repeating this cycle over and over (sometimes thousands of times per second), the DNA polymerase synthesizes a complete new strand of DNA by linking together the nucleotides that bind at each template position.

DNA, with its simple repetitive structure, is ideal for copying, but proteins are another story. Like DNA, they are linear polymers of similar subunits, but proteins have 20 kinds of subunits (amino acids), with no simple structural affinities relating them to each other. A protein cannot be replicated by synthesizing its complement, because no such complementarity rules exist. Furthermore, proteins adopt highly complex folded shapes, often hiding most of their amino acids deep in the interior of a globular core, completely accessible to the outside. Even if there were a simple way of reading off the sequence of amino acids, it would require navigating the convoluted twists and turns of the protein backbone. For these reasons, proteins are recognized as being "informational sinks" from which precise sequence information is effectively irretrievable. In order to make additional copies of a protein, it's not possible to just "bring one in for duplication" like making a copy of a key at the hardware store: instead, the only way to do it (at least in nature) is by going back to the DNA blueprint, and building a new one from scratch.

Evident already from just this simple overview is the highly interconnected and interdependent nature of the modern cellular machinery. A fundamental "chicken-or-egg" dilemma haunts the interrelationship of protein and DNA: proteins can be synthesized only if the DNA-encoded blueprints are available, and DNA can be synthesized only if the replication proteins are available. Which came first?

Often when confronted with difficult evolutionary questions like this, we can find our answers through comparative analysis of different organisms. For example, the human brain is arguably the most complex object yet discovered anywhere in the universe, yet we can observe

8

many of the intermediate stages it passed through during its evolution. Its humble beginnings can be seen in the decentralized neural webs of jellyfish, which allow them to coordinate their swimming motions. From there, its evolution can be traced to the first simple neuronal organizing centers in worms—to the tiny brains of insects that allow them to process complex behaviors and social cues—to the brains of fish, amphibians, and reptiles, with their accreted cerebellum—to mammals, with their cerebral cortex layered onto that—finally to humans, whose cortex has ballooned to such ridiculous proportions that it dwarfs and almost completely hides the interior, ancient parts of the brain.[7]

Comparative analysis has traditionally been the most fruitful approach to difficult biological questions, but in the case of the chicken-or-egg problem of DNA and proteins, we're simply out of luck, because there's no comparison. All organisms—from the most alien hyperthermophilic archaebacteria lurking at the bottom of the ocean, to lonely wildflowers in the rarefied alpine air, to cities full of human beings—all known living things use the full-fledged DNA+Protein system. (This fact in itself is a astonishing testament to the unity of life on Earth.)

So we are left with two possibilities: either the DNA+Protein system was part of life from the very beginning, having arisen spontaneously in an ancestral protocell from which all modern organisms descended; or else life was instead once based on a simpler system—just proteins, for example, or just DNA, or just something else. A compelling argument for the latter possibility is statistical parsimony: the chances of even one sophisticated biopolymer emerging by chance from the "prebiotic soup" can been seen as dauntingly remote,[8] whereas the perfectly coordinated appearance of two at once—pre-woven into their exquisite replicative interdependence—would have required a miracle. This does not entail its categorical exclusion, but we shall reserve our primary attention for those explanations invoking the fewest miracles. That said, the task left to us is to describe a plausible living system that is simpler than the modern-day Protein+DNA one.

RNA

To further complicate matters, we must not neglect to mention RNA, yet a third universal component of modern biochemistry. When the role of RNA was first being elucidated, and it was recognized as an essential intermediary between DNA and proteins, responsible for interpreting the "genetic code" that relates nucleotide codons to amino acids, the discovery generated a lot of excitement, in part because it gave a whiff of a solution to the chicken-or-egg problem. Some saw in tRNA a possible vestige of an early evolutionary time in which RNA was the central biomolecule.[9–12] RNA, with its highly regular structure and simple set of components, very similar to those of DNA, made it clearly plausible as an alternative genetic polymer (a role it was already known to play in some viruses).[13] At the same time, its surprising ability to fold into complex shapes and perform complicated tasks showed it was no mere passive bystander in the cellular scheme, and led Crick to remark that "tRNA looks like Nature's attempt to make RNA do the job of protein."[14] The intimate involvement of RNA in protein synthesis, as well as the ubiquity of nucleotide moieties among the coenzymes of central metabolism,[39,40] led some theorists to speculate that RNA catalysts might have been far more important in very early stages of evolution, an idea that became known as the "RNA World" hypothesis.

The speculation fell into relative obscurity until the discovery of catalytic RNAs just over a decade later.[15,16] The discovery of RNA enzymes ("ribozymes") that catalyzed intron splicing and tRNA processing was a momentous milestone in the history of the RNA World hypothesis,

immediately reinvigorating widespread interest in the idea.[17–25] It inaugurated the era of *in vitro* selection of functional nucleic acids,[26–30] which has since yielded a panoply of artificial ribozymes catalyzing a tremendous variety of chemical reactions,[31–36] some of them particularly interesting because of their potential relevance to the transition from the RNA World to the Protein/DNA world of today.[37,38] Subsequently, the three-dimensional structure of the ribosome revealed RNA lying at the very heart of modern biochemistry and provided what is arguably the most compelling evidence yet for the RNA World.[41,42]

However, one critical piece of evidence is still missing. The RNA World hypothesis has many variations, and researchers debate its specifics,[43] but the central defining feature of the idea is that RNA at one point was responsible for its own replication. Today there is no known RNA molecule that can do this. For that reason, anyone who seriously believes in a long-ago RNA World must blindly accept as an article of faith that RNA can make RNA. In order to rescue the question from the realm of faith and reclaim it for science, it would be exceedingly useful to discover an RNA molecule that can replicate RNA. Any one will do—it need not have anything in common with RNA replicases from the long-ago RNA World. It probably will look nothing like them, and furthermore we could almost certainly never find out either way.

There are a few general strategies you could follow in trying to obtain an RNA polymerase ribozyme. First, you could rent a deep-sea submersible and go off in search of undisturbed arcane microenvironments where the RNA World never went extinct, capture and culture the ribo-organisms, and clone their replicase. This approach has not been discussed at length in the literature. A second general strategy would be to take a naturally occurring ribozyme (possibly one that does something at least partly similar to polymerization) and use a combination of engineering and *in vitro* evolution to subvert it from its natural function and turn it into a polymerase.

Polymerization using the *Tetrahymena* self-splicing intron

Just such an approach was pioneered by Szostak and collaborators, using the *Tetrahymena* self-splicing intron. The choice of this ribozyme as a starting point was due in part to its being the first catalytic RNA discovered, but also in part to the chemical similarities that the exon-joining phase of its reaction bears to a single step of primer extension. Fig. 1 lists some of the reactions that this ribozyme can perform. Its natural task is self-excision from an RNA molecule, followed by splicing together of the exons (Fig. 1A). In this reaction, the similarity to RNA polymerization can be understood by considering the orange strand as analogous to a primer, the red strand as the template, and the pink U as the incoming nucleotide (although encumbered by the RNA chains extending from both its 3' and 5' termini). The reaction proceeds by attack of the primer 3' hydroxyl upon the alpha-phosphate of the pink U, resulting in the formation of a phosphodiester bond. In this case, the leaving group is an RNA strand rather than pyrophosphate, as would happen during polymerization. However, it was found that both of the reagents could be shortened to dinucleotides (Fig. 1B).[44,45] Next, by inducing the ribozyme to iterate the process, it was demonstrated to catalyze addition of 6 pyrimidines to a primer, by shifting the growing primer along the template (Fig. 1C).[46] In even closer analogy to RNA polymerization, the ribozyme binding site was engineered to allow for more general polymerization, albeit with fidelity of only 65%, and still with a nucleotide as a leaving group (Fig. 1D, where 2AP stands for the adenine analogue 2-aminopurine). The fidelity problem was partly resolved by switching to longer oligonucleotides as substrates, and it was demonstrated that the ribozyme could assemble
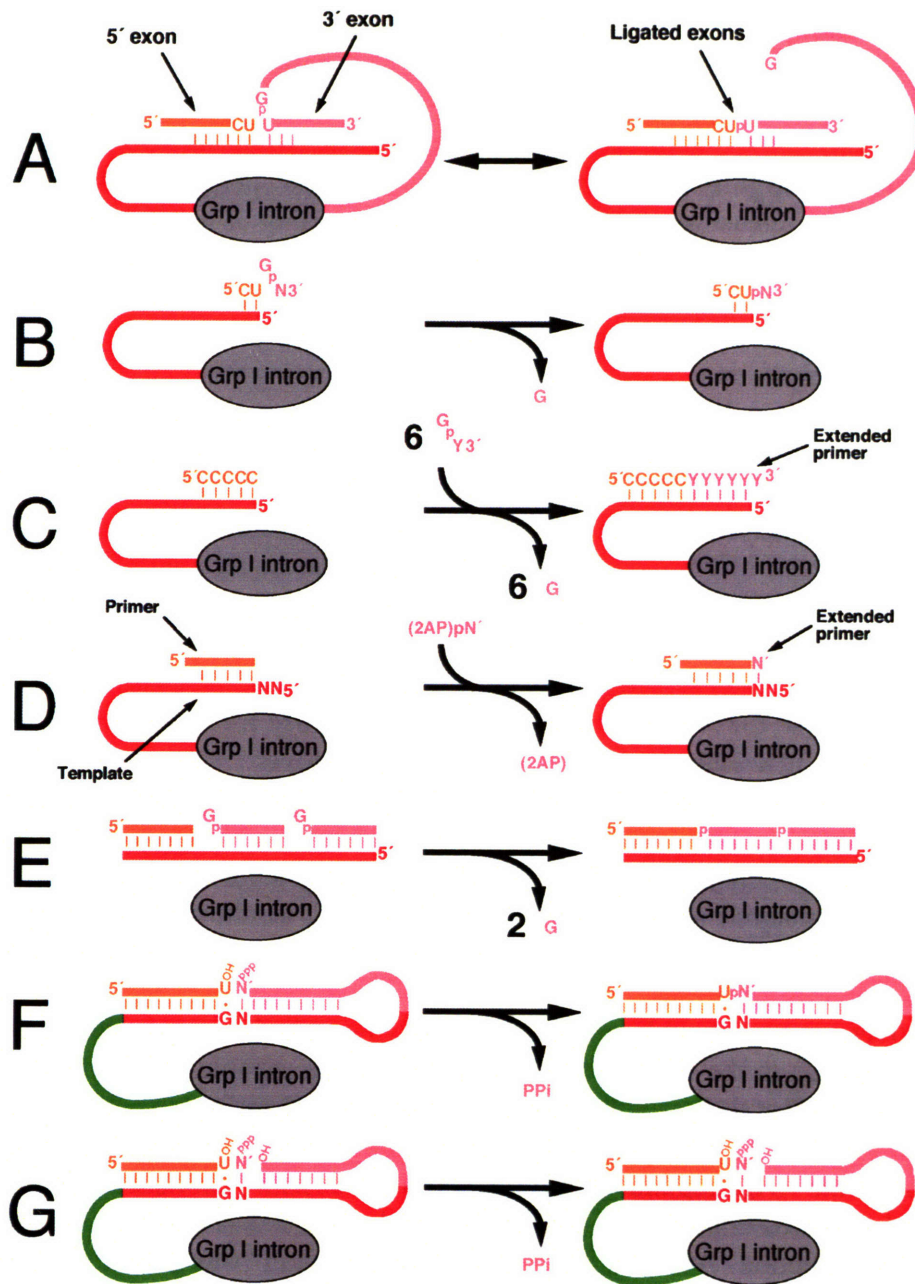
Fig. 1. Reactions catalyzed by the *Tetrahymena* Group I intron (adapted from refs. 56 and 49). **A.** Reaction catalyzed by the intron in its natural context. **B–G.** Reactions catalyzed by intron variants in engineered settings.

oligos aligned on a complementary template (Fig. 1E).[47] More recently, the Group I intron has been converted to a general ligase ribozyme (Fig. 1F),[48] and thence to a polymerase (Fig. 1G)[49] capable of adding G or C (but not A or U) and producing a pyrophosphate leaving group.

Thus, the first starting point for polymerase ribozyme evolution was the first-discovered natural ribozyme. A second new path, also blazed in the prodigious Szostak laboratory, began with something quite different: the first completely artificial ribozyme, isolated from a pool of completely random RNA sequences.

Polymerization using the Class I RNA ligase ribozyme

Consider the ligation reaction shown in Fig. 2A, in which a short oligonucleotide base-pairs to a partial stem-loop, and the resultant nicked hairpin is sealed by formation of a phosphodiester bond. The reaction is slightly exothermic, due to the chemical energy available in the triphosphate at the 5' ligation site, releasing pyrophosphate, and the binding of the two reactants, mediated by base-pairing, is very efficient. However, it may come as a surprise that under the seemingly very permissive reaction conditions of 100 mM $MgCl_2$ and pH 9, the uncatalyzed reaction is so slow that it would take 25 years for 50% of the complex to be ligated.[50,51] Szostak and Bartel decided to look for RNA catalysts to accelerate the reaction, constructing the pool shown in Fig. 2B. Using *in vitro* evolution, they isolated from this pool 65 such catalysts, including the one shown in Fig. 2C, which they named the Class I RNA ligase. Besides being the first ribozyme ever selected from a pool of random sequences, it remains one of the fastest ribozymes known.[50,52–54]

The reaction catalyzed by the Class I ligase also bears several crucial resemblances to primer extension, as illustrated in Fig. 2, where the red strand can be thought of as the polymerization template, providing a place for the primer (orange) and incoming nucleotide (pink) to anneal by Watson-Crick pairing. Also, the leaving group in the reaction is pyrophosphate, directly analogous to polymerization using NTPs. The polymerase-like character of the Class I ligase was developed further as follows: first, the 3' end of the template was detached from the catalytic core, converting the ribozyme into a true enzyme capable of multiple turnover catalysis (Fig. 2D). Second, the ribozyme was shown to be capable of using an NTP as a substrate, using the unpaired nucleotide in the template strand (red) to specify which nucleotide gets added to the primer (Fig. 2E), and achieving an overall fidelity of 92%.[55] Addition of a few more nucleotides could be achieved by expanding the unpaired section of the template strand, as shown in Fig. 2F; but this began to challenge the limits of the ribozyme, as the interposition of more and more template residues inevitably distorted its active site.

Moreover, it was troubling that the ribozyme relied on specific base-pairing to the template. A replicase in the RNA World would need to be capable of copying a wide variety of RNA sequences, not just the particular one it specifically base-paired to. When the ribozyme was modified as in Fig. 2G, by detaching the template and replacing the other half of the essential P2 stem with a separate RNA heptamer (purple), activity was sadly abolished. The ligase had come as far as simple-minded engineering could take it; what it needed next was the awesome healing power of *in vitro* selection. Accordingly, a new "helper domain" was grafted onto the ailing ligase, as shown in Fig. 2H, and from there our story unfolds in earnest.

Summary of thesis research

The pool shown in Fig. 2H was known as the Ligase+N_{76} pool, and it proved to be a fruitful source of polymerase ribozymes. The first of these to be isolated, known as Pol 1, was optimized to the point where it could catalyze extension of an RNA primer-template (PT) by as many as 14 nucleotides, with an average fidelity of nearly 97%. It was able to extend PTs of very different sequences and lengths, showing that it was indeed a *general* RNA polymerase ribozyme, unlike its earlier incarnations. Included here as an Appendix is a report detailing the procedures that led to the isolation of Pol 1 and its improved variant Evolved Pol 1. My contribution to this work was primarily in making detailed measurements of its fidelity; I also worked out a few refinements
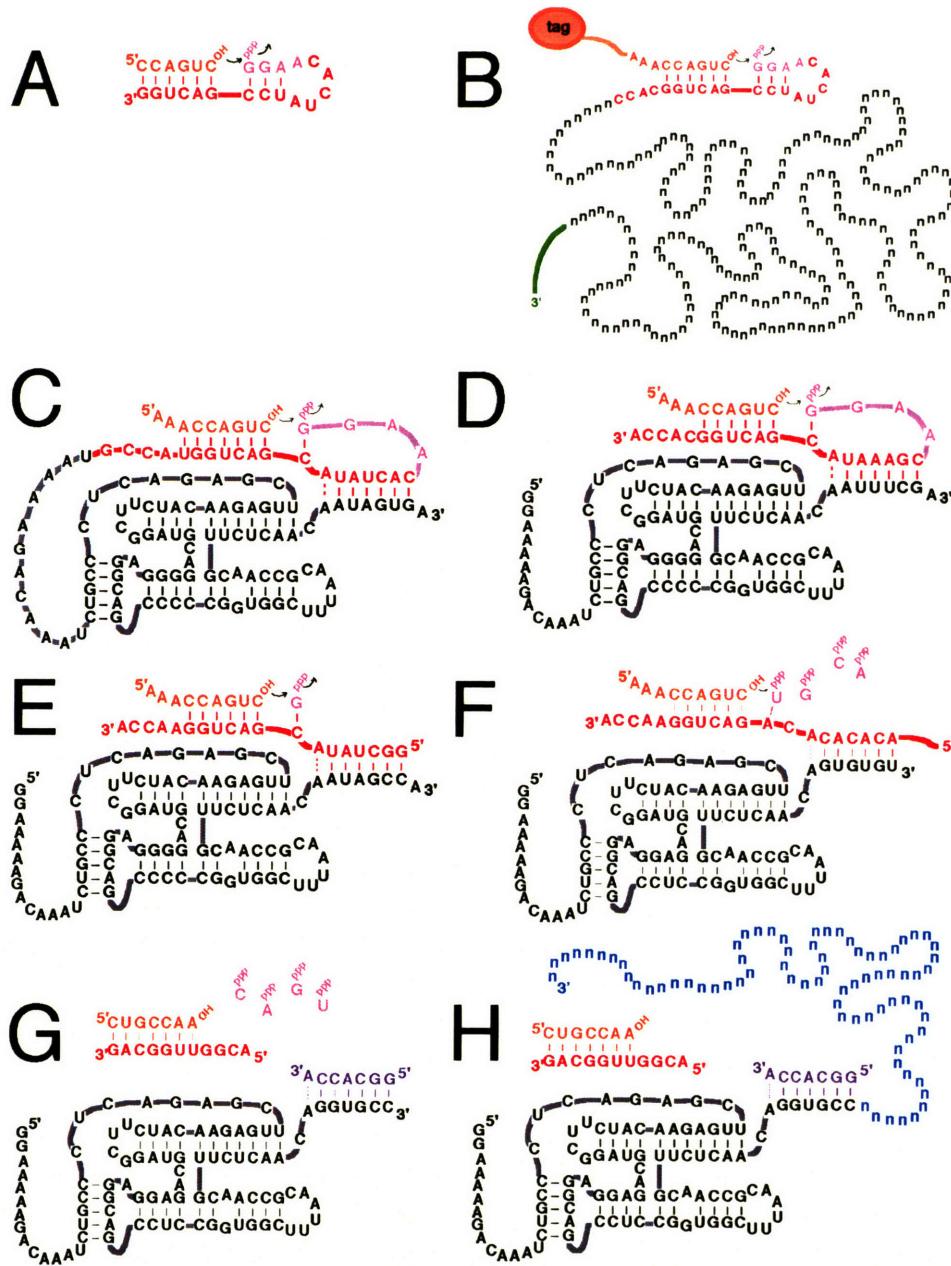
12

Fig. 2. **A.** Uncatalyzed RNA ligation, which is very slow. **B.** Pool of random sequences for ligase selection. **C.** Class I RNA ligase and the reaction it was selected to catalyze. **D–F.** Other ligation and polymerization reactions catalyzed by the Class I ligase. **G.** Polymerization reaction it cannot catalyze, because the base-pairing between ribozyme and substrate has been removed. **H.** Ligase+N_{76} pool for polymerase selection.

of the selection technique, by determining that the biotin selection step had not been working as planned, and by extending the mercury-gel selection technique to allow for selection of two nucleotide additions instead of just one (as discussed in Chapter 1).

The other polymerases isolated from the Ligase+N_{76} pool were named Pols 2–9, and they are the subject of Chapter 1. Each of them is a potential starting point for further polymerase evolution.

In the meantime, while polymerase selections were going on, I undertook a study of the detailed kinetic properties of Evolved Pol 1, including its processivity. Because of the very weak affinity of this polymerase for its PT substrate, a novel method of measuring processivity had to be developed that did not rely on the ability to saturate the PT binding site. This new method, and the results showing that Evolved Pol 1 is, perhaps surprisingly (given such weak PT binding), a partially processive polymerase.

Chapter 3 addresses the doped selection that I performed on Pol 2. This allowed determination of its secondary structure, by comparative sequence analysis. It also uncovered a variant with improved polymerization activity relative to the Pol 2 parent sequence. This improved variant was named Pol 2+. Through a series of site-directed mutagenesis experiments, I was able to identify a few critical areas of the Pol 2 structure, allowing formulation of an intriguing hypothesis about the evolutionary strategy it had exploited during the doped selection, namely an accumulation of preadaptive diversity in its tail region.

Chapter 4 reports the development of a novel capture-oligonucleotide technique for polymerase selection which allows the experimenter to select directly for addition of 10–12 nt to a primer-template. This is a significant improvement over the previous limit of 2 nt which, though seemingly meager, produced Evolved Pol 1, which is capable of adding up to 14 nt. If we could select for a dozen nucleotides, might we uncover a polymerase that could add hundreds? Although the possibility remains, its demonstration will not be reported here. Instead, the capture-oligo technique produced a different success: it improved the activity of Pol 2+ to such a level that it can fairly be called an equal of Evolved Pol 1; accordingly, it was renamed Evolved Pol 2.

In summary, the focus of the present report is a collection of RNA polymerase ribozymes which have been isolated through *in vitro* selection. Together, they constitute the beginnings of a proof for that fundamental conjecture underlying the RNA World hypothesis, namely that RNA is an inherently good enough catalyst to support its own replication.

References

1.   Strohman, R. (2000) We need a metaphor to explain life's mystery. *Nature* **408**:767–8.
2.   Deamer, D. W. (1997) The first living systems: a bioenergetic perspective. *Microbiol. Mol. Biol. Rev.* **61**:239–61.
3.   Stevenson, D. J. (1983) "The nature of the Earth prior to the oldest known rock record: the Hadean Earth," in *Earth's Earliest Biosphere, its Origin and Evolution*, (J. W. Schopf, Ed.), pp. 14–29, Princeton Univ. Press, Princeton, NJ.
4.   Lazcano, A. and Miller, S. L. (1994) How long did it take for life to begin and evolve to cyanobacteria? *J. Mol. Evol.* **39**:546–54.
5.   Wachtershauser, G. (1988) Before enzymes and templates: Theory of surface metabolism. *Microb. Rev.* **52**:452–84.
6.   Watson, J. D. and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**:964–7.
7.   Cziko, G. (1995) "Brain evolution and development," in *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*, MIT Press, Cambridge, MA.
8.   Shapiro, R. (2000) A replicator was not involved in the origin of life. *IUBMB Life* **49**: 173–6.

9.	Rich, A. (1962) "On the problems of evolution and biochemical information transfer," in *Horizons in Biochemistry*, (M. Kasha and B. Pullman, Eds.), pp. 103–26, Academic Press, New York.

10.	Woese, C. R. (1967) *The Origins of the Genetic Code*, Harper & Row, New York.

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

12.	Crick, F. H. C. (1968) The origin of the genetic code. *J. Mol . Biol.* **38:**367–79.

13.	Gierer, A. and Schramm, G. (1956) Infectivity of ribonucleic acid from tobacco mosaic virus. *Nature* **177:**702–3.

14.	Crick, F. H. C. (1966) The genetic code – yesterday, today, and tomorrow. *Cold Spring Harb. Symp. Quant. Biol.* **31:**1–9.

15.	Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschiling, D. E. and Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31:**147–57.

16.	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. *Cell* **35:**849–57.

17.	Pace, N. R. and Marsh, T. L. (1985) RNA catalysis and the origin of life. *Origins Life* **16:** 97–116.

18.	Sharp, P. A. (1985) On the origin of RNA splicing and introns. *Cell* **42:**397–400.

19.	Gilbert, W. (1986) The RNA world. *Nature* **319:**618.

20.	Cech, T. R. (1986) A model for the RNA-catalyzed replication of RNA. *Proc. Natl. Acad. Sci.* **83:**4360–3.

21.	Orgel, L. E. (1986) RNA catalysis and the origins of life. *J. Theor. Biol.* **123:**127–49.

22.	Westheimer, F. H. (1986) Polyribonucleic acids as enzymes. *Nature* **319:**534.

23.	Joyce, G. F., Schwartz, A. W., Miller, S. L. and Orgel, L. E. (1987) The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl. Acad. Sci.* **84:**4398–402.

24.	Westheimer, F. H. (1987) Why nature chose phosphates. *Science* **235:**1173–8.

25.	Orgel, L. E. (1987) Evolution of the genetic apparatus: a review. *Cold Spr. Harb. Symp. Quant. Biol.* **52:**9–16.

26.	Joyce, G. F. (1989) Amplification, mutation and selection of catalytic RNA. *Gene* **82:** 83–7.

27.	Horwitz, M. S., Dube, D. K. and Loeb, L. A. (1989) Selection of new biological activities from random nucleotide sequences: evolutionary and pratical considerations. *Genome* **1:**112–7.

28.	Ellington, A. D. and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346:**818–22.

29.	Robertson, D. L. and Joyce, G. F. (1990) Selection in vitro of an RNA enyme that specifically cleaves single–stranded DNA. *Nature* **344:**467–8.

30.	Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA Polymerase. *Science* **249:**505–10.

31.	Szostak, J. W. (1992) In vitro genetics. *Trends Biol. Sci.* **17:**89–93.

32.	Joyce, G. F. (1992) Directed molecular evolution. *Scientific American* **267:**90–7.

33.	Ellington, A. D. (1993) Experimental testing of theories of an early RNA World. *Methods. Enzymol.* **224:**646–64.

34.	Joyce, G. F. (1994) In vitro evolution of nucleic acids. *Curr. Opin. Struct. Biol.* **4:**331–6.

35.	Wilson, D. S. and Szostak, J. W. (1999) In vitro selection of functional nucleic acids. *Annu.*

Rev. Biochem. **68**:611–47.

36. Breaker, R. R. (2004) Natural and engineered nucleic acids as tools to explore biology. *Nature* **432**:838–45.

37. Illangasekare, M. and Yarus, M. (1999) Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci.* **96**:5470–5.

38. Illangasekare, M. and Yarus, M. (1999) A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* **5**:1482–9.

39. White III, H. B. (1976) Cofactors as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**: 101.

40. Benner, S. A., Ellington, A. D. and Tauer, A. (1989) Modern metabolism as a palimpset of the RNA World. *Proc. Natl. Acad. Sci.* **86**:7054–8.

41. Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**:905–20.

42. Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**:920–30.

43. Dworkin, J. P., Lazcano, A. and Miller, S. L. (2003) The roads to and from the RNA world. *J. Theor. Biol* **222**:127–34.

44. Zaug, A. J. and Cech, T. R. (1986) The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**:470–5.

45. Kay, P. S. and Inoue, T. (1987) Catalysis of splicing-related reactions between dinucleotides by a ribozyme. *Nature* **326**:343–6.

46. Been, M. D. and R., C. T. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* **239**:1412–6.

47. Doudna, J. A. and Szostak, J. W. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**:519–22.

48. Jaeger, L., Wright, M. C. and Joyce, G. F. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci.* **96**:14712–7.

49. McGinness, K. E. and Joyce, G. F. (2002) RNA-catalyzed RNA ligation on an external RNA template. *Chem & Biol* **9**:297–307.

50. Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261**:1411–8.

51. Rohatgi, R., Bartel, D. P. and Szostak, J. W. (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.* **118**:3332–9.

52. Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**:364–70.

53. Ekland, E. H. and Bartel, D. P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**:3231–8.

54. Bergman, N., Johnston, W. and Bartel, D. P. (2000) Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**:3115–23.

55. Ekland, E. H. and Bartel, D. P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**:373–6.

56. Bartel, D. P. (1999) "Re-creating an RNA Replicase," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., Cold Spring Harbor Laboratory Press, New York.

CHAPTER ONE

New RNA Polymerase Ribozymes

Abstract

The search is underway for a catalytic RNA molecule capable of self-replication. Finding such a ribozyme would lend crucial support to the RNA World hypothesis, which holds that very early lifeforms relied on RNA for both replicating and storing genetic information. We previously reported an RNA polymerase isolated from a pool of variants of an existing RNA ligase ribozyme.[1] Here we report eight additional ligase-derived polymerase ribozymes isolated from this pool. Because each of them is a new potential starting point for further *in vitro* evolution and engineering, together they substantially enrich the pool of candidates from which an RNA replicase ribozyme might eventually emerge.

Introduction

In discussing the origins of life, current editions of many biology textbooks now sketch a description of the RNA World, a very early and hypothetical period of evolution during which it is imagined that ribo-organisms ruled the earth.[2–5] These ancestral creatures had only one kind of encoded polymer: both their enzymes and their genome were made of RNA.[6–8] The appeal of the RNA World hypothesis is rooted in its simplicity: although the evolutionary innovation of such a sophisticated molecule as RNA would have been no small feat,[9,10] (but see also refs. 11, 12), and it might have been preceded by a simpler genetic polymer during the "pre-RNA World",[13,14] and it has even been suggested that the earliest lifeforms lacked genes of any kind,[15] the idea of a transitional RNA-only phase at some period during early evolution seems comfortably plausible in comparison to the alternative scenario in which oligonucleotides and coded protein synthesis emerged together in miraculous concert. Since the discovery that RNA, outside its well established role as an information carrier, can also act as an enzyme to catalyze chemical reactions,[16,17] and in fact catalyzes protein synthesis in all modern organisms,[18] RNA has been the leading candidate for a polymer that "did it all" in our very early ancestors.[19–22]

Despite the popularity of the RNA World hypothesis, some of its crucial tenets remain assumptions. Chief among them is the notion that RNA can be replicated without proteins. In the early stages of the RNA World, nonenzymatic polymerization must have played a major role,[23-26] but in most versions of the RNA World hypothesis, RNA at some point became genetically self-sufficient, taking over responsibility for its own synthesis. In order to substantiate the notion that RNA can be a good enough enzyme to catalyze RNA replication, considerable effort has been devoted to the search for an RNA polymerase ribozyme.[27,28]

The search has so far been dominated by efforts to extend the functionality of existing ribozymes. Some progress has been made using natural self-splicing introns as a starting point,[29–34] while other approaches have focused on an artificial ribozyme called the Class I RNA ligase.[35,36] Derivatives of this ligase are able to extend an RNA primer by several nucleotides, but they require specific base-pairing to the RNA template strand.[37,38] This dependence on specific sequence elements prevents them from being general polymerases, capable of copying any template. In an attempt to convert the ligase into a general polymerase, we previously constructed a pool of more than 10^{15} ligase variants, to each of which was appended a 76-nt segment of random sequence.[1] This was called the Ligase+N_{76} pool (Fig. 1). It was hoped that some of these chimeric RNA molecules would be able to bind and extend a primer-template RNA duplex without base-pairing to it. The pool was enriched for molecules with the desired activity by repeated selection

18

and amplification, and after ten rounds of this *in vitro* selection, a single ribozyme was found showing robust template-dependent polymerization activity. This ribozyme, referred to here as Pol 1, was then subjected to eight further rounds of optimizing selection, improving its activity to the point where it could polymerize a whole turn of an RNA helix. The improved polymerase, referred to here as Evolved Pol 1, has been characterized with respect to its secondary structure and fidelity,[1] kinetics and processivity,[39] and substrate recognition.[40]

Here we report a new Ligase+N_{76} polymerase selection (called branch B), in which the selection protocol was modified

Fig. 1. Ligase+N_{76} starting pool, consisting of an RNA ligase ribozyme (black), plus P2-completing heptamer (purple) and two randomized loops (blue), concatenated to a 76-nt random region (blue), and flanked by primer-binding sites (green).

using lessons learned from the original experiment (called branch A). The new selection led to the isolation of seven additional ligase-derived polymerase ribozymes, designated Pol 2 through Pol 8, each with another unique "auxiliary domain" derived from the randomized N_{76} segment. Additionally, a weak polymerase from the original selection was re-examined and named Pol 9. The whole collection of polymerase ribozymes was compared head-to-head in polymerization assays using a variety of primer-templates (PTs). Pols 1–4 had approximately equivalent levels of activity, while the activity of Pols 5–9 was less robust. In general, Evolved Pol 1 showed the strongest polymerization activity, although with some PTs it was nearly matched by Pols 1–4.

Results

New selection. Having found Pol 1 as the only robust polymerase from the branch A selection, yet suspecting that the Ligase+N_{76} starting pool might still hold undiscovered polymerases, we designed a new selection strategy to look for ribozymes that the first experiment might have missed. We returned to an early point in the original selection and branched off along a new path to look for more polymerases. This new selection path, called branch B (Fig. 2), incorporated several modifications that were expected to alter the course of the *in vitro* evolution.

In branch A, ribozymes had been incubated with 4-thioUTP, and ribozymes that tagged themselves with 4-thioU were isolated on the basis of their decreased mobility in a mercury gel. Pol 1 was later confirmed to utilize 4-thioUTP and unmodified UTP with comparable efficiency in polymerization assays (data not shown). However, in later rounds of branch A, an additional constraint had been imposed: ribozymes were incubated with biotin-ATP, with the idea that ribozymes that tagged themselves with both 4-thioU and biotin-A would be isolated by successive purification using mercury gels then streptavidin-coated magnetic beads (Table 1). Although branch A was eventually successful, Pol 1 was later found to reject biotin-ATP as a substrate, despite its efficient use of unmodified ATP. Moreover, the parental RNA ligase ribozyme was also shown to use ATP, but not biotin-ATP (data not shown), as a substrate in the single-nucleotide addition reaction it catalyzes using an internal template.[37] These findings suggested in hindsight that the use of biotin-ATP in some rounds of branch A had been ineffective in imposing additional
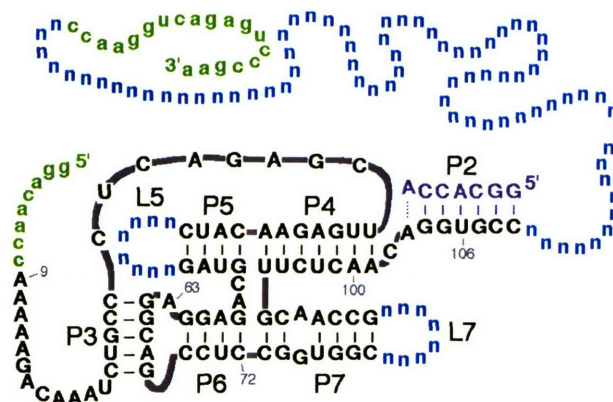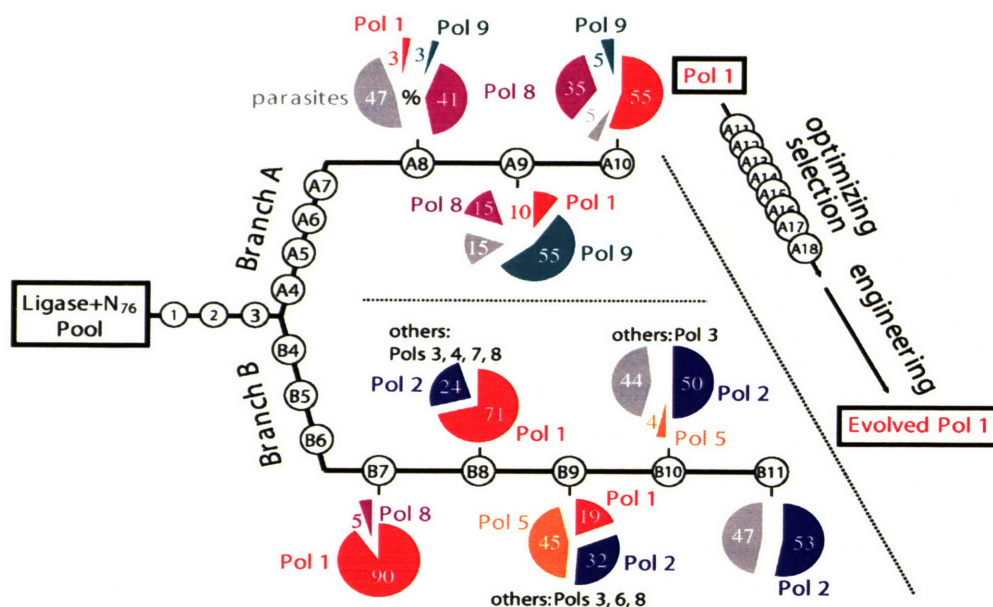
Pol 1 Pol 9   Pol 9   Pol 1

parasites 47 % 41 Pol 8   35 5 55

Branch A
A7
A6
A5
A4

Pol 8 15 10 Pol 1
15 55 Pol 9

Ligase+N$_{76}$ Pool — 1 — 2 — 3

A8 — A9 — A10

A11
A12
A13
A14
A15
A16
A17
A18

optimizing selection engineering

Branch B
B4
B5
B6

others: Pols 3, 4, 7, 8   others: Pol 3
Pol 2 24 71 Pol 1   44 50 Pol 2
                          4 Pol 5

Evolved Pol 1

B7 — B8 — B9 — B10 — B11

5 Pol 8
Pol 1 90

Pol 5 19 Pol 1
45 32 Pol 2

47 53
Pol 2

others: Pols 3, 6, 8

Fig. 2. Two branches of the polymerase selection, and breakdowns of pool populations. Pool evolution is represented as a line extending rightward, with each selection round indicated as a circled number. Pie graphs report the predominant families at certain rounds (as percent of total population). Families detected at low levels are reported as "others".

Round	Template	NTPs (mM)			Time (hr)	Selectiv criteria
		^{4S}U	^{B}A	A,C,G		
Shared initial rounds						
1	GGUCAGAUU	2	0	0	36	^{4S}U
2	GGUCAGAACC	2	0	0	20	^{4S}U
3	GGUCAGAA	2	0	0	20	^{4S}U
Branch A						
A4	CUUAGUUCAUU	2	0	0	19	^{4S}U
A5	CUUAGUUCAUU	2	0	0	1	^{4S}U
A6	GGUCAGAUU	1	1	0	14	^{B}A,^{4S}U
A7	CUUAGUUCAUU	1	1	0	17	^{B}A,^{4S}U
A8	GGUCAGAUU	1	1	0	17	^{B}A,^{4S}U
A9	GGUCAGAUU	1	1	0	4	^{B}A,^{4S}U
A10	CUUAGUUCAUU	1	0	0	20	^{4S}U
Branch B						
B4	ACAUACGGAUAUU	2	0	0	24	^{4S}U
B5	UCGACGGAACC	2	0	0	18	2 x ^{4S}U
B6	ACCUGAGaaCC	0.5	0	0	18	2 x ^{4S}U
B7	CAAGUCCaaGG	0.1	0	2 ea.	14	2 x ^{4S}U
B8	ACCUGAGaaCC	0.1	0	2 ea.	15	2 x ^{4S}U
B9	CAAGUCCaaGG	0.1	0	2 ea.	1	2 x ^{4S}U
B10	CAAGUCCaaUGAUCGUA	0.1	0	2 ea.	5	2 x ^{4S}U
B11	UCGACGGaaCCUGCGUC	0.1	0	2 ea.	0.4	2 x ^{4S}U

Table 1. Polymerization templates and selection parameters used in polymerase evolution. Polymerization primer was complementary to the underlined portion of each template (lowercase "a" indicates the adenine isomer 2-aminopurine.) ^{B}A indicates N6-biotin-A, and ^{4S}U indicates 4-thioU.

selective pressure, and that the survival of ribozymes at this step in those rounds was due merely to background binding to the streptavidin-coated beads. Moreover, some valuable polymerase ribozymes could have been lost or disfavored when the population was forced through this potential bottleneck. Therefore, in designing branch B to look for additional polymerases in the ligase-based pool, we eliminated the use of biotin-ATP (Table 1).

In branch A, ribozymes that added one nucleotide had been lumped together with ribozymes that added two or more. At each step of the selection, all ribozymes that managed to add the first 4-thioU were recovered and amplified, regardless of how much additional polymerization they had succeeded in catalyzing. Thus, Pol 1 was isolated on the basis of its ability to add merely a single nucleotide. During the Pol 1

optimization selection (rounds A11–18, Fig. 2), we learned that RNA molecules containing two 4-thioUs halted in a mercury gel, rather than merely slowing down, and that we could therefore increase the stringency of the mercury-gel technique to select for addition of two 4-thioUs by collecting only those molecules caught at the mercury interface. In branch A, this two–4-thioU technique had been used in the Pol 1 optimization phase (rounds A11–18), but not in the original polymerase discovery phase (rounds 1–A10). In branch B, the two–4-thioU technique was implemented much earlier (Table 1), in hopes that the more stringent selection criterion (demanding addition of two 4-thioUs instead of just one) would shift the course of the *in vitro* evolution and reveal new polymerases.

Branch B was further enhanced by earlier application of several other selection techniques that had been employed only in the Pol 1 optimization phase (rounds 11–18) of branch A. These modifications included the use of competitor NTPs (unlabeled ATP, CTP, and GTP), to select for high-fidelity polymerization; the use of 2-aminopurine in the polymerization template, to improve Watson-Crick pairing geometry with 4-thioU (by avoiding steric clash of the 4-thioU sulfur atom and the adenine 6-amino group); the use of longer templates, to favor ribozymes that could accommodate them; the reduction of 4-thioUTP concentration, to favor ribozymes with stronger NTP binding; and the use of mutagenic PCR, to increase pool diversity and optimize active ribozymes. Finally, whereas branch A had essentially alternated one pair of primer-templates, sometimes repeating a PT in subsequent rounds (Table 1), branch B employed a greater diversity of PT sequences in order to enhance its selective power for polymerases that catalyze general (as opposed to sequence-specific) template-directed RNA polymerization.

Branch B began with pool 3 of the original selection (Fig. 2), a pool that had already undergone three rounds of selection for the ability to add a single 4-thioU but did not yet have detectable polymerization activity. This pool was subjected to eight new rounds of selection, incorporating the changes described above (Table 1). Pool activity was detected after the first new round of selection (round B4), with about 2% of the pool adding a single 4-thioU in 24 hours. In every subsequent round, the selection criterion was addition of two 4-thioUs, and this activity was first detected after round B6. Following round B8, pool activity was robust, and was shown to require a correctly paired primer and template. After round B10, the pool was shown to catalyze polymerization in the untethered format (with no covalent linkage between pool and primer) using all three different PTs tested. After one more round, selection was stopped and pool populations were analyzed.

Dramatic population shifts. 139 clones from pools B7–11 were isolated and sequenced, then grouped into 33 families, with the members of each family having nearly identical sequences, by virtue of descent from a single ancestral sequence in the starting pool. Immediately obvious was the early dominance of the Pol 1 family in the branch B selection: 90% of the isolates from pool B7 were identical or nearly identical to Pol 1 (Fig. 2). Had this result been apparent immediately after round B7, it might have led to the early abandonment of the branch B selection; despite our efforts to improve the selection protocol and uncover new polymerases, evolution seemed stuck in a rut. Fortunately, however, branch B had already been carried forward several more rounds, thereby revealing the first new polymerase family, represented by Pol 2. This new family gradually but completely displaced the Pol 1 family during rounds B8–10, apparently indicating its superior fitness under branch B selection conditions. Several additional families were also detected during the population shift: the Pol 5 family dominated pool B9 but disappeared in the

21

next round, while Pols 3, 4, 6, 7, and 8 were detected at small numbers in pools B7–9. At round B10, about half of the clones had multiple lesions in their ligase domains. Clones such as these always failed in polymerization assays and were provisionally classified as "parasites," because their garbled catalytic domain and consequent inactivity meant that they had been able to prosper during the selection by alternative, uncharacterized mechanisms. Most of the parasites in branch B belonged to a few large families.

New polymerases. A consensus clone was chosen to represent each new polymerase family (Table 2 and Fig. 3). The new clones were evaluated alongside Pol 1 in polymerization assays using a variety of PTs (Fig. 4). On the basis of these comparisons, the new clones were named Pols 2–8 in order of their approximate overall activity. All eight polymerases extended PT A, which codes for addition of a single C. Pols 1–5 went further by adding an additional, untemplated nucleotide, as do most proteinaceous polymerases.[41,42] With a longer template (PT D), extension by at least 3 nt was detected using all eight polymerases. With other PTs, the performance of Pols 1–4 was consistently strong, and their activity levels were generally comparable. Pols 5 and 6 were usually much weaker, and Pols 7 and 8 extended only a few of the PTs tested.

Polymerization requires both domains. Without an auxiliary domain, the ligase core was completely inactive in all polymerization assays (lanes "L" in Fig. 4). Pol 2 activity was dependent on the P2-completing heptamer (data not shown), as observed previously with Pol 1,[1]
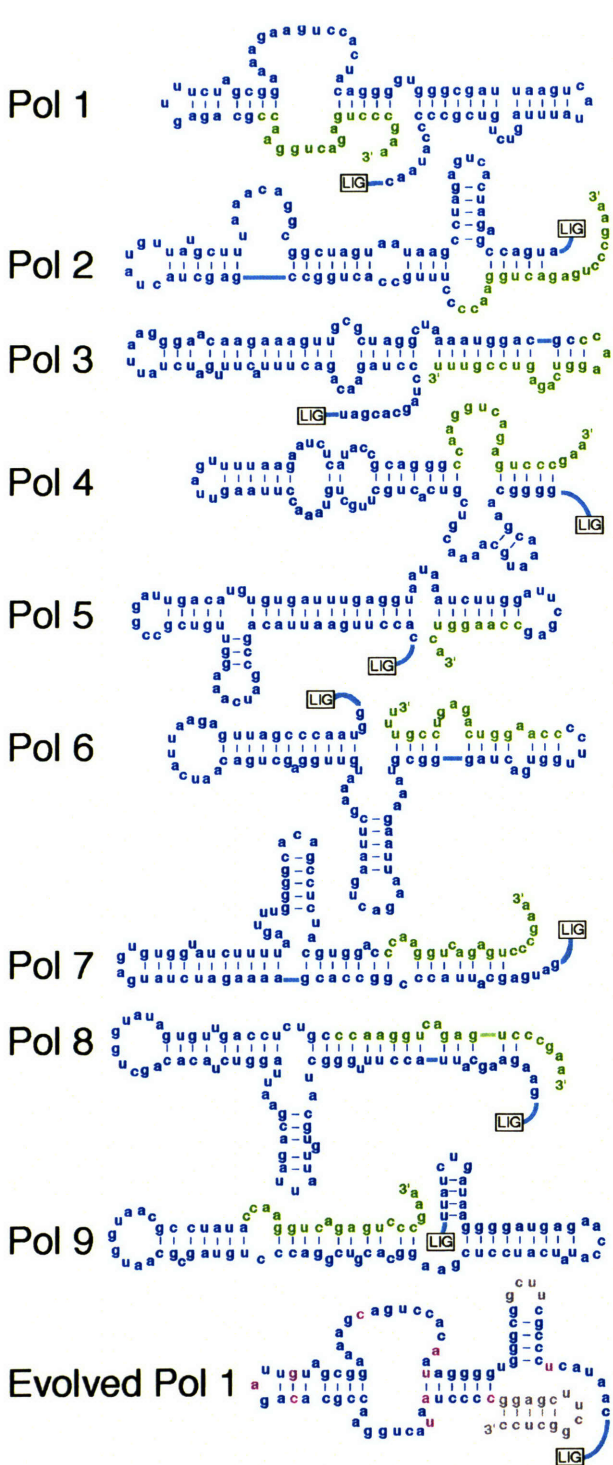
Fig. 3. Polymerase ribozymes. Each polymerase contains a ligase domain (shown as "LIG"–see Table 2 for details) concatenated to an auxiliary domain derived from the pool N_{76} region (blue) and 3' primer-binding site (green). Secondary structures of Pols 2–9 are speculative. Evolved Pol 1 is shown with changes from Pol 1 indicated in pink and grey.

Polymerase		Loop 5	Loop 7	Ligase changes
Pol 1	A10.2	UGUGAAUU	GCGAUUGC	
Pol 2	B11.78	ACUCAUAA	CAUCAUAA	A9C,A63C
Pol 3	B9.41	UAGUAUCG	AAUCUCUC	A63U
Pol 4	B8.36	ACAUUGGU	CUAAGUUG	
Pol 5	B9.02	AGUCCCAA	UCCGCUAA	
Pol 6	B9.50	AGUCCCAA	UCCGCUAA	A100G
Pol 7	B8.64	GCGUAUGU	ACGUGCCU	C72U
Pol 8	B8.38	CAUAUUCGG	GGGGUGCC	
Pol 9	A9.1	GCAGUAGC	UGAUACUA	
Ev Pol 1	A18.12.23	UUCG	GCGAUAGC	Δ3–8,U106C

Table 2. Ligase loop sequences and other changes in each polymerase.
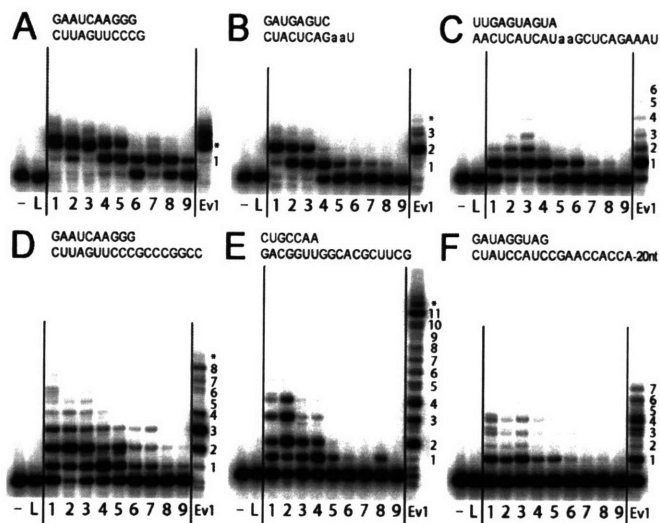
Fig. 4. Ribozyme-catalyzed primer extension. Each panel shows polymerization using a different primer-template (PT), shown at top (lowercase "a" indicates 2-aminopurine.) In each panel, the first two lanes are negative controls, showing primer incubated with no ribozyme (–) or with ligase core only (L). Lanes 1–9 and Ev1 show the activity of Pols 1–9 and Evolved Pol 1. Extension products are numbered at right, with stars indicating full-length molecules extended by an additional, untemplated nucleotide.

indicating that the ligase P2 stem remained essential for polymerization. Clones with disrupted ligase domains (parasites) were always inactive in polymerization assays (not shown). However, many families with intact ligase domains were also found to be inactive, indicating that not just any auxiliary domain will do.

Attempts were made to trim some of the new polymerases. Earlier work that had produced the ligase ribozyme itself[36] showed that ribozymes isolated from *in vitro* selection experiments could often tolerate terminal truncations. We suspected that the ribozymes emerging from the branch B selection would not require their very 3′ ends, because the sequence of the pool's 3′ terminus had been systematically varied at each round of the selection in a deliberate effort to discourage the pool from relying on its 3′-terminus (for instance by subversively using it as the polymerization primer in place of the primer ligated to the 5′ end of the pool at each round). However, in the case of Pol 2, deleting just 8 nt from its 3′ end was enough to nearly abolish its activity. Attempts to trim Pol 5 were slightly more successful: this polymerase tolerated the deletion of 10 nt from its 3′ terminus. (The truncated version of Pol 5, depicted in Fig. 3, was used in the Fig. 4 assays.) Further truncation was not tolerated, however, with Pol 5 completely inactivated by deletion of 20 or more nt from its 3′ end.

A closer look at Pol 1. During the original branch A selection, 74 variants were cloned and sequenced from pools A8–10, then grouped into 23 families.[1] One family, represented by isolate 10.2, showed robust template-dependent polymerization and was renamed the round-10 ribozyme, referred to here as Pol 1. It was the starting point for further optimization (rounds A11–18) and site-directed sequence engineering, yielding eventually the round-18 ribozyme or Evolved Pol 1. For one of the primer-templates examined, this polymerase can accurately synthesize a full turn of an RNA helix, extending this primer by 14 nt.[1] It also showed strong activity with a 3-nt–shorter version of this template (PT E, Fig. 4). However, with other PTs, the activity of Evolved Pol 1 was

more modest. PTs B and C are analogous to the PTs used during the selection, in that they code first for the addition of two Us. With PT B, Evolved Pol 1 managed to add the first two Us to all of the PT molecules but added the final encoded A to only a tiny fraction of them (Fig. 4). With PT C, which codes for the addition of 12 nt, Evolved Pol 1 was able to add the first U to most of the PTs, but it usually stalled there; it added the second U to only a small fraction of the molecules; addition of nt 3–6 was barely detectable, and no polymerization was observed beyond 6 nt. With PT F, which codes for 30 nt, Evolved Pol 1 added at most 7 nt; similar results were seen with two other PTs of the same length (data not shown).

Race against hydrolysis. Evolved Pol 1 typically extends its PT by 4–8 nt, and never by more than 14 nt, even when the template codes for many more. Why doesn't it do better? The reason is its very weak primer-template binding. The polymerization reaction uses micromolar concentrations of RNA, but the polymerase binds its PT substrate with only millimolar affinity, and simply increasing the PT concentration fails to improve the situation, because higher RNA concentrations actually inhibit the ribozyme.[39] This extremely poor affinity makes PT binding the rate-limiting step. As the ribozyme slowly polymerizes NTPs, it suffers the constant hydrolytic onslaught of the high-magnesium, high-pH reaction buffer, and this causes the reaction to decelerate. We measured how long Evolved Pol 1 survives under polymerization assay conditions and found that its half-life was in close agreement with the halving time of reaction rate; after a day-long exposure to 200 mM magnesium and a pH of 8.5, only about half of the polymerase remained full-length.

Weak additional ribozymes from branch A. Pol 1 was the only robust polymerase discovered during branch A of the selection. However, a second polymerase family was also reported, represented by isolate 9.1 and comprising 55% of the pool after round A9 (Fig. 2). It had an unmutated ligase domain and was shown to catalyze weak but template-dependent extension of a tethered primer.[1] Here we renamed this ribozyme Pol 9 (Fig. 3) and tested it with several untethered PTs, confirming that its polymerization activity is very weak, in fact undetectable with most PTs (Fig. 4). A third family that rose to prominence during branch A comprised 41% of pool A8 and 35% of pool A10 (Fig. 2), but despite its evolutionary success and intact ligase domain, it showed no activity in the initial tethered PT assay, and so was not reported. After the branch B selection, this family was recognized as including Pol 8 (Fig. 3), a very weak polymerase present at low frequency in several branch B pools (Fig. 2).

Discussion

The Class I RNA ligase is one of the fastest ribozymes known,[43] and it catalyzes a reaction similar in many crucial respects to a single step of RNA polymerization: the reactants are aligned by Watson-Crick pairing to an RNA template, the 3′-hydroxyl of one reactant attacks the 5′-triphosphate of the other reactant, pyrophosphate is released, and a new internucleotide linkage is created. When assisted by base-pairing, the ligase can extend an RNA primer by as many as 6 nt.[37] However, with PTs that don't base-pair to it, the ligase can't add even a single nucleotide (Fig. 4, lanes "L"). The Ligase+N_{76} pool (Fig. 1) was constructed in hopes of finding an auxiliary domain that, when appended to the ligase, would somehow help it bind the PT sequence-nonspecifically, thus converting the ligase into a general polymerase.[1] Now nine such auxiliary domains have been identified.

A bouquet of long stems. The secondary structures of the auxiliary domains of Pols 2–9 are not yet known. Speculative structures based on the predictions of the m-fold algorithm[44] are shown in Fig. 3. In many cases, long paired regions (for example, a putative 13-bp stem in Pol 3) lend credibility to the structures, because these long stems would be extremely stable in the presence of 200 mM magnesium, and they significantly constrain the rest of the fold. Moreover, m-fold correctly predicted the structure of the Evolved Pol 1 auxiliary domain; the predicted structure includes all the base-pairs shown in the Fig. 3 structure, which was modeled independently by comparative analysis of 25 diverse variants.[1] It is tempting to interpret these highly base-paired, largely linear structures as a general feature of successful polymerase auxiliary domains. The auxiliary domains of the most active polymerases, Pols 1–5, were each compared to their permuted cohort, a set of sequences in which the N_{76}-derived sequence was randomly reordered, neutralizing the contributions of evolution but preserving base composition.[45] In each case, the evolved structure was more stable than at least 70% of its permuted cohort (95% in the cases of Pols 1 and 5). Thus, the predicted stability trend reflects conformational order accumulated during selection, and not merely the innate folding behavior of RNA. It would be interesting to explore whether this property directly aids polymerization, or whether it was favored during the selection for more general reasons, such as compactness or resistance to hydrolysis.

Polymerase ecology. The defining feature of *in vitro* evolution is the use of iterated selection and amplification to move from a starting pool in which sequences are distributed randomly, to an evolved pool in which they occur in proportion to their aptitude for the desired activity. In our ribozyme selections, however, we have often been frustrated by a perversely weak correlation between activity and multiplicity. For instance, the Pol 8 and Pol 9 families dominated the branch A pools numerically but showed only weak activity, whether assayed in the tethered format (analogous to the selection context) or untethered (requiring the activity ultimately desired). Conversely, Pols 3 and 4, two of the most active polymerases to emerge from the Ligase+N_{76} pool, never multiplied beyond a few percent of total population. This weak correlation exposes a fundamental inefficiency in the selection protocol. A related issue is the emergence of "parasites", clones with ruined catalytic domains and no polymerization activity whatsoever, which emerge consistently in later rounds of our polymerase selections. We speculate that these inactive clones survive by efficiently inserting their primers into the active sites of working polymerases, thus exploitatively ensuring their own propagation while contributing nothing to the overall level of pool activity. This hypothesis explains their absence during early selection rounds, when there are not yet enough working polymerases around to exploit. Similar phenomena have been observed in other *in vitro* selection experiments.[46]

On the path to RNA self-replication. Each of these nine rudimentary polymerases is a potentially promising evolutionary intermediate between ligase and replicase. In the case of Pol 1, the promise has already borne fruit: after eight rounds of optimizing selection and a little site-directed tinkering, it gave rise to Evolved Pol 1, the strongest polymerase ribozyme yet reported.[1] Evolved Pol 1 can add 14 nt to one particular PT, but more typically it adds 4–8 (Fig. 4). Previous work demonstrated its sensitivity to PT sequence: a change as slight as adding or subtracting a single nucleotide from the starting primer altered the observed extension rate by as much as an order of magnitude.[39] Such sequence-specific variation is not surprising, having been observed as

well with proteinaceous polymerases.[47,48] Nonetheless, without exception, Evolved Pol 1 has extended every PT tested: it is truly a general RNA polymerase.

Pols 2–9 have not yet had the benefit of optimization, and it is interesting to imagine what Evolved Pols 2–9 might be like. One could imagine constructing a pool of variants of each ribozyme, then mixing them all together into one super-pool and letting them all compete against each other during selection. By this approach, the problem of "declaring wild type"[49] would be softened slightly: instead of forcing the population through the bottleneck of a single sequence, an increased genetic diversity of forms would be retained, and with it the possibility of productive recombination between unrelated structures. In any case, it will be interesting to see if an RNA replicase ribozyme can eventually be isolated, whether it descends from one of these ligase-derived polymerases, or from any other branch of the ribozyme family tree.

Acknowledgments

Materials and methods

In vitro **selection.** The starting pool for the branch B selection (Fig. 2) was an aliquot of pool 3 DNA archived during the original selection experiment.[1] RNA was transcribed using T7 RNA polymerase, purified by PAGE (polyacrylamide gel electrophoresis), and ligated to an RNA primer using T4 DNA ligase and a DNA splint, then PAGE-purified again and annealed to an RNA template. Primer and template sequences were changed in each round (Table 1). The pool was also annealed to the RNA heptamer GGCACCA, which completed the P2 stem of the ligase domain. The polymerization reaction was initiated by addition of selection buffer (60 mM $MgCl_2$, 200 mM KCl, 50 mM EPPS, pH 8.0) plus 4-thioUTP and competitor NTPs, then allowed to proceed at 22°C for 0.4–24 hrs (Table 1). The reaction was stopped by addition of 80 mM EDTA, and excess 4-thioUTP was removed by Centricon filtration. Molecules containing 4-thioU were isolated by mercury PAGE. From round B4 onward, RNA was excised from the interface between the –Hg and +Hg regions, enforcing selection for addition of two 4-thioUs. RNA was eluted from the gel slice, then used to seed the next round of amplification–selection.

Cloning and sequencing. Molecules were cloned from pools B7–11 using TOPO-TA cloning (Invitrogen), then 72 were sequenced and grouped into 9 families of nearly identical sequences. The population statistics shown in Fig. 2 were computed from this first set of clones (For rounds 7–11, N = 18, 12, 15, 20, and 19 respectively) Additional minor families were then uncovered using a two-step colony hybridization screen. In the first step, thousands of cloned colonies from pools B8–10 were transferred to nitrocellulose membranes and screened with 70-nt DNA probes corresponding to the auxiliary domains of the known major families. Clones that hybridized to any of these probes were ruled out as novel candidates. In the second step, the clones were re-screened using a probe that corresponded to the ligase domain. Clones that hybridized only in this

second step were candidates of interest, because they had an intact ligase domain and an unknown auxiliary domain. Sixty-seven such clones were sequenced, bringing the total number of sequence families to 33.

Polymerization assays. Polymerase ribozymes (5 μM final concentration) were transcribed from cloned DNA templates, PAGE-purified, and annealed to a 1.25-fold molar excess of P2 oligo. Primer (0.1 μM, 5'-radiolabeled) and template (0.5 μM) were annealed separately, then mixed with assay buffer (200 mM $MgCl_2$, 50 mM Tris-HCl, pH 8.5) and NTPs (4 mM each). Polymerization reactions were initiated by mixing annealed ribozyme with annealed PT/buffer/ NTPs. After a 24-hour incubation at 22°C, reactions were mixed with 4 volumes of gel loading buffer (8M urea, 25 mM EDTA), template was separated from primer by addition of 5 molar equivalents of a competitor RNA identical to fully extended primer, and extension products were resolved by denaturing PAGE.

Measurement of polymerase survival time during polymerization. Three identical polymerization assays were carried out as above, varying only the position of the 5'-radiolabel: in one reaction the primer was end-labeled (as normally), whereas in the other two either the template or the ribozyme (Evolved Pol 1) was end-labeled. Aliquots were withdrawn and quenched at a series of timepoints, and RNAs of different sizes were separated by denaturing PAGE. At each timepoint, amounts of full-length ribozyme and degradation fragments were quantitated.

References

1.   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. *Science* **292**:19–25.
2.   Audesirk, T., Audesirk, G. and Byers, B. E. (2005) *Biology: Life on Earth*, 7th ed., Pearson Prentice Hall.
3.   Campbell and Reece (2002) *Biology*, 6th ed., Benjamin Cummings.
4.   Purves, W. K., Sadava, D., Orians, G. H. and Heller, H. C. (2001) *Life: The Science of Biology*, 6th ed., Sinnauer Assoc.
5.   Freeman (2002) *Biological Science*, Prentice Hall.
6.   Pace, N. R. and Marsh, T. L. (1985) RNA catalysis and the origin of life. *Origins Life* **16**: 97–116.
7.   Orgel, L. E. (1986) RNA catalysis and the origins of life. *J. Theor. Biol.* **123**:127–49.
8.   Gilbert, W. (1986) The RNA world. *Nature* **319**:618.
9.   Shapiro, R. (1988) Prebiotic ribose synthesis: a critical analysis. *Orig. Life Evol. Biosph.* **18**:71–85.
10.   Larrade, R., Robertson, M. P. and Miller, S. L. (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc. Natl. Acad. Sci.* **92**:8158.
11.   Ricardo, A., Carrigan, M. A., Olcott, A. N. and Benner, S. A. (2004) Borate minerals stabilize ribose. *Science* **303**:196.
12.   Springsteen, G. and Joyce, G. F. (2004) Selective derivitization and sequestration of ribose from a prebiotic mix. *J. Am. Chem. Soc.* **126**:9578–83.
13.   Nelson, K. E., Levy, M. and Miller, S. L. (2000) Peptide nucleic acids rather than RNA

may have been the first genetic molecule. *Proc. Natl. Acad. Sci* **97**:3868–71.

14. Ichida, J. K., Zou, K., Horhota, A., Yu, B., McLaughlin, L. W. and Szostak, J. W. (2005) An in vitro selection system for TNA. *J. Am. Chem. Soc.* **127**:2802–3.

15. Shapiro, R. (2000) A replicator was not involved in the origin of life. *IUBMB Life* **49**: 173–6.

16. Cech, T. R. (1990) Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci. Rep.* **10**:239–61.

17. Altman, S. (1990) Enzymatic cleavage of RNA by RNA. *Biosci. Rep.* **10**:317–37.

18. Steitz, T. A. and Moore, P. B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trens Biochem Sci* **28**:411–8.

19. Joyce, G. F. (2002) The antiquity of RNA-based evolution. *Nature* **418**:214–21.

20. Dworkin, J. P., Lazcano, A. and Miller, S. L. (2003) The roads to and from the RNA world. *J. Theor. Biol.* **222**:127–34.

21. Hughes, R. A., Robertson, M. P., Ellington, A. D. and Levy, M. (2004) The importance of prebiotic chemistry in the RNA World. *Curr. Opin. Chem. Biol.* **8**:629–33.

22. Orgel, L. E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**:99–123.

23. Luther, A., Brandsch, R. and von Kiedrowski, G. (1998) Surface-promoted replication and exponential amplification of DNA analogues. *Nature* **396**:245–8.

24. 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**:13734–40.

25. Huang, W. and Ferris, J. P. (2003) Synthesis of 34–50 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun. (Camb)* **12**:1458–9.

26. Franchi, M. and Gallori, E. (2005) A surface-mediated origin of the RNA world: biogenic activities of clay-adsorbed RNA molecules. *Gene* **346**:205–14.

27. Bartel, D. P. (1999) "Re-creating an RNA Replicase," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., , Cold Spring Harbor Laboratory Press, New York.

28. McGinness, K. E. and Joyce, G. F. (2003) In search of an RNA replicase ribozyme. *Chem & Biol* **10**:5–14.

29. Been, M. D. and R., C. T. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* **239**:1412–6.

30. Doudna, J. A. and Szostak, J. W. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**:519–22.

31. Bartel, D. P., Douda, J. A., Usman, N. and Szostak, J. W. (1991) Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Molec. Cell. Biol.* **11**:3390–4.

32. Chowrira, B. M., Berzal-Herranz, A. and Burke, J. M. (1993) Novel RNA polymerization reaction caalyzed by a group I ribozyme. *EMBO J.* **12**:3599–605.

33. Doudna, J. A., Usman, N. and Szostak, J. W. (1993) Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry* **32**:2111–5.

34. McGinness, K. E. and Joyce, G. F. (2002) RNA-catalyzed RNA ligation on an external RNA template. *Chem & Biol* **9**:297–307.

35. Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261**:1411–8.

36.   Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**:364–70.

37.   Ekland, E. H. and Bartel, D. P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**:373–6.

38.   McGinness, K. E., Wright, M. C. and Joyce, G. F. (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.* **9**:585–96.

39.   Lawrence, M. S. and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* **42**:8748–55.

40.   Müller, U. F. and Bartel, D. P. (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem & Biol* **10**:799–806.

41.   Bausch, J. N., Kramer, F. R., Miele, E. A., Dobkin, C. and Mills, D. R. (1983) Terminal adenylation in the synthesis of RNA by Q-Beta replicase. *J. Biol. Chem.* **258**:1978–84.

42.   Clark, J. M., Joyce, C. M. and Beardsley, G. P. (1987) Novel blunt-end addition reactions catalyzed by DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.* **198**:123–7.

43.   Bergman, N., Johnston, W. and Bartel, D. P. (2000) Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**:3115–23.

44.   Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–15.

45.   Schultes, E. A., Hraber, P. T. and LaBean, T. H. (1999) Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.* **49**:76–83.

46.   Hanczyc, M. M. and Dorit, R. L. (1998) Experimental evolution of complexity: In vitro emergence of intermolecular ribozyme interactions. *RNA* **4**:268–75.

47.   Echols, H. and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60**:477–511.

48.   Kunkel, T. A. (1992) Biological asymmetries and the fidelity of eukaryotic DNA replication. *Bioessays* **14**:303–8.

49.   Joyce, G. F. (2004) Directed evolution of nucleic acid enzymes. *Ann. Rev. Biochem.* **73**:791–836.

CHAPTER TWO

Processivity of Ribozyme-Catalyzed RNA Polymerization

[1] Abbreviations: NTP, nucleoside triphosphate; EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane; PT, primer-template duplex.

Abstract

The "RNA world" theory proposes that early in the evolution of life, before the appearance of DNA or protein, RNA was responsible both for encoding genetic information and for catalyzing biochemical reactions. Ribo-organisms living in the RNA world would have replicated their RNA genomes by using an RNA polymerase ribozyme. Efforts to provide experimental support for the RNA world hypothesis have focused on producing such a polymerase, and *in vitro* evolution methods have led to the isolation of a polymerase ribozyme that catalyzes primer extension which is accurate and general, but slow. To understand the reaction of this ribozyme, we developed a method of measuring polymerase processivity that is particularly useful in the case of an inefficient polymerase. This method allowed us to demonstrate that the polymerase ribozyme, despite its inefficiency, is partially processive. It is currently limited by low affinity for primer-template, but once it successfully binds the primer-template in the productive alignment, it catalyzes an extension reaction that is so rapid, it can occur multiple times during the short span of a single binding event. This finding contributes to the understanding of one of the more sophisticated activities yet to be generated *de novo* in the laboratory and sheds light on the parameters to be targeted for further optimization.

Introduction

One part of the mystery of life's origins is the emergence of catalyzed information propagation. Little is known about the evolutionary precursors of modern DNA and RNA polymerases, those data-copying enzymes which, together with the ribosome, make up the core of contemporary molecular biology. Information replication, in the view of many theorists, was achieved by even the earliest, most primitive forms of life (*1, 2*). According to the RNA world hypothesis, information replication was occurring in RNA-based "ribo-organisms" before the emergence of proteins or DNA (*3–6*). This idea flows from the unique dual nature of RNA as both an inherent information carrier and a versatile biological catalyst. One ribozyme indispensable to the RNA world would have been a polymerase, responsible for copying all the ribozymes of a ribo-organism by catalyzing the templated polymerization of mono- or oligonucleotides. The RNA world hypothesis thus presupposes that some RNA sequences can fold into RNA-directed RNA polymerases.

Experimental support for this presumption is starting to accumulate. *In vitro* evolution methods have yielded an RNA polymerase ribozyme that catalyzes an accurate and general primer extension reaction (*7*). This polymerase (Fig. 1A) uses ribonucleoside triphosphates and the coding information of an RNA template to successively extend an RNA primer. It accepts primers and templates of any sequence or length, provided the 3′ terminus of the primer pairs to the template, and it catalyzes polymerization with an average fidelity of 0.97. In its present form, however, the polymerase is slow. It requires at least six hours to extend a primer-template by one helical turn (Fig. 1B, C). Furthermore, because the buffer conditions that maximize its polymerization rate also promote its own hydrolysis, the polymerase is substantially inactivated by the long incubation. With its current limitations, this polymerase ribozyme is incapable of synthesizing the long stretches of RNA that would be needed in the RNA world. To generate a better polymerase, it would be useful to understand which steps of the polymerization reaction are most in need of improvement.

The polymerase ribozyme can be compared with its proteinaceous counterparts, which are generally many orders of magnitude faster, adding up to 750 bases per second (*8, 9*). Some protein polymerases, notably those responsible for replicating chromosomal DNA, catalyze highly processive catalysis, adding many thousands of nucleotides without releasing the nascent chain (*10, 11*). This can be contrasted with completely non-processive, or distributive polymerization, in which a primer-template duplex is iteratively bound, extended by a single nucleotide, and then released. Modern replicative DNA polymerases owe much of their efficiency to their high processivity, which allows them to minimize time lost in repeatedly releasing and re-binding the template.

Given the importance of processivity in protein-catalyzed polymerization, we set out to examine the processivity of the RNA-

Fig. 1. Ribozyme-catalyzed RNA polymerization. **A.** Stylized representation of the polymerase ribozyme secondary structure (blue, black, purple), with primer (orange) indicated by "P", template (red) indicated by "T", and dots (pink) representing NTPs. **B.** Example of an RNA primer-template (PT1). Primer is radiolabeled (*) at its 5′ end. **C.** Sequencing gel showing the results of a polymerization reaction using the primer-template from panel B. Arrow indicates fully-extended product.

catalyzed polymerization reaction. The pattern of products that the polymerase ribozyme has been observed to generate already suggests some degree of processivity. In the case of completely distributive polymerization, by the time fully-extended primer had begun to accumulate, short primers would be nearly used up, and the distribution of products would peak at some intermediate length. However, cursory inspection of Fig. 1C shows the inverse product distribution: as the reaction proceeds, it becomes dominated by fully extended product and very short products. The underrepresentation of intermediate-length products is not straightforward to reconcile with a completely distributive mechanism. On the other hand, the inefficiency of the ribozyme-catalyzed polymerization seems at odds with the notion of processivity. In order to shed some light on this apparent paradox, we investigated the polymerase reaction more closely.

Classical methods of determining polymerase processivity rely on exclusion of multiple binding events. Experiments are designed such that all observed primer extension comes from a single round of processive polymerization. For instance, in a study of yeast polymerase η, a substrate trap was used to achieve this purpose (*12*). The polymerase was preincubated with radiolabeled primer-template, and then the polymerization reaction was initiated by simultaneous addition of dNTPs and a saturating amount of herring sperm DNA as a trap. As the reaction proceeded, any primer-template that fell off the polymerase was replaced by herring sperm DNA. Therefore, all observed extension of the labeled primer was the result of a single binding event. Earlier studies of *E. coli* DNA polymerase I ensured a single binding event by adding primer-template in large excess over polymerase so that the polymerases were statistically extremely unlikely ever to re-bind the same primer-template molecule (*10, 13*). Subsequent studies of *E. coli* DNA polymerase I and mouse DNA polymerase α used a large excess of primer-template over labeled dNTPs (*14, 15*). Under those conditions, the predominant way for two labeled dNTPs to

end up in the same product molecule was through addition during a single round of processive polymerization.

The experimental approaches used in the above studies could not be adapted to our situation. This was because of the low affinity of our polymerase for its primer-template. Using a substrate trap (such as herring sperm DNA) was impossible, because saturating levels of RNA could not be achieved, and moreover high concentrations of RNA appeared to inhibit the polymerase. Lowering polymerase and NTP concentrations in order to restrict observed polymerization to single binding events would have had the unacceptable consequence of abolishing the signal available for measurement. Faced with the limitations imposed by our polymerase ribozyme, chiefly its low affinity for primer-template, we devised a strategy for measuring its processivity that did not rely on being able to exclude multiple binding events. Our approach involved constructing computational kinetic models of distributive and processive polymerization and determining which of the two models better mirrored the experimentally observed polymerization.

Results

Affinity for primer-template. Processivity depends on two competing processes: extension and release. A polymerase is processive only to the extent that it can hold onto its primer-template substrate longer than it takes to extend it. Therefore, in order to determine whether the polymerase is processive or distributive, it was informative to measure its affinity for primer-template. When a primer-template (PT2, Table 1) was titrated into a polymerization reaction, measured reaction rates were linearly proportional to primer-template concentration over the entire testable range, revealing a Michaelis-Menten parameter (K_m) of at least 400 μM, and possibly much higher. However, titration of ribozyme into a similar reaction showed that polymerization rates decline at ribozyme concentrations above 15 μM, suggesting that high concentrations of RNA could inhibit the polymerase.

Other primer-templates had saturable titration curves, which on the surface suggested tighter binding to the polymerase. For instance, in a titration of PT4, reaction rates leveled off quickly, with half-maximal rate observed at 6 μM. A similar result was seen with the unrelated primer-template PTX (5′GAAUCAAG / 3′CUUAGUUCCCG), which showed half-saturation at 7 μM. However, doubt was cast on this interpretation of these low apparent K_m values by experiments attempting to detect competitive or noncompetitive inhibition. If PT4 truly bound to the polymerase active site with micromolar affinity, then at concentrations above 7 μM it would interfere with polymerization of PTX. However, only a very slight inhibitory effect was observed when excess PT4 was titrated as a potential inhibitor into the polymerization reaction of PTX; at an inhibitor concentration of 75 μM, polymerization rate dropped only 5%. In the reciprocal experiment, no decline in reaction rate was observed when PTX was titrated into the reaction of PT4. Based on these experiments, the dissociation constants (K_d) of the polymerase and these primer-templates must be about 1 mM or higher. The low apparent K_m values measured in the titration experiments could be explained by substrate self-sequestration: the homodimers of PT4 and PTX are stabilized by potential base-pairing and stacking interactions at the monomer junctions (Fig. 2), whereas there is no obvious way for PT4 and PTX to form stable heterodimers.

Fig. 2. PT4 and PTX homodimers.

A pre-binding and dilution experiment was performed to ascertain directly how much of the primer-template binds to the polymerase. The polymerase and primer-template PT6 were pre-incubated together in reaction buffer, allowing them time to bind to one another. Then, at the moment of NTP addition, the reactants were diluted 40-fold in order to minimize later binding of polymerase and primer-template. A small "burst" of rapid extension was observed, with 0.17% of the primer-template being extended within the first 10 seconds of the reaction. After that burst, extension proceeded at a much lower rate (Fig. 3). In a control reaction performed without the pre-incubation of polymerase and primer-template, no burst phase was observed; extension proceeded only at the slow association-limited rate, identical to that measured after the burst in the first-experiment.

The burst phase observed in the pre-binding experiment can be quantified in two ways: its size and its duration. The size of the burst phase correlates with how much primer-template was pre-bound to the polymerase, and the duration correlates with how long the pre-bound complex survived in the polymerization reaction. The pre-bound complex has two possible fates; either it successfully reacts to yield extension product (with rate constant k_{cat}) or else it simply dissociates (with rate constant k_{off}). Because the burst was complete within 10 seconds, a lower bound of 6 min^{-1} can be placed on the sum of $k_{cat} + k_{off}$. Interpreting the size of the burst requires knowing what fraction of the pre-bound complex reacted to yield product

Fig. 3. Pre-binding and dilution experiment, showing single-nucleotide extension of PT6 (Table 1). In the "prebound" reaction (red, closed circles), 5 μM ribozyme and 0.5 μM primer-template were incubated together before the reaction was initiated. In the "non-prebound" reaction (blue, open circles), ribozyme and primer-template were kept separate until the moment the reaction was initiated. Amplitude of burst phase (0.0017) is equal to the distance between lines fit to the linear portion of each timecourse. The dotted red line indicates the lower bound on the trajectory of the burst phase.

instead of simply dissociating. At one extreme, if one assumes that much more than 0.17% of the primer-template was productively pre-bound to the polymerase (implying a low K_d), but only a very small fraction of it reacted, then this would imply very fast dissociation from the polymerase, with a k_{off} of at least 6 min^{-1}. At the other extreme, if one assumes that only 0.17% of the primer-template was pre-bound to the polymerase in the productive alignment and all of it was extended during the burst, then this would imply that the K_d of the polymerase/primer-template interaction is approximately 3 mM. Given the earlier measurement suggesting a K_d in the millimolar range or higher, this second extreme seemed closer to reality.

Taken together, the above experiments suggested that the affinity of the polymerase for its RNA substrate is weak, probably in the millimolar range, shedding light on the slow polymerization observed in Fig. 1C. In that experiment, the polymerase and its primer-template were both present at low-micromolar concentrations—a thousandfold below saturation. Moreover, significant increases in rate could not be achieved simply by raising the ribozyme concentration, because polymerization was inhibited by high concentrations of ribozyme. These limitations confined our experiments to sub-saturating conditions for all analyses of the ribozyme and its reactions.

Measuring processivity. We devised a method of measuring processivity that did not rely on being able to saturate the polymerase with substrate trap or otherwise exclude multiple binding events. First, a set of RNA primers was chemically synthesized, all of which could anneal to a single template. Each primer in the series was one nucleotide longer than the primer preceding it. This set of primer-templates (Table 1) can be viewed as a series of intermediates in an RNA polymerization reaction. Each primer-template in the series has two possible histories: either it can be generated from its one-nucleotide-shorter precursor by the ribozyme, or else it can be added to the reaction mixture before the start of the reaction. Each primer-template was used as the starting material for a ribozyme-catalyzed polymerization reaction. For each reaction, the observed rate constant (k_{obs}) was measured for the binding and extension of that primer-template. These observed rate constants were then combined into a model of distributive polymerization, in which every step of polymerization started from unbound primer-template. This model was tested by comparison to the multiple-nucleotide polymerization observed experimentally. The observed polymerization was too fast to be explained by the distributive model, and so a processive model was developed to replace it.

Primer-template	18-nt template 0.1 mM each NTP		18-nt template 4 mM each NTP		21-nt template 4 mM each NTP	
	k_{obs} (hr^{-1})	P	k_{obs} (hr^{-1})	P	k_{obs} (hr^{-1})	P
PT1 5'CUGCCAAC 3'GACGGUUGGCACGCUUCG(CAG)^a	0.16	–	1.4	–	0.08	–
PT2 5'CUGCCAACC 3'GACGGUUGGCACGCUUCG(CAG)	0.0036	0.01	0.09	0.03	0.09	0.00
PT3 5'CUGCCAACCG 3'GACGGUUGGCACGCUUCG(CAG)	0.08	0.07	0.56^b	0.15	1.8	0.30
PT4 5'CUGCCAACCGU 3'GACGGUUGGCACGCUUCG(CAG)	0.014	0.16	0.16^b	0.45^b	1.5	0.47
PT5 5'CUGCCAACCGUG 3'GACGGUUGGCACGCUUCG(CAG)	0.15	0.00	1.3	0.07	11	0.9
PT6 5'CUGCCAACCGUGC 3'GACGGUUGGCACGCUUCG(CAG)	0.22	0.04	1.4	0.25	3.8	0.9
PT7 5'CUGCCAACCGUGCG 3'GACGGUUGGCACGCUUCG(CAG)	0.048	0.01	0.41	0.02	6.0	0.33

Table 1. Primer-templates, observed extension rate constants, and processivity coefficients.
^{a} Template bases shown in parentheses are absent in the 18-nt template.
^{b} These measurements, carried out once already during the procedure illustrated in Fig. 3, were repeated as part of the expanded set of experiments reported in this table; this explains their slight deviations from the values reported in Fig. 3. These differences of 0–20% provide an indication of the experimental variability associated with these measurements.

The distributive model. We began with an extension reaction using PT3 of the Table 1 series as the starting material (Fig. 4A gel). For each timepoint, extension was quantitated by summing together the ladder of products accumulating above the starting band and dividing by the sum of all bands including the starting band, thereby yielding the fraction of primer extended by at least a single nucleotide. The extension was modeled as a single irreversible chemical reaction obeying the observed rate constant of extension $k_{obs(3)}$, determined by best fit to be 0.46 hr^{-1} (Fig. 4A graph). The reaction was first-order with respect to the primer-template concentration, because the experiment was performed under sub-saturating conditions, well within the concentration regime where polymerization rate showed this simple linear dependence. The polymerization rate also depended on the concentration of NTPs, magnesium, hydroxide, and the polymerase itself, but these did not vary appreciably over the timecourse of the reaction, and so they were treated as part of the constant standard reaction conditions. Therefore, throughout this work, observed first-order rate constants (k_{obs}) of polymerization are reported, with respect to these standard conditions.

Next, the analogous polymerization assay was performed, starting with PT4 instead of PT3 (Fig. 4B gel). This primer-template was identical to the single-extension product of the first experiment. The observed rate constant of extension for PT4, $k_{obs(4)}$, was fitted as 0.14 hr^{-1} (Fig. 4B graph).

The two observed rate constants thus determined were then combined into a model of distributive polymerization. The two-step process, starting with PT3 and yielding PT5, was considered as a system of two consecutive irreversible reactions (Fig. 4C scheme). The timecourse predicted by this kinetic scheme was computed and plotted (Fig. 4C graph). This is the timecourse of extension of PT3 by at least two nucleotides, as it would be observed if the polymerase catalyzed completely distributive polymerization. In such a scenario, the starting material PT3 diffuses through the reaction medium until it encounters the polymerase, at which point it is bound, extended by one nucleotide to become PT4, and promptly released back into solution. It then diffuses until binding again to a polymerase molecule, when it undergoes another nucleotide addition and is released back into solution as PT5.

To evaluate how well this modeled timecourse predicted the actual behavior of the polymerase in extending PT3 by two nucleotides, the results of the first experiment starting with PT3 (Fig. 4A gel) were re-quantitated in a different way, this time summing together all product bands at least two nucleotides larger than the starting material, thereby yielding the fraction of primer extended at least twice. This data was compared to the distributively modeled timecourse (Fig. 4C graph). Clearly, the data and distributive model were in disagreement. The accumulation of PT5 occurred too fast to be explained by a fully distributive mechanism. Furthermore, the distributive model assumed product release to occur immediately after nucleotide addition. Any delay due to product release would shift the modeled distributive timecourse even lower. The curve shown in Fig. 4C therefore reflects the best-case scenario for the distributive model. Moreover, the distributive model could not be redeemed by simply increasing the individual rate constants. Doing so yielded incorrectly-shaped curves that undershot early data points and overshot late ones (not shown).

The processive model. Having demonstrated that the polymerase catalyzed addition of two nucleotides faster than a fully distributive mechanism allowed, it remained to be explained just how it accomplished this feat. One reasonable hypothesis was that the polymerase could hold onto its primer-template long enough to add a second nucleotide some fraction of the time. We defined
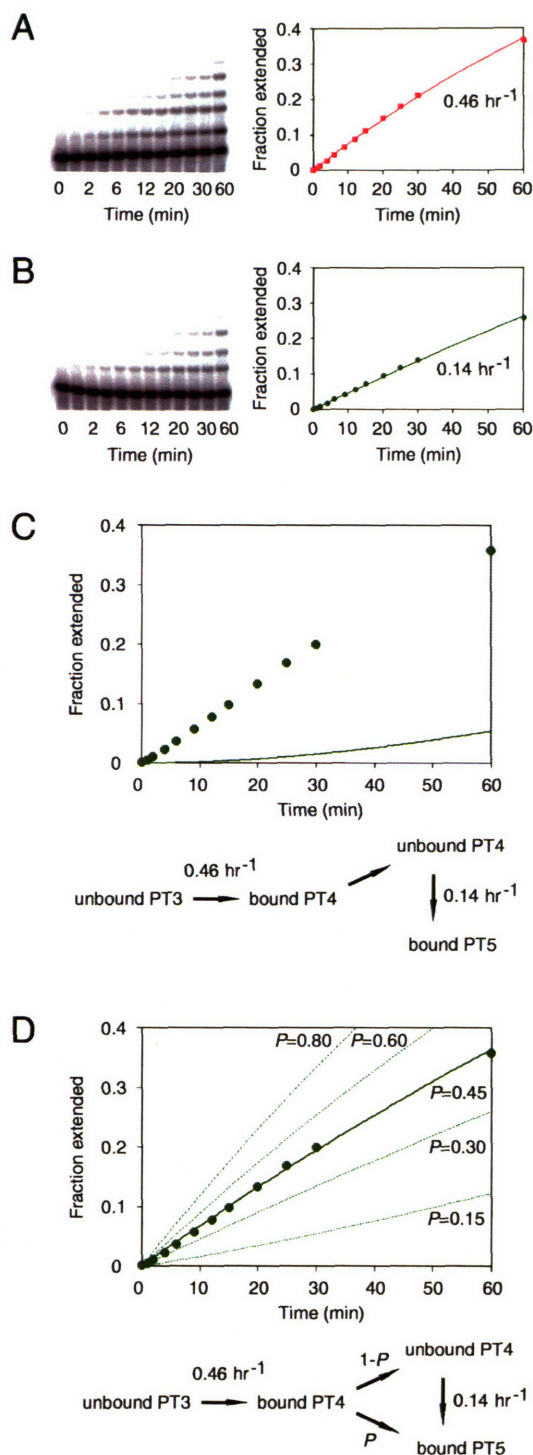
Fig. 4. Development of the processive model. **A.** Single-nucleotide extension of PT3. Single-nucleotide extension was quantitated by summing together all bands above the lowest band (the starting material), and dividing by the sum of all bands in the lane, to yield the fraction of starting material extended by at least one nucleotide. This fraction is plotted (red squares) along with a timecourse (red line) fitted to the observed rate constant $k_{obs(3)}$ of 0.46 hr^{-1}. **B.** Single-nucleotide extension of PT4, quantitated as in panel A and plotted (green circles) along with a timecourse (green line) fitted to the observed rate constant $k_{obs(4)}$ of 0.14 hr^{-1}. **C.** Two-nucleotide extension of PT3 and a minimal scheme for the distributive model. Plotted (green circles) is the fraction of starting material that has been extended by at least two nucleotides, quantitated from the gel in panel A by summing together all bands except the lowest two bands, and dividing by the sum of all bands in the lane. Distributive scheme reflects conversion of unbound PT3 to bound PT4 with $k_{obs(3)}$, followed by obligate release of PT4, then conversion of unbound PT4 to bound PT5 with $k_{obs(4)}$. Superimposed on the data is the modeled timecourse of distributive polymerization. Distributive model (green line) is a poor fit to the observed data. **D.** Two-nucleotide extension of PT3 and a minimal scheme for the processive model. Data points (green circles) are as in panel C. Processive scheme reflects conversion of unbound PT3 to bound PT4 with $k_{obs(3)}$, followed by one of two possible paths: either further extension of bound PT4 to bound PT5, occurring with probability P; or else release of bound PT4, occurring with probability $1-P$, followed by conversion of unbound PT4 to bound PT5 with $k_{obs(4)}$. Modeled processive timecourses (solid and dotted green lines) superimposed upon the data show the effect of varying P in the processive model. Best fit to data is achieved with $P=0.45$ (solid green line).

the processivity coefficient P as the probability of adding a second nucleotide before primer-template release. This parameter ranges from zero (completely distributive polymerization, already ruled out) to one (completely processive polymerization, in which the polymerase holds onto its substrate and extends it all the way to the end of the template before letting go.)

In the processive model of polymerization (Fig. 4D scheme), the starting material PT3 diffuses to and binds to the polymerase, where it is extended by a single nucleotide, yielding

PT4. This step was modeled with $k_{obs(3)} = 0.46$ hr^{-1}, as before. The distributive model required that PT4 then be released from the polymerase and re-bound before it could be extended again. At this juncture however, the processive model diverged, allowing PT4 to be extended a second time (yielding PT5) while still bound to the polymerase. This occurs with a probability of $P_{(4)}$. The other possible outcome, with a probability of 1-$P_{(4)}$, is for PT4 to fall off the polymerase and continue along the distributive pathway by binding again to a polymerase molecule and being extended to PT5. As before, the binding and extension of PT4 were modeled with $k_{obs(4)} = 0.14$ hr^{-1}. The Fig. 4D graph shows the effect of varying $P_{(4)}$ in the model so described. Each curve represents the partially processive timecourse of extending PT3 to PT5, as calculated for a given value of $P_{(4)}$. A close match to the observed timecourse was obtained with $P_{(4)} = 0.45$. In other words, the polymerase binds PT3 and extends it a first time, and then with nearly equal probability, either releases it or extends it a second time.

Analysis was extended to other positions along the template by iteration of the process just described. For example, a polymerization assay was performed using PT5 (not shown). From this additional data, $k_{obs(5)}$ and $P_{(5)}$ were determined. Each subsequent polymerization assay yielded an additional pair of parameters. For the final data set (Table 1), the measurements obtained in Fig. 4 were repeated as part of an expanded set of experiments encompassing seven consecutive positions along the template. A strong dependence on sequence context was observed, with both k_{obs} and P varying over 20 fold, but in nearly all cases, some degree of processivity was detectable.

Quantitation of the processivity of the polymerase shed additional light on the pre-binding experiment (Fig. 3), allowing a clearer interpretation of its results. The burst phase observed in that experiment had established a lower bound of 6 min^{-1} for the sum of $k_{cat} + k_{off}$, but because it was unknown what fraction of bound primer-template reacted during the burst instead of falling off, it was impossible to establish lower bounds on the individual rate constants k_{cat} and k_{off}. Measuring the processivity at that position as 0.25 (PT6 in Table 1) revealed that one out of four bound primer-templates had been extended, and the other three had dissociated from the polymerase before they could be extended. This in turn implied that k_{off} had three times the magnitude of k_{cat} and allowed us to establish lower bounds of 4.5 min^{-1} for k_{off} and 1.5 min^{-1} for k_{cat} (Fig. 5). Furthermore, measuring a processivity of 0.25 implied that the extension observed during the burst phase (0.17%) represented only a fourth of the prebound primer-template; the total amount of prebound primer-template was 0.7%. This value (along with the ribozyme concentration, 5 μM) allowed us to refine our estimate of the K_d of this interaction to 700 μM. As polymerization is more efficient in the context of PT6 than in some other contexts, some primer-templates are likely to have higher K_d values than PT6.

$k_{cat} \geq 1.5$ min^{-1}

bound PT7

bound PT6

unbound PT6

$k_{off} \geq 4.5$ min^{-1}
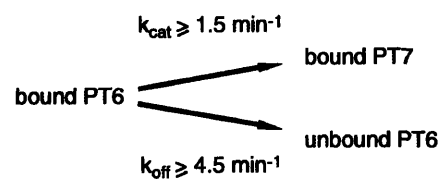
Fig. 5. Lower bounds on k_{cat} and k_{off}. Polymerase-bound PT6 is either extended or released, at rates no lower than the indicated limits.

Validating the processive model. In order for the processive model to capture the behavior of two-nucleotide extension, it required two parameters, k_{obs} and P. At this point, the crucial question arose: would these parameters, determined locally for pairs of successive primer-templates, be sufficient to model long-range polymerization? This question was addressed by challenging the processive model to predict polymerization of three, four, five nucleotides, and beyond. As hoped, the results of the simulation agreed with all observed data. No further refinement or extra parameters were required in order to successfully predict the timecourse of many-nucleotide extension. Fig. 6

shows the results of this test. In each panel A through E, a different primer-template was used as starting material, and the observed polymerization data is plotted in both the upper and lower graphs. In the upper graph, the predictions of the distributive model are superimposed on the data. In the lower graph, the predictions of the processive model are superimposed. In all cases, both models correctly predicted the first extension (uppermost curve), but only the processive model succeeded in predicting subsequent extensions.
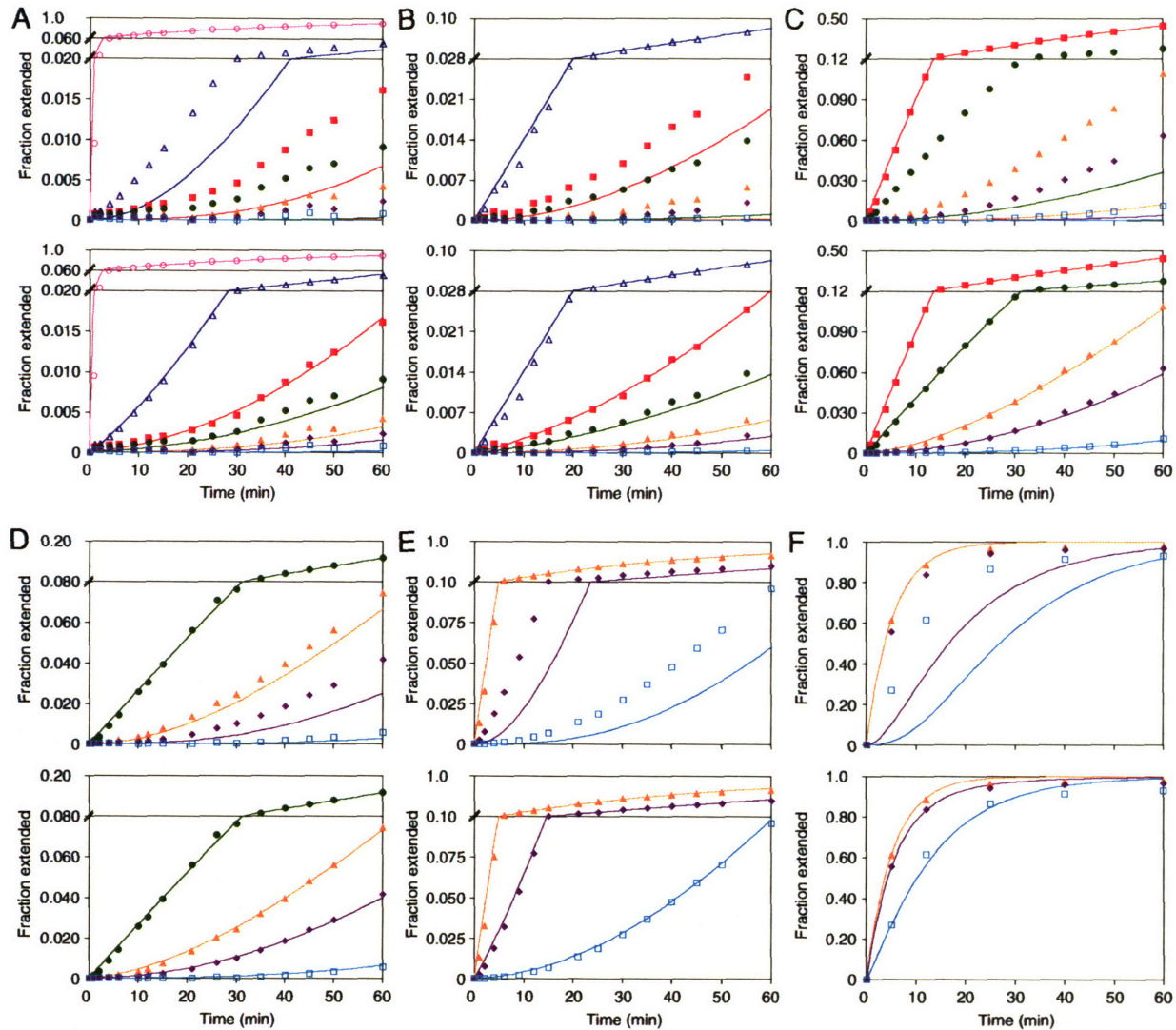
Fig. 6. Polymerization of many nucleotides. **A.** Starting from PT1. **B.** Starting from PT2. **C.** Starting from PT3. **D.** Starting from PT4. **E.** Starting from PT5. In panels A through E, the 18-nt template was used. **F.** Starting from PT5 and using the 21-nt template. In panels A through F, upper and lower plots show the same data, measured from polymerization assays. Upper plot superimposes the predictions of the distributive model, and lower plot superimposes the predictions of the processive model. Plotted is the fraction of starting material that has been extended to at least PT2 (pink lines/open circles), at least PT3 (blue lines/open triangles), at least PT4 (red lines/closed squares), at least PT5 (green lines/closed circles), at least PT6 (yellow lines/closed triangles), at least PT7 (purple lines/closed diamonds), and at least PT8 (light blue lines/open squares). Upper regions of vertical axes are compressed (except in panel F).

39

Effect of lowered NTP concentrations. At the core of processive model is the partitioning of the bound primer-template along two pathways: extension and release. If some change in reaction conditions were to affect one of these two pathways disproportionately, then it should also cause a change in the observed processivity. We tested this prediction by lowering the concentration of NTPs in the reaction. This change was expected to reduce the observed extension rate without dramatically affecting the rate of primer-template release. According to the processive model, it should also cause a drop in processivity. All parameters were re-measured using 1/40 the original NTP concentration, and as expected, all observed extension rate constants and processivity coefficients fell (Table 1).

Effect of lengthened template. As an additional test of the model's robustness, polymerization was re-examined using a template with three additional nucleotides at its 5′ end. This slight change was found to influence the behavior of the polymerase dramatically (Table 1). In almost all cases, the observed extension rate constants and the processivity coefficients both increased, in some cases by more than an order of magnitude. Nevertheless, when these markedly different parameters were combined into a new processive simulation, it yielded correct predictions of the multiple-nucleotide extension observed using the lengthened template. The analysis illustrating the greatest processivity detected thus far ($P_{(5)}$ = 0.9 when using the 21-nt template) is shown in Fig. 6F. In this context, the processivity is high enough to confer a "running start" advantage: the polymerase generates PT7 from PT5 by two extension reactions faster than it generates PT7 from PT6 by a single extension reaction (Fig. 7). This result implies that the polymerase binds PT6 considerably slower than it binds PT5, and that the easier route to productively bound PT6 is through the binding of PT5 and its extension to PT6.
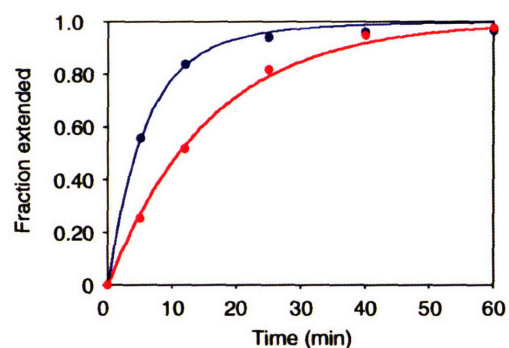
Fig. 7. Example of a "running start" advantage conferred by processivity. Two timecourses of polymerization are shown, using the 21-nt template (see Table 1), one starting from PT5 (blue circles, upper curve), the other starting from PT6 (red circles, lower curve). Plotted for each timecourse is the fraction of starting material extended to at least PT7. Processive-model timecourses are superimposed. Accumulation of PT7 is faster when the more distant starting point (PT5) is used.

Discussion

The kinetic parameters of the processive model were all determined on a strictly local basis, yet in aggregate they succeeded in capturing the experimentally observed global behavior of the system, thereby significantly validating the model. This finding provides strong evidence that the polymerase ribozyme can catalyze the processive addition of multiple nucleotides in a single binding event.

The processivity of this polymerase depends in large part on the particular sequence context of polymerization. In two contexts, its processivity reached 0.9, placing it among low-processivity proteinaceous polymerases like yeast polymerase η (12). In other contexts, its processivity was weak or undetectable. Sequence-dependent variation in k_{obs} was also seen, a phenomenon

observed with proteinaceous polymerases (*16, 17*). There was no systematic relationship between the k_{obs} and P values in different contexts, implying dramatic differences in rate constants of primer-template release (k_{off}) in different sequence contexts. Such variation must also explain the differences in P when measured using the 18- and 21-nt templates.

Given the weak affinity of the polymerase for primer-template, the detectable processivity of the polymerase came as a surprise. Although the polymerase has great difficulty in achieving productive primer-template binding and alignment, once it succeeds, the extension reaction occurs with great speed – so fast, in fact, that the reaction can occur multiple times before the complex dissociates. This finding makes sense considering that the catalytic core of the polymerase is a ligase ribozyme capable of promoting a self-ligation reaction with a k_{cat} of 500 per minute (*18*). It appears that much of the innate catalytic ability of the ligase was retained when its core was moved into the context of the polymerase pool. Indeed, when the polymerase was given the advantage of a covalent tether to its primer, its polymerization rate increased 300-fold (data not shown).

For a ribo-organism in the RNA world, processivity could have been instrumental in overcoming the problem of strand displacement. Because of the tremendous stability of long RNA duplexes, dislodging a newly synthesized RNA strand from its template is a formidable task. One way of achieving it is to wait until the subsequent round of synthesis, letting the next nascent RNA strand gradually displace the previous one as it is synthesized (*19*). This strategy requires that the polymerase be processive, for if the polymerase fell off the template early in the synthesis, the previous full-length product would simply re-anneal, displacing the much shorter nascent RNA strand.

The finding that the polymerase is currently limited by its weak affinity for primer-template suggests that future *in vitro* selection experiments with this polymerase ribozyme should target primer-template binding rather than only chemical proficiency and NTP binding. Doing so could increase its efficiency to the point where it would be able to synthesize a complementary strand of its own length. The polymerase is nearly 200 nt in length, but with its current limitations, it can polymerize only about 14 nt in six hours. Longer incubations yield little further extension because the incubation conditions degrade the polymerase. However, computational modeling can provide an estimate of how long it would take the polymerase to synthesize its complementary strand if it were freed from the constraints imposed by the incubation conditions. We modeled polymerization starting from an 8-nt primer, using a 200-nt template. We assigned values of k_{obs} and P along the template by reiterating the succession of parameters from Table 1. This was done so that our model would take into account the large fluctuations in both parameters observed along a given template. Using the values for the 18-nt template, we found that the polymerase would require over three weeks to fully extend 50% of the starting material. For the 21-nt template values, the polymerase would require just over two weeks. Next we modeled the polymerization that would be expected if the primer-template affinity of the polymerase could be improved 100-fold (decreasing the K_d of PT6, for example, to about 7 μM). We modeled this improvement as a 100-fold decrease in k_{off} at each position along the template. This decrease in k_{off} was transferred to the parameters of both the processive and the non-processive modes of extension as follows: First, because P is determined by the k_{cat}/k_{off} ratio, P was increased at each position according to the 100-fold increase in that ratio. Second, because the observed rates of non-processive extension starting from unbound material (k_{obs}) were limited by primer-template affinity, k_{obs} at each position was increased 100-fold. The model using the 18-nt template parameters predicted that the improved polymerase would require just over 1.5 hours to fully extend 50% of the primer. Using the 21-nt template parameters, it

would require just over 4 hours. These timescales are within the polymerase lifetime imposed by the incubation conditions. Thus, sufficient gains in primer-template affinity may eventually yield the efficient polymerization that would be required of an RNA replicase.

Acknowledgements

Materials and methods

Polymerization reactions. The polymerase was the round-18 ribozyme from Johnston et al., prepared as described therein (*7*). Polymerization reactions contained 2 μM ribozyme (with 2.5 μM GGCACCA oligoribonucleotide), 1 μM RNA template, and 0.5 μM 5'-radiolabeled RNA primer. These RNAs were mixed together in water, heated (80°C, 2 min), and then incubated at 22°C for at least 5 min before starting the reaction by simultaneous addition of 4 mM each nucleoside triphosphate (NTP),[1] 200 mM $MgCl_2$, and 100 mM Tris pH 8.5 (final concentrations.) Typical reaction volume was 30 μL. Reactions were incubated at 22°C. At selected timepoints, reaction aliquots were withdrawn and quenched by addition of 4 volumes of 50 mM EDTA/8M urea. Quenched aliquots were heated (80°C, 2 min) in the presence of competitor RNA designed to hybridize to the template RNA, then analyzed on sequencing gels. Variations of this protocol were as described in the text.

Pre-binding experiment. Two parallel polymerization reactions were performed, each containing 5 μM ribozyme (with 6.25 μM GCCACCA oligo), 0.5 μM radiolabeled primer and 2.5 μM template (PT6, 18-nt template), 200 mM $MgCl_2$, 100 mM Tris pH 8.5, and 4 mM GTP. In the "prebound" reaction, all RNAs were heated together in water (80°C, 2 min) and cooled 5 min to 22°C, then salt and buffer were added, and the mixture was incubated another 5 min at 22°C. Reaction was started by 40-fold dilution into GTP, salt, and buffer. In the "non-prebound" reaction, ribozyme and primer-template were kept separate during the heating, cooling, and salt/buffer incubation. Pre-incubated primer-template was diluted into GTP, salt, and buffer, and the reaction was started by addition of pre-incubated ribozyme. Upon this step, final concentrations of all reaction components became exactly as in the "prebound" reaction, and timecourses proceeded as described above.

Kinetic modeling. Polymerizations were modeled computationally by numerical integration of a set of equations that formalize the chemical reactions occurring in the polymerization. All numerical integration was performed using a Microsoft Excel spreadsheet in which each column tracked one variable, and each row recomputed all the variables a small time interval (Δt) after the preceding row. For example: in a timecourse of single-nucleotide addition, such as the irreversible extension of PT3 to yield PT4, only a single reaction needs be considered. This reaction is first-order with respect to the concentration of PT3 and obeys the observed rate constant $k_{obs(3)}$. Two complementary equations are required to formalize this reaction: one to represent the disappearance of PT3, and another to represent the concomitant appearance of PT4. The two equations are (1) $\Delta[PT3] = -k_{obs(3)}[PT3]\Delta t$; and (2) $\Delta[PT4] = +k_{obs(3)}[PT3]\Delta t$. The spreadsheet for

42

this timecourse had three columns, one each for t (time), [PT3], and [PT4]. All columns were initialized to zero in the first row of the spreadsheet, except for [PT3], which was initialized to 0.5 μM, the starting concentration of PT3 in the reaction. Each subsequent row recomputed the three columns as follows. The new value of t was the old value of t, plus the constant Δt. The new value of [PT3] was the old value of [PT3] (found in the previous row) minus the product of the constant $k_{obs(3)}$ times the old value of [PT3] times Δt. The new value of [PT4] was the old value of [PT4], plus that same product. The best fit for $k_{obs(3)}$ was determined by adjusting $k_{obs(3)}$ until the root-mean-squared discrepancy between the modeled timecourse and the experimental data was minimized. As long as Δt was sufficiently small, the results of the simulation did not depend on its value. All simulations reported herein used a Δt of 2 sec. Simulations using a Δt of 10 sec produced indistinguishable results.

Distributive model. The completely distributive model shown in Fig. 4C was constructed by generalizing the system described above. For each extension step in the polymerization, one chemical reaction was added to the model, reflecting the single-nucleotide extension of unbound $PT_{(n)}$ to yield bound $PT_{(n+1)}$, obeying rate constant $k_{obs(n)}$. Subsequent release of $PT_{(n+1)}$ was modeled as occurring instantaneously. The parameters of the distributive model are the set of $k_{obs(n)}$ values for each $PT_{(n)}$. These parameters were fit independently of one another, one from each timecourse of single-nucleotide extension.

Processive model. For each extension step in the processive model shown in Fig. 4D, three chemical reactions were included in the model. The first was for single-nucleotide extension of unbound $PT_{(n)}$ to yield bound $PT_{(n+1)}$, as described above. The second was for processive extension of bound $PT_{(n+1)}$ to yield bound $PT_{(n+2)}$, obeying rate constant $k_{cat(n+1)}$. The third was for release of bound $PT_{(n+1)}$, obeying rate constant $k_{off(n+1)}$. The processivity coefficient P was defined as the ratio $k_{cat} / (k_{cat} + k_{off})$. Two versions of the processive model were implemented, one with bound $PT_{(n)}$ partitioning instantaneously, and one with the partitioning rate set equal to 6 min^{-1}, the lower bound on $(k_{cat} + k_{off})$ calculated from the burst duration in the pre-binding experiment. These two implementations gave indistinguishable results. The parameters of the processive model are the set of $k_{obs(n)}$ and $P_{(n)}$ values for each $PT_{(n)}$. Values of $k_{obs(n)}$ were inherited from the distributive model. Values of $P_{(n)}$ were then fit independently of one another, one from each timecourse of two-nucleotide extension. Extension beyond two nucleotides was not examined during the parameter fitting. It was reserved for testing the model's predictive power.

References

1.   Orgel, L. E. (1968) Evolution of the genetic apparatus, *J. Mol. Biol. 38*, 381–393.

2.   Eigen, M. (1971) Self-organization of matter and the evolution of biological macromolecules, *Natürwissenschaften 58*, 465–523.

3.   Pace, N. R., and Marsh, T. L. (1985) RNA catalysis and the origin of life, *Origins Life 16*, 97–116.

4.   Gilbert, W. (1986) The RNA world, *Nature 319*, 618.

5.   Orgel, L. E. (1986) RNA catalysis and the origins of life, *J. Theor. Biol. 123*, 127–149.

6.   Joyce, G. F., and Orgel, L. E. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 49–77, Cold Spring Harbor Laboratory Press.

7.   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, *Science 292*, 19–25.

8.   Fay, P. J., Johanson, K. O., McHenry, C. S., and Bambara, R. A. (1981) Size classes of products synthesized processively by DNA polymerase III and DNA polymerase III holoenzyme of *Escherichia coli*, *J. Biol. Chem. 256*, 976–983.

9.   Kornberg, A., and Baker, T. A. (1991) *DNA Replication*, W. H. Freeman, New York.

10.   McClure, W. R., and Chow, Y. (1980) The kinetics and processivity of nucleic acid polymerases, *Meth. Enz. 64*, 277–297.

11.   Xiang-Peng, K., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp, *Cell 69*, 425–437.

12.   Washington, M. T., Johnson, R. E., Prakash, S., and Prakash, L. (1999) Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase η, *J. Biol. Chem. 274*, 36835–36838.

13.   Bambara, R. A., Uyemura, D., and Choi, T. (1978) On the processive mechanism of *Escherichia coli* DNA polymerase I, *J. Biol. Chem. 253*, 413–423.

14.   Detera, S. D., Becerra, S. P., Swack, J. A., and Wilson, S. H. (1981) Studies on the mechanism of DNA polymerase *a*, *J. Biol. Chem. 256*, 6933–6943.

15.   Detera, S. D., and Wilson, S. H. (1982) Studies on the mechanism of *Escherichia coli* DNA polymerase I large fragment, *J. Biol. Chem. 257*, 9770–9780.

16.   Echols, H., and Goodman, M. F. (1991) Fidelity mechanisms in DNA replication, *Annu. Rev. Biochem. 60*, 477–511.

17.   Kunkel, T. A. (1992) Biological asymmetries and the fidelity of eukaryotic DNA replication., *Bioessays 14*, 303–8.

18.   Glasner, M. E., Bergman, N. H., and Bartel, D. P. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme, *Biochemistry 41*, 8103–12.

19.   Bartel, D. P. (1999) Re-creating an RNA replicase, *The RNA World, Second Edition*, 143–162.

CHAPTER THREE

The secondary structure and sequence optimization of a second ligase-derived RNA polymerase ribozyme

Abstract

In modern cells, proteins catalyze RNA synthesis, but earlier in evolution—according to the RNA World hypothesis—RNA itself was the catalyst. Experimental efforts to demonstrate RNA-catalyzed RNA polymerization have yielded several rudimentary RNA polymerases, the first of which (Pol 1) was optimized to the point where it could polymerize a full turn of an RNA helix. Here we report the optimization of a second polymerase ribozyme (Pol 2), using a "doped selection" procedure. Covarying bases revealed by this procedure indicated paired regions, supporting the proposed secondary structure for Pol 2. The identities of several nucleotides in the 3' tail of the ribozyme were found to dramatically influence its polymerization activity with certain primer-templates, suggesting that base-pairing between the ribozyme and the polymerization template may still be a strategy employed by these polymerases, while other primer-templates were used regardless of the 3' tail sequence. One variant of Pol 2 isolated in the doped selection (clone 702) showed improved activity with all primer-templates tested. Another isolate (clone 713), differing from clone 702 by a single point mutation in the tail, catalyzed extension of a primer by 8 nucleotides, and it was designated Pol 2+ to indicate its improvement over the parent. Despite this improvement, Evolved Pol 1 remains the most active RNA polymerase ribozyme isolated to date.

Introduction

In order to substantiate the popular theory of an extinct ancestral RNA World,[1–4] there is an ongoing experimental effort to identify an RNA molecule that can fold into an RNA polymerase efficient enough to synthesize its own complement strand, and therefore capable of providing the replicase activity that would have been essential in the RNA World scenario.[5, 6] Several RNA polymerase ribozymes have been isolated by *in vitro* selection from a starting pool containing RNA ligase ribozymes attached to random 76-nt auxiliary domains, called the Lig+N_{76} pool (Fig. 1A). The first polymerase isolated from this pool (Pol 1) is reported in the Appendix, along with its improved variant (Evolved Pol 1). The other polymerases were reported in Chapter 1. One of these (Pol 2) showed about the same level of activity as Pol 1 when assayed with a variety of primer-templates. It was chosen for optimization. Here we report the "doped selection" performed on Pol 2, and the secondary structure information that emerged from it, as well as an improved variant (Evolved Pol 2) which shows improved activity (relative to Pol 2) with all primer-templates tested.

Results

Choosing a starting point. The original *in vitro* polymerase selection experiment starting from the Lig+N_{76} pool contained two branches. The main result of branch A was Pol 1, reported in the Appendix. The main result of branch B was Pol 2, the focus of the present report. As reported in Chapter 1, Pol 2 first emerged in round 8, and its abundance increased steadily through the next few rounds, comprising the majority of the pool after round 11. Clone 11.22 had the consensus sequence of the Pol 2 family. We assayed clone 11.22 and fourteen other Pol 2 variants from pools 8–11, using two different primer-templates. The most active Pol 2 variant was clone 11.78 (Fig. 1B), which differed from the consensus clone by four point mutations (A165U, G166C, U175G, A176G). Clone 11.78 was chosen as the starting point for further optimization. Attempts to truncate this clone were unsuccessful, as it required its 3'-terminal 8 nt for activity (Chapter 1). It also required the P2-completing heptamer.

46
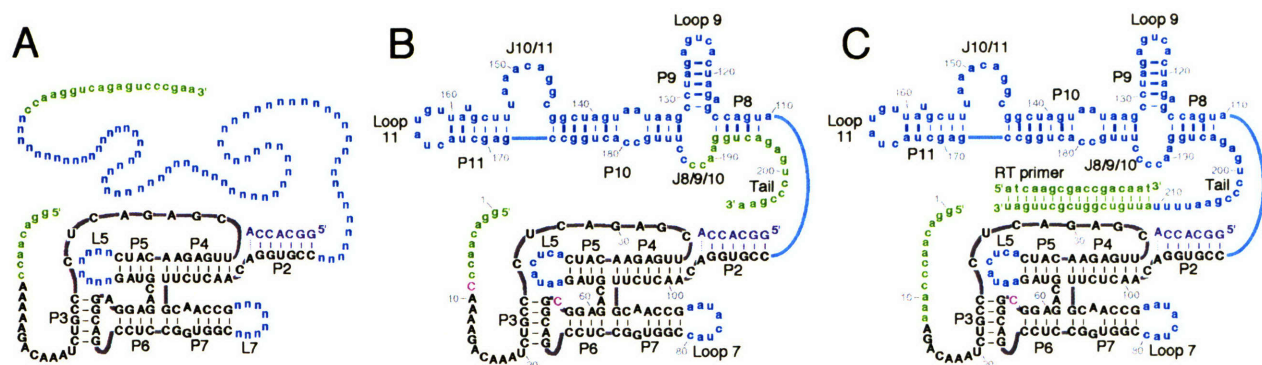
Fig. 1. Secondary structures of *in vitro* selection pools and ribozyme Pol 2. **A.** Structure of Lig+N$_{76}$ pool (adapted from Johnston et al. 2001–see Appendix). Ligase ribozyme is shown in black, with P2 stem-completing heptamer supplied in polymerization reaction shown in purple. Randomized 76-nt auxiliary domain is shown in blue, as are the two ligase loops, also randomized in the pool. Green regions indicate fixed-sequence primer-binding sites. **B.** Structure of Pol 2 (clone 11.78), which was isolated from the pool shown in A. Thick dashes indicate base-pairing supported by comparative sequence analysis. Two pink nucleotides indicate mutations that arose in the ligase region. **C.** Structure of 11.78 doped pool. Blue residues were mutagenized at a level of 20%. A new RT primer binding site was designed and appended to the 3' end of 11.78, allowing mutagenesis throughout the entire auxiliary domain. RT primer was included in the polymerization reaction, rendering that region of the pool double-stranded.

During the optimization of Pol 1 (see Appendix), that polymerase acquired a point mutation in its ligase domain (U106C), disrupting a base pair in the P2 stem. Reversing this point mutation diminished polymerization activity, raising speculation that the Evolved Pol 1 might prefer a perturbed geometry in that region of its ligase domain. We wondered if Pol 2 might also benefit from this adjustment. To test the idea, a variant of Pol 2 was constructed containing the U106C point mutation. However, instead of improving its activity of Pol 2, the mutation caused a slight decrease in activity (data not shown). Therefore, the U106C mutation was not considered further as a possibility for Pol 2 optimization.

Design of doped pool. The optimization of Pol 2 was carried out by means of a partially-randomized ("doped") selection, as done before with Pol 1. A doped pool of Pol 2 variants was constructed (Fig. 1C), based on the most active Pol 2 variant found above (clone 11.78). Each member of the pool was mutagenized at a level of 20% in the auxiliary domain. In order to allow mutagenesis of the entire auxiliary domain, a new fixed-sequence primer-binding site was appended to the 3' end of the polymerase, separated from the ribozyme by four Us, also 20% mutagenized. In order to prevent the polymerase from further elaborating itself by incorporating the new single-stranded 3' end into its already complex structure, this new primer-binding site was rendered double stranded (and hopefully inert) during the selection protocol by annealing the 3' amplification primer to it. The ligase core was mutagenized at 3% in half of the pool, while in the other half it was not intentionally mutagenized. This low level of ligase mutagenesis was chosen on the basis of prior observations that the ligase does not benefit from additional point mutations, its optimization appearing already largely complete. The ligase loops, which seem able to tolerate almost any sequence, were mutagenized at 20% throughout the pool. The pool was constructed using methods described in the Appendix, and its composition was verified by cloning and sequencing. The complexity of the starting pool ("Pool Zero") was 5×10^{14}.

47

A new method of primer attachment. The central feature of our polymerase selection technique is enrichment on the basis of primer extension. After the selection incubation, primers that got extended are physically separated from primers that did not get extended. In order for the pool to evolve, there must be some means of identifying the active pool molecules that catalyzed primer extension, so that they can be selectively amplified. This can be done by covalently attaching an RNA primer to each member of the pool, via a long, flexible linker, and keeping the pool dilute enough during the polymerization incubation, so that the intramolecular reaction (extension of one's own primer) dominates over the intermolecular reaction (extension of someone else's primer). In this manner, when we purify the extended primers, the active polymerases that did the extension automatically copurify with them.

During the Lig+N_{76} selection, the primer was attached to the pool by Moore-Sharp ligation,[7] by using T4 DNA ligase and a splint oligo that paired to the fixed-sequence 5'-leader of the pool and to a 9-nt fixed-sequence DNA linker at one end of the primer (Fig.2A). By the end of the Pol 1 optimization selection, it became apparent that Evolved Pol 1 had developed a sequence preference for the DNA linker (data not shown). This was undesirable in light of our goal of general RNA polymerization, as it represented a reliance on specific sequence elements.

The problem could be solved by simply varying the sequence of the DNA linker, but we reasoned that DNA would always be an attractive ligand for RNA, and so we instead we developed a completely different method of attaching the primer, using a polyethylene glycol (PEG) linker
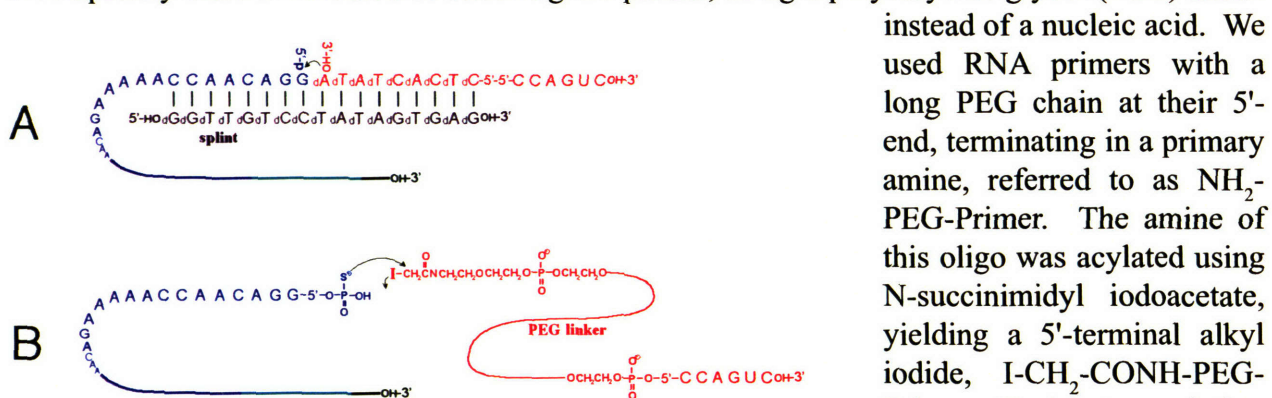
Fig. 2. Two ways of ligating the primer to the pool. **A.** Primer is synthesized with a 5'–5' linkage joining it to a 9-nt stretch of linker DNA (sometimes called "reverse DNA" because of its synthesis in the reverse direction on the DNA synthesizer.) Pool is transcribed in the presence of excess GMP, causing most of the pool to be initiated with a monophosphate (rather than a triphosphate). Primer and pool are aligned on a DNA splint oligo as shown, and they are covalently linked by the action of T4 DNA ligase. Ligated pool is then gel-purified to remove splint. **B.** Primer is synthesized with a long polyethylene glycol (PEG) linker at its 5' end, terminating in a primary amine, converted to an alkyl iodide by treatment with N-succinimidyl iodoacetate (not shown). Pool is transcribed with excess GMPS, so that most of the pool has a 5'-monophosphothioate. Pool and primer are ligated nonenzymatically by the nucleophilic displacement of iodide, as shown.

instead of a nucleic acid. We used RNA primers with a long PEG chain at their 5'-end, terminating in a primary amine, referred to as NH_2-PEG-Primer. The amine of this oligo was acylated using N-succinimidyl iodoacetate, yielding a 5'-terminal alkyl iodide, I-CH_2-CONH-PEG-Primer. During transcription of the ribozyme pool or clone to be attached to the primer, GMPS was included in excess over GTP, with the result that most transcribed molecules had a 5'-phosphorothioate (S-Ribozyme). The I-CH_2-CONH-PEG-Primer and S-Ribozyme were mixed together and allowed to react, with nucleophilic substitution yielding the desired ligated construct Ribozyme-S-CH_2-CONH-PEG-Primer

(Fig. 2B). The efficiency of this ligation reaction was improved by use of a splint to bring together the two reaction nucleic acids, and by use of high salt (1M NaCl). The ligation product was gel-purified before polymerization reactions. The polymerization activity of Pol 2 was observed to be unaffected by switching from the DNA linker to the PEG linker (data not shown), and so the new PEG linker technique was employed in rounds 1–3 and 7–10 of the Pol 2 doped selection (Table 1).

Pool evolution. Ten rounds of selection were carried out using the protocol described in Chapter 1 and the Appendix, and using the primer-templates and conditions listed in Table 1. Robust pool activity was detected after round 6. At that point, the pool was mutagenized again using error-prone PCR, with expected levels of 0, 2, 4, and 6 percent mutagenesis.[8] In this way we hoped to explore potentially beneficial mutations in the auxiliary domain, but we were aware that most of the mutagenesis in the auxiliary domain would be lost in the subsequent rounds due to linked deleterious mutations in the ligase domain. We sought a way to achieve mutation in the auxiliary domain while maintaining a "wild type" ligase. To do this, we employed recombination,

Rd.	primer-template		NTPs (mM)		time (hr)	reacted (%)	
			^{4S}U	A,C,G		+U	+UU
1	11.23 17.68	PEG–AGCUGCC UCGACGGaaCCUGCGUC	0.1	2 ea.	44	—	—
2	14.21 22.35	PEG–AGUCAGUCGU UCAGUCAGCAaaUGCCCUGGUC	0.1	2 ea.	6	—	—
3	14.20 25.19	PEG–GACCUCGCCA CUGGAGCGGUaaCCUCGUUUGGCUG	0.1	2 ea.	26	—	—
4	16.11 17.69	DNA–UGGACUC ACCUGAGaaGUGUAUGU	0.1	2 ea.	9	—	—
5	19.35 22.38	DNA–CUGACCAGCG GACUGGUCGCaaUCCGUCUGAA	0.1	2 ea.	24	—	—
6	19.34 22.35	DNA–AGUCAGUCGU UCAGUCAGCAaaUGCCCUGGUC	0.1	2 ea.	27	—	0.25
7	13.16 21.75	PEG–GUCCAGUAG CAGGUCAUCaaGUUCUCGCAG	0.1	2 ea.	1	0.2	0.02
8	14.22 22.65	PEG–GAUAGGACAU CUAUCCUGUAaaCGGUCCCAUU	0.1	2 ea.	2	1.6	0.17
9	12.8 20.120	PEG–AGGCAUAG UCCGUAUCaaUUCGCACUUC	0.05	2 ea.	0.25	0.83	0.04
10	13.16 21.75	PEG–GUCCAGUAG CAGGUCAUCaaGUUCUCGCAG	0.05	2 ea.	0.08	1.3	0.04

Table 1. Doped selection conditions. For each round of the doped selection, the primer-template used is shown, with primer on top (5'–3') and template on bottom (3'–5'). Primer was linked to pool through either a polyethylene glycol linker ("PEG") or a reverse-DNA linker ("DNA") as illustrated in Fig. 2. Lowercase "a" in template strands indicates the adenine isomer 2-aminopurine. Also listed are the NTP concentrations used in each round (including competitor oligos included to select for polymerization fidelity), the time allowed during the reaction incubation, and the observed percentage of pool adding either one or two Us, with "–" indicating no reaction detected.

taking advantage of the Ban I restriction site that divides the two domains. We mixed together mutagenized pool and unmutagenized pool, digested with Ban I, and re-ligated with DNA ligase. This procedure was expected to yield some molecules with a "wild type" ligase attached to a mutagenized auxiliary domain. Another recombination was performed at the Apo I site within in the auxiliary domain. Re-ligation yields were good for both recombinations, and the recombined pools were combined for round 7. Pool activity, as observed by the percent of pool reacting during the selection incubation, increased during rounds 7–9 and appeared to plateau with round 10 (Table 1). Pool activity was also monitored using the untethered (*trans*) configuration, with no linker connecting the polymerase and primer-template (Fig. 3). In agreement with the tethered (*cis*) observations, there was an abrupt increase in pool activity at round 6, followed by a leveling off of activity during rounds 7–10.

A

Primer-template A   Primer-template D   Primer-template F

B

	Primer-template		
Sample	A	D	F
Pool 3	0.0	0.0	0.0
Pool 4	0.0	0.0	0.0
Pool 5	0.1	0.1	0.0
Pool 6	1.0	1.7	0.2
Pool 7	1.0	1.0	0.6
Pool 8	1.4	1.2	0.9
Pool 9	1.4	1.2	1.0
Pool 10	1.6	1.0	1.1
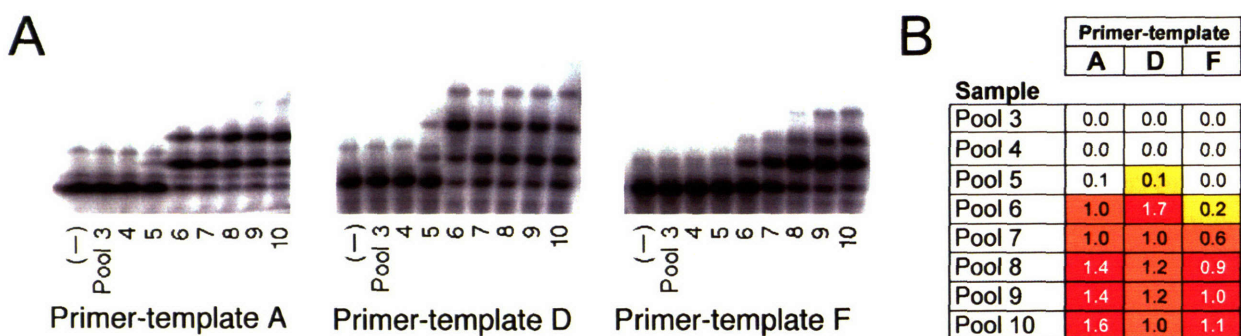
Fig. 3. Progress of the Pol 2 doped selection. **A.** Bulk polymerization assays of pools 3 through 10, using three different primer-templates (see Table 2 for PT sequences). Standard polymerization assays contained 2 μM ribozyme (pool in this case), 2.5 μM P2-completing heptamer, 2.5 μM RT primer, 0.1 μM 5'-radiolabeled polymerization primer, and 0.5 μM polymerization template. RNA was annealed for 2 min at 80°C and allowed to cool to 22°C. Reactions were initiated by addition of 200 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.5), and 4 mM each NTP. Reactions were incubated ~24 hr at 22°C, then quenched by addition of 4 M urea and 200 mM EDTA, separated on 20% PAGE (3 hr at 78W), and visualized by phosphorimaging. Lane (–) is unreacted primer, showing trace amounts of a longer contaminant in the cases of PTs A and F, due to incomplete purification of the radiolabeled primer. **B.** Quantitation of assays shown in A. Average number of nucleotides added per primer was computed from relative band intensities. Results are color-coded to indicate high activity (red), moderate activity (orange), weak activity (yellow), or inactive (white), following the cutoff values in Table 2.

Emergent families. The selection experiment was stopped after round 10, with pool activity having plateaued. At total of 115 clones were isolated and sequenced from rounds 6, 7, 8, and 10. (A note on clone nomenclature: clones from round 6 were named 601, 602, 603, etc.; clones from round 7 were named 701, 702, and so on.) The complete RNA sequences of all clones are listed in the Supplemental Figure.

Greatest sequence diversity was observed at round 6, which was also the point when pool activity first emerged. Eight sequence families (defined by a set of shared mutations) and eight other unique clones were isolated from this pool (Table 3). Family I was the largest constituent, accounting for 17 of the 45 clones. The consensus sequence of Family I was cloned eight times (clone 603 and seven others). Despite the early success of Family I, it was rapidly displaced by

Table 2. Primer-templates (PTs) used in polymerization assays. The sequences of PTs A–F are shown, along with the lab names of these oligos. Primer strand is shown on top, written 5' to 3', with the star indicating the position of the radiolabel, and template strand is shown on bottom, written 3' to 5'. Lowercase "a" in oligo 17.68 represents the adenine isomer 2-aminopurine. Listed at right are the cutoff values used in evaluating polymerase activity. Values express average nucleotides added per primer, and reflect the minimum primer extension required for a given polymerase clone to be considered highly active (red), active (orange), or weakly active (yellow), when assayed with that PT.

primer-template	oligo names	sequence	activity level cutoffs +	++	+++
A	7.5 / 17.68	*AGCUGCC / UCGACGGaaCCUGCGUC	0.1	0.6	1.2
B	7.6 / 18.99	*CUGCCAA / GACGGUUGGCACGCUUCG	0.1	0.6	1.2
C	8.9 / 18.90	*GAAUCAAG / CUUAGUUCCCGCCCGGCC	0.1	0.5	1.0
D	10.7 / 18.90	*GAAUCAAGGG / CUUAGUUCCCGCCCGGCC	0.1	0.8	1.4
E	11.26 / 21.30	*CUGCCAACCGU / GACGGUUGGCACGCUUCGCAG	0.1	0.7	1.2
F	14.19 / 21.30	*CUGCCAACCGUGCG / GACGGUUGGCACGCUUCGCAG	0.1	0.3	0.8

Round	6	7	8	10
Family I	17	1	2	
Family II	3	6	10	21
Family III	2			
Family IV	3			
Family V	3		1	
Family VI	2		1	
Family VII	2		1	
Family VIII	2			
Unique clones	8	3	1	
"Junk"	3	9	8	6
Total	45	19	24	27

Table 3. Summary of families isolated during Pol 2 doped selection. "Junk" clones contained two or more lesions (disrupted base-pairs) in their ligase domains.

Family II (consensus clone 702) over the course of the following rounds. In the round 10 pool, Family II accounted for 21 of the 27 clones, with not a single clone of Family I isolated. Takeovers of this sort have been observed during previous *in vitro* selection experiments[9] (also see Chapter 1) and presumably reflect a response to changing pressures as the experimenter increases the selective stringency (Table 1).

Polymerization assays. Individual clones were tested in *trans* format (untethered primer) in standard polymerization assays, performed as described previously (Chapter 1). A series of different primer-templates (PTs) was used in these assays (Table 2), in order to obtain a more complete picture of activity level than would be possible using only a single PT. We previously measured large differences in polymerization activity resulting from changes in PT sequence (Chapter 2), and we wished to avoid relying exclusively on data from a single PT. The goal of the polymerization assays was to determine the maximum polymerization output of each clone during its lifetime under the reaction conditions. Accordingly, we considered only the endpoint (24-hour timepoint) of the polymerization reactions. High magnesium concentration (200 mM) and high pH (8.5) were chosen to maximize ribozyme activity. Throughout this study, polymerization assays are quantitated by calculating the average number of nucleotides added per primer molecule. For ease of comparison, these measurements are color-coded red (high activity), orange (moderate activity), yellow (weak activity), or white (inactive), according to the fixed cut-off values listed in Table 1. Results of polymerization assays for all clones are listed in the Supplemental Figure.

Examples of standard polymerization assays are shown in Fig. 4. Six clones from the doped selection, representing four different sequence families, were tested alongside the parent ribozyme (clone 11.78). These assays revealed that the doped selection had succeeded in isolating improved polymerization activity. For example, the parent ribozyme had no detectable activity using PTs C or E, whereas doped-selection isolate 818 (representing Family II) showed activity with each of them. A second notable result was the contrasting sequence preferences of different clones. For instance, clone 615 (Family III) showed strong activity with PT C but weak activity with PT E, whereas clone 1018 (Family II) showed the opposite pattern. This phenomenon was observed even among clones of the same family: within Family II, clone 818 was superior to clone 1018 at using PT D, whereas with PT E, the trend was reversed. Also shown for comparison is the activity of Evolved Pol 1, isolated in previous work.[10]

A.

Primer-template C   Primer-template D   Primer-template E

B.

	Primer-template		
	C	D	E
(–)	0.0	0.0	0.0
EvPol1	4.1	2.5	4.1
Parent			
Pol2 (11.78)	0.0	0.8	0.0
Clone **Family**			
615 III	0.7	2.2	0.0
625 III	0.0	0.8	0.1
634 IV	0.0	0.8	0.1
818 II	0.0	1.8	0.6
843 VI	0.0	1.4	1.5
1018 II	0.0	0.7	1.4
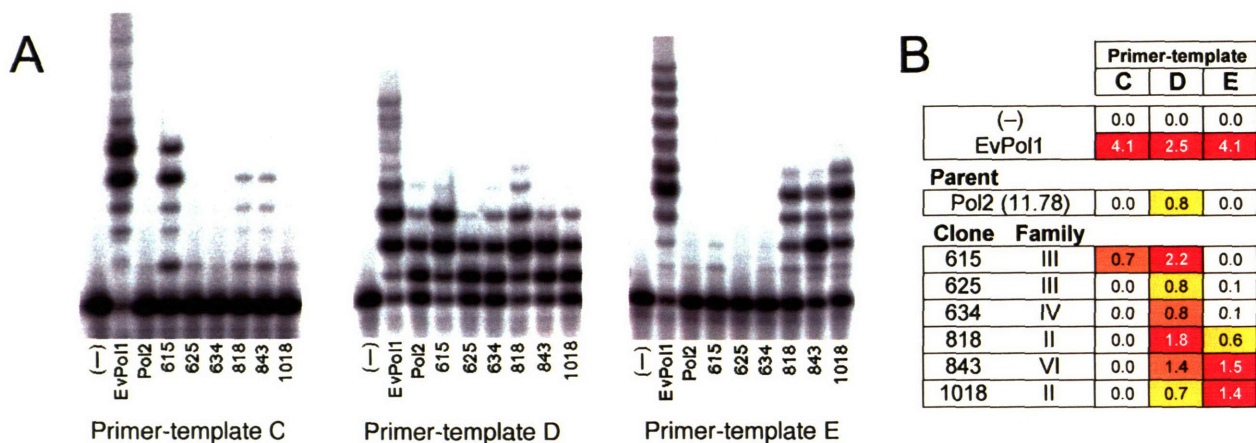
Fig. 4. Polymerization assays of selected clones. **A.** Polymerization assays of Pol 2 (11.78) and six clones from the doped selection, with activity of Evolved Pol 1 (EvPol1) shown for comparison. Assay conditions were as in Fig. 3. Lane (–) is unreacted primer. Three different PTs are used (Table 2). **B.** Assay quantitation. Results are expressed as average nt added per primer, with colors as in Fig. 3.

In previous *in vitro* selection experiments (Chapter 1) that used a starting pool containing random sequence, several families of "parasites" emerged. These families sometimes accounted for the majority of clones isolated from late rounds of the selection, but they were completely inactive in polymerization assays, an observation easily rationalized by their highly disrupted ligase domains. We speculated that these sequences had the fortuitous ability to insert their own primers into the active sites of working polymerases, thereby subverting the selection process and surviving despite their inactivity. In the present experiment, which used a doped pool instead of a random-sequence pool, no such parasite families emerged. Instead, small numbers of "junk" clones were isolated at each round of the selection (Table 3). These clones superficially resembled the parasites from past experiments, insofar as having numerous disrupted ligase base-pairs and no polymerization activity. However, they were all different from each other and did not cluster in families. Furthermore, they did not show the characteristic steady rise in numbers seen with the parasites. Instead, we consider them merely to reflect the level of inactive sequences maintained in the pool due to nonspecific selection background. A sharp rise in the number of these clones was observed at round 7, coinciding with sequence damage due to PCR mutagenesis, but the subsequent rounds efficiently selected them out, and the incidence of cloned junk sequences receded. In the present study, we define junk sequences as those clones containing two or more

ligase lesions (disrupted base-pairs). These clones are marked "J" in the Supplemental Figure, and their number of ligase lesions is indicated under "LL".

Covariations reveal Pol 2 secondary structure. Comparative sequence analysis is a proven technique for obtaining information about ribozyme secondary structure.[10, 11] (We applied this technique to the results of the doped selection in order to deduce the secondary structure of Pol 2. Fig. 5 shows an alignment of the fifteen independent sequence groups that displayed polymerization activity. These include the consensus clones of Families I through VIII, as well as seven additional unique clones. We identified a secondary structure that all fifteen of these sequences were able to adopt; this is the structure shown in Fig. 1 for Pol 2 and the doped pool. It contains 27 base-pairs, 21 of which are conserved in 15/15 sequences (the other six are conserved in 14/15.) Clones from the winning Family II contained four additional base-pairs, bringing their total to 31 (and raising speculation that this may at least in part explain their increased fitness.) Evidence for the proposed Pol 2 secondary structure is provided by the observation of covarying base-pairs: for example, nucleotides 159 and 168 form a base-pair in putative stem P11. This base-pair is conserved in 15/15 independent active sequences, but the identity of the pairing nucleotides is variable; three out of the four possible Watson–Crick base-pairs are observed. Of the 27 base-pairs in the proposed Pol 2 secondary structure, 15 are supported by such covariation. For comparison, the secondary structure of Pol 1 (containing 24 base-pairs) was proposed on the basis of only five covarying base-pairs.[10]
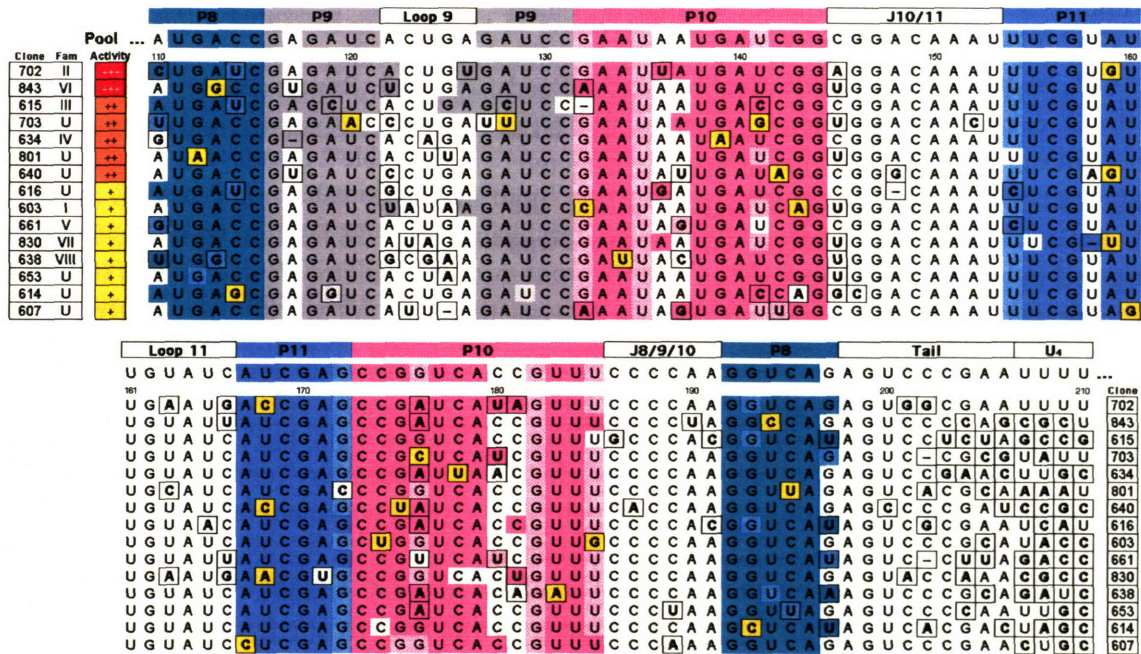
Fig. 5. Covariations reveal base-paired regions of Pol 2. Listed are the 15 independent sequence families that showed polymerization activity (Families I through VIII, plus seven unique clones.) Only the sequence of the auxiliary domain is shown; the ligase domain and fixed-sequence RT primer binding site are omitted (See Supp. Fig. for complete sequences.) Paired regions are shown in color; wobble pairs are indicated by lighter shading. Boxed residues indicate mutations with respect to the parental (pool) sequence. Highlighted in yellow are covariations, where both nucleotides of a base pair mutated in the same clone but preserved Watson-Crick pairing. Clones are listed in order of overall activity (see Supp. Fig. for individual assay results.)

Family II is the best. During the course of the Pol 2 doped selection, Family II emerged as the most successful sequence family, accounting for all active clones from the round 10 pool (Table 3). Polymerization assays revealed that this numerical dominance correlated with robust activity, a gratifying concordance that contrasts with some of our previous results (Chapter 1). Fig. 6 lists fourteen of the most active clones from the doped selection. Two clones from Family II (702 and 1037) stood out sharply, showing polymerization activity with all six assay PTs. Furthermore, clone 702 (the Family II consensus sequence) was better than the parent ribozyme at using all six PTs.

Mutational analysis. We investigated which of the Family II mutations were responsible for the improved polymerization activity (Fig. 7). Of course, the most obvious difference between the doped selection isolates and the parental ribozyme was the appended RT primer-binding site (and the RT primer annealed to it), shown in green in Fig. 7. When this region, plus the U_4 linker and three additional nt, was deleted from the 3' end of the polymerase (mutant 818.6T), polymerization activity was not adversely affected; instead, the polymerase mysteriously acquired the ability to use PT C efficiently. This effect was investigated in greater detail later.

Most of the other mutations in Family II were also shown to be without effect: for example, the two ligase loops could be replaced by stable tetraloops[12] without affecting activity (mutants 818.7 and 818.8) The G199A point mutation that distinguishes clone 818 from clone 702 also had little effect. Repairing the P8 wobble pair (mutant 818.13) had no effect.

	Primer-template					
	A	B	C	D	E	F
Parent						
Pol2 (11.78)	0.7	0.2	0.0	0.7	0.0	0.3
Clone Family						
702 II	1.4	1.4	0.2	1.7	0.3	0.9
1037 II	1.9	0.2	0.1	1.5	3.2	1.2
818 II	1.4	1.9	0.0	1.8	0.6	0.8
713 II	0.7	3.3	0.0	0.9	1.6	1.1
843 VI	1.5	0.1	0.0	1.4	1.5	0.8
1020 II	1.2	1.1	0.0	1.7	0.3	0.8
703 U	1.2	0.2	0.0	1.5	0.0	0.5
615 III	0.4	0.0	0.7	2.2	0.0	0.1
801 U	1.6	0.4	0.0	0.6	0.1	0.2
634 IV	1.2	0.1	0.0	0.8	0.1	0.4
640 U	1.1	0.1	0.0	1.0	0.1	0.1
653 U	1.0	0.1	0.0	0.4	0.1	0.1
614 U	0.9	0.2	0.0	0.3	0.0	0.0
603 I	0.8	0.0	0.0	0.7	0.0	0.1
For comparison:						
Pol 1	0.4	n.d.	1.0	0.5	0.1	0.1
Evolved Pol 1	1.3	4.1	2.5	2.6	2.3	4.1

Fig. 6. Best clones from the Pol 2 doped selection. Polymerization activity of the parental ribozyme and 14 doped-selection isolates, measured using all six PTs from Table 2. Assay conditions were as in Fig. 3. Results are expressed as average nt added per primer, and color-coded using the Table 2 cutoffs. Clones are ordered by approximate overall activity level. Five sequence families and five unique clones are represented. Five different Family II isolates are included in the list, due to their diverse activity patterns. (Same-family isolates with equivalent activity patterns are omitted from the list but can be found in the Supp. Fig.) Activities of Pol 1 and Evolved Pol 1 are shown for comparison. (n.d., not determined.)

Reopening the parental double bulge in P10 had no effect (mutant 818.35). Shortening hairpin P9 by a base-pair and completely changing the sequence of its loop (clone 814) had no effect. This finding, coupled with the observation that Family III has a completely remodeled P9 hairpin (due to an interesting deletion+mutation coincidence—see Fig. 5 alignment and Fig. 16A structure), suggested that the sequence of the P9 hairpin had no bearing upon polymerase function. An attempt to entirely delete the hairpin, replacing nucleotides 116–131 with either CC (mutant 818.18), AA (mutant 818.19), or simply a phosphodiester bond (mutant 818.17), proved hamfisted, with all three changes completely abolishing activity (data not shown).

The increased polymerization activity of Family II was found to result from only a few changes—those highlighted with large pink boxes in Fig. 7. One of them, the mutation of

Fig. 7. Analysis of Family II mutations. **A.** Shown is the structure of clone 818, with magenta nucleotides showing mutations with respect to the parental (pool) sequence. The effect of reversing each mutation individually (arrows) was examined. The four mutations that were shown to have a large effect on polymerization are indicated by large pink boxes. **B.** Polymerization assay results for the reversals shown in A. (Assay conditions and quantitation as in previous figures.) The four reversals that significantly reduced activity are indicated again with pink boxes. (Note: clone 817 is marked by a star to point out that instead of a precise reversal to the parental mismatch, it further mutated the base-pair to a different mismatch, with little effect on activity.) Grey boxes indicate values not determined.

nucleotides 201–202 from CC to GG in the 3' tail of the polymerase, was especially dramatic: reversing this change (mutant 818.T2) caused the loss of all gains from the doped selection. At the other end of the auxiliary domain, two point mutations (U163A and C166G) in Loop 11 were also found to be necessary for the improvement in activity; reversing either of these changes (mutants 818.3 and 818.28, respectively) was enough to erase the gains from the doped selection.

In an effort to improve the activity of Family II, we constructed the mutants shown in Fig. 8; however, deletion of the bulged P9 nucleotide A117 (mutant 818.1), or conversion of various wobble pairs to Watson–Crick pairs (mutants 818.4, 818.33, and 818.34) had no effect on polymerization activity. Next we explored structural minimization via truncation of the 5' leader, as shown in Fig. 9, and we found that the polymerase tolerated deletion of 9 nt there (mutant 818.6). Finally, we demonstrated that the P2-completing heptamer can be incorporated into the contiguous structure of the ribozyme, as shown in Fig. 10.
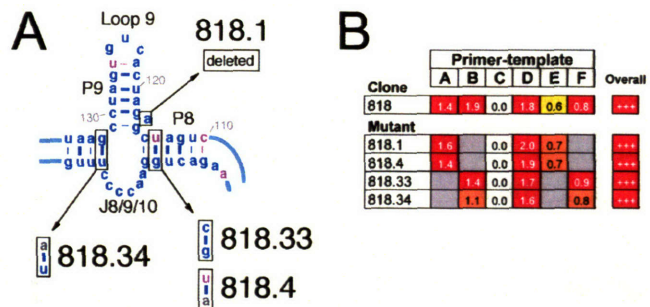
Fig. 8. Attempts to improve polymerization activity by optimizing helical elements. **A.** Region of clone 818 centered on the P9 hairpin, showing four mutations that were tested for potential improvement of polymerization activity. **B.** Results of mutant assays: all of the mutations were neutral.

55
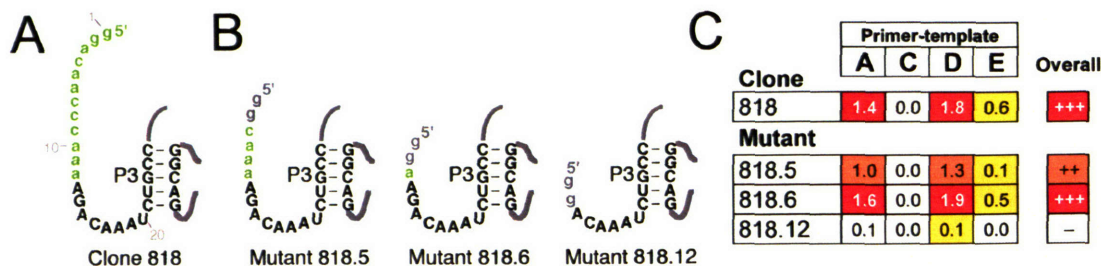
Fig. 9. Truncation of the 5' leader of Pol 2. **A.** Structure of the clone 818 leader sequence (green), with P3 stem shown for orientation. **B.** Progressive truncation of the 5' leader, with initial two Gs (grey) conserved for purposes of *in vitro* transcription. **C.** Polymerization assay results for the ribozymes in A and B. Activity decreased with deletion of 6 nt (mutant 818.5), but was completely restored by deleting 3 additional nt (mutant 818.6). Deletion beyond that point (mutant 818.12) inactivated the polymerase.
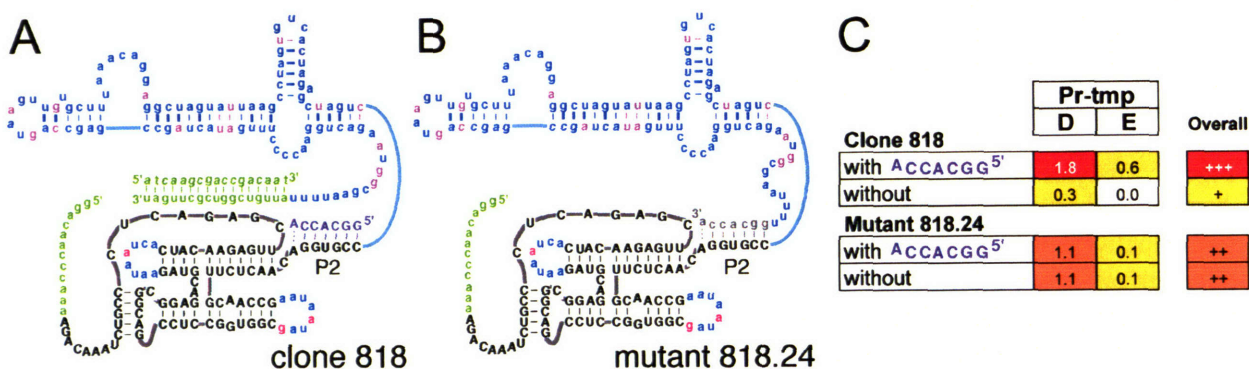
Fig. 10. Converting Pol 2 to single-RNA-chain format. **A.** Structure of clone 818, including the P2-completing heptamer (purple) supplied in polymerization reactions. **B.** Structure of variant 818.24, in which the ribozyme's own 3' terminus (grey) is designed to complete the P2 stem. **B.** Polymerization activity of clone 818 and mutant 818.24, each tested with and without addition of the P2-completing heptamer (purple). Clone 818 required the heptamer, with activity severely reduced in its absence. Clone 818.24, which contains both sides of the P2 stem within its own sequence, showed polymerization activity independent of the exogenous heptamer.
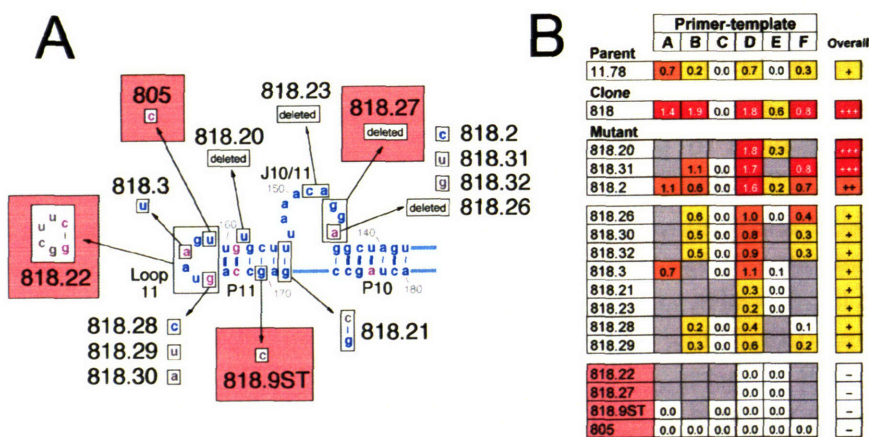
Fig. 11. Mutations in the vicinity of P11 have dramatic effects. **A.** Shown is a region of clone 818 (nt 138–180), with magenta nucleotides indicating mutations that occurred during the selection, and grey nucleotides indicating site-directed mutations. Four mutations found to completely inactive the ribozyme are indicated by large pink boxes. **B.** Polymerization assay results for the mutations shown in A. Grey boxes indicate values not determined.

Critical structural elements. The foregoing mutational analysis identified the P11 hairpin region as a critical element of the Pol 2 structure. Additional evidence came from the mutants shown in Fig. 11: for instance, replacement of Loop 11 by a stable tetraloop (mutant 818.22) inactivated the ribozyme. Even more strikingly, the single point mutation U161C (clone 805) was sufficient for this inactivation, suggesting that Loop 11 plays an important role in the polymerization reaction, and that the loop geometry may be distorted by the addition of an extra P11 base-pair. Mutating the Loop 11 nucleotide G166 to any of the other three possible bases also severely impaired activity (mutants 818.28, 818.29, and 818.30). Disruption of a P11 base-pair by the mutation G170C (mutant 818.9ST) caused complete inactivation, as did conversion of the wobble pair U154:G172 to a C:G Watson–Crick pairing (mutant 818.21). Interestingly, deletion of the P11 bulged nucleotide U158 (mutant 818.20) had no effect on activity. The single-stranded J10/11 region was intolerant of deletions (mutants 818.23, 818.26, and 818.27); however, at the end farthest from P11, nucleotide A145 was partially tolerant of mutation, with activity unaffected by mutating it to U.

The analysis in Fig. 7 identified a second region of the Pol 2 structure that was particularly sensitive to mutation: the 3' tail, lying between the P8 stem and the RT primer-binding site. This sensitivity was next investigated in greater depth. It was noted earlier that some of the Family II members had sharply different patterns of activity with different PTs. Fig. 12 lists seven isolates from Family II and their activity with the six assay PTs (Table 2). The chief differences between these clones were their tail sequences, and with two of the PTs (B and E), the activity of these clones was found to depend strongly on their tail sequences. For instance, the tail mutation A205U, which occurred in clones 1017 and 1037 (the latter as a consequence of a 5-nt deletion), causes a loss of activity with PT B and gain of activity with PT E. The effect of this mutation was confirmed in isolation (mutant 814.1). The tail mutation G199A, very common in Family II clones, was found to slightly elevate activity with both PTs B and E, while the tail mutation G201A, which occurred in isolation in clone 713, caused a pronounced improvement with both PTs.

	Primer-templates						P8 + Tail + U₄ (nt 192–210)	RT pb site (220–227)
	Tail-independent				Tail-dependent			
	A	D	F	C	B	E		
Family II consensus clone							192 194 196 198 200 202 204 206 208 210	220 227
702	1.4	1.7	0.9	0.2	1.4	0.3	G G U C A G A G U G G C G A A U U U U	AUUGUCGGUCGCUUGAU
Other Family II clones								
820	1.4	1.7	1.0	0.2	1.3	0.3	G G U C A G A G U G G C G A A U U U U	AUUGUCGGUCGCUUGAU
818	1.4	1.8	0.8	0.0	1.9	0.6	G G U C A G A [A] U G G C G A A U U U U	AUUGUCGGUCGCUUGAU
814	1.4	1.8	1.0	0.0	1.5	0.5	G G U C A G A [A] U G G C G A A U U U U	AUUGUCGGUCGCUUGAU
1037	1.9	1.5	1.2	0.1	0.2	3.2	G G U C A G A G U G G C G [U] – – – – –	AUUGUCGGUCGCUUGAU
1017	1.7	1.3	0.8	0.0	0.1	2.0	G G U C A G A [A] U G G C G [U] A U U U U	AUUGUCGGUCGCUUGAU
713	0.7	0.9	1.1	0.0	3.3	1.6	G G U C A G A G U [A] G C G A A U U U U	AUUGUCGGUCGCUUGAU
Clone 814 and mutants								
814	1.4	1.8	1.0	0.0	1.5	0.5	G G U C A G A [A] U G G C G A A U U U U	AUUGUCGGUCGCUUGAU
814.1	2.0	1.9	1.3	0.1	0.1	3.7	G G U C A G A [A] U G G C G [U] – – – – –	AUUGUCGGUCGCUUGAU
814.2	2.0	1.8	1.8	0.0	0.7	4.4	G G U C A G A G U [A] G C G [U] – – – – –	AUUGUCGGUCGCUUGAU

Fig. 12. Importance of the 3' tail sequence. Shown are polymerization assay results and 3'-terminal ribozyme sequences (nt 192–227, with P8 half-stem included for orientation) of several Family II variants differing chiefly in the identity of their 3' tails. Boxed residues indicate differences from the Family II consensus sequence (clone 702). Activity with four of the six assay PTs (A, D, F, C) is largely independent of tail sequence, whereas activity with the other two (B, E) shows a strong dependence on tail sequence; deletion of nt 205–209 is sufficient to reverse the preferences of clone 814, yielding those of mutant 814.1.

B

primer 5'-CUGCCAA-3'
template 3'-GACGGUUGGCACGCUUCG-5'

Tail | U₄ → Tail | U4

Clone	B	198	200	202	204	206	208	210
713	3.3	A G U	A G C G A A	U U U U
1018	2.6	A G U	A G C G A A	U U U U
1032	2.0	A G U	A G C G A A	U U U U
1026	1.9	A G U	A G C G A A	U U U U
649	1.9	A G U	G G C G A A	U U U U
818	1.9	A A U	G G C G A A	U U U U
814	1.5	A A U	G G C G A A	U U U U
702	1.4	A G U	G G C G A A	U U U U
820	1.3	A G U	G G C G A A	U U U U
833	1.1	A G U	G G C G A A	U U U U
1020	1.1	A A U	G G C G A A	U U U U
631	1.0	A G U	G G C G A A	U U U U
817	1.0	A G U	G G C G A A	U U U U
814.2	0.7	A G U	A G C G U	- - - - -
628	0.6	A G U	G G C G A A	U U U U
1037.1	0.6	A G U	A G C G U	- - - - -
1027	0.4	A G A	G G C G A A	U U U U
1037	0.2	A G U	G G C G U	- - - - -
814.1	0.1	A A U	G G C G U	- - - - -
1037.2	0.1	A A U	G G C G U	- - - - -
1017	0.1	A A U	G G C G U	A U U U U

E

primer 5'-CUGCCAACCGU-3'
template 3'-GACGGUUGGCACGCUUCGCAG-5'

Tail | U₄

Clone	E	198	200	202	204	206	208	210
814.2	4.4	A G U	A G C G U	- - - - -
1037.1	3.9	A G U	A G C G U	- - - - -
814.1	3.7	A A U	G G C G U	- - - - -
1037.2	3.4	A A U	G G C G U	- - - - -
1037	3.2	A G U	G G C G U	- - - - -
1017	2.0	A A U	G G C G U	A U U U U
713	1.6	A G U	A G C G A A	U U U U
1018	1.4	A G U	A G C G A A	U U U U
1032	0.9	A G U	A G C G A A	U U U U
1026	0.8	A G U	A G C G A A	U U U U
818	0.6	A A U	G G C G A A	U U U U
814	0.5	A A U	G G C G A A	U U U U
1027	0.5	A G A	G G C G A A	U U U U
817	0.4	A G U	G G C G A A	U U U U
820	0.3	A G U	G G C G A A	U U U U
702	0.3	A G U	G G C G A A	U U U U
1020	0.3	A A U	G G C G A A	U U U U
833	0.2	A G U	G G C G A A	U U U U
631	0.2	A G U	G G C G A A	U U U U
649	0.1	A G U	G G C G A A	U U U U
628	0.1	A G U	G G C G A A	U U U U

Fig. 13. Potential base-pairing between templates and ribozyme tails. Each panel shows the same list of twenty-one Family II variants and their tail sequences. Boxed nucleotides indicate differences from the Family II consensus (clone 702). In panel **B**, Clones are listed in order of activity with PT B. Template B contains the sequence 3'-CGCUU-5' in its single-stranded region (pink). The most active clones all contain the complementary sequence 5'-GCGAA-3' (pink) in their tails; the potential pairing to the template is indicated with vertical dashes. In panel **E**, Clones are listed in order of activity with PT E. Template E contains the sequence 3'-UCGCA-5' (aqua). The most active clones contain the complementary sequence 5'-AGCGU-3' (aqua) in their tails; the potential pairing is indicated as in B. The next-most-active clones have a wobble pair (lighter shading).

Prehensile tails. In some cases, these PT-specific tail effects could be rationalized by base-pairing between the polymerization template and the ribozyme tail, as illustrated in Fig. 13. With each of the two tail-dependent PTs (B and E), the most active clones contained in their tails a 5-nt stretch (pink or aqua nucleotides) complementary to the single-stranded region of the polymerization template, in principle allowing them to adhere to the template through base-pairing. Notably, both of the potential pairings are located 4 nt upstream of the primer; perhaps this pairing register allows the polymerase to position the primer-template adjacent to its active site.

As mentioned above, clone 713 has a tail sequence that increases its activity with both PTs; this is reflected in Fig. 13E, where clone 713 is the top clone without the aqua binding sequence. The tail sequence of clone 713 can be viewed as a hybrid between the pink and aqua tails, affording it the ability to base-pair to either template. It should be noted that PTs B and E are actually related to each other in sequence (primer E is primer B plus 4 nt, and template E is template B plus 3 nt). Fig. 14 shows the results of polymerization assays using a PT constructed by pairing primer B with template E. This PT (named 7.6/21.30) was previously shown to be fully extended by Evolved Pol 1 (Johnston et. al 2001), when using 5 µM polymerase in the polymerization reaction. However, all of the polymerization assays in the present report used 2 µM ribozyme; under these conditions, extension by Evolved Pol 1 was detected only to 10 or 12 nt. (Fig. 14A, first set of lanes). Extension by Pol 2 was extremely weak, with only a hint of the first extension product seen after an 18-hour incubation. In sharp contrast, two of the clones from the doped selection showed robust activity with this PT. Clone 702 (the Family II consensus sequence) added 4 nt during
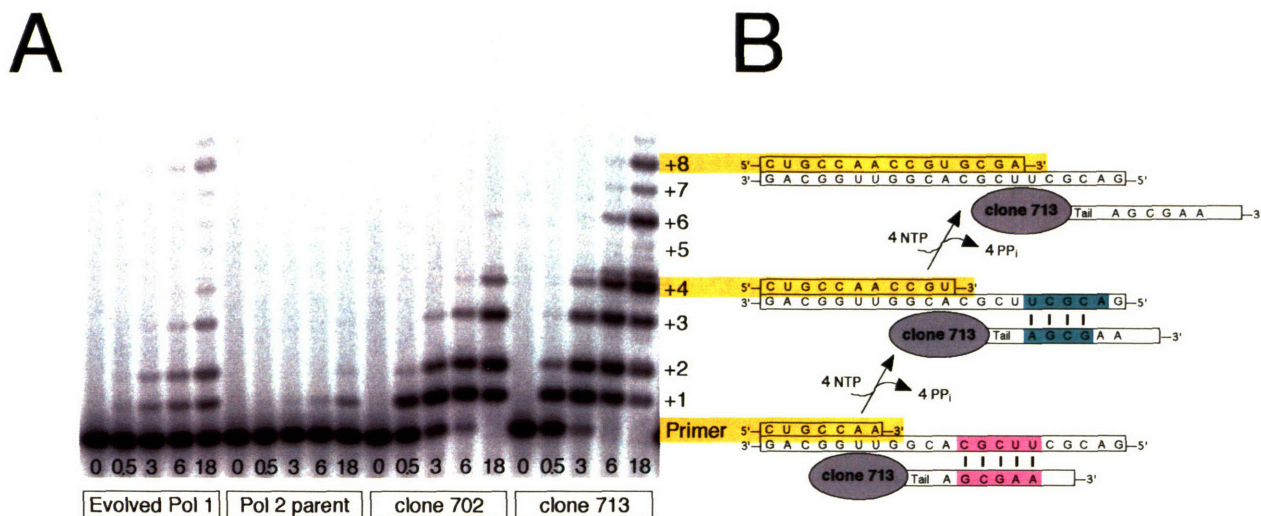
Fig. 14. Polymerization with a template coding for 14 nucleotides. **A.** Polymerization timecourses using PT 7.6/21.30, comparing the activity of Evolved Pol 1 with three variants of Pol 2: clone 11.78 (the Pol 2 parent) and the two doped-selection Family II isolates 702 and 713. Clone 713 shows especially robust polymerization using this PT. **B.** Possible mechanism of clone 713 polymerization using PT 7.6/21.30 (starting sequence shown at bottom). Clone 713 can anneal to this template with its pink tail sequence, as shown in Fig. 13. After extension by 4 nt, the PT becomes identical to PT E (Table 2), and clone 713 then can shift to a new tail pairing register, making four of the five base-pairs in the aqua region identified in Fig. 13.

robust activity with this PT. Clone 702 (the Family II consensus sequence) added 4 nt during this time period, potentially employing its "pink" tail sequence (Fig. 13B) in order to base-pair to the template as shown in Fig. 13. However, no extension beyond 4 nt was detected using this clone. This observation can be rationalized by noting that the 4-nt extension product in question is identical to PT E (Table 2), a poor substrate for clone 702 (Fig. 13E). Clone 713 (which differs from clone 702 by single point mutation in the tail) also showed strong polymerization of 4 nt during the first 3 hours of the polymerization reaction (Fig. 14A, rightmost set of lanes); however, during the next 15 hours, clone 713 surpassed clone 702, adding four additional nucleotides. A possible mechanism is suggested in Fig. 14B: during the first four additions, clone 713 base-pairs to the template using its "pink" pairing, but after that it shifts registers to make way for further primer extension, switching over to its partial (4/5) "aqua" pairing during addition of the next several nucleotides. In this manner, by successively base-pairing to the template in two different registers, clone 713 is able to outperform the other members of Family II (and Evolved Pol 1) with this PT.

Evolved Pol 1 also has template 21.30 as one of its "favorites"; this is the template that supported polymerization of 14 nt (see Appendix), a feat so far unexcelled by any polymerase ribozyme. We noted that Evolved Pol 1 has the potential to use its 3' tail to base-pair to template 21.30, forming the eight base-pairs shown in Fig. 15A; this pairing might in part explain the exceptional activity seen with that template. However, when the 3' tail was deleted from Evolved Pol 1, eliminating the possibility of this base-pairing to the template (variant 8.12.13, Fig. 15B), polymerization of 14 nt was still detectable, albeit greatly reduced (Fig. 15C). The deleterious effect of the 3'-truncation seemed milder with two other PTs (Fig. 15C). Thus, while Evolved Pol has the potential to pair with some templates, this pairing is not required for activity.
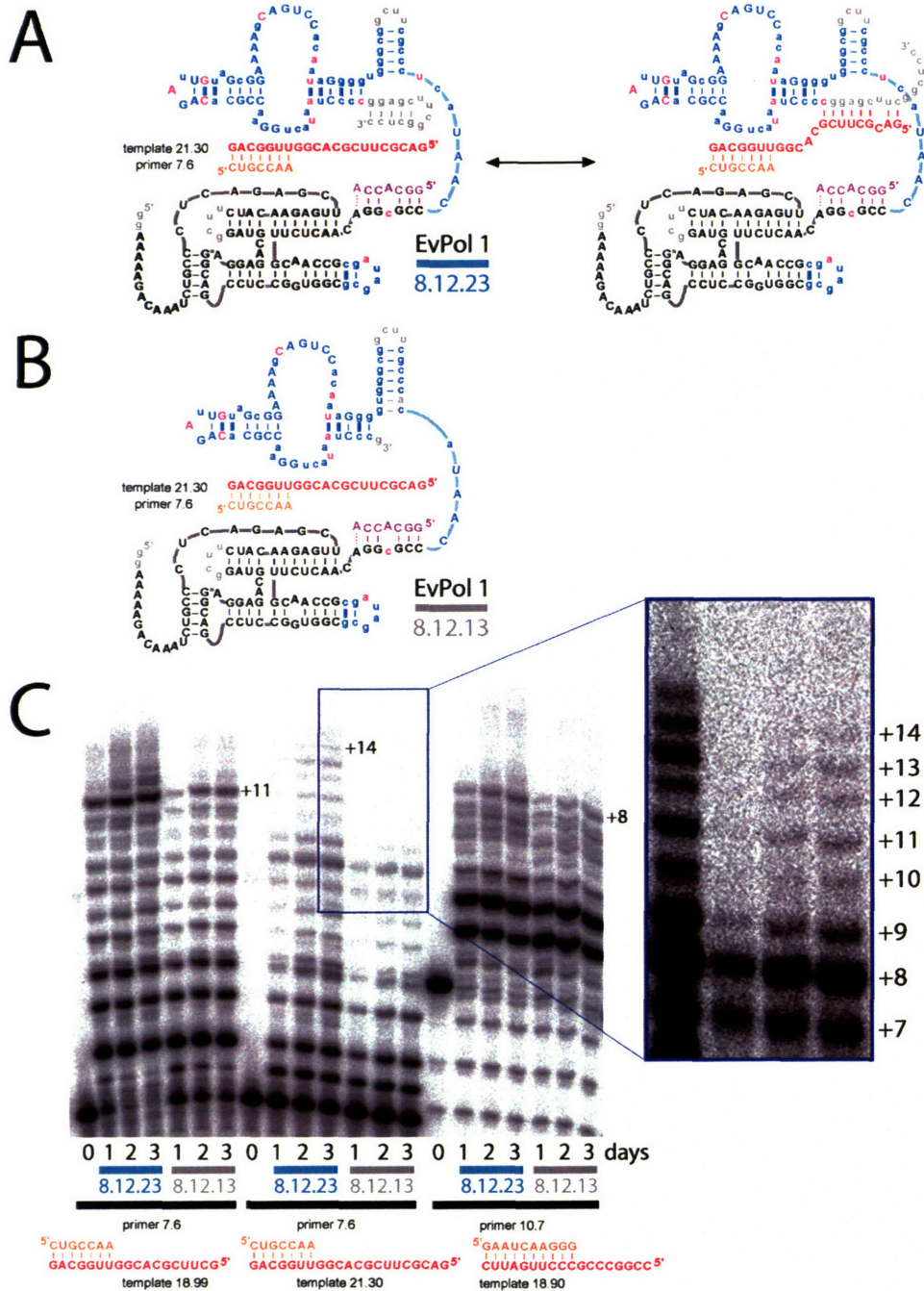
Fig. 15. Evolved Pol 1 *can* base-pair to some templates, but it doesn't *need* to. **A.** Potential base-pairing between Evolved Pol 1 and its "favorite" template, 21.30. The ribozyme 3' terminus (grey) had been designed to fold into a 5-bp hairpin (see Appendix), but if it instead paired with the polymerization template (red dashes), that might in part explain its particularly robust polymerization using that template (up to 14 nt, the maximum seen so far). **B.** Truncated variant 8.12.13 cannot form this pairing to template 21.30. **C.** Polymerization assays (under standard conditions, with ribozyme concentration increased to 5 μM.) Evolved Pol 1 (8.12.23) and truncated variant (8.12.13) were compared using three different PTs (sequences shown at bottom.) Polymerization reaction timepoints were taken at 1, 2, and 3 days. Both ribozymes were able to fully extend all three PTs, but the fraction of fully-extended PT was lower using ribozyme 8.12.13, especially with template 21.30 (inset shows longer gel exposure).

60
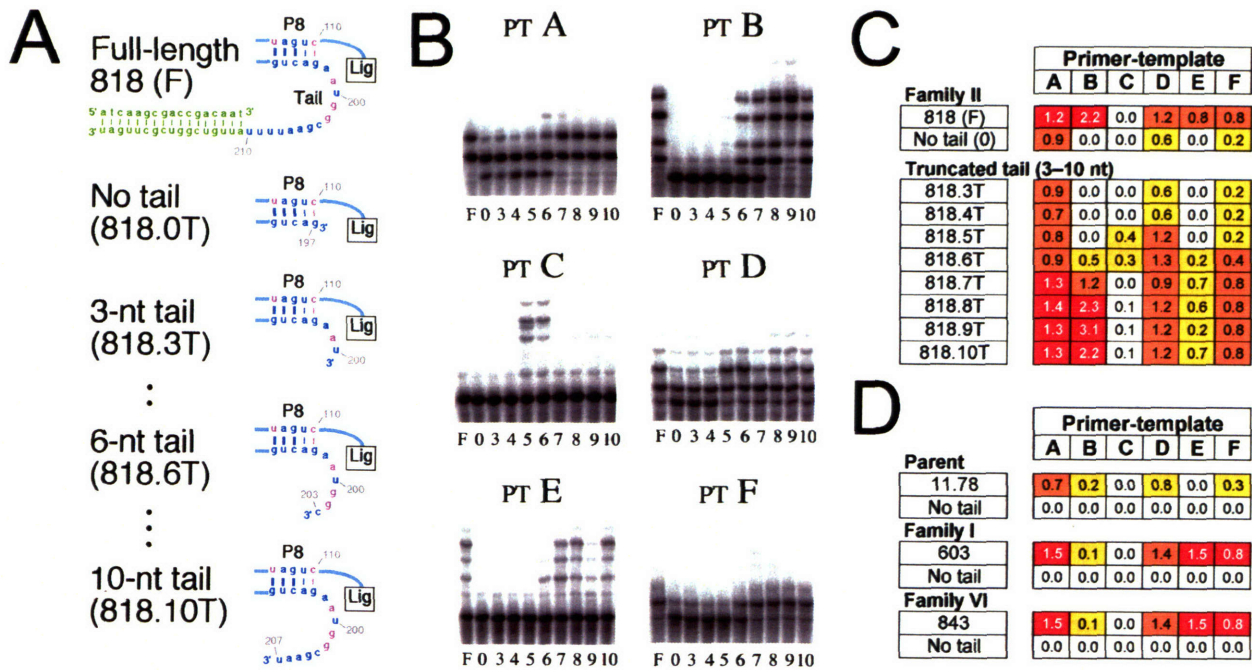
Fig. 16. Tail truncation analysis. **A.** Partial structure of clone 818 (Family II), with P8 stem shown for orientation. Full-length clone 818 has a 13-nt tail (including U_4 linker), followed by the RT primer/binding site duplex (green). Deleted-tail mutant 818.0T terminates with the last nucleotide of the P8 helix (G197). Truncated-tail mutants 818.3T–818.10T have tails between 3- and 10-nt long, as indicated. **B.** Primer extension gels showing activity of clone 818 and the tail-truncation series. **C.** Quantitation of gels in B. Clone 818 was significantly impaired without its tail but still showed activity with some PTs. Restoring 3–10 nt of the tail gradually returned activity to the level of the full-length clone. Intriguingly, clone 818 showed activity with PT C only when its tail was exactly 5 or 6 nt in length. **D.** Results of complete tail deletion of the parental Pol 2 sequence (clone 11.78) and representatives of two other doped-selection families (clones 603 and 843). In all three cases, tail deletion completely abolished activity.

Fig. 16C

Family II

	A	B	C	D	E	F
818 (F)	1.2	2.2	0.0	1.2	0.8	0.8
No tail (0)	0.9	0.0	0.0	0.6	0.0	0.2

Truncated tail (3–10 nt)

	A	B	C	D	E	F
818.3T	0.9	0.0	0.0	0.6	0.0	0.2
818.4T	0.7	0.0	0.0	0.6	0.0	0.2
818.5T	0.8	0.0	0.4	1.2	0.0	0.2
818.6T	0.9	0.5	0.3	1.3	0.2	0.4
818.7T	1.3	1.2	0.0	0.9	0.7	0.8
818.8T	1.4	2.3	0.1	1.2	0.6	0.8
818.9T	1.3	3.1	0.1	1.2	0.2	0.8
818.10T	1.3	2.2	0.1	1.2	0.7	0.8

Fig. 16D

Parent

	A	B	C	D	E	F
11.78	0.7	0.2	0.0	0.8	0.0	0.3
No tail	0.0	0.0	0.0	0.0	0.0	0.0

Family I

	A	B	C	D	E	F
603	1.5	0.1	0.0	1.4	1.5	0.8
No tail	0.0	0.0	0.0	0.0	0.0	0.0

Family VI

	A	B	C	D	E	F
843	1.5	0.1	0.0	1.4	1.5	0.8
No tail	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 17. Tail-swap experiment. **A.** Structure of active doped-selection clones 818 (Family II) and 615 (Family III), with mutations from the parental sequence indicated in magenta. Tails are boxed, including U_4 linker and RT primer/binding site (green). Tails were swapped (arrows) between the two clones, yielding mutants 818.T1 and 615.T1. **B.** Polymerization assay results for clones 615 and 818 and their tail-swapped mutants: the tail swap completely abolished the activity of clone 615 and severely reduced that of 818. **C.** Results of a similar tail-swap experiment between the Pol 2 parent (clone 11.78) and the Family II consensus (702).

Fig. 17B

Clone	Family	A	B	C	D	E	F
818	II	1.4	1.9	0.0	1.8	0.6	0.8
615	III	0.4	0.0	0.7	2.2	0.0	0.1
818.T1	tail-swap	0.2	0.0	0.1	1.4	0.0	0.1
615.T1	tail-swap	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 17C

Clone	Family	A	B	C	D	E	F
11.78	(parent)	0.7	0.2	0.0	0.7	0.0	0.3
702	II	1.4	1.4	0.2	1.7	0.3	0.9
11.78/T702	tail-swap	0.0	0.9	0.0	0.1	0.0	0.1
702/T11.78	tail-swap	0.2	0.0	0.0	0.2	0.0	0.1

Other tail effects. Similarly with Pol 2, not all of the PT-specific tail effects could be rationalized by specific base-pairing. For instance, as noted above, 818 mysteriously acquired the ability to extend PT C when its RT primer-binding site was deleted (mutant 818.6T). We probed this effect more closely by constructing a single-nucleotide-resolution tail deletion series (Fig. 16). The analysis revealed that clone 818 was active with PT C only when its tail was exactly 5 or 6 nt in length. The other PTs showed a simpler dependence on tail length, with clone 818 showing full activity until its tail was truncated to 7 nt or shorter. With its tail fully deleted (mutant 818.0T), clone 818 showed about the level of the parent ribozyme (clone 11.78). In contrast, the parent ribozyme was completely inactivated by removal of its tail (Fig. 16D), as were clones 603 and 843, representing two other families from the doped selection.

The ability of mutant 818.6T to use PT C was of special interest because most Pol 2 variants had trouble using that PT as a substrate. The only other exception was clone 615, representing Family III (Fig. 4). The structures of clones 615 and 818 are compared in Fig. 17. Notably, they have no mutations in common within the polymerase auxiliary domain. We wondered what properties of clone 615 conferred its exceptional ability to use PT C. One possible explanation was that its remodeled P9 hairpin was somehow involved. Another explanation, suggested by the pattern of tail-template pairing established for some of the Family II variants, was that clone 615 was using its tail to base-pair to PT C, perhaps by using the sequence GCCG just before the RT primer-binding site. In order to test this hypothesis, we swapped the tails of clones 818 and 615, as shown in Fig. 17. However, the predictions of this simple model were not borne out experimentally; the tail swap failed to transfer appreciable PT C activity to clone 818. Instead, it drastically reduced the clone's activity with other PTs (mutant 818.T1). Even more strikingly, clone 615 was completely inactivated by the tail swap (mutant 615.T1). We performed an analogous tail-swap experiment between the Pol 2 parent (clone 11.78) and clone 702 (the Family II consensus). When the Family II tail (nt 197–206) was transplanted to the Pol 2 parent (mutant "11.78/T702"), it abolished activity with all PTs except B, which saw a modest increase (Fig. 17C). Similarly, in the reciprocal experiment, when the parental tail sequence was transplanted to clone 702 (mutant "702/T11.78"), activity was almost completely abolished. In summary, the effect of swapping tails between Pol 2 variants was generally deleterious.

Pol 2 is almost completely nonprocessive. None of Pol 2 variants isolated in the doped selection was as efficient as Evolved Pol 1 (Fig. 4). To begin to understand why, we examined the polymerization kinetics and processivity of Pol 2, an experiment analogous to our previous study of Evolved Pol 1 (Chapter 2). First, we needed to select a model clone from among the best isolates of the doped selection. Family II was the most successful family, and the consensus sequence of this family (clone 702) showed improved polymerization activity with all six assay PTs. However, as discussed above, clone 713, with its tail point mutation, catalyzed addition of 8 nt to the PT shown in Fig. 14, a property that made it amenable to the processivity analysis described in Chapter 2. In recognition of its improvement over the parental Pol 2 sequence, we renamed clone 713 as Pol 2+ and refer to it by this designation henceforth.

Using Template B and a series of primers 7–14 nt in length (Table 4), we tested Pol 2+ in standard polymerization assays and measured its initial rates of first-nucleotide addition (k_{obs}) as described in Chapter 2. These rates are shown in Table 4. Pol 2+ exhibited wide variation in rate, depending on the length of the starting primer, similar to that seen previously with Evolved Pol 1. Its slowest rate was with PT 4, with no extension detected within 2 hours. Its fastest rate was with

Primer-template		Evolved Pol 1		Pol 2+	
		k_{obs} (hr^{-1})	P	k_{obs} (hr^{-1})	P
PTB	5'CUGCCAA 3'GACGGUUGGCACGCUUCG	0.26	n.d.	1.4	n.d.
PT1	5'CUGCCAAC 3'GACGGUUGGCACGCUUCG	1.4	n.d.	0.07	0.2
PT2	5'CUGCCAACC 3'GACGGUUGGCACGCUUCG	0.09	0.03	0.18	0.00
PT3	5'CUGCCAACCG 3'GACGGUUGGCACGCUUCG	0.56	0.15	0.16	0.00
PT4	5'CUGCCAACCGU 3'GACGGUUGGCACGCUUCG	0.16	0.45	<0.004	n.d.
PT5	5'CUGCCAACCGUG 3'GACGGUUGGCACGCUUCG	1.3	0.07	4.5	n.d.
PT6	5'CUGCCAACCGUGC 3'GACGGUUGGCACGCUUCG	1.4	0.25	0.06	0.02
PT7	5'CUGCCAACCGUGCG 3'GACGGUUGGCACGCUUCG	0.41	0.02	0.13	0.02

Table 4. Comparison of Evolved Pol 1 and Pol 2+ kinetics and processivity. The series of primer-templates (PTs) is the same as that in Chapter 2, supplemented by PT B. Each PT has the same template, and a primer one nucleotide longer than the preceding one. Initial extension rates (k_{obs}) and processivity coefficients (P) were measured as in Chapter Two. Values measured for Evolved Pol 1 were determined previously; values for Pol 2+ (clone 713) were determined in the present study. Parameters marked "n.d." were not determined. Extension of PT 4 by Pol 2+ was too slow to be observed during the 2-hr timecourse of the experiment; only an upper bound was determined.

PT 5, which it extended at an initial rate of 4.5 hr^{-1}. In general, Pol 2+ showed a similar range of rates to Evolved Pol 1; in some cases Pol 2+ was faster and in others slower.

However, when the processivity of Pol 2+ was compared to that of Evolved Pol 1, clearer differences became evident: Pol 2+ showed almost no detectable processivity. In only one instance did its processivity coefficient P exceed 0.02. This was with PT 1, where it had a P value of 0.2, indicating a 20% chance of adding the second nucleotide before falling off. Evolved Pol 1, in contrast, had shown processivity as high as 0.45 (using the same template), indicating roughly equal odds at that position of adding a second nucleotide or falling off the primer-template after the first extension.

Discussion

Evolved Pol 1 is still the best. Although the Pol 2 doped selection succeeded in isolating clones such as Pol 2+ with activity improved relative to that of the parent, it did not raise Pol 2 to the level of polymerization activity displayed by Evolved Pol 1, which remained the most active polymerase developed so far in the lab. The low processivity of Pol 2 (relative Pol 1) may in part explain this disparity.

Virtues masked by magnesium? Many of the mutations that changed Pol 2 into Pol 2+ could be reversed with impunity, perhaps most notably the remarkable sealing of the P10 double bulge. This finding cast doubt on the speculation that the increase in base-pairing (from 27 to 31 bp) in the Family II isolates was responsible for their increased polymerization activity. It is possible, however, that mutations such as these exerted a greater effect under the selection conditions (60 mM MgCl$_2$), whereas the effect was mostly masked under the assay conditions (200 mM MgCl$_2$).

Results of previous selections using the Class I Ligase are consistent with this idea: in one case, the improvements from the selection manifested only at lowered magnesium concentrations.[13] In an initial examination of this possibility in the case of Pol 2, we re-measured the activity of Pol 2+ and several of its key revertant mutants in polymerization assays (data not shown) where $MgCl_2$ was decreased to 20 mM. However, this decrease proved too drastic, and none of the polymerases showed even a glimmer of activity, despite the use of PT C, a consistently "easy" PT. Even Evolved Pol 1 was completely inactive at this Mg concentration.

Similar PTs not as similar as they might seem. The close sequence similarity of PTs B and E might have suggested at the outset that using of both of them as assay PTs would be redundant; however, the opposite tendency was observed, with many of the polymerase clones showing a pronounced preference for one or the other. This underscores the utility of examining the successive steps of a single polymerization reaction, because each new addition highlights a unique sequence context. This insight has figured prominently in studies of the processivity[14] and substrate binding[15] of Evolved Pol 1.

A tale of tail tailoring. The clearest result of the Pol 2 doped selection was the evolution of the tail sequence. By the end of the selection (round 10), the pool was composed almost entirely of Pol 2+ variants (Family II) which differed primarily in the contents of their tails. In some cases, we found evidence that the polymerase was using its tail to base-pair to the single-stranded portion of the PT. Interestingly, the most convincing example of such tail-template pairing was the case of PTs B and E, the sequences of which the polymerase had never seen during its *in vitro* selection! These observations together suggest an evolutionary model in which Pol 2+ won out by developing a tolerance for genetic diversity in its tail (where reverse transcription is highly error-prone), allowing at least some members of the family to base-pair to the template and extend their primers at each round of selection, later to have their tail diversity regenerated by random errors during amplification. This storehouse of diversity in the tail can be thought of as a kind of pre-adaptation, for instance in the case of PTs B and E, whose sequences the polymerase did not encounter during *in vitro* selection.

It should not come as too much of a surprise that the polymerase is taking advantage of base-pairing to the template; that's what it was born to do.[9,16] It looks as though the polymerase is reverting to its old ways. It will be useful in designing future selections to consider how this tendency can be discouraged.

Although some of the tail effects could be rationalized by base-pairing to the template, other effects resisted simple explanation; for example, the disastrous consequences of the tail-swap between clones 615 and 818, which seemed to rule out a simple role for the tails in that instance, suggesting instead that the tails might have co-evolved with another portion of the polymerase, perhaps forming a tertiary contact, an interaction that the tail swap disrupted. Further experiments would be required to confirm this possibility.

Interestingly, this is not the first time that tail effects have been the dominant result of a selection involving the Class I Ligase. When the ligase was subjected to a continuous evolution protocol, the most salient change to its structure was mutation and enlargement of the 3' tail.[17] In that experiment, the tail expansion might have served to delay inactivation of the ribozyme by reverse transcription, thereby conferring a marginal kinetic advantage in the continuous evolution scheme. In our experiment, the tail mutation seemed to fulfill a different function, namely the

maintenance of a collection of potential base-pairing elements for use in extending whatever PTs the polymerase encountered.

The ultimate goal of the polymerase selection project is the production of a general RNA replicase, capable of copying any RNA template. This goal has typically been construed as a prohibition on template binding via base-pairing. However, this restriction is not logically necessary: if the ribozyme had a diverse enough collection of potential pairing sites that allowed it to base-pair to any primer-template and thereby use it as a polymerization substrate, then the ribozyme would still meet all the formal requirements of a general polymerase.

References

1.   Rich, A. (1962) "On the problems of evolution and biochemical information transfer," in *Horizons in Biochemistry*, (M. Kasha and B. Pullman, Eds.), pp. 103–26, Academic Press, New York.

2.   Crick, F. H. C. (1968) The origin of the genetic code. *J. Mol. Biol.* **38**:367–79.

3.   Woese, C. R. (1967) *The Origins of the Genetic Code*, Harper & Row, New York.

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

5.   Bartel, D. P. (1999) "Re-creating an RNA Replicase," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., Cold Spring Harbor Laboratory Press, New York.

6.   McGinness, K. E. and Joyce, G. F. (2003) In search of an RNA replicase ribozyme. *Chem & Biol* **10**:5–14.

7.   Moore, M. J. and Sharp, P. A. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**:992–7.

8.   Cadwell, R. C. and Joyce, G. F. (1992) "Randomization of genes by PCR mutagenesis," in *PCR Methods and Applications*, pp. 28–33, Cold Spring Harbor Laboratory Press, New York.

9.   Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261**:1411–8.

10.   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. *Science* **292**:19–25.

11.   Ekland, E. H. and Bartel, D. P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**:3231–8.

12.   Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA Polymerase. *Science* **249**:505–10.

13.   Bergman, N. H. (2001) The reaction kinetics and three-dimensional architecture of a catalytic RNA, *Ph. D. thesis* in Biology, MIT, Cambridge, MA.

14.   Lawrence, M. S. and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* **42**:8748–55.

15.   Müller, U. F. and Bartel, D. P. (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem & Biol* **10**:799–806.

16.   Ekland, E. H. and Bartel, D. P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**:373–6.

17.   Wright, M. C. and Joyce, G. F. (1997) Continuous in vitro evolution of catalytic function. *Science* **276**:614–7.

Supplemental Figure. (Next three pages.) Polymerization assay results and polymerase sequences of all Pol 2 relatives. **A.** Sequence of the $Lig+N_{76}$ pool, with clones from round 11 listed below in order from least active to most active. Boxed residues in clone sequences indicate mutations with respect to the pool, or (in degenerate regions of the pool) mutations with respect to the Pol 2 consensus sequence, clone 11.22. Other conventions are as in Fig. 5. **B.** Sequence of the 11.78 doped pool, with clones from rounds 6–10 listed below. Next to the name of each clone is its family number (or "U" for unique clones, or "J" for 'junk' clones), and then its number of ligase lesions ("LL"), i.e. disrupted base-pairs in the ligase domain. Polymerization assays using primer-templates A—F (as described in Fig. 3) are reported as average nucleotides added per primer. Results are color-coded to indicate highly active (red), active (orange), weakly active (yellow), inactive (white), or not assayed (grey). An composite activity level is assigned to each clone, taking all available assay data into account ("overall") and is color-coded similarly. Clones are listed in order of overall activity category. Within categories, active clones are ordered by clone number. Inactive clones are ordered by number of ligase lesions. Clones with stars following their names were isolated twice (each in two different rounds). Clone 603, with two stars, was isolated eight times (all in round 6).

69

CHAPTER FOUR

A new tool for polymerase ribozyme evolution: selecting for addition of many nucleotides

Abstract

RNA-catalyzed RNA polymerization is a central feature of the RNA World hypothesis. Several RNA polymerase ribozymes have previously been isolated, using the Class I RNA Ligase as a catalytic core. The mercury-gel selection technique used in these experiments allowed the experimenters to demand extension of a primer by two successive nucleotides. Despite this seemingly meager stringency, the selection ultimately produced a polymerase ribozyme that could add as many as 14 nucleotides to a primer. Although this result is a pleasant surprise, there is still much improvement required: the polymerase ribozyme has a length of ~200 nt, and in order to be considered as a potential replicase in any minimal biotic model, it would need to be able to synthesize RNA on the order of its own length. Therefore, a remaining experimental challenge is the isolation of an RNA polymerase ribozyme that is fast enough to extend a primer by hundreds of nucleotides during its lifespan. Here we develop a capture-oligo-based method that allows the experimenter to select directly for addition of ~12 nucleotides of primer extension, a substantial improvement over the previous limit of two. Having seen that we could ask for two and receive fourteen, we wondered how many we might get if we asked for 12. To test the capture-oligo method, we constructed a pool of potential polymerases based on Evolved Pol 1 and Pol 2+, two previously isolated ribozymes. After several rounds of selection combining the prior mercury gel technique and the new capture-oligo technique, active ribozymes were isolated, including one of them, a variant of Pol 2+, which equals the activity of Evolved Pol 1, and exceeds it in a few cases, but we did not isolate any clones that polymerized more than 14 nucleotides. Analysis of the *in vitro* evolution experiment revealed that the older (mercury gel) selection technique is apparently counterproductive in some cases, suggesting that future experiments employing the capture-oligo technique as the sole selection criterion might produce even more active polymerases.

Introduction

The RNA World hypothesis states that RNA was the primary catalytic biomolecule during an early period of evolution before DNA and proteins took over.[1-3] One of the chief tasks facing RNA during that time, known as the "RNA World" was its own replication.[4-6] There had to be an RNA polymerase ribozyme to keep replenishing all the enzymes of the ribo-organism, including, of course, by producing new replicase copies.[7-13] Because no such ribozyme is known in modern biology, experimental efforts have focused on trying to produce an artificial RNA polymerase ribozyme, by applying the technique of *in vitro* selection, an enormously successful tool for obtaining novel ribozyme activities.[14-16]

Some of the earliest work on trying to make an RNA polymerase ribozyme used natural self-splicing introns as a starting point.[17-21] The Group I intron of *Tetrahymena* was grudgingly converted into a polymerase, but it used oligonucleotides as substrates instead of NTP monomers, and it had prohibitively low fidelity. A more recent study using the same starting point tried to overcome some of these limitations by replacing about half of the enzyme with random sequence, and using that as a starting point for selecting ligases.[22] This selection produced the class hc ligase, which was subsequently converted into a polymerase that could bind a primer-template duplex sequence-non-specifically, but which was limited to the kinds of nucleotides it could add, as well as requiring a G-U wobble pairing at the end of the primer.[23]

Other approaches have used for their starting point a completely artificial ligase ribozyme,

called the Class I RNA which was isolated in an *in vitro* selection experiment on the basis of its ability to catalyze phosphodiester formation between two nucleic acid species, the same chemistry required for RNA polymerization ligase.[24–27] This ligase was converted into a polymerase by adding on a large domain of random sequence and selecting for polymerization activity.[28] The selection step is shown in greater detail in Fig. 1A. The power behind the selection is the ability to attach the primer to the pool via a long flexible linker (a technique proven in previous work[22]), and then to use a sulfur-tagged nucleotide (4-thioUTP), which the polymerase is challenged to add across from a template A. Those polymerases in the pool that do succeed in tagging themselves with the sulfur-containing nucleotide will migrate slowly in a mercury gel, due to the extremely strong interaction between mercury and sulfur. Those pool molecules that added two (or more) halt completely at the mercury interface, and these can be excised by cutting out that area of the gel and eluting the RNA into water (Fig. 1A, red box), thereby isolating polymerases that were active during the reaction incubation. This mercury gel selection technique was introduced in our lab during a different *in vitro* selection experiment, which resulted in the discovery of artificial pyrimidine synthetase ribozymes.[29]

Several other ligase-based polymerases have been isolated using the mercury gel technique (see Chapter 1), and in a completely different approach, another group has successfully converted the Class I ligase to a polymerase by applying the technique of continuous *in vitro* evolution, yielding a ribozyme that can catalyze three successive nucleotide additions to a primer.[30] All of these ligase-derived polymerases are promising candidates that may eventually be developed to the point where they could conceivably take over the job of a replicase in the RNA World. However, right now, none of them is good enough yet. The most nucleotides that have ever been observed to by polymerized by an artificial RNA polymerase ribozyme is fourteen ,[28] using Evolved Pol 1. The ribozyme itself is nearly 200 nt long, and so there is a great deal of improvement required before it could in principle make full-length copies of itself from an appropriate template.

One rather surprising result of the polymerase selection that produced Evolved Pol 1 is that we got considerably more than we asked for. Our selection technique (Fig. 1A) asked only for addition of two nt. Our scheme didn't care if the polymerase stopped there or kept going; everything that added at least two Us was treated exactly the same, with no selective advantage for addition beyond two. Nevertheless, the result was a polymerase (Evolved Pol 1) that can add 14 nucleotides to a particular primer-template, a remarkable surpassing of our demands.

Exciting though it was to see polymerization of 14 nt, it is not enough, and we need to design a way of selecting for polymerases that add more than that. In this article, we report the development of a hybridization-based selective step (Fig. 1B) that will allow selection directly for addition of 10–12 bases of high-fidelity polymerization. The method is conceptually analogous to the enrichment of polyA-containing mRNAs by using a biotinylated poly-dT oligo immobilized on streptavidin-coated beads. However, because we are looking for general polymerases, not just poly-A polymerases, the scheme is slightly different.

We evaluated our new selection procedure by building a new starting pool for polymerase selections, based on the two best ribozymes developed so far in the lab, Evolved Pol 1 (see Appendix) and Pol 2+ (see chapter 3). Using the capture-oligo selection technique, we were able to select from this pool a variant of Pol 2 that matches the high activity and generality of Evolved Pol 1. Accordingly, the new variant (clone N704) was designed Evolved Pol 2. Although the capture-oligo selection strategy succeeded brilliantly in optimizing Pol 2, so far it has yet to raise the bar on the maximum length of ribozyme-synthesized RNA. We explore some possible
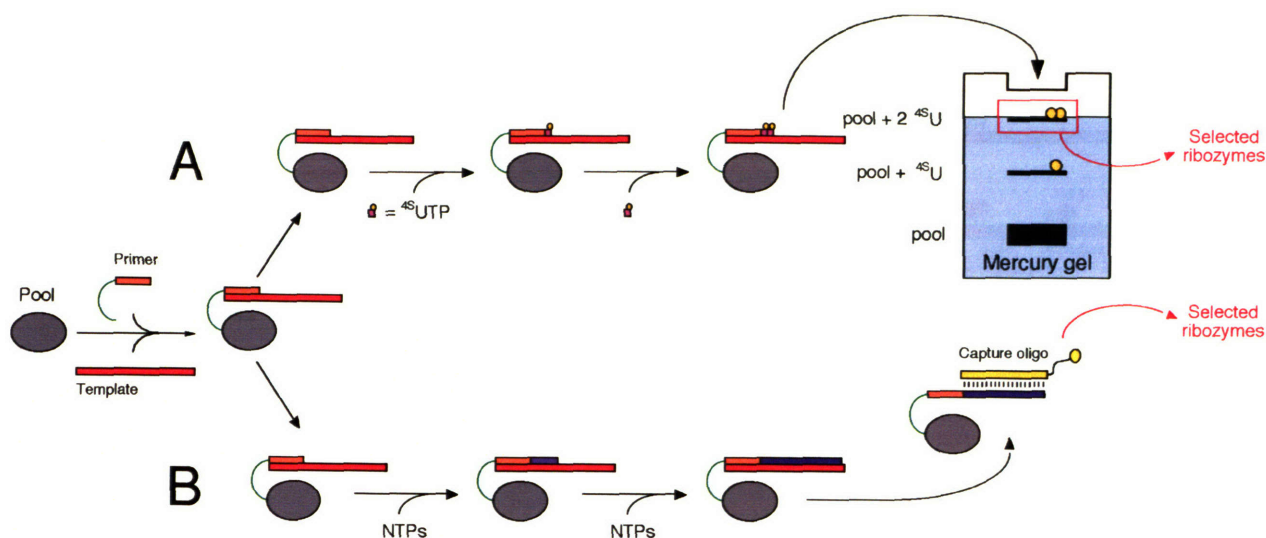
Fig. 1. Two polymerase selection techniques. Ribozyme pool (grey oval) is ligated to primer (orange) and annealed to template (red), then subjected to either **A**, incubation with 4-thioUTP (pink box), a nucleotide containing sulfur (yellow dot), followed by fractionation on a mercury gel, where molecules containing two sulfurs halt at the interface of the mercury-containing region (light blue zone), and excision of the interface region (red box), to yield ribozymes that successfully added 2 nt to their primers; or **B**, incubation with all four unmodified NTPs, followed by template removal (by gel purification) and annealing of pool to a capture oligo (yellow), to which only those pool molecules that were able to sufficiently extend their primers (dark purple) can hybridize. After washing away poorly hybridized ribozymes (those which did not extend their primers very far), the successful ribozymes are eluted.

inefficiencies in our selection technique and identify the old selective step (mercury gels) as a likely handicap to the selection process. Future work may rely exclusively on the capture-oligo method to produce a "next-generation" polymerase ribozyme.

Results and Discussion

Turning to hybridization. Development of the capture-oligo selection method began with an examination of how well we could discriminate between DNA oligos that had different levels of potential hybridization to a capture oligo. Fig. 2 shows the experimental design. An 20-nt DNA capture oligo was biotinylated at its 5'-end during automated DNA synthesis, by using biotin-dT phosphoramidite (Glen Research). Three target DNA oligos were synthesized, each shorter than the next. The full-length oligo had an 18-nt region of complementarity to the capture oligo, the intermediate-length oligo had only 10 nt of that complementary region, and the shortest oligo had none of the complementary region; it had no more than 3 nt of chance complementarity to the capture oligo. These three DNA target oligos were mixed together and annealed to a large stoichiometric excess of capture oligo, in a solution of 6X standard sodium citrate (SSC), and added to a solution of streptavidin-coated paramagnetic beads (Promega). (Throughout this report, all beads manipulations were performed in a volume of 0.3 mL, using about 0.5 mg of beads per experiment--this was measured to provide at least 0.5 nmol of binding capacity for biotinylated oligo, in agreement with published specifications.) The binding mixture was allowed 15 min at room temperature, to allow capture of the biotinylated capture oligo by the streptavidin. Following
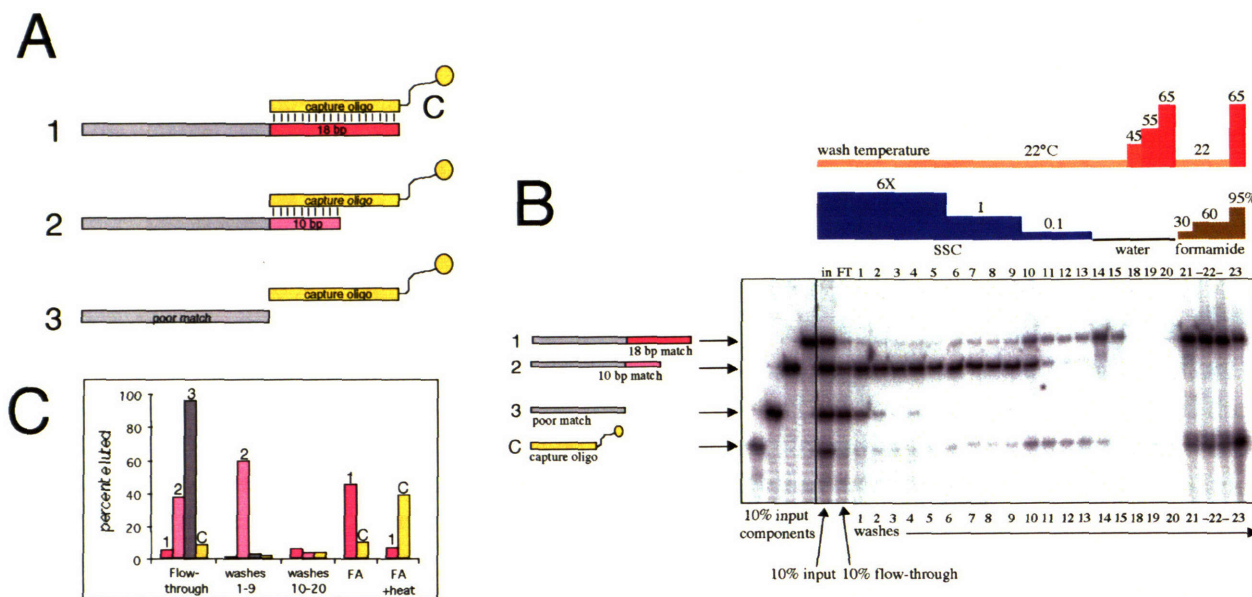
Fig. 2. Selection pilot experiment using DNA oligos. **A.** Target oligo 1 (grey + dark pink) has 18 bp of complementarity to a biotinylated capture oligo (yellow). Target oligo 2 (grey + light pink) has only 10 bp complementarity. Target oligo 3 (grey) is a poor match to the capture oligo, with no more than 3 bp. **B.** Specific retention of hybridized targets. The capture oligo and three target oligos were all radiolabeled; their lengths distinguished them on the gel). 25 pmol of each target oligos and 100 pmol of the capture oligo were mixed together in a total volume of 20 μL and annealed in 6X SSC buffer (standard sodium citrate; 1X equals 150 mM NaCl, 15 mM sodium citrate, pH 7.2). The input solution (shown diluted ten-fold in the lane marked "in") was applied to streptavidin-coated magnetic beads (in a total volume of 300 μL), and after 15 minutes, the beads were concentrated by exposure to a magnet, and the supernatant was removed as "flow-through" (shown diluted ten-told in the lane marked "FT"). The beads were then washed with decreasing concentrations of SSC (lanes 1–15), followed by heating in water (lanes 18–20). Next, the beads were washed with increasing concentrations of formamide (lanes 21, 22), and finally heated to 65°C (7 min) in 95% formamide (lane 23). All wash volumes were 300 μL. **C.** Quantitation of results in B, showing percentage recovery of each oligo during five phases of the experiment. Target oligo 3 (grey, poor match) was not bound; 95% of it eluted in the flow-through. Target oligo 1 (dark pink, 18 bp complementarity) was bound completely, with about 90% retained until the application of formamide. Target oligo 2 (light pink, 10 bp complementarity) showed intermediate behavior; 40% of it eluted in the flow-through, and the other 60% eluted gradually during washes 1–9. The capture oligo (yellow) resisted elution until heating in formamide.

the incubation period, the beads were concentrated by exposure to a hand-held magnetic device, and the supernatant was withdrawn, carefully avoiding disturbing the beads pellet. In this "flow-through" volume, almost all of the non-hybridizing DNA was recovered, but almost none of the strongly-hybridized one. As the beads were washed successively with decreasing concentrations of salt, the medium-length oligo gradually fell off the beads, because it did not have enough complementarity to withstand the decreased ionic conditions. The full-length oligo, however, with 18 bp of complementarity, remained bound much longer, with most of it staying bound until the nonspecific denaturation of the complexes by addition of formamide. The capture oligo, retained via the near-covalent-strength between of streptavidin and biotin, required heading in addition to formamide, in order to remove it from the beads. This first experiment demonstrated that the capture-oligo system is sensitive enough to distinguish between 10 and 18 bp of complementarity. The next step was to investigate its resolution in greater detail.
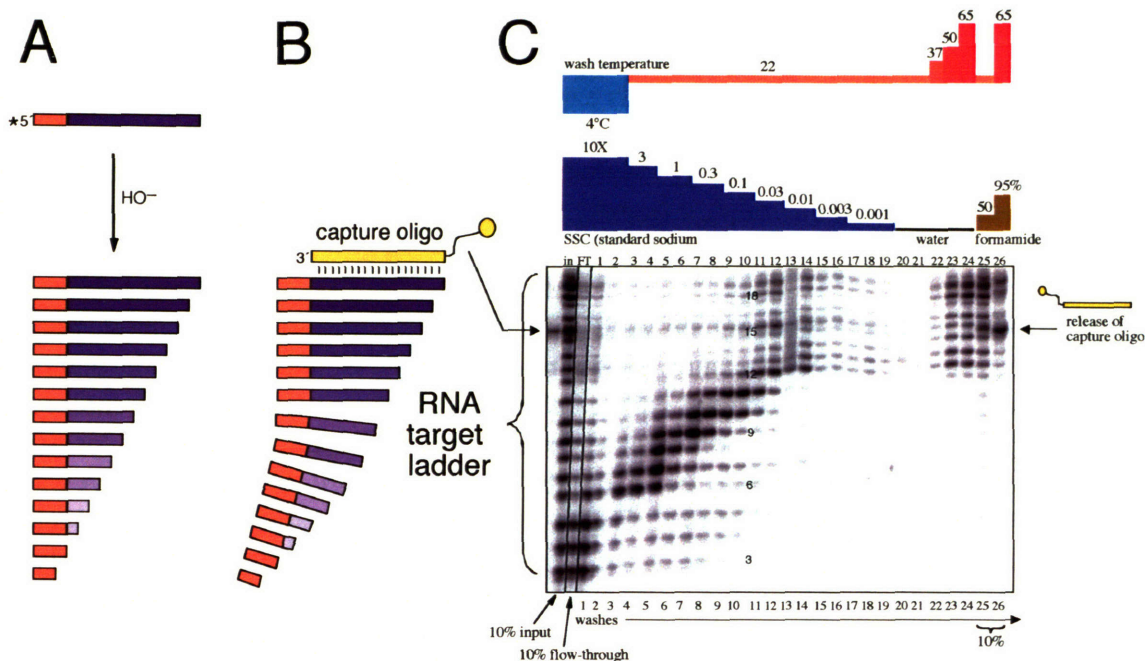
Fig. 3. Separation of RNA according to extent of hybridization to a capture oligo. **A.** A 23-nucleotide RNA (green + dark purple) is 5'-radiolabeled and subjected to partial base hydrolysis, yielding the ladder of fragments shown. **B.** The RNA target ladder generated in A is mixed with a large stoichiometric excess of biotinylated capture oligo (yellow). Full-length target RNA (dark purple), with 18 bp of complementarity to the capture oligo, hybridizes strongly to it. Intermediate-length fragments (lighter shades of purple) have between 1 and 17 bp of complementarity and hybridize with accordingly reduced strength. **C.** Fractionation of the RNA target ladder according to hybridization ability. Lane marked "in" shows a ten-fold dilution of the input mixture of RNA target ladder and capture oligo (also radiolabeled for ease of tracking, and loaded separately in the first gel lane). Input mixture was incubated with streptavidin-coated beads for 1 hour at 4°C, and unbound material ("flow-through") was withdrawn (10-fold dilution shown in lane "FT".) The beads caused a depletion of the upper bands of the ladder, corresponding to retention of strongly hybridized fragments. Beads were washed with progressive dilutions of salt, leading to the sequential release of longer and longer fragments (lanes 1–12). (Numbers next to gel bands indicate number of base-pairs of complementarity to the capture oligo.) As before, capture oligo was retained on beads until heating (7 min) with formamide (lane 26). (Note: lanes 25 and 26 are loaded as 10-fold dilutions.)

Making an RNA ladder to simulate polymerization products. Fig. 3 shows the results of the next pilot experiment. A 23-nt RNA was transcribed from a template by *in vitro* transcription using T7 RNA polymerase, and then 5'-radiolabeled. This RNA contained a 5-nt leader sequence (orange in Fig. 3), followed by an 18-nt sequence complementary to the same capture oligo used earlier. Based on the results of the first experiment (Fig. 2), which showed that 18 bp of complementarity was sufficient for long-lasting retention on the capture oligo, we expected that this 23-nt RNA would also be efficiently retained on the capture oligo. We wanted to see what would happen to shorter fragments of it, and so we subjected it to mild base hydrolysis (50 mM Na_2CO_3, 5 min at 90°C, followed by neutralization with HCl). The resulting ladder of 5'-end-labeled products (Fig. 3A) was analogous to a ladder of polymerization intermediates starting from a 5'-labeled primer. Each progressively shorter fragment had less and less potential to hybridize to the capture oligo (Fig. 3B), and we expected that this would correlate with degree of retention on the beads. This

75

expectation was confirmed experimentally: we observed a very sensitive dependence between the length of eluted fragments and the salt concentration of the wash solution (Fig. 3C). After washing with 3X SSC, nothing remained on the beads except fragments with at least 6 bp of hybridization. After reduction to 0.3X SSC, 9 bp was required for retention. At 0.03X, only those fragments with at least 12 bp of hybridization remained. At this point, further dilution had no effect: the behavior of fragments with 12–18 bp of complementarity tended to be the same; these fragments could only be eluted by heating in water, or addition of formamide.

A better way of eluting. Keeping in mind the ultimate purpose of this method, polymerase ribozyme selection, we wished to ensure that our new technique would not invite the emergence of selection parasites. One such potential class of selection artefacts would be RNA molecules that ignored the RNA primer-template and ignored the capture oligo, and which bound specifically to streptavidin, but only tight enough to withstand the wash steps until the selected RNA was eluted by heating in water, or by addition of formamide. At that point, the parasite RNA could release the streptavidin from its treacherous embrace and gleefully enter into heaven among the truly worthy polymerase ribozymes, thereby ensuring its propagation in the selection. Wishing to prevent this mischief, we devised a gentle method of specifically eluting the captured RNA, one that did rely on an abrupt shift in temperature or ionic conditions (Fig. 4)

As a consequence of the experimental design, all of the fragments in the RNA target ladder happened to contain the 5'-leader sequence 5'–GGACA (orange in Figs. 3 and 4), which was unrelated to the sequence of the capture oligo. When the capture-oligo/target was immobilized on the beads, this short single-stranded tag would dangle off into solution, where it could readily be employed as an "elution handle." We constructed an "elutor" oligo which was complementary to the target RNAs (and so conceivably it could displace them by competitive elution if supplied in sufficient excess), but it was one step better than that: it contained the complement sequence 3'–CCTGT as a "handle grabber" whereby it could pair to the dangling "handle" sequence and initiate active strand invasion of the capture-oligo/target duplex, displacing the capture oligo and releasing the target into solution as an elutor:target duplex (Fig. 4C). The utility of this method was confirmed in Fig. 4B, which shows an experiment in which the DNA target oligos and the RNA target ladder, both discussed previously, were all mixed together with the single capture oligo to which they are all complementary (capture oligo was present in large stoichiometric excess) and capture again on streptavidin beads. Washing with low salt removed all fragments that hybridized with less than 12 bp. Upon addition of the elutor oligo (Fig. 4B, lane 12), the whole ladder of remaining fragments was suddenly eluted from the beads, in dramatic confirmation of the elutor principle. Just as importantly, DNA target oligo 1 (grey + dark pink), which also binds with 18 bp of complementarity, did not respond to the elutor oligo; this is because it lacks the elutor handle (this result demonstrates that active "handle-grabbing" elution is much faster than passive competitive elution.) Again, the 18-bp-hybridized DNA oligo remained bound to the capture oligo until more aggressive denaturation, this time using mild alkali (Fig. 4B, lanes 17–20), and the capture oligo was only released after heating in formamide (lane 21).

Putting it all together. Having established a technique for purifying RNAs that have a certain minimum number of base-pairs of complementarity to a capture oligo, we looked at the behavior of the system using an actual polymerase ribozyme. This was complicated by the fact that a ribozyme (or pool) contains hundreds of additional RNA nucleotides that could in principle interfere with
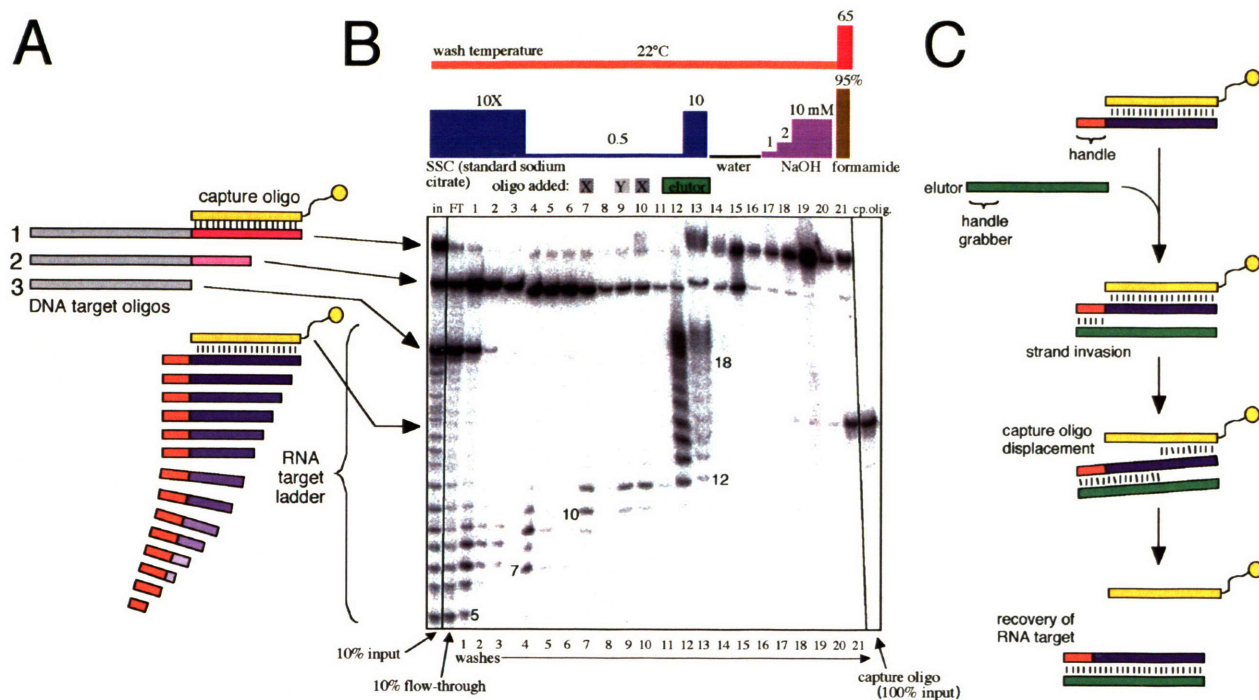
Fig. 4. Use of an elutor oligo to release captured RNA. **A.** This experiment combined the inputs shown in Figs. 2 and 3. A large stoichiometric excess of biotinylated capture oligo (yellow) was mixed with the three DNA oligos from Fig. 2. Oligo 1 (grey + dark pink) had 18 bp of complementarity to the capture oligo, oligo 2 (grey + light pink) had 10 bp, and oligo 3 was a poor match. Also in the input mixture was the RNA target ladder from Fig. 3 (green + purple). These fragments contained up to 18 bp of complementarity. Each fragment also contained the 5-nt leader sequence 5'GGACA (orange), which was not complementary to the capture oligo, and which served as a "handle" for specific elution during the fractionation procedure. **B.** Fractionation of input pool. Input mixture (capture oligo, DNA target oligos, and RNA target ladder, all radiolabeled, shown in lane "in" diluted 10-fold. Capture oligo had lower specific radioactivity and is shown undiluted in final lane, marked "cp. olig") was captured on streptavidin-coated beads, and unbound material was removed as flow-through (lane "FT", diluted 10-fold). Beads were washed with decreased salt, leading to the elution of fragments with 10 bp of complementarity or less. (Numbers next to bands indicate bp of complementarity to capture oligo.) In lane 12, elutor oligo (green), fully complementary to the RNA target, including the "handle grabber" sequence 3'CCTGT, was added to the wash solution, successfully eluting all remaining bands of the target RNA ladder. DNA oligo 1 (grey + dark pink, top band on gel) lacked the elutor handle and was unresponsive to the elutor oligo, staying annealed until treatment with mild alkali (lane 19). Capture oligo stayed bound until heating in formamide (lane 21). Pseudo-elutor oligos (X and Y in lanes 7, 9, 10) with a mutated handle grabber, 3'CCTTG, were ineffective at eluting the RNA targets. **C.** Mechanism of elutor oligo action. Strongly hybridized RNA fragment (orange + purple) resists low-salt elution from the capture oligo (yellow); however, the capture oligo is efficiently displaced by addition of elutor oligo (green), which invades the pairing starting at the "elutor handle" (orange) of the bound RNA.

the hybridization of the extended primer and the capture oligo. Another technical hurdle that we had to overcome was how to evaluate the result of the experiment: when working with the short oligos of Figs. 2–5, we were able to achieve single-nucleotide electrophoretic resolution, revealing exactly which fragment lengths eluted in the successive washes. However, with a typical ribozyme, the unextended species (ribozyme + primer) is typically on the order of 250 nt, and a primer extended by 12 nt (the maximum that has been observed in *cis*, as will be discussed further later) would increase that length from 250 to 262 nt, a range in which it is very difficult to achieve single-nucleotide electrophoretic resolution. (With a pool of ribozymes, it would be practically impossible, given the electrophoretic variability caused by differential base composition.)

Nevertheless, we were able to efficiently resolve these issues, and Fig. 5 demonstrates the successful application of our capture-oligo technique to the purification of highly extended ribozymes. We used Evolved Pol 1 (see Appendix) as the model ribozyme for this final pilot experiment; it is important to realize that Fig. 5 shows results from using a single clone, not a pool. The polymerase was attached to its primer-template using the usual method, and then it was reacted
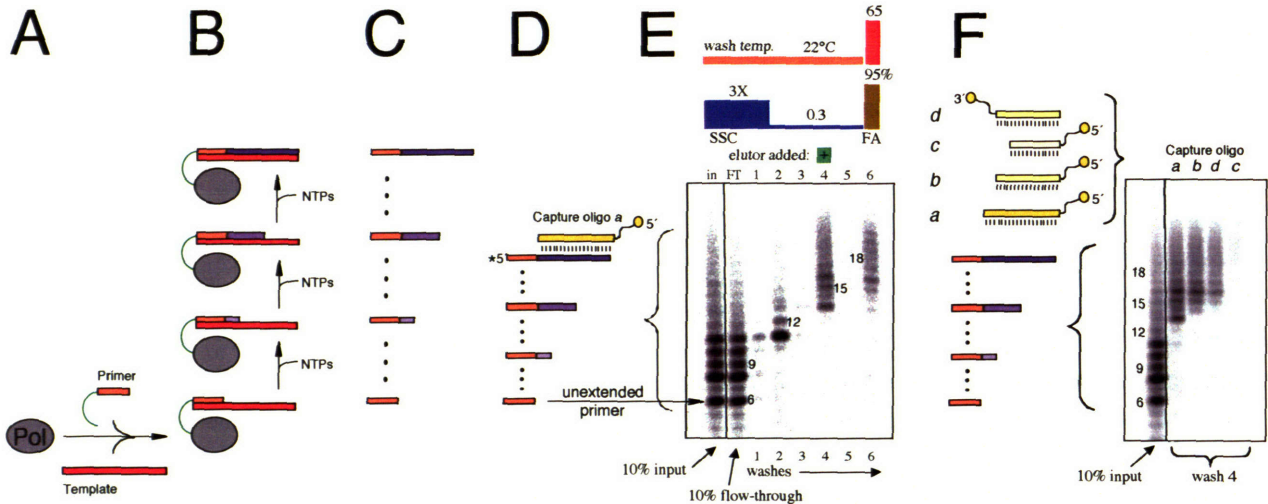
Fig. 5. Selection of highly extended polymerase ribozymes. **A.** Polymerase ribozyme clone (grey) was ligated to primer (orange) and annealed to template (red). **B.** Ligated polymerase was incubated with NTPs, resulting in primer extension (purple) to varying degrees. **C.** Template was removed by gel purification, and primer was cleaved from ribozyme using a deoxyribozyme. Result is a ladder of extended primers. **D.** Extension products were hybridized to capture oligo (yellow). **E.** RNA mixture shown in D was incubated with streptavidin-coated beads (input shown diluted 10-fold in lane "in"), and then flow-through was removed (lane FT, shown diluted 10-fold.) Beads were washed with low salt (lanes 1–3), and then elutor oligo (green) was added, resulting in the elution of RNA hybridized with 13 or more bp (lane 4). Mechanism of elutor action was as shown in Fig. 4C.). Heating in formamide (lane 6) released residual bound material. **F.** Modulation of selective stringency by altering length of capture oligo. Capture oligo *a*, used in panel E, had 18 bp of complementarity to fully extended primers. Capture oligo *b*, shortened by 2 nt, had only 16 bp of complementarity; capture oligo *c* had only 12 bp of complementarity. Capture oligo *d* was the same as *b*, except that the biotin was attached at the 3' end instead of the 5' end. The fractionation experiment shown in E was repeated, substituting different capture oligos. The contents of the elutor-release lane (wash 4) are shown in F for each of the four capture oligos, with the input (10-fold diluted) included for comparison. Shortening the capture oligo was found to systematically increase the selective power of the protocol.

with NTPs under standard polymerization conditions. Following the reaction, we added 2 μM of a deoxyribozyme (DNAzyme)[31,32] designed to cleave the primer from the ribozyme as in Fig. 6. This allowed us to observe the products of the *cis* polymerization reaction at single-nucleotide resolution, as shown in Fig. 5E. Ribozyme polymerization reactions typically produce a ladder of extension products (Chapters 1–3), and this reaction was no exception, as seen in the lane marked

"in," which shows the products of the ribozyme extension reaction. The band corresponding to unextended primer is marked with an arrow (confirmed by comparison to the DNAzyme cleavage product of ligated ribozyme that had not been reacted with NTPs, not shown.) The reaction shows strong bands corresponding to addition of 5 nt, and a blurrier fuzz of additional extension products above that. When the extension mixture was annealed to a capture-oligo and bound to streptavidin-coated beads, there was little apparent depletion of the input mixture (as judged

Fig. 6. Cleavage of primer from ribozyme by using a DNAzyme. **A.** Ribozyme or pool (blue) is shown ligated via a reverse-DNA linker (green) to a polymerization primer (orange) and annealed to a template (red). Reaction of ligated ribozyme with NTPs allows primer extension (purple). To visualize the primer extension at single-nucleotide resolution, the primer must be cleaved from the ribozyme (because the ribozyme's sheer bulk prevents single-nucleotide resolution on sequencing gels). This cleavage can be performed as shown using a DNAzyme (grey–see refs. 31, 32.) Cleavage occurs with retention of the 5'-radiolabel (red star) on the primer strand. **B.** Cleavage monitored by electrophoresis on a 10% gel. Ligated ribozyme (unreacted with NTPs) is shown after a 10-minute incubation at 37°C (50 mM MgCl$_2$, 50 mM Tris-HCl pH 8.5) that either omitted (−) or included (+) the DNAzyme. About 70% of the ribozyme was cleaved after 10 min, with little further cleavage observed after 45 min (not shown). Note: when the DNAzyme is used in polymerization assays, it is added directly to the reaction mixture at the end of the polymerization reaction, with no need for an intermediate purification or annealing step.

by flow-through lane "FT", loaded at the same concentration as input lane "in", Fig. 5E) There was a slight depletion of the higher-MW "fuzz", suggesting capture of highly extended ribozymes. This suggestion was borne out: after washing the beads several times with low salt, we added elutor oligo (in this case, the polymerization template, red strand in Fig. 5A), and this caused the elution of the highly extended products seen in lane 4 (Fig. 5E), which is loaded at 10x the concentration of the input lane. Notably, the elution contains no detectable material hybridized with 12 bp or less, corresponding to a very robust selection for the addition of at least 7 nt. Heating with formamide released some additional highly extended material; ideally we would like to optimize the elution step so as to deliver *all* highly extended ribozymes, because for reasons already discussed, we are reluctant to rely on gross environmental perturbations such as heating or solvent swap, for elution of selected ribozymes.

Capture oligo *a*, used in Fig. 5E corresponded to a low-stringency selection. We experimented with increasing selective stringency simply by shortening the 3' end of the capture oligo, as shown in Fig. 5F. For instance, capture oligo *b* was the same as *a*, except 2 nt shorter, effectively raising the bar for the RNA targets, requiring of them two additional nucleotides to achieve the same total number of base-pairs of complementarity to the capture oligo. This change

was reflected in the cut-off on oligos present in the elution (wash 4), as shown in the Fig. 5F gel (lane *b*); the eluted RNA was 1–2 nt longer than when using the first capture oligo (lane *a* in Fig. 5F, same as lane 4 in Fig. 5E). Interestingly, moving the biotin from the 5' end of the capture oligo to the 3' end (capture oligo *d*, constructed by using a biotin-CPG column during automated DNA synthesis) caused a similar slight increase in stringency, adding about 1 more nt to the length of the selected material (lane *d*). This effect may be attributable to increased steric interference between the bulky ribozyme and the magnetic particles; capture oligos *b* and *d* have identical pairing potential to the target RNA, but with the 5'-biotinylated capture oligo (*b*), the ribozyme and biotin occupy opposite ends of the capture duplex. When stringency was increased even further, by removing 4 more nt from capture oligo *b* (yielding capture oligo *c*), we appeared to reach the detectable limits of the polymerization products; only a very slight haze of material was observed in the uppermost product region (lane *c*). The ability to modulate selective stringency by simply varying the length of the capture oligo should prove a particularly useful feature of the new technique.

A next-generation polymerase pool. Having demonstrated that the capture-oligo technique can be employed to purify ribozymes that have extended their primers by many nucleotides, we turned finally to planning an *in vitro* selection experiment employing the new method. In order to do that, we needed to decide what kind of starting pool to use. The most ambitious plan would be to begin from a random-sequence pool and trust that the new selection method would be able to pull out highly efficient polymerases instead of "bizzarozymes" that simply decorate their internal 2' hydroxyls with the tagged nucleotide.[33] However, we chose a more conservative approach, banking on the demonstrated efficiency of polymerases that have already been evolved in the lab. We hoped to make them even better, by applying the new selection technique and asking directly for polymerization of many nucleotides.

Fig. 7 shows the five-subpool organization of the "Pol+N_{100}" pool that we designed. Its name reflects the fact that every member of the pool contained a polymerase ribozyme, plus at least 100 nucleotides of totally random sequence. The polymerase ribozyme was either Evolved Pol 1 (see Appendix) in subpools A–C, or Pol 2+ (see Chapter 3) in subpools D and E. In each case, the polymerase consisted of the ligase core, which was not mutagenized in this pool, and an auxiliary domain, which was mutagenized at a level of 3%. The random sequence regions were interspersed among the polymerase secondary-structural elements as shown in Fig. 7. In subpools B–E, ligase Loop 5 was replaced by random sequence; this was expected not to be detrimental to the activity of the ligase, because the ligase was originally isolated with a 70-nt loop here that could be replaced with much smaller loops without any loss of activity.[26] Similarly, in subpools D and E, Loop 9 of the polymerase auxiliary domain was replaced by random sequence, because this hairpin had been found highly tolerant of sequence change (Chapter 3). Part of the rationale for including much of the random sequence in loops of the polymerase, which may be located far from the active site, was a hope that it might provide a chance for the evolution of stabilizing tertiary contacts (perhaps between new elements in loops 5 and 9 in subpool D, for example). We imagined the possibility of evolving a peripheral clamp-like structure that would increase the rigidity of the polymerase core, potentially refining its binding contacts to the primer-template.

In subpools D and E, the Pol 2 "tail" region (between the P8 stem and the RT-primer/ binding site) was replaced with random sequence, because we did not wish to propagate any tail-template pairing ability that may have evolved in the Pol 2 doped selection (Chapter 3). Subpool
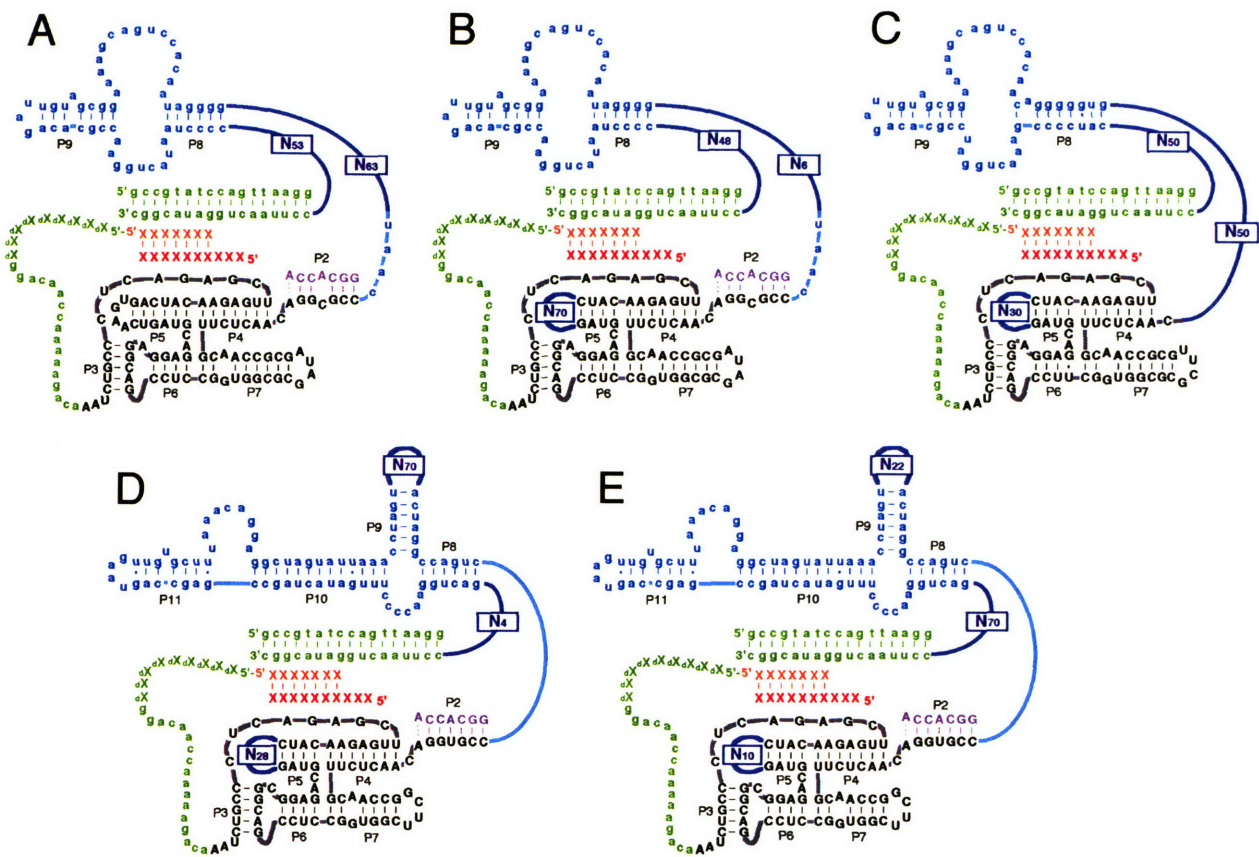
Fig. 7. Design of Pol+N$_{100}$ pool for next-generation polymerase selection. Pool is comprised of five subpools A–E, which were constructed separately and then mixed together in approximately equal stoichiometry before commencing *in vitro* selection. Each subpool contains an unmutagenized Class I RNA ligase (black, uppercase), an 3%-mutagenized polymerase auxiliary domain (light blue, lowercase), and 100 or more random nucleotides (dark blue, "N") arranged as indicated. Green nucleotides are fixed primer-binding sequences. RT primer, included in polymerization reactions to render the ribozyme 3' terminus double-stranded, is also shown in green. Purple strand is the P2 stem-completing heptamer. Pool is shown in the primer-ligated format, with RNA primer strand (orange, "X") attached through a DNA linker (green, "dX") as described in the text, and complementary template strand (red, "X") annealed to the primer by base-pairing. The total length of each subpool (excluding primer and linker) is 300 nt. **A.** Subpool A is based on Evolved Pol 1 and has 116 nt of random sequence inserted as two portions into the auxiliary domain as shown. **B.** Subpool B, also based on Evolved Pol 1, has 124 nt of random sequence, with 70 nt of that included as a large randomized loop in the ligase domain. **C.** Subpool C is based on a derivative of Evolved Pol 1 known in the lab as 11C18, which has improved usage of some primer-templates and is observed to longer require the P2-completing oligo. This subpool contains 130 nt of random sequence, organized in the three chunks as shown. **D, E.** Subpools D and E are based on Pol 2+ and each contain 102 nt of random sequence. In both cases, a portion of the random sequence was used to replace an auxiliary domain loop found to tolerate variation in sequence and length.

C was based on a variant of Evolved Pol 1 which had lost the requirement for the P2-completing heptamer; consequently it is shown as not binding to it. The P2-completing heptamer was included in all selection reactions, however, for the benefit of the other four subpools, whose parents did still require it for activity.

All five subpools were 300 nt in length (excluding primer and linker). This is the longest pool we have ever constructed in our lab, and possibly one of the longest RNA pools used in an *in vitro* selection. Accordingly, it was not possible to synthesize the pool DNA template as a single oligo. Previous methods of constructing long pools for *in vitro* selection have used restriction digests and ligations to concatenate the segments of the pool.[24] We used the method of nested mutually-primed PCR illustrated in Fig. 8. The five subpools were constructed separately, and the success of the construction procedure was confirmed by sequencing 12 isolates cloned from each subpool. The subpools were mixed together in approximately equal ratio (no initial PCR amplification was performed), and the composite pool was transcribed to yield Pool Zero RNA.
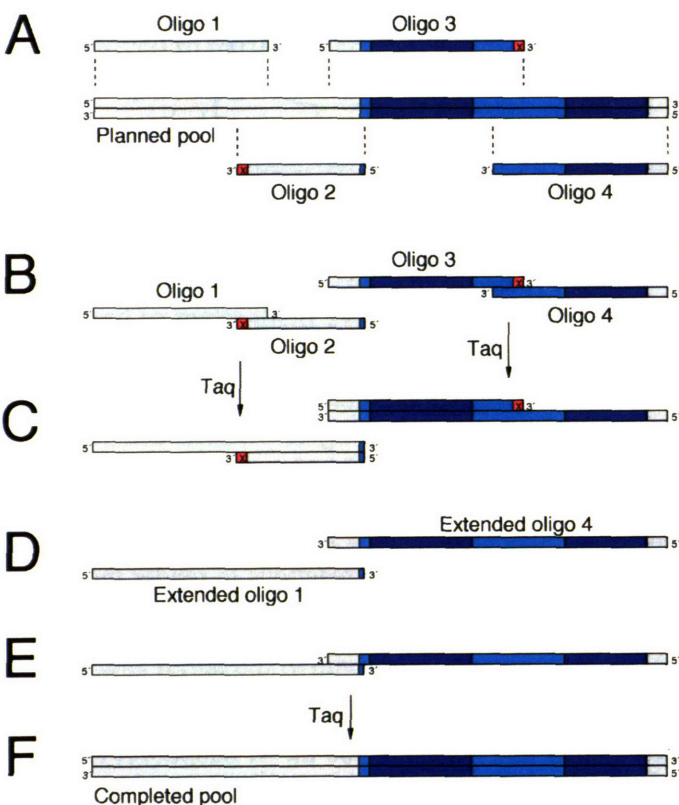
Fig. 8. Pool construction scheme. Full-length template DNA (320 bp) for each subpool is constructed according to the following protocol, using subpool A as an example. Grey indicates fixed sequence, light blue indicates 3%-mutagenized sequence, and dark blue indicates random sequence. **A.** Four single-stranded DNA oligos (60–120 nt each) are synthesized. Oligos 1 and 3 are fragments of the pool top strand as shown, and oligos 2 and 4 are fragments of the bottom strand. Oligos 2 and 3 are blocked at their 3' ends (red, "X") by using a C3 CPG column during DNA synthesis. **B.** Oligos 1 and 2 are annealed to each other in their 20-nt region of complementarity. Oligos 3 and 4 are similarly annealed. **C.** Oligo 1 is extended by Taq DNA polymerase, using oligo 2 as a template strand. Oligo 4 is extended in a separate reaction. **D.** Extended oligo 1 is purified on the basis of its length. Extended oligo 4 is purified separately. **E.** Extended oligos 1 and 4 are annealed to each other and extended by Taq DNA polymerase, yielding full-length double-stranded pool, **F.**

***In vitro* evolution.** Table 1 shows the conditions of the 10 rounds of selection that were performed starting from the Pol+N$_{100}$ starting pool. At each round, the pool was ligated to a polymerization primer, in order to select for self-extending polymerases, as has been described previously (Appendix and Chapters 1 and 3). Only the "reverse-DNA linker" ligation method was used in this selection (Chapter 2, Fig. 2A), because we had seen previously that PEG-linked primers (Chapter 2, Fig. 2B) sometimes caused difficulty for Evolved Pol 1 (data not shown). However,

Rd.	primer-template	NTPs (mM) 4SU	A	C	G	MgCl₂ (mM)	time (hr)	selection applied Hg	CptO	capture oligo
1	19.88 Pool-tccaacaac-5'5'-UUGAGUAGUA 22.120 3'-AACUCAUCAU&&GCUCAGAAAU	0.3	0	0	0	30	25	+	−	
2	18.174 Pool-tcttacatt-5'5'-GAUAGGUAG 17.138 3'-CUAUCCAUC&&CCUGGA	0.3	0	0	0	30	22	+	−	
3	17.133 Pool-tctaaacat-5'5'-GAUGAGUC 11.32 3'-CUACUCAG&&U	0.3	0	0	0	30	38	+	−	
4	16.109 Pool-ttacctaac-5'5'-CUACCAC 22.118 3'-GAUGGUG&&UCCUUAACAGUAA	0.3	0	0	0	30	45	+	−	
5	19.88 Pool-tccaacaac-5'5'-UUGAGUAGUA 22.120 3'-AACUCAUCAU&&GCUCAGAAAU	0.3	0	0	0	30	25	+	−	
6	16.109 Pool-ttacctaac-5'5'-CUACCAC 22.118 3'-GAUGGUG&&UCCUUAACAGUAA	1	1	1	1	50	21	+	+	3'BGATGGTGAATCCTTAACAGTAA 23.78
7	19.88 Pool-tccaacaac-5'5'-UUGAGUAGUA 22.120 3'-AACUCAUCAU&&GCUCAGAAAU	1	1	1	1	50	3	+	+	3'BATCATAAGCTCAGAAAT 18.181
8	16.109 Pool-ttacctaac-5'5'-CUACCAC 22.118 3'-GAUGGUG&&UCCUUAACAGUAA	1	1	1	1	50	0.5	+	+	3'BTGAATCCTTAACAGTAA 18.180
9	19.88 Pool-tccaacaac-5'5'-UUGAGUAGUA 22.120 3'-AACUCAUCAU&&GCUCAGAAAU	1	1	1	1	30	0.5	+	+	3'BATCATAAGCTCAGAAAT 18.181
10	17.133 Pool-tctaaacat-5'5'-GAUGAGUC 11.32 3'-CUACUCAG&&U	0.3	1	1	1	30	0	+	−	

Table 1. Primer-templates and selection parameters used during the Pol+N₁₀₀ polymerase selection. Listed for each of the 10 rounds is the sequence of the RNA primer and the "reverse-DNA" linking it to the pool. Templates are shown annealed to the primer. "&" in template strands indicates the adenine isomer 2-aminopurine. NTP concentrations (4-thioUTP was always substituted for UTP), magnesium concentrations, and time allowed for incubation are listed for each round. The selection applied at each round is indicated with a "+" at each round under the column marked "Hg", because a mercury gel was employed at each step (as in Fig. 1A). In rounds 6–9, the beads/capture-oligo procedure (Fig. 1B) was applied as a second selective step, following the mercury gel step. The capture oligo sequence is shown for these rounds (B indicates biotin.)

in order to avoid the dependence on fixed linker sequence that had developed during the Evolved Pol 1 optimizing selection (as described in Chapter 1), we varied the sequence of the reverse-DNA linker at every round.

In early rounds of the selection, only the mercury-gel selection technique was used. Pool activity was detected as early as round 2, and during round 3, a very strong signal was detected, with 0.9% of the pool being observed to add two 4-thioUs, as judged on the mercury selection gel (not shown). Starting in round 6, the capture-oligo technique was integrated into the selection protocol as a second selective step (in series), following the mercury gel step. This was made possible by reacting the pool with 4-thioUTP and the other three unmodified NTPs in equal concentration, allowing extension along a long template. During round 7, five percent of the pool added two Us in 3 hours. Incubation times were shortened to increase the selection stringency. During round 10, two percent of the pool added two Us in one minute.

Observations of pool activity during the selection provided information about how it was evolving with respect to the *cis* reaction (tethered PT), but the ultimate goal of the selection is

a polymerase that uses an external PT. Accordingly, we wished to supplement our observations by also monitoring the pool in *trans* (untethered PT). Fig. 9 shows polymerization assays of pool RNA during the late rounds of the selection (round 6–10). RNA was tested from each component selective step to evaluate the relative contributions of the mercury-gel and capture-oligo selective techniques. This provided the first direct validation of the capture-oligo technique as

B *CUGCCAA
GACGGUUGGCACGCUUCG

round 6   7   8   9   10

E *CUGCCAACCGU
GACGGUUGGCACGCUUCGCAG

round 6   7   8   9   10

○ after Hg gel
● after capture oligo

Fig. 9. Pool evolution during late rounds of Pol+N$_{100}$ selection. Pool activity was monitored in the *trans* configuration (untethered PT), using the two PTs shown (PTs B and E, from Table 3). Pool RNA from rounds 6–10 was assayed, with unreacted primer shown in lane "–". In rounds 6–9, two selective steps were performed in series: first, mercury-gel fractionation (Fig. 1A), with results shown in lanes above yellow circles; and second, capture-oligo fractionation (Fig. 1B), lanes above blue circles. A slight increase in *trans* activity was seen following capture-oligo steps, but mercury-gel treatment appeared consistently to cause a decrease in *trans* activity.

a viable way of selecting for polymerase ribozymes: for instance, in Fig. 9 (top panel), the round 6 pool showed addition of barely 1 nt to PT B after the mercury-gel selective step (rd 6, yellow-circle lane), but following further selection using the capture-oligo technique (rd 6, blue-circle lane), addition of 3 nt could be detected. Similarly, application of the capture-oligo selective step in round 7 caused a 2–3 nt increase in pool-catalyzed *trans* polymerization using PT E (Fig. 9, lower panel)

Although it was encouraging to see evidence that the capture-oligo technique was a success as applied during the Pol+N$_{100}$ selection, we were troubled by a separate finding: it appeared that the mercury-gel selective step was actually counterproductive during the late rounds of the selection. In each consecutive round during rounds 6–9, the pool's activity actually declined (as measured in *trans* using PTs B and E, Fig. 9), instead of increasing, as would normally be expected. This decline was offset by the subsequent improvement caused by the capture-oligo step. The reason for this apparent counterproductivity was not immediately clear.

New clones. Seeing that pool progress (as measured in *trans*) had apparently plateaued during rounds 7–10, we ended the selection experiment after the mercury-gel step of round 10. We then cloned and sequenced individual isolates from pools 6, 7, and 10. (The round-6 and -7 pools were post-capture-oligo-step.) A total of 24 clones were sequenced. A note on clone nomenclature: Isolates from the Pol+N$_{100}$ selection were named with prefix "N" to distinguish them from the isolates of other selections being discussed. The first 1 or 2 digits in the clone name tell what round it was isolated from.

Table 2 lists the clones from the Pol+N$_{100}$ selection. (Complete RNA sequences are reported in the Supplementary Figure.) Four of the five component subpools (Fig. 7) were represented, as well as both parental polymerases (Evolved Pol 1 and Pol 2+). Five of the eight clones from subpool B were grouped as "Family I" due to the shared sequence in their random-derived regions; the other three were unrelated. The seven clones from subpool A were unrelated to each other.

Clone	Subpool Family	LL	C	E	G	H	Z	Overall
Evolved Pol 1		clear	4-10	11	1-3	1-2	1-6	+++
N704	E II	clear	4-10	4-11	1-10	1-7	4	+++
N714	E III	clear	0	w 1-2	1-5	1-4	0	++
N717	E III	clear	1-4	w 1-2	1-2	1-3	0	++
N707	E II	P7 les	1-5	4-7	1	1-3	1-4	++
N604	E	clear	0	w 1	w 1	1	1	+
N705	E II	P7 les	0	w 1-2		1	0	+
N701	E II	P7 les	0	w 1-2		0	w	−
N1003	B	P5 les	1-5	2-4	1-5	1-7	1-4	+++
N1002	B I	P7 les	w 1-4	0	1	1-2	5-10	++
N1005	B I	clear	w 1-4	w 1-2	1	1-2	2-10	++
N1006	B I	clear	w 1-4	1-2	1	1-2	1-8	++
N703	B	clear	3-5	4-5	1	1	1-4	++
N1007	B I	clear	w 1-4	w 1-2		1-2	w 1	+
N702	B	P4 les	w 1	w 1-2	w 1	1	1-4	+
N601	B I	P4 les	0	0		0	0	−
N602	A	clear	3-10	w 1-5	1-3	1-2	8	+++
N711	A	clear	1-5	w 1-2	1-5	1-7	2-10	+++
N716	A	clear	1-5	1-2	1-5	1-7	5-10	+++
N603	A	P3 les	2-5	4-5	0	1	0	++
N605	A	L7 les	3-5	4-5	w 1	1	0	++
N712	A	P3 les	w 1	1-2	1	1	w 1	+
N713	A	clear	w 1-4	4	w 1	1	w 1	+
N706	C	many les	0	0		0	w	−
N715	C	many les	0	0		0	w	−

Table 2. Clones isolated from Pol+N$_{100}$ polymerase selection. Table shows clone name, followed by which supbool it belonged to, an which family (if any) of similar-sequence isolates from a given subpool. "LL" (ligase lesions) reports disrupted base-pairs in the ligase domain, or "clear" if none. Polymerization assays (in *trans*) were carried out under the standard assay conditions described in Chapter Three, with 5 µM ribozyme. Results of assays with five different PTs (sequences given in Table 3) are reported under columns showing the PT letter. Values indicate observed ranges of extension products, with "w" indicating "weak". Values are color-coded according to arbitrary cutoffs to indicate strong (red), moderate (orange), weak (yellow), or no (white) polymerization observed. Grey boxes indicate values not determined. Each clone is assigned an approximate composite level of activity, listed in column marked "Overall". Clones are shown in groups corresponding to subpool of origin, and then listed within those groups by overall activity level.

Table 3. Primer-templates (PTs) used in polymerization assays. (PTs A–F are the same ones used in Chapter 3.) The sequences of PTs A–M and Z are shown, along with the lab names of these oligos, and the number of coding (unpaired) nucleotides in each template. Primer strand is shown on top, written 5' to 3', with the star indicating the position of the radiolabel, and template strand is shown on bottom, written 3' to 5'. Lowercase "a" in certain templates indicates the adenine isomer 2-aminopurine; lowercase "t" in primer Z indicates a biotinylated dT residue (this modification was unrelated to the primer's use in polymerization assays and did not appear to interfere with them.)

primer-template	oligos	primer-template sequence	coding nt
A	7.5 17.68	*AGCUGCC UCGACGGaaCCUGCGUC	10
B	7.6 18.99	*CUGCCAA GACGGUUGGCACGCUUCG	11
C	8.9 18.90	*GAAUCAAG CUUAGUUCCCGCCCGGCC	10
D	10.7 18.90	*GAAUCAAGGG CUUAGUUCCCGCCCGGCC	8
E	11.26 21.30	*CUGCCAACCGU GACGGUUGGCACGCUUCGCAG	10
F	14.19 21.30	*CUGCCAACCGUGCG GACGGUUGGCACGCUUCGCAG	7
G	7.16 22.118	*CUACCAC GAUGGUGaaUCCUUAACAGUAA	15
H	10.26 22.120	*UUGAGUAGUA AACUCAUCAUaaGCUCAGAAAU	12
I	9.37 39.56	*GAUAGGUAG CUAUCCAUCCGAACCACCAACAAUAGCCAGCACUCACGC	30
J	9.37 39.58	*GAUAGGUAG CUAUCCAUCAGCCCCAAUAAACAUAAACAAGAACCCAGC	30
K	9.37 39.60	*GAUAGGUAG CUAUCCAUCCUACUAUAUACAUUCGCCCACUUAAUCAUC	30
L	7.6 21.30	*CUGCCAA GACGGUUGGCACGCUUCGCAG	14
M	7.6 24.26	*CUGCCAA GACGGUUGGCACGCUUCGCAGAGG	17
Z	14.28 110Tmp	*UtGGAGCAAAACGA CCUCGUUUUGCUCCGCAUCCCAAAUUUUCCGCCAACG... AGGGACAAGGCUUCAACUUCAACUUGCGUGUUGCCCUACC... AACAAAUCGGUCAAACAUCUCGCAGGAGCGAGG	98

85

Two families could be identified among the seven clones from subpool E: Family II (four clones), and Family III (two clones). All of the clones from subpools A, B, and E had fairly intact ligase domains, with at most one or two disrupted base-pairs. The column marked "LL" (ligase lesions) lists whether the clones have an undisrupted ("clear") ligase domain, or else lists the ligase stem that was disrupted. The two clones from subpool C both had severely disrupted ligase domains and were shown to be inactive.

Pol 2, your day has finally come!   A set of five PTs was chosen from the list in Table 3, and the new clones were tested for polymerization activity in standard *trans* assays. The results are reported in Table 2. The activity of Evolved Pol 1 (one of the parental ribozymes) is listed at the top for comparison. Four of the five best clones (N1003, N602, N711, and N716) were descended from Evolved Pol 1, but they failed to improve on its abilities. It was interesting to note, however, that most of the selected Pol 1-related clones had repaired their P2 mismatch to a Watson-Crick pair. Subpools A and B were constructed with the P2 stem containing a mismatch (the U106C mutation) that had been found to confer increased activity on Evolved Pol 1 (see Appendix; also discussed in Chapter 3).

The other clone (N704) from the top five was a descendant of Pol 2+, and it showed activity of nearly identical calibre to that of Evolved Pol 1. This was an exciting result, given the meager gains from the Pol 2 doped selection (Chapter 3). It seemed for a while that all Pol 2 variants were generally inferior to Evolved Pol 1. Nevertheless, the Pol+N_{100} selection finally succeeded in raising Pol 2 to the level of Evolved Pol 1. For this reason, clone N704 was renamed "Evolved Pol 2" and is referred to by that name henceforth. Fig. 10 shows the structure of Evolved Pol 2. The secondary structure assigned to random-sequence-derived regions should be regarded as speculative, awaiting a more rigorous determination by comparative sequence analysis.
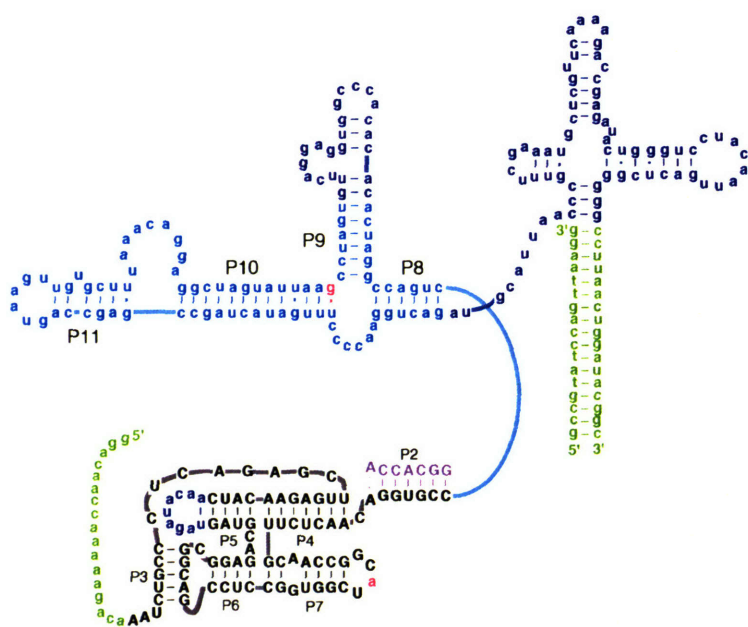
Fig. 10. Evolved Pol 2. Secondary structures of ligase (black) and Pol 2 auxiliary domain (light blue) were determined previously by comparative sequence analysis. New domains added during the Pol+N_{100} selection (dark blue) have not yet been subjected to comparative analysis; the secondary structures shown for them are speculative. The pink "g" in P10 is a mutation with respect to the sequence of subpool E (Fig. 6) In designing the subpools based on Pol 2+, this proximal wobble pair in P10 was "corrected" to an A–U pair; nevertheless in the Pol+N_{100} selection, the most successful clone (Evolved Pol 2) restored it to a wobble pair. It is unknown what effect, if any, this has on polymerization activity. The other pink nucleotide, an "a" in ligase Loop 7, was the only other mutation in the defined-sequence segments of the pool.
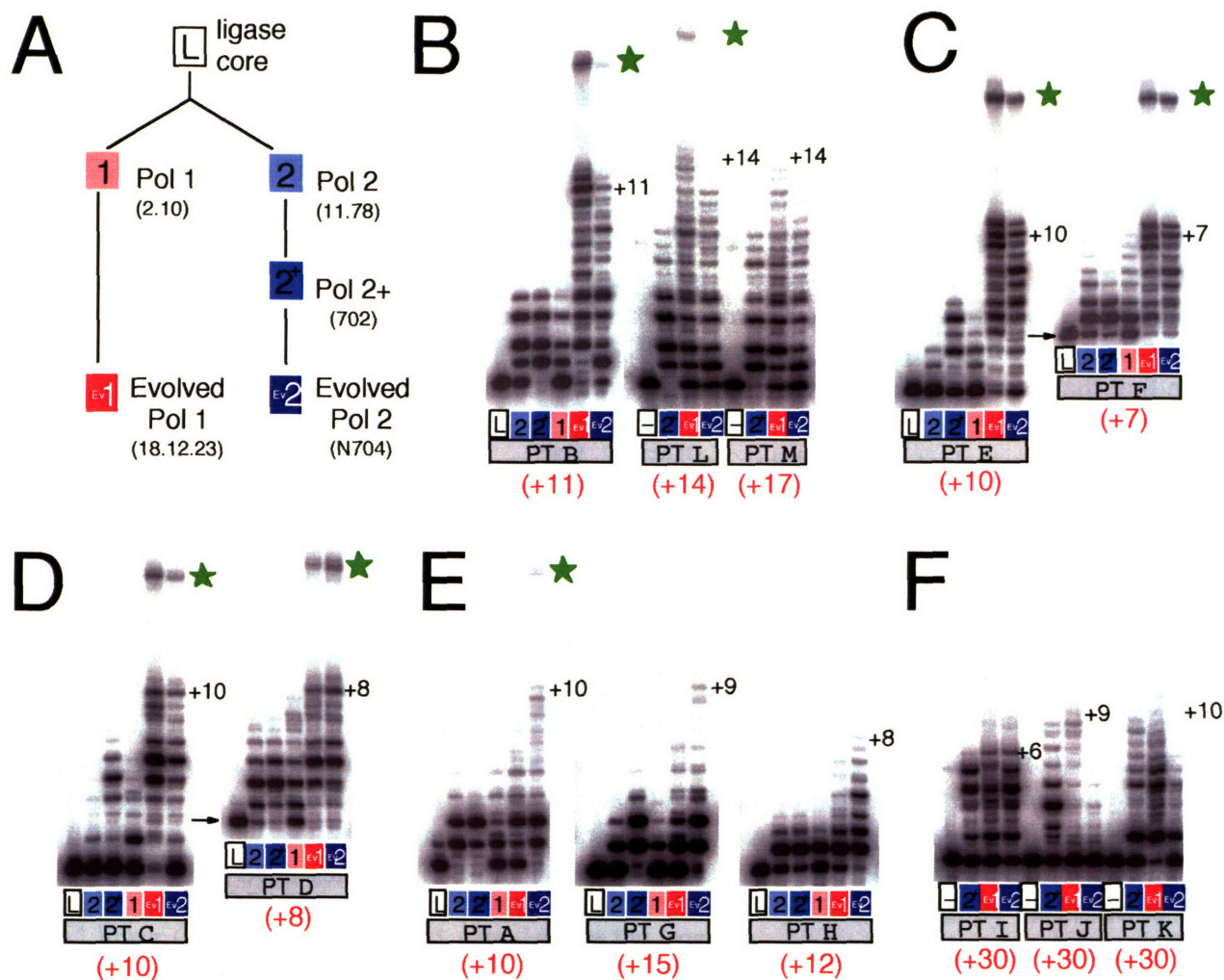
Fig. 11. Comparison of milestone polymerases. **A.** Polymerase family tree and key to gel lanes. The Class I Ligase ("L") is marked with a white box. The ligase-derived polymerases from branch A of the original Lig+N_{76} selection (see Appendix) are marked with pink (Pol 1) and red (Evolved Pol 1) boxes. Polymerases from branch B (see Chapter 1) are marked with light blue (Pol 2), medium blue (Pol 2+), and dark blue (Evolved Pol 2) boxes. **B–E.** Primer extension gels showing endpoint (2–3 days incubation) of each polymerization reaction (standard assay conditions, with 5 μM ribozyme.) Lanes marked "–" contained no ribozyme. PT names are indicated; sequences are shown in Table 3. Numbers in parentheses (orange) indicate number of template coding nucleotides. Number next to product bands (black) indicate number of nucleotides added. Stars (green) draw attention to anomalous large products in certain reactions, discussed in text. **B.** Same primer, three templates of increasing length. Middle PT (PT L) is the celebrated fourteen-nucleotide example. Evolved Pol 2 shows a very faint +14 band. **C.** Same template, two primers of different lengths (Arrow emphasizes the fact that the PT E +3 product is identical to unextended PT F.) **D.** Another example of same template, two primers of different length (arrow analogous to C). **E.** Three unrelated PTs, all with 2-aminopurine as the first two coding residues (coding for +UU), analogous to all PTs used in the selections that produced these polymerases. Ironically, they are poor substrates for the polymerases. Evolved Pol has superior activity in these cases. **F.** Same primer, three different templates, all coding for 30 nt. These substrates are a problem for all of the polymerases; none can extend them past +10 nt.

This calls for a family reunion. In order to survey the collection of polymerases isolated so far in the Bartel Lab, we decided to collect together all the major ligase-derived polymerases and evaluate their polymerization behavior in side-by-side assays. Fig. 11 shows the results of this experiment. A total of 13 different PTs was used for the assays (sequences are listed in Table 3), with the number of template coding residues varying from seven (PT F) to 30 (PTs I, J, K). For 6/13 of these PTs, both Evolved Pol 1 and Evolved Pol 2 synthesized fully extended product. In one additional case (PT A), only Evolved Pol 2 gave fully extended product. In the other six cases (including the ones with the 30-nt coding regions), no polymerase was able to fully extend the PT. One interesting class of PTs (A, G, and H) had 2-aminopurine as their first two coding residues (specifying addition of U), analogous to the PTs used during *in vitro* selection (Table 1). Despite the fact that these templates are of the exact sort encountered by the polymerases in *cis* during selection, they are ironically the ones that cause them the most difficulty in *trans*, as shown in Fig. 11E. Nevertheless, Evolved Pol 2 showed a pronounced improvement with all three of these troublesome PTs, outperforming Evolved Pol 1 by several nucleotides in each case.

Ligases will be ligases. The green stars in Fig. 11 direct the reader's attention to an unusual side product we have often observed in our standard polymerization assays. An extra band often appears high in the gel, corresponding to a length of ~30 nt, but with its exact position depending on the specific PT used. This anomalous band always manifests in direct proportion to full-length product; in reactions where the polymerase fails to make it to the end of the PT, the high-MW band likewise fails to appear. A quick glance at Fig. 11 confirms this correlation: the green stars pick out the places where the number of nt added (black) equaled that encoded in the template (orange). The high-MW band is probably a ligation product, but curiously it forms only after the PT duplex has been rendered fully double-stranded. Our polymerase ribozymes have a tendency to add an untemplated extra nucleotide (like most proteinaceous polymerases),[34,35] and the results of this untemplated activity can be seen in many of the Fig. 11 gels. Perhaps once the untemplated nucleotide is added, it can act as a 3'-overhanging "sticky end" to attract a ligation partner, which the ligase core then happily joins onto the template strand. This model would explain the small increase in apparent length of the anomalous product that accompanied the 3-nt increase in template size in changing from PT B to PT L (Fig. 11B).

How can the selection be improved? The Pol+N_{100} selection was a success in that it yielded Evolved Pol 2, but it did not achieve the goal of uncovering a polymerase that could add dozens or hundreds of nucleotides. We take some time now to analyze what some of the inefficiencies in the selection protocol might have been, and how these might be corrected in the future.

A maxim of *in vitro* selection states, "You get what you select for." Accordingly, the experimenter must always stop to ask, "Am I selecting for what I want?" Lamentably, there are some fundamental physical constraints that usually preclude a resounding "Yes." Chief among these is the need to cast the selection task as a self-modifying reaction. When seeking a ribozyme catalyst for a reaction involving two small substrates, in practice at least one of them must be tethered to the pool. For instance, in the selection for pyrimidine synthetase ribozymes, the ribose reagent was covalently attached to the end of the pool molecules,[29] and the tagged pyrimidine was free in solution. In a selection for Diels-Alderase ribozymes, the diene was tethered, and the tagged dieneophile was free.[36] The selection for polymerase ribozymes is no exception to this constraint, and consequently the PT is tethered to the end of the pool, with *in vitro* selection performed solely

in this *cis* context (Fig. 12A, *top*). In contrast, as soon as cloned polymerases are in hand, we switch over immediately to assaying them in *trans* (Fig. 12A, *bottom*), with the PT supplied free in solution. The *trans* reaction is the true goal of the project; we want a polymerase ribozyme that catalyzes general RNA polymerization by recognizing the general features of an RNA double-helix (such as phosphate residues and minor-groove moieties), not a polymerase that works only when its substrate is provided on a leash.

This disparity between the selective task and the sought activity is always a possible source of frustration, as confirmed by the analysis shown in Fig. 12B. Here, several pools and clones were tested in the two formats illustrated in panel A. The same primer-template was used in both formats (Primer H from Table 3); the only difference between the two experiments was the *cis/trans* distinction. This allowed us to ask whether *cis* activity (what we asked for) correlated with *trans* activity (what we wanted). Pool 6 showed very low activity in both formats (Fig. 12B), because activity was just starting to emerge at the point the selection. During the next round, there was a big jump in activity, both *cis* and *trans*. So far so good. But over the next three rounds, *trans* activity declined, with pool 10 showing half the activity of pool 7. This corroborated our earlier observations of declining *trans* activity during the later rounds of the selection (Fig. 9).

Fig. 12. Selecting in *cis* does not consistently yield improvement in *trans*, **A.** Illustration of *cis/trans* distinction. In the *cis* configuration, ribozyme (grey oval) is covalently tethered to the polymerization primer (orange) via a "reverse-DNA" linker (green) as in Chapter 3, Fig. 2A. In the *trans* configuration, the PT is untethered. **B.** Comparison of polymerization in *cis* and *trans*. Three samples of pool RNA (from after round 6, 7, and 10), plus four clones (two from round 7, two from round 10) were each tested in *cis* and *trans* using PT H (Table 2). (Technical note: the *cis* version of Primer H is named 18.174 and was used in round 2 of the selection—see Table 1.) Relative polymerization activity of pools and clones is shown on the graphs. Pool activity in *cis* consistently increased during the selection, whereas pool activity in *trans* began to decline after round 7.

The surprising observation was that the pool's *cis* activity, in contrast, had actually increased; pool 10 had about twice the *cis* activity of pool 7 (Fig. 12B, top). This meant that we were indeed getting what we were selecting for, albeit not what we ultimately wanted. Fig. 12B also shows the results of the corresponding *cis/trans* assays performed on two clones from round 7 and two from round 10. Clone activity generally resembled that of the source pool, with one notable standout: clone N1003 retained high *trans* activity despite the decline in *trans* activity of pool 10. This observation was reassuring: despite the setbacks incurred during the late rounds of the selection, the pool still had some good polymerases in it.

Having established that the *cis* selection as designed is working just fine, we must look elsewhere for the reason behind the declining *trans* activity seen between rounds 7 and 10. Fig. 9 showed that the decline was due solely to the mercury-gel selective steps; the reason for this effect is unclear. One possible explanation is that by lumping together all those molecules that can add at least 2 nt (as discussed in the introduction of this chapter), the mercury-gel step somehow ends up disproportionately rewarding those molecules that add 2 nt and then stop. Whatever the mechanism, there is no question about the conclusion, and the mercury-gel step must either be debugged or abandoned.

Is the leash keeping us in the backyard? Given the seeming likelihood that the capture-oligo selection technique may take on an increased importance relative to the mercury-gel method, we wondered how far the technique could be pushed. Specifically, how many nucleotides could we select for? This limit is determined in part by the length of the template coding region, but also in part by geometric constraints imposed by the *cis* polymerization format, as illustrated in Fig. 13. We considered whether the polymerase could fully extend a long PT in *cis*. Fig. 13A shows

Fig. 13. Attempt to extend the range of polymerization in *cis* (with tethered PT). **A.** Polymerase clone N1003 is shown ligated to PT J (Table 3), which codes for addition of 30 nt. However, only 10–12 nt of polymerization is achieved in a 22-hr incubation, as shown in panel E, lane "0". One possible explanation for this apparent limit is suggested in **B**: after adding 12 nt, the polymerase active site has shifted forward a full turn of an RNA helix along the PT, necessarily introducing extra strain to the linker (green). Could the polymerase be at the end of its leash? **C.** An attempt to relax the situation by introducing an unstructured polyU region (pink) into the linker. A polyU linker 3–18 Us in length was inserted into the polymerase leader at the position indicated. **D.** It was hoped that this extra slack in the tether would allow the polymerase to proceed farther along the template. **E.** However, the simple engineering attempt did not succeed. The gel shows results of *cis* polymerization assays (following DNAzyme treatment as in Fig. 6). Lane marked "–" shows DNAzyme cleavage product of ligated but unreacted ribozyme, allowing identification of the band corresponding to unextended primer. Numerous background bands on this gel reflect typical low-level degradation of un-DNAzyme-cleaved RNA (uncleaved RNA is seen at the very top of the gel, not shown here.) Ribozyme without polyU linker (lane 0) added 10–12 nt. Insertion of three or six Us into the leader sequence (at the position indicated in C, D) had no discernable effect on polymerization behavior (lanes 3, 6). Addition of twelve Us severely reduced activity level without altering the distribution of product lengths. Eighteen Us completely inactivated the ribozyme.

90
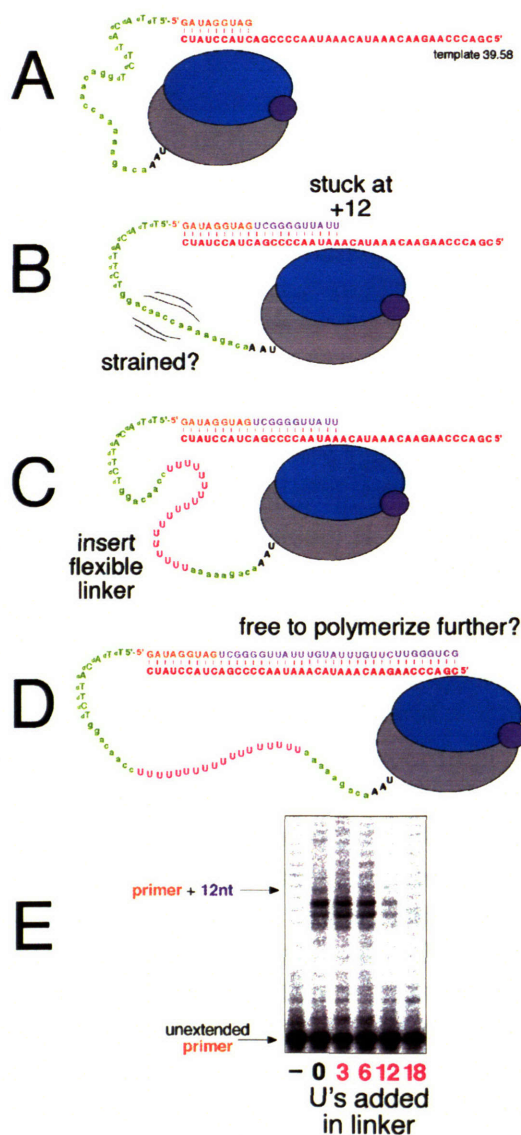
polymerase clone N1003 ligated to PT J (Table 3), which codes for 30 nt. This clone can add 10–12 nt to the PT in *cis*, as shown in the polymerization gel (Fig. 13E, lane 0), but it goes no further than that (Fig. 13B). One possible explanation is the development of strain in the linker, due to the extrusion of the lengthened PT duplex. We tried to relieve this strain by inserting a poly-U stretch into the linker (Fig. 13C). Poly-U behaves as a random coil,[37] and our hope was that the additional flexibility would free the polymerase to proceed further down the template, surpassing the apparent 10–12-nt barrier. However, the predictions of our simple model were not borne out by experiment; Fig. 13E shows that the addition of three or six U residues into the polymerase linker/leader sequence had no detectable effect on the polymerization reaction. Moreover, addition of a 12-U linker was inhibitory to the reaction without yielding any increase in length of extension products. An 18-U linker completely inactivated the polymerase.

Conclusions

Notwithstanding the challenging questions raised by the Pol+N$_{100}$ selection, it can be considered a genuine success for having validated the new capture-oligo selection technique, and for having brought Evolved Pol 2 to the light of day. We now have two different ligase-derived polymerases that catalyze robust polymerization activity with a wide variety of primer-template sequences and lengths (all tested so far, in fact). Further selections are now required to refine the activity of these two polymerases, and this will require new selection methods, such as the capture-oligo technique, which allow the experimenter to select directly for accurate and extensive polymerization, rather than addition of just one or two tagged nucleotides. Let us not forget, also, Pols 3 through 9 (Chapter 1): with a little optimization and a few extra nucleotides of random sequence, one of them could one day evolve into a *bona fide* RNA replicase, capable of synthesizing its full-length complement, and viable—at least in principle—in an RNA World organism.

Acknowledgement

References

1.   Woese, C. R. (1967) *The Origins of the Genetic Code*, Harper & Row, New York.
2.   Orgel, L. E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.* **38**:381–93.
3.   Crick, F. H. C. (1968) The origin of the genetic code. *J. Mol. Biol.* **38**:367–79.
4.   Pace, N. R. and Marsh, T. L. (1985) RNA catalysis and the origin of life. *Origins Life* **16**: 97–116.
5.   Gilbert, W. (1986) The RNA world. *Nature* **319**:618.
6.   Orgel, L. E. (1986) RNA catalysis and the origins of life. *J. Theor. Biol.* **123**:127–49.
7.   Benner, S. A., Ellington, A. D. and Tauer, A. (1989) Modern metabolism as a palimpset of the RNA World. *Proc. Natl. Acad. Sci.* **86**:7054–8.
8.   Bartel, D. P. and Unrau, P. J. (1999) Constructing an RNA world. *Trends Cell Biol* **9**: M9–M13.
9.   Bartel, D. P. (1999) "Re-creating an RNA Replicase," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., Cold Spring Harbor Laboratory Press, New York.
10.   Cech, T. R. (1986) A model for the RNA-catalyzed replication of RNA. *Proc. Natl. Acad. Sci.* **83**:4360–3.
11.   Joyce, G. F. (1991) The rise and fall of the RNA World. *New Biol.* **3**:399–407.
12.   Joyce, G. F. and Orgel, L. E. (1999) "Prospects for understanding the origin of the RNA World," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., Cold Spring Harbor Laboratory Press, New York.
13.   McGinness, K. E. and Joyce, G. F. (2003) In search of an RNA replicase ribozyme. *Chem & Biol* **10**:5–14.
14.   Breaker, R. R. (2004) Natural and engineered nucleic acids as tools to explore biology. *Nature* **432**:838–45.
15.   Wilson, D. S. and Szostak, J. W. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**:611–47.
16.   Joyce, G. F. (2004) Directed evolution of nucleic acid enzymes. *Ann. Rev. Biochem.* **73**: 791–836.
17.   Been, M. D. and R., C. T. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* **239**:1412–6.
18.   Doudna, J. A. and Szostak, J. W. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**:519–22.
19.   Doudna, J. A., Usman, N. and Szostak, J. W. (1993) Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry* **32**:2111–5.
20.   Bartel, D. P., Douda, J. A., Usman, N. and Szostak, J. W. (1991) Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Molec. Cell. Biol.* **11**:3390–4.
21.   Chowrira, B. M., Berzal-Herranz, A. and Burke, J. M. (1993) Novel RNA polymerization reaction caalyzed by a group I ribozyme. *EMBO J.* **12**:3599–605.
22.   Jaeger, L., Wright, M. C. and Joyce, G. F. (1999) A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. *Proc Natl Acad Sci* **96**:14712–7.
23.   McGinness, K. E. and Joyce, G. F. (2002) RNA-catalyzed RNA ligation on an external RNA template. *Chem & Biol* **9**:297–307.
24.   Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261**:1411–8.

25.   Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**:364–70.

26.   Ekland, E. H. and Bartel, D. P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**:3231–8.

27.   Ekland, E. H. and Bartel, D. P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**:373–6.

28.   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. *Science* **292**:19–25.

29.   Unrau, P. J. and Bartel, D. P. (1998) RNA-catalysed nucleotide synthesis. *Nature* **395**: 260–3.

30.   McGinness, K. E., Wright, M. C. and Joyce, G. F. (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.* **9**:585–96.

31.   Santoro, S. W. and Joyce, G. F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci.* **94**:4262–6.

32.   Cruz, R. P., Withers, J. B. and Li, Y. (2004) Dinucleotide junction cleavage versatility of 8–17 deoxyribozyme. *Chem & Biol* **11**:7–8.

33.   Unrau, P. J., *pers. comm.*

34.   Bausch, J. N., Kramer, F. R., Miele, E. A., Dobkin, C. and Mills, D. R. (1983) Terminal adenylation in the synthesis of RNA by Q-Beta replicase. *J. Biol. Chem.* **258**:1978–84.

35.   Clark, J. M., Joyce, C. M. and Beardsley, G. P. (1987) Novel blunt-end addition reactions catalyzed by DNA polymerase I of *Escherichia coli*. *J. Mol. Biol.* **198**:123–7.

36.   Seelig, B. and Jaschke, A. (1999) A small catalytic RNA motif with Diels-Alderase activity. *Chem & Biol* **6**:167–76.

37.   Cantor, C. R. and Schimmel, P. R. (1995) *Biophysical Chemistry: The Behavior of Biological Macromolecules*, Freeman, New York, Part 3

Supplemental Figure. RNA sequences of clones isolated in the Pol+N$_{100}$ selection, organized by subpool of origin. Boxed residues indicate mutations with respect to the pool. Sequences derived from randomized regions are printed in white on a blue background. Omitted are the two subpool C clones, which had highly disrupted ligase domains and were completely inactive in polymerization assays.
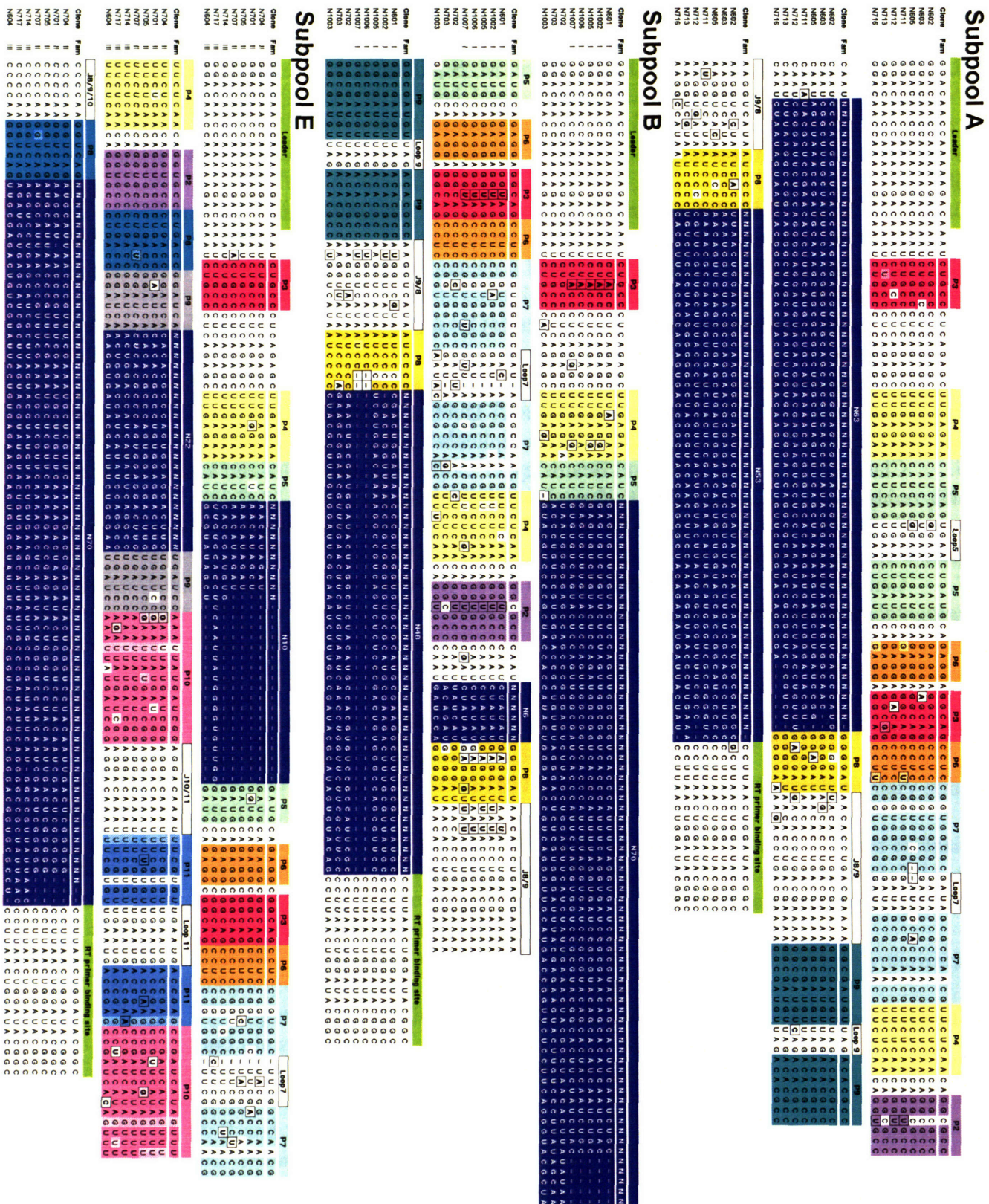
FUTURE DIRECTIONS

We are currently living through a time of rapid expansion in the list of roles for nucleic acids. RNA is now recognized at the heart of modern biochemistry as the catalytic component of the ribosome,[1–3] new natural ribozymes are suddenly being discovered again,[4–6] and previously undemonstrated RNA-catalyzed activities are being produced by *in vitro* selection.[7] New selection techniques are being developed to explore the catalytic properties of plausible prebiotic precursors of RNA.[8] The ability of RNA to exert biochemical control via allosteric processes both in nature[9] and in engineered settings[10] is becoming increasingly appreciated. A whole new class of tiny natural RNAs have been discovered to act in gene regulation pathways conserved throughout vast lineages of animals.[11–13] The templating properties of DNA are being explored in the context of *in vitro* selection and combinatorial chemistry in ways never before imagined.[14,15]

Against this dramatic backdrop, we can consider the project of producing an RNA replicase. Besides providing crucial support for the RNA World hypothesis,[16–18] and significantly enhancing the prestige of RNA as a catalytic polymer, an RNA replicase could be employed as the central component of the first artificial lifeforms.[19,20] Constructing these minimal biotic systems in the laboratory would open whole new avenues of evolutionary research.

With it now well established that RNA is capable of general RNA-templated RNA synthesis, we can emphasize a few fruitful avenues of enquiry that may finally yield a true RNA replicase ribozyme. First:

Learn more about what we've got. At the heart of all the polymerase ribozymes described herein is an RNA ligase ribozyme, one of the fastest ribozymes known.[21–27] Although a tentative model of its structure has been proposed,[28] there can be no substitute for the detailed structural and mechanistic insights that will emerge once we finally get a look at its atomic-resolution structure. A breakthrough in this area is considered imminent, thanks to the tireless efforts of David Shechner in the Bartel Lab.

Another line of work currently underway involves the use of nucleotide analog interference mapping[29,30] to investigate how the two modules of the polymerase (ligase and auxiliary domain) might interact with each other. Insights from this work by Sarah Bagby may yield crucial clues about how we can possibly improve the efficiency of the polymerase.

Longer polymerization. At the moment, the limiting factor of our polymerase ribozymes is their low affinity for primer-template.[31] This affinity must be improved, if we are ever to demonstrate the long polymerization that would be required of an RNA replicase.

One way to imagine doing this would be by designing an *in vitro* selection experiment to isolate a double-stranded-RNA-binding activity. This could be done in the straightforward manner of selecting aptamers[32,33] that involves immobilization of the target (in this case dsRNA) on a column, and selective retention of those pool members with affinity for the target. If a strong aptamer for dsRNA that recognized its target in a sequence-nonspecific manner could be isolated, it would be an attractive candidate for combining with the polymerase to produce a chimeric molecule with greater PT affinity. (This might involve doing a selection to optimize the interaction of the two domains.)

A second approach would be more along the lines of the experiment reported in Chapter 4, wherein random sequence was added to the existing polymerase, and a selection for long polymerization was performed. If we are correct in our assumption that the limiting factor for the polymerase is currently its weak affinity, then maybe the best approach is just to press on with

the selections, because the only way the ribozyme can achieve improvement is by improving its substrate binding properties. It remains to be seen how much patience this approach will require.

Another possible strategy for achieving longer polymerization would be to try to find a polymerase that works under milder reaction conditions–say pH 7 instead of pH 8.5. At neutral pH, with decreased base hydrolysis, the polymerase would survive longer and have more time to complete its task. However, the current polymerase has a pH optimum around 8.5, and its polymerization at pH 7.0 is considerably slower (the average rate fell by a factor of 16, when polymerization was measured using the series of seven PTs used in Chapter 2). If this loss of activity at neutral pH could be prevented, then the polymerase might be able to polymerize longer stretches of RNA under those conditions. One possible way of doing this would be to replace its current Class I RNA ligase core with an ligase variant that was evolved to function at reduced pH.[34]

Strand dissociation. One problem that would have faced any replicase in the RNA World is the difficulty of dissociating a newly synthesized RNA strand from the strand that templated it.[35] Before a complete RNA replication system can be demonstrated, this problem will have to be solved. One possible approach would be to conduct an *in vitro* selection experiment for a helicase ribozyme. That's one activity that still hasn't yet been shown to fall within the catalytic repertoire of RNA, and it would constitute an impressive addition. One might imagine doing this by requiring pool molecules to unwind a section of themselves in order to be propagated in the selection, perhaps by adhering to a mineral surface with specific affinity for single-stranded nucleic acids.

"How to get the Darwinian ball rolling." Even when a fully competent RNA replicase is eventually produced, there will still remain the question of how a collection of them could evolve by natural selection. If the best replicases copied their unrelated neighbors just as avidly as copying themselves, then they would have no selective advantage. Evolution relies on the coupling between fitness and amplification. Several possible scenarios have been suggested to explain how replicators might have evolved before the emergence of cellular enclosures. For example, they might have lived on a two-dimensional surface with limited lateral dispersal, a scenario that has been simulated computationally[36] and seems plausible, but which would be far more valuable in the form of an experimental demonstration.

A second proposed mechanism for the preferential replication of "self" molecules invokes the involvement of a "genetic tag".[37] Once we have replicators of a certain minimal efficiency, it would be interesting to see if we could make them recognize their own kind by employing this genetic tag approach.

A third avenue of exploration into this topic has been opened by Uli Müller in the Bartel Lab, whose experiments with ribozyme polymerization on hydrophobic assemblies have raised interesting questions about possible mechanisms of replicase evolution. These hydrophobically linked associations can be seen as a kind of "inside-out cell," and they serve a similar function to enclosure by a cell membrane, in increasing the average length of time that ribozymes related by descent spend with each other, thereby increasing the proportion of instances in which a replicase copies one of its own descendants rather than an unrelated RNA. This system provides an attractive paradigm for possible mechanisms of early evolution in the RNA World.

The ultimate archaeological find. Throughout the origin-of-life literature there runs a theme of muted hopeful longing, almost sheepishly confessed, of one day uncovering actual surviving examples of our most ancient ancestors. This cautious optimism transcends the ideological boundaries that have traditionally divided the field into opposing "metabolism first" and "replication first" camps; we hear it when Wächtershäuser remarks that "surface metabolists might still live in regions unperturbed by cellular forms of life, and they could be detected, for example, by staining,"[38] and we hear it when Joyce suggests that "there even is a chance that a remnant of the RNA world will be found lurking in some special contemporary microenvironment."[39] Such a find would be a spectacular advance for all of science, and recent startling discoveries highlighting the existence of unusual lifeforms very deep beneath the surface of the earth suggests that might be a good place to look.[40-44]

Skeptics may find a kindred soul in Orgel, who admitted, "It is just possible that evidence concerning the early stages in the origins of life is still accessible on earth. Such evidence might consist either of preserved organic chemicals which predate biosynthesis or of the fossils of transitional organisms which can be shown to have possessed a primitive genetic system." He concluded, however, "I think the prospects of making such finds are poor."[45] Let those optimists among us weigh his assessment alongside another of his predictions from the same manuscript: "Could polynucleotide chains with well-defined secondary structures act as primitive enzymes? I doubt that they alone could exhibit extensive catalytic activity, although one cannot be quite sure."

References

1.   Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**:905–20.

2.   Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**:920–30.

3.   Steitz, T. A. and Moore, P. B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trens Biochem Sci* **28**:411–8.

4.   Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**:281–6.

5.   Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N. J. and Akoulitchev, A. (2004) Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**:526–30.

6.   Knudsen, S. M. and Ellington, A. D. (2004) Ribozyme deja vu. *Nature Struct. Biol.* **11**: 301–3.

7.   Tsukiji, S., Pattnaik, S. B. and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nature Struct. Biol.* **10**:713–7.

8.   Ichida, J. K., Zou, K., Horhota, A., Yu, B., McLaughlin, L. W. and Szostak, J. W. (2005) An in vitro selection system for TNA. *J. Am. Chem. Soc.* **127**:2802–3.

9.   Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L. and Breaker, R. R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**:275–9.

10.   Zivarts, M., Liu, Y. and Breaker, R. R. (2005) Engineered allosteric ribozymes that respond to specific divalent metal ions. *Nucleic Acids Res.* **33**:622–31.

11.   Lau, N. C., Lim, L. P., Weinstein, E. G. and Bartel, D. P. (2001) An abundant class of tiny RNAs with probably regulatory roles in Caenorhabditis elegans. *Science* **294**:858–62.

12.   Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**:281–97.

13.   Baskerville, S. and Bartel, D. P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**:241–7.

14.   Gartner, Z. J., Tse, B. N., Grubina, R., Doyon, J. B., Snyder, T. M. and Liu, D. R. (2004) DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**:1601–5.

15.   Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. and Liu, D. R. (2004) Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **431**.

16.   Bartel, D. P. and Unrau, P. J. (1999) Constructing an RNA world. *Trends Cell Biol* **9**: M9–M13.

17.   Levy, M. and Ellington, A. D. (2001) RNA world: Catalysis abets binding, but not vice versa. *Curr. Biol.* **11**:R665–7.

18.   McGinness, K. E. and Joyce, G. F. (2003) In search of an RNA replicase ribozyme. *Chem & Biol* **10**:5–14.

19.   Szostak, J. W., Bartel, D. P. and Luisi, P. L. (2001) Synthesizing life. *Nature* **409**:387–90.

20.   Chen, I. A., Roberts, R. W. and Szostak, J. W. (2004) The emergence of competition between model protocells. *Science* **305**:1474–6.

21.   Bartel, D. P. and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* **261**:1411–8.

22.   Ekland, E. H., Szostak, J. W. and Bartel, D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**:364–70.

23.   Ekland, E. H. and Bartel, D. P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**:3231–8.

24.   Ekland, E. H. and Bartel, D. P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**:373–6.

25.   Bergman, N., Johnston, W. and Bartel, D. P. (2000) Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**:3115–23.

26.   Glasner, M. E., Yen, C. C., Ekland, E. H. and Bartel, D. P. (2000) Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry* **39**:15556–62.

27.   Glasner, M. E. (2003) Nucleotide recognition and metal ion requirements of an RNA ligase ribozyme, *Ph. D. thesis* in Biology, MIT, Cambridge, MA.

28.   Bergman, N. H., Lau, N. C., Lehnert, V., Westhof, E. and Bartel, D. P. (2004) The three-dimensional architecture of the class I ligase ribozyme. *RNA* **10**:176–84.

29.   Strobel, S. A. and Shetty, K. (1997) Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping. *Proc. Natl. Acad. Sci* **94**: 2903–8.

30.   Soukup, J. K., Minakawa, N., Matsuda, A. and Strobel, S. A. (2002) Identification of A-minor tertiary interactions within a bacterial group I intron active site by 3-deazaadenosine interference mapping. *Biochemistry* **41**:10426–38.

31.   Lawrence, M. S. and Bartel, D. P. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* **42**:8748–55.

32.   Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment:

RNA ligands to bacteriophage T4 DNA Polymerase. *Science* **249**:505–10.

33.   Ellington, A. D. and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**:818–22.

34.   Kuhne, H. and Joyce, G. F. (2003) Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH. *J. Mol. Evol.* **57**:292–8.

35.   Bartel, D. P. (1999) "Re-creating an RNA Replicase," in *The RNA World*, (R. F. Gesteland, T. R. Cech and J. F. Atkins, Eds.), 2nd ed., Cold Spring Harbor Laboratory Press, New York.

36.   Szazbo, P., Scheuring, I., Czaran, T. and Szathmary, E. (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**: 340–3.

37.   Weiner, A. M. and Maizels, N. (1987) tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci.* **84**:7383–7.

38.   Wächtershäuser, G. (1988) Before enzymes and templates: Theory of surface metabolism. *Microb. Rev.* **52**:452–84.

39.   Joyce, G. F. (2002) The antiquity of RNA-based evolution. *Nature* **418**:214–21.

40.   Cowen, J. P., Giovannoni, S. J., Kenig, F., Johnson, H. P., Butterfield, D., Rappe, M. S., Hutnak, M. and Lam, P. (2003) Fluids from aging ocean crust that support microbial life. *Science* **299**:120–3.

41.   Webster, G., Parkes, R. J., Fry, J. C. and Weightman, A. J. (2004) Widespread occurrence of a novel division of bacteria identified by 16S rRNA gene sequences originally found in deep marine sediments. *Appl. Environ. Microbiol.* **70**:5708–13.

42.   D'Hondt, S., Jorgensen, B. B., Miller, D. J., et al. (2004) Distributions of microbial activities in deep subseafloor sediments. *Science* **306**:2216–21.

43.   Todo, Y., Kitazato, H., Hashimoto, J. and Gooday, A. J. (2005) Simple foraminifera flourish at the ocean's deepest point. *Science* **307**:689.

44.   Schippers, A., Neretin, L. N., Kallmeyer, J., Ferdelman, T. G., Cragg, B. A., Parkes, R. J. and Jorgensen, B. B. (2005) Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* **433**:861–4.

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

APPENDIX

RNA-catalyzed RNA polymerization:
accurate and general RNA-templated primer extension

RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

Wendy K. Johnston, Peter J. Unrau,* Michael S. Lawrence, Margaret E. Glasner, David P. Bartel†

The RNA world hypothesis regarding the early evolution of life relies on the premise that some RNA sequences can catalyze RNA replication. In support of this conjecture, we describe here an RNA molecule that catalyzes the type of polymerization needed for RNA replication. The ribozyme uses nucleoside triphosphates and the coding information of an RNA template to extend an RNA primer by the successive addition of up to 14 nucleotides—more than a complete turn of an RNA helix. Its polymerization activity is general in terms of the sequence and the length of the primer and template RNAs, provided that the 3' terminus of the primer pairs with the template. Its polymerization is also quite accurate: when primers extended by 11 nucleotides were cloned and sequenced, 1088 of 1100 sequenced nucleotides matched the template.

The RNA world hypothesis states that early life forms lacked protein enzymes and depended instead on enzymes composed of RNA (1). Much of the appeal of this hypothesis comes from the realization that ribozymes would have been far easier to duplicate than proteinaceous enzymes (2–5). Whereas coded protein replication requires numerous macromolecular components [including mRNA, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, and the ribosome], replication of a ribozyme requires only a single macromolecular activity: an RNA-dependent RNA polymerase that synthesizes first a complement, and then a copy of the ribozyme. If this RNA polymerase were itself a ribozyme, then a simple ensemble of molecules might be capable of self-replication and eventually, in the course of evolution, give rise to the protein-nucleic acid world of contemporary biology. Finding a ribozyme that can efficiently catalyze general RNA polymerization would support the idea of the RNA world (1, 6, 7) and would provide a key component for the laboratory synthesis of minimal life forms based on RNA (8, 9).

Although progress has been made in finding ribozymes capable of template-directed RNA synthesis, none of these ribozymes has the sophisticated substrate-binding properties needed for general polymerization (7). Derivatives of self-splicing introns are able to join oligonucleotides assembled on a template (10–12). However, the templates that can be copied are limited to those that match the oligonucleotide substrates, and it has not been possible to include sufficient concentrations of all the oligonucleotide substrates needed for a general copying reaction. More recently, efforts have shifted to derivatives of an RNA-ligase ribozyme that was isolated from a large pool of random RNA sequences (13–15). Some derivatives are capable of template-directed primer extension using nucleoside triphosphate (NTP) substrates, but their reaction is also limited to a small subset of possible template RNAs (15). These ligase derivatives recognize the primer-template complex by hybridizing to a particular unpaired segment of the template (Fig. 1A). In using a short segment of a special template to direct primer extension, these ribozymes resemble telomerases more than they resemble the enzymes that replicate RNA and DNA by means of general polymerization.

Polymerase isolation. We have used the catalytic core of the ligase ribozyme (14, 16) as a starting point for the generation of a ribozyme with general RNA polymerization activity. The new polymerase ribozyme was isolated from a pool of over 10^{15} different RNA sequences. Sequences in the starting pool contained a mutagenized version of the parental ligase (Fig. 1B). To sample a broad distribution of mutagenesis levels, the starting pool comprised four subpools in which the core residues of the ligase domain were mutagenized at levels averaging either 0, 3, 10, or 20% (17). Two loops within the ligase domain, both unimportant for ligase function, were replaced with 8-nucleotide (nt) random-

sequence segments (Fig. 1B). The 5' terminus of the ligase domain was covalently attached to an RNA primer so that molecules able to catalyze primer extension could be selected by virtue of their attachment to the primer that they extended.

In contrast to the parental ribozyme, which hybridizes to the template, a ribozyme capable of general polymerization must recognize the primer-template complex without relying on sequence-specific interactions. Therefore, the template RNA was designed to be too short for extensive hybridization with the ribozyme (Fig. 1B). For the parental ribozyme (Fig. 1A), the pairing between the ribozyme and the template also comprised a stem known to be necessary for ligase function (16). To restore this stem, a short RNA, GGCACCA (purple RNA in Fig. 1B), was introduced to hybridize to the segment of the ligase domain that formerly paired with the template. Finally, because a more general mode of primer-template recognition might require the participation of an additional RNA domain, a 76-nt random-sequence segment was appended to the 3' terminus of the degenerate ligase domain (Fig. 1B).

Sequence variants able to recognize the primer-template in this new configuration and then extend the primer with tagged nucleotides were enriched by repeated rounds of in vitro selection and amplification (Table 1). RNAs that extended their primer by using 4-thioUTP were isolated on APM gels (urea denaturing gels cast with a small amount of N-acryloyl-aminophenylmercuric acetate, which impedes migration of RNA containing 4-thioU) (18). To favor variants that recognize generic rather than sequence-specific features of the primer-template, different primer-template sequences and lengths were used in different rounds of selection (Table 1). To favor the more efficient ribozyme variants, the stringency of the selection was increased in later rounds by requiring addition of two tagged nucleotides, such as biotinylated A and 4-thioU (19), and by decreasing the time of incubation with the tagged NTPs (Table 1).

After eight rounds of selection and amplification, desirable variants had increased in abundance to the point that a detectable fraction of the pool molecules could extend their primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion (Fig. 2). Other variants able to tag themselves were detected as early as round four, but most of these ribozymes added tagged nucleotides in the absence of the template oligonucleotide, or they decorated themselves at sites other than the 3' terminus of the primer (20). Seventy-four variants from rounds 8 through 10 were cloned and found to represent 23 sequence families, each family having descended from a different ancestral sequence of the starting pool. Ribozymes from two families extended their

Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*Present address: Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada.
†To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

102

primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion. These two families are represented by isolates 9.1 and 10.2 (Fig. 2). Isolate 10.2 (Fig. 1C), from round 10, was from the more prevalent and active of these two families and was chosen for further study.

Polymerization with multiple turnover. Having isolated a ribozyme that did not rely on forming base pairs with the template RNA during primer extension, we next determined whether it instead recognized the particular sequence used to link the primer to the ribozyme. We were pleased to discover that the round-10 ribozyme did not require this sequence. In fact, it did not require any covalent attachment to the primer. When incubated with a 10-fold excess of both a 6-nt primer and a 9-nt template, as well as the appropriate

nucleoside triphosphate, the round 10 polymerase fully extended the primer, with multiple turnover (Fig. 3). The primer and template sequences in this experiment were designed to differ from those used most frequently during the selection (Table 1, aligning the sequences relative to the primer 3' termini). Despite the complete change in the sequence of the primer-template complex, the polymerase isolate was able to recognize the complex and extend the primer (21). This indicates that the ribozyme binds the primer-template without relying upon recognition of a particular sequence.

The new mode of primer-template recognition appears to be conferred by the accessory domain derived from the 76-nt random-sequence segment and the 3'-terminal segment that binds the primer used for the reverse tran-

scription–polymerase chain reaction (RT-PCR). Without the accessory domain, the ligase domain of the round-10 ribozyme, like the parental ligase itself (Fig. 3B), displayed no activity in polymerization assays requiring general primer-template recognition. Indeed, deleting only 9 nt from the 3' terminus of the round-10 ribozyme severely diminished activity (20). It is interesting that the core ligase residues emerged unchanged in this round-10 isolate (compare Fig. 1, B and C), and the GGCACCA oligonucleotide designed to complete the ligase domain proved to be necessary for polymerase function (20). This suggests that the parental ligase did not need to adapt in order to accommodate the more general primer-template recognition afforded by the accessory domain.

The round-10 ribozyme was tested with numerous other primer-template pairs. In all cases

Fig. 1. Secondary structure models of the ribozymes. Short dashes indicate base pairs. (**A**) A ribozyme (black strand) able to promote limited RNA polymerization (15). It extends an RNA primer (orange strand) by using nucleoside triphosphates and coding information from an appropriate RNA template (red strand). The ribozyme can accommodate any of the four RNA nucleotides at residues indicated by an X, provided that the primer pairs with the template. However, the 5' portion of the template must pair with the ribozyme. The depicted ribozyme was derived from an RNA ligase ribozyme; black uppercase residues and defined residues of the template comprise the core of the ligase ribozyme. (**B**) A pool of RNA sequences based on the ligase ribozyme (17). Colors differentiate residues representing the ligase core (black, purple), random sequence (blue), primer (orange), template (red), and RT-PCR primer-binding sites (green). Residues prefixed by "d" are DNA; all others are RNA. The 5' end of the RNA primer is covalently joined to the 5' end of each pool molecule via a phosphodiester linkage (-5'-5'-) (38). The sequence of the primer-template (X) in a given round usually differed from that of the previous round (Table 1). (**C**) The round-10 ribozyme (isolate 10.2). Residues derived from the random-sequence segments or the 3' RT-PCR primer-binding site of the starting pool are colored blue; other drawing conventions are as in (B). Comparative sequence analysis of improved isolates from rounds 14 and 18 (23) supports the importance, as well as the proposed secondary structure, of the accessory domain (residues 110 to 204), particularly within the 3' region of this domain (residues 150 to 201). Blue uppercase residues were invariant among all 22 improved isolates. Because the chance conservation of a residue not important for activity is low ($P = 0.0074$ for conservation in 22 of 22 isolates), nearly all 29 of these residues must be important for ribozyme function. Thick blue dashes mark covarying pairs, five of which (G151:C200, A153:U198, C154:G197, U175:A183, and C176:G182) support the proposed pairing within the 3' region of the accessory domain. (**D**) The round-18 ribozyme, a shortened derivative of an improved isolate from round 18. Nucleotide changes from the round-10 isolate that arose from combinatorial mutagenesis are in pink; changes engineered when reducing the ribozyme's size are in gray (23). The four changes consistently found among the improved round-14 and round-18 isolates are in uppercase pink. Other drawing conventions are as in (C).

it was able to recognize the primer-template complex and to extend the primer by the Watson-Crick match to the template. Extension across from C or G template residues was usually more efficient and accurate than extension across from A or U. Extension was also much more efficient when the unpaired portion of the template was shorter than 5 nt (Fig. 4B).

Sequence optimization of the polymerase. To find improved polymerase variants, the in vitro selection procedure was continued for another eight rounds of selection and amplification, starting with a newly synthesized pool of variants based on the round-10 isolate (22). In this pool, the accessory domain, as well as the two 8-nt segments at the

loops within the ligase domain, were mutagenized at an average level of 20%. Because this mutagenesis included the former RT-PCR primer-binding sequence, a new RT-PCR primer-binding sequence was appended to each pool molecule. In half the pool molecules, most of the ligase core residues were mutagenized at an average level of 3%, whereas in the other half, they were not intentionally mutagenized.

The additional rounds of selection-amplification were performed with three noteworthy modifications to the protocol of the first 10 rounds (Table 1). First, longer template RNAs were used to favor those variants better able to recognize primer-template complexes with long, unpaired template segments. Second, selection was based on the ability to add two tagged U's rather than one tagged U and one tagged A. This change was implemented after learning that the round-10 ribozyme uses biotinylated ATP much less efficiently than unmodified ATP. Third, high concentrations of unmodified ATP, CTP, and GTP were included to disfavor those variants prone to incorporating these competitor nucleotides instead of the tagged Watson-Crick match to the template.

Isolates from rounds 14 to 18 were screened for the ability to fully extend a 10-nt primer (GAAUCAAGGG) on an 18-nt template (3'-CUUAGUUCCCGCCCGGCC, underline indicates segment that pairs with the primer). Most isolates from round 18 had disrupted ligase domains and showed no sign of polymerase activity when assayed individually. They presumably were selected because of a parasitic ability to deliver their primer to the active site of a different molecule. Other isolates had polymerase activity and were much more active than the round-10 parental ribozyme. Comparative sequence analysis of the 22 most active isolates (23) identified conserved residues and structural features, which clustered in the 3'-terminal half of the accessory domain (Fig. 1C) and are likely to be critical for its function. This analysis also suggested a model for the secondary structure of the accessory domain (Fig. 1C) and identified four residues in the domain that consistently differed from the round-10 ribozyme (23). These mutations likely conferred increased polymerase activity.

One isolate from round 18 was particularly adept at using long templates. To investigate features of the ribozyme needed for activity, derivatives of this isolate were constructed and tested (23). A 189-nt derivative (Fig. 1D) that retains the polymerization activity of the full-length round-18 isolate has been most extensively characterized. This derivative (hereafter referred to as the round-18 ribozyme) has all the features of the accessory domain that were conserved among the 25 most active isolates, including the four mutations thought to confer improved activity (Fig. 1D). Additionally, it has
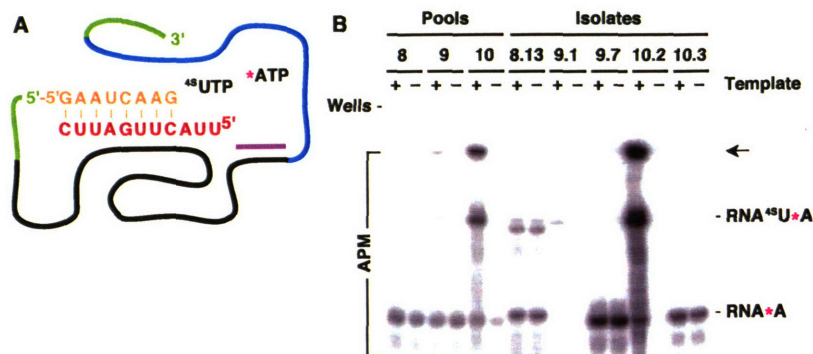
Fig. 2. Detection of ribozymes able to extend an attached primer by two nucleotides in a template-dependent manner. (**A**) Schematic of the RNAs in this experiment. Ribozymes attached to an RNA primer (orange) were incubated with 1 mM 4-thioUTP (^{4S}UTP) and trace [α-^{32}P]ATP (*ATP), in the presence or absence of an RNA template (red) that codes for the addition of U and A. (**B**) Activities of ribozyme pools and isolates after 8 to 10 rounds of selection. Extension reactions were for 12 hours, under the conditions used during the rounds of selection (18). Shown is a PhosphorImager scan of an APM denaturing gel separating RNAs extended with radiolabeled A (RNA*A) from those extended with both 4-thioU and radiolabeled A (RNA^{4S}U*A). The arrow points to RNA^{4S}U*A extended by a second 4-thioU, which did not migrate into the APM portion of the gel. Note that addition of the second ^{4S}U could not have been directed by an A in the template because only one of the template coding residues is an A; some misincorporation of ^{4S}U was expected in this experiment because of the very large excess of ^{4S}UTP over ATP. The sequence families represented by 9.1 and 10.2 add both 4-thioU and radiolabeled A in a template-dependent manner. The round-10 isolate (10.2) was chosen for further analysis and is shown in Fig. 1C.

Table 1. Parameters and substrates for in vitro selection. For each round of selection, pool RNA with covalently attached primer (38) was incubated with the indicated template RNA and NTPs for the indicated time (18). Nucleotide analogs 4-thioU, biotin-N6-A, and 2-aminopurine (39) are abbreviated ^{4S}U, ^{B}A, and ^{2N}P, respectively. The primer attached to the pool molecules was complementary to the underlined segment of the template. Variants with polymerase activity were selected based on their primer being extended with the tagged nucleotides indicated in the Selection criteria column (18). In late rounds, 2 mM ATP, 2 mM CTP, and 2 mM GTP were included as competitor NTPs (Comp. NTPs). Pool mutagenesis was either during chemical synthesis (Synthesis) or during error-prone amplification (PCR) of the template DNA (17, 22, 40).

Round	Mutagenesis	Template RNA	NTPs	Time (hour)	Selection criteria
1	Synthesis	3'-GGUCAGAUU	^{4S}UTP (2 mM)	36	^{4S}U
2	None	3'-GGUCAGAACC	^{4S}UTP (2 mM)	20	^{4S}U
3	None	3'-GGUCAGAA	^{4S}UTP (2 mM)	20	^{4S}U
4	None	3'-CUUAGUUCAUU	^{4S}UTP (2 mM)	19	^{4S}U
5	None	3'-CUUAGUUCAUU	^{4S}UTP (2 mM)	1	^{4S}U
6	None	3'-GGUCAGAUU	^{4S}UTP, ^{B}ATP (1 mM each)	14	^{B}A, ^{4S}U
7	None	3'-CUUAGUUCAUU	^{4S}UTP, ^{B}ATP (1 mM each)	17	^{B}A, ^{4S}U
8	None	3'-GGUCAGAUU	^{4S}UTP, ^{B}ATP (1 mM each)	17	^{B}A, ^{4S}U
9	None	3'-GGUCAGAUU	^{4S}UTP, ^{B}ATP (1 mM each)	4	^{B}A, ^{4S}U
10	None	3'-CUUAGUUCAUU	^{4S}UTP (1 mM)	20	^{4S}U
11	Synthesis	3'-UCGACGGAACC	^{4S}UTP (1 mM)	4	2 ^{4S}U
12	None	3'-ACCUGAGAAGG	^{4S}UTP (1 mM)	4	2 ^{4S}U
13	None	3'-CAAGUCCAACC	^{4S}UTP (1 mM)	0.2	2 ^{4S}U
14	None	3'-UCGACGGAACC	^{4S}UTP (1 mM)	0.2	2 ^{4S}U
15	PCR	3'-UCGACGG^{2N}P^{2N}PCCUGCGUC	^{4S}UTP (0.1 mM), Comp. NTPs	20	2 ^{4S}U
16	PCR	3'-CAAGUCC^{2N}P^{2N}PUGAUCGUA	^{4S}UTP (0.1 mM), Comp. NTPs	4	2 ^{4S}U
17	PCR	3'-ACCUGAG^{2N}P^{2N}PGUGUAUGU	^{4S}UTP (0.1 mM), Comp. NTPs	2	2 ^{4S}U
18	None	3'-UCGACGG^{2N}P^{2N}PCCUGCGUC	^{4S}UTP (0.1 mM), Comp. NTPs	0.1	2 ^{4S}U

104

a U-to-C mutation within the ligase domain in the segment designed to pair with the 7-nt RNA that completes the ligase domain (Fig. 1D). Reversing this point mutation diminished activity, and omitting the 7-nt RNA abolished activity altogether. We therefore speculate that, although the 7-nt RNA must still pair to this segment, a non–Watson-Crick distortion of the helix better accommodates long template RNAs. It is noteworthy that the four other active round-18 isolates also had point mutations within the segment designed to pair with the 7-nt RNA (23).

Extensive and accurate RNA polymerization. Although the round-18 ribozyme was only marginally improved over the round-10 ribozyme when short templates were used (Fig. 3), it was much better when longer templates were used (Fig. 4). With templates coding for 4, 8, 11, and 14 nucleotides, the round-18 ribozyme extended the primer by the corresponding number of residues (Fig. 4B). Normal RNA linkages were synthesized, as determined by nuclease analysis of the extension product (23). Furthermore, extension was predominantly by the Watson-Crick match to the template. When primers that were fully extended using the template coding for 11 nt were cloned and sequenced, 89 of 100 sequences precisely matched the template. Of the 1100 residues sequenced, only 12 were mismatches (Fig. 4C), implying an overall Watson-Crick error rate of 0.011 per nucleotide. Thus, the round-18 ribozyme can accurately use information from an RNA template and all four nucleoside triphosphates to extend an RNA primer by a complete turn of an RNA helix.

To examine the accuracy of polymerization more systematically, we measured the efficiency of matched and mismatched extension using four templates that differed only at the first coding nucleotide. For each template, the Watson-Crick match was added most efficiently (Table 2). The best fidelity was with the -C- template, for which the error rates ranged from 0.00004 to 0.0002. Fidelity was lower for the -G- and -U- templates, primarily because extension by the two G:U wobble mismatches had error frequencies of 0.044 and 0.085. The overall fidelity was 0.967 per residue. In other words, with all four NTPs supplied at equimolar concentrations, extension by the matched nucleotide typically would be 96.7% of the total extension.

A fidelity of 96.7% (Table 2) is somewhat lower than the 99% fidelity inferred from sequencing fully extended primer molecules (Fig. 4C). Two factors contribute to the higher fidelity observed in Fig. 4C. The first is the influence of sequence context on fidelity (24). The second arises from the fact that, after a mismatch was incorporated, further extension of the growing chain was less efficient because the 3' terminus of the primer no longer paired with

the template (25). Thus, the full-length product of Fig. 4C was enriched in molecules with few misincorporated nucleotides. Mismatch incorporation also reduces the extension efficiency of proteinaceous polymerases, a property particularly important for certain DNA polymerases because it facilitates exonucleolytic proofreading (26).

Polymerase fidelity is most simply expressed by assuming that all four NTPs are present at equal concentration, even though cellular NTP concentrations are not equimolar (27). For the round-18 ribozyme, certain asymmetric NTP ratios would produce observed fidelities significantly greater than 0.967. For example, lowering the GTP concentration to one-tenth that of the other NTPs would decrease G misincorporation by 10-fold, while

only lowering the fidelity of extension across from C from 0.9996 to 0.996. Because G misincorporation was the major source of error, this would increase the overall fidelity from 0.967 to 0.985 with the templates in Table 2.

A Watson-Crick fidelity of 0.985 is still lower than the ≥0.996 fidelity seen with viral polymerases that replicate RNA by using RNA templates (28, 29), and it is much lower than that seen for polymerases that replicate DNA (30). Nevertheless, the Watson-Crick fidelity of the round-18 ribozyme compares favorably to that of other ribozymes. Previously, the best ribozyme fidelity had been obtained with the engineered ligase derivative (Fig. 1A), which has an overall fidelity of 0.85 with equimolar NTPs and observed fidelities of 0.88 to 0.92 with more favorable
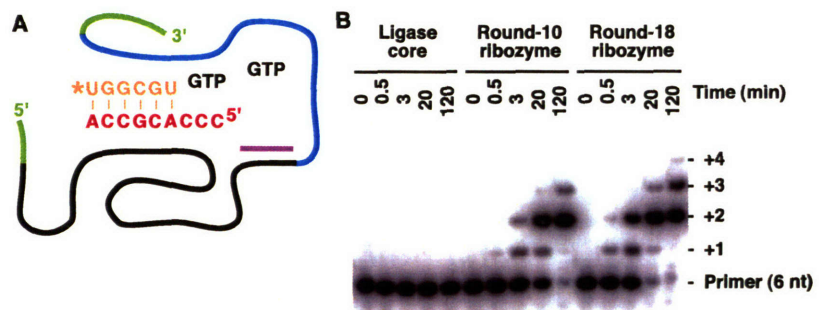
Fig. 3. Intermolecular primer extension using a short primer-template. (**A**) Schematic of the RNAs used in these polymerization reactions. Drawing conventions are as in Fig. 2A. Note that the primer is ^{32}P end-labeled and that neither the primer RNA nor the template RNA is tethered or hybridized to the ribozyme. (**B**) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 1 μM ribozyme, 10 μM primer, 10.5 μM template, and 4 mM GTP, and were incubated for the indicated time in polymerization assay conditions (33). "Ligase core" refers to a ribozyme identical to that of Fig. 1A (black strand), except that its 3' terminus was modified to pair with the 7-nt RNA (GGCACCA) that completes the ligase core; no extension was observed with this ribozyme. The round-10 and round-18 ribozymes are depicted in Figs. 1C and D, respectively. After long incubation times, some of the primer was extended with three templated residues plus one nontemplated residue. Many proteinaceous polymerases, including Qβ replicase (42) and Taq DNA polymerase (43), also tend to add an extra, nontemplated residue.

Table 2. Watson-Crick fidelity of RNA polymerization. For each template-NTP combination, the efficiency of extension by at least 1 nt was determined. For each template, the four efficiencies were normalized to that of the matching NTP, yielding the relative efficiencies of extension. The relative efficiency of extension for a mismatch is the same as its error rate (27) and misinsertion ratio (26). Fidelities were calculated as the efficiency for the match, divided by the sum of the efficiencies for all four NTPs. The average fidelity is the geometric average of the fidelities for each template (41). For each Watson-Crick match, the absolute efficiency (per molar per minute) is also shown in parentheses. It is reported as the observed rate constant of primer extension divided by NTP concentration, from polymerase assays (33) using 5 μM ribozyme, 2 μM primer (CUGCCAACCG), and 2.5 μM template (3'-GACGGUUGGCXCGCUUCG, where X is the indicated template residue). In these assays, NTPs were supplied at concentrations well below half-saturating.

Template	Relative efficiency of extension				Fidelity
	ATP	CTP	GTP	UTP	
-A-	0.0034	0.0014	0.0043	1.0 (5.3)	0.991
-C-	0.0002	0.0002	1.0 (5.4)	0.00004	0.9996
-G-	0.0002	1.0 (41)	0.0006	0.044	0.957
-U-	1.0 (87)	0.0001	0.085	0.0002	0.921
				Average =	0.967

105

NTP ratios (15). Moreover, the fidelity of the round-18 ribozyme approaches that of one proteinaceous polymerase, pol η, a eukaryotic polymerase needed for accurate replication of UV-damaged DNA (31). Yeast pol η has an overall fidelity of 0.984, which would increase to 0.989 with an optimal NTP ratio (32).

General RNA-dependent RNA polymerization. The round-18 ribozyme worked with every primer-template tested. As the primer-templates have no sequence features in common, the ribozyme does not rely on any sequence-specific contacts. Additionally, because the primer-template complex must shift in register relative to the polymerase active site each time another nucleotide is added, every polymerization experiment actually examines primer extension in a series of differing sequence contexts, demonstrating further that the polymerization is general with respect to nucleotide sequence. Granted, the efficiency of nucleotide addition varied depending on the sequence context, as evidenced by an uneven distribution of extension intermediates (Fig. 4), but this phenomenon is also observed with protein polymerases (26, 27).

All templates used heretofore were less than 21 nt long, leaving open the question of whether the ribozyme could accommodate longer primer-template substrates, as would be required of an RNA replicase. To address this question, three related substrates were tested. The first was a short substrate, with a 10–base pair primer-template duplex and a 10-nt template coding region. The second substrate was the same, except its template coding region was lengthened from 10 to 100 nt. The ribozyme extended this substrate by as many as 9 nt in 23 hours, although somewhat less efficiently than it extended the short version (Fig. 5). The third substrate was the same as the second, except that its primer-template duplex was lengthened from 10 to 60 base pairs. The ribozyme extended this substrate just as efficiently as the second substrate. Thus, the ribozyme is free from steric constraints that would preclude polymerization using long templates or long primer-template helices.

Given this general recognition of primer-templates, the range for primer extension, currently just beyond one helical turn, is limited merely by the ribozyme's efficiency. Polymerization is too slow for more extension to be observed within 24 hours, and longer incubations yield limiting returns, because buffer and ionic conditions optimal for polymerization (33) also promote ribozyme and template degradation. Reactions would have to be supplemented periodically with fresh ribozyme to achieve polymerization substantially beyond one helical turn. Nonetheless, in initiating extension of a long primer hybridized to a long template, the round-18 ribozyme demonstrates polymerization that is
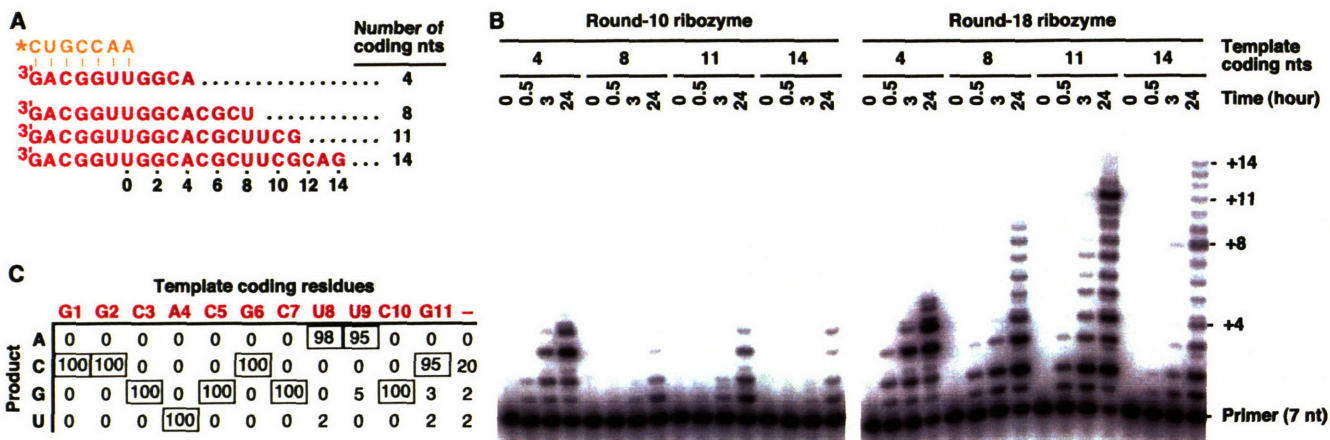
Fig. 4. Improved RNA polymerization. (A) Schematic of the 7-nt primer (orange) and the templates (red) with 4 to 14 coding residues used in these polymerization reactions. The ribozymes (not depicted) were as in Fig. 3. (B) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 5 μM ribozyme, 2 μM primer, 2.5 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (33). (C) Tabulation of nucleotides inserted across from template residues during primer-extension by the round-18 ribozyme. Full-length primer-extension products encoded by the template with 11 coding residues were cloned, and 100 full-length clones were sequenced (23). For each template residue (red), the number of clones that had an A, C, G, or U at the corresponding position was tabulated. Tallies representing Watson-Crick matches to the template are boxed. The column below the red dash reports the identity of the nontemplated nucleotide (Fig. 3B, legend) added to the 3′ terminus of 24 of the sequenced primer-extension products. Coding residues are numbered as in (A).
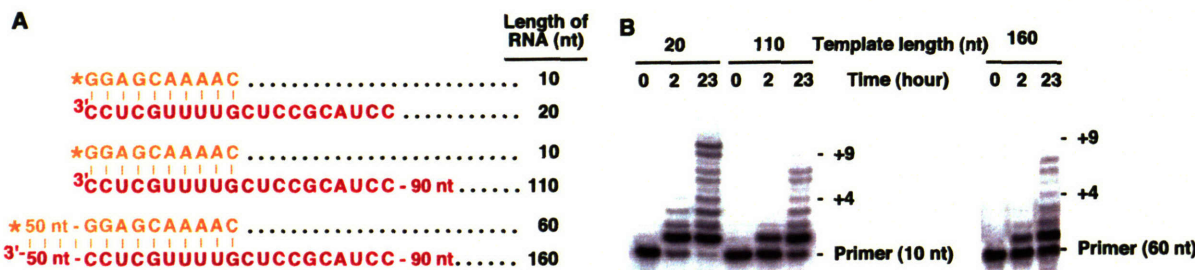
Fig. 5. RNA polymerization using long primers and long templates. (A) Schematic of the three primer-template combinations examined (23). Each combination included a 10- or 60-nt primer (orange) and a 20-, 110-, or 160-nt template (red). All three combinations have the same sequence near the site of polymerization (residues defined). (B) PhosphorImager scan of denaturing gels separating primer-extension products. Reactions included 5 μM round-18 ribozyme, 0.5 μM primer, 1 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (33).

27. T. A. Kunkel, *Bioessays* **14**, 303 (1992).
28. D. A. Steinhauer, J. J. Holland, *J. Virol.* **57**, 219 (1986).
29. C. D. Ward, J. B. Flanegan, *J. Virol.* **66**, 3784 (1992).
30. L. A. Loeb, T. A. Kunkel, *Annu. Rev. Biochem.* **51**, 429 (1982).
31. M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *J. Biol. Chem.* **274**, 36835 (1999).
32. When using k_{cat}/K_m values from (*31*), the optimal ratio of dNTP concentrations for yeast pol η would be 100:23:90:35 for dATP, dCTP, dGTP, and TTP, respectively. When using the values from Table 2 for the round-18 ribozyme, a ratio of 50:100:4:25 for ATP, CTP, GTP, and UTP would result in an observed fidelity of 0.990.
33. Ribozyme RNAs were transcribed from PCR-generated DNA templates, which were produced by using plasmid DNA and the appropriate primers. The DNA template for the round-18 ribozyme had penultimate 2'-methoxy substitutions to reduce the addition of nontemplated residues at the ribozyme 3' terminus (*46*). Standard polymerization assays were in 200 mM $MgCl_2$, 50 mM tris, pH 8.5, at 22°C. Although polymerization using an intermolecular primer was readily observed in the conditions of the selection (60 mM $MgCl_2$, 200 mM KCl, 50 mM EPPS, pH 8.0, 22°C), the standard assay conditions were more optimal. Ribozyme, primer, template, and NTPs were included at the concentrations indicated, and the GGCACCA RNA that *completes the ligase domain* was in 1.25-fold molar excess over the ribozyme concentration. These RNAs were mixed together in water, incubated at 80°C for 3 min, then incubated at 22°C for at least 5 min before starting the reaction by simultaneous addition of buffer, salt, and NTPs. Heating the ribozyme separately from the primer and template RNAs did not affect the polymerization reaction kinetics. Reactions were stopped by addition of 1.6 volumes of 145 mM EDTA/6 M urea, heated (90°C, 5 min) in the presence of competitor RNA designed to hybridize to the template RNA, then analyzed on sequencing gels. Incubation with competitor RNA, added in 10-fold molar excess over template RNA, led to better gel resolution because it prevented reassociation of the extended primer with the template.
34. N. H. Bergman, W. K. Johnston, D. P. Bartel, *Biochemistry* **39**, 3115 (2000).
35. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel, *Biochemistry* **39**, 15556 (2000).
36. P. J. Unrau, D. P. Bartel, unpublished results.
37. Consider hypothetical replicases about the size of the round-18 ribozyme (~200 nt) that synthesize an average of five 200-nt RNA strands within their lifetime. For replication through complementary-strand then second-strand synthesis, two of the five strands produced by each polymerase must have the correct residues at all of the positions that contribute to function. If the identities of effectively 80% of the 200 nucleotides contribute to function, then these replicases would require a fidelity of at least 0.995 ($0.995^{160} = 0.45 \approx 2/5$). Note that this scenario does not account for the dilution of productive ribozyme that would occur as active ribozyme variants replicate an increasing number of inactive variants; compartmentalization and selection would be needed to achieve sustainable replication (*7*, *9*).
38. The 5' terminus of gel-purified pool RNA was ligated (*47*, *48*) to the DNA portion of a DNA-RNA chimeric oligonucleotide, 3'-ATATCACTC-5'-5'-$X_{(6-8)}$ (residues defined, deoxynucleotides synthesized by using "reverse" phosphoramidites from Glenn Research; 5'-5', phosphodiester linkage joining the 5' terminus of the DNA with the 5' terminus of the RNA; $X_{(6-8)}$, RNA primer segment 6 to 8 nt in length complementary to the template indicated in Table 1). The pool was then gel-purified to remove the splint oligonucleotide needed for ligation.
39. An A analog, 2-aminopurine pairs with U and 4-thioU, but unlike A, its Watson-Crick pairing with 4-thioU does not involve the sulfur atom. Because the sulfur atom subtly distorts pairing geometry, templates with 2-aminopurine were used in later rounds of selection.
40. Repeated error-prone PCR was performed as described (*13*), except serial dilutions were after every 6 cycles of PCR. RNA from four subpools with expected mutagenesis levels of 0, 2, 4, and 6%, was mixed and used for round 15. RNA from two subpools with expected mutagenesis levels of 0 and 1% was used for round 16. RNA from two subpools with expected mutagenesis levels of 0 and 2% was used for round 17.
41. Because the net accuracy of RNA polymerization is represented by the product of the fidelities for each added nucleotide, geometric averages are reported throughout this study when describing overall fidelity.
42. J. N. Bausch, F. R. Kramer, E. A. Miele, C. Dobkin, D. R. Mills, *J. Biol. Chem.* **258**, 1978 (1983).
43. J. M. Clark, *Nucleic Acids Res.* **16**, 9677 (1988).
44. P. J. Unrau, D. P. Bartel, *Nature* **395**, 260 (1998).
45. Sequence composition of the template explains practically none of the higher fidelity in Fig. 4C. The template of Fig. 4C has one A, four C, four G, and two U coding residues. Thus, the expected fidelity per nucleotide, calculated by using the values of Table 2, is $[0.991 \times (0.9996)^4 \times (0.957)^4 \times (0.921)^2]^{(1/11)} = 0.969$, a value very close to the generic overall fidelity of 0.967 calculated in Table 2.
46. C. Kao, M. Zheng, S. Rudisser, *RNA* **5**, 1268 (1999).
47. M. J. Moore, P. A. Sharp, *Science* **256**, 992 (1992).
48. T. Tuschl, M. P. Ng, W. Pieken, F. Benseler, F. Eckstein, *Biochemistry* **32**, 11658 (1993).
49. We thank members of the laboratory and F. Solomon for helpful comments on this manuscript. Supported by grants from the NIH (D.P.B.), a Medical Research Council (Canada) postdoctoral fellowship (P.J.U.), and a Howard Hughes Medical Institute predoctoral fellowship (M.E.G.).

4983-46