The Reaction Kinetics and Three-Dimensional Architecture of a Catalytic RNA

by

Nicholas H. Bergman

B.S., Microbiology
University of Michigan, 1995

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 2001

© 2001 Nicholas H. Bergman. 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 par.

Signature of Author:

By

Department of Biology
(May 18, 2001)

Certified by:

David P. Bartel
Associate Professor of Biology
Thesis Supervisor

Accepted by:

Professor of Biology
Chair, Biology Graduate Committee

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARCHIVES

MAY 21 2001
LIBRARIES
The Reaction Kinetics and Three-Dimensional Architecture
of a Catalytic RNA

by

Nicholas H. Bergman

Submitted to the Department of Biology
on May 18, 2001 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ABSTRACT

The Class I ligase ribozyme was isolated previously from random sequences based on its
ability to promote a reaction similar to a single step in RNA polymerization: attack of a
primer 3'-hydroxyl on a 5'-triphosphate, with formation of a new 3'-5' bond and release of
pyrophosphate. Derivatives have been shown to catalyze general primer extension
reactions, making the ligase a useful paradigm for RNA self-replication and RNA
polymerase biochemistry as well as RNA catalysis in general. In order to establish the
ligase as a model system, we have characterized both the reaction and tertiary
architecture of the ribozyme.

The reaction kinetics of both multiple- and single-turnover ligation were examined, and
from these data minimal kinetic frameworks were constructed. These frameworks
provide a basis for the interpretation of future mechanistic work, and suggest strategies
by which individual steps in the ligation reaction might be targeted for future
improvement.

In order to test whether the chemical step of Class I ligation could be further optimized,
an in vitro selection was performed under conditions that specifically isolated chemistry.
Selected variants had a slightly improved chemical step, and substantially improved
Mg"+-dependence, such that at 0.5 mM Mg"+ a composite improved ligase was 50-fold
faster than the parent ribozyme.

The tertiary architecture of the ligase was examined using hydroxyl radical probing,
which provided a measure of the solvent accessibility at each position in the RNA
backbone. In collaboration with another group, these data were used to model the tertiary
architecture of the ligase in three dimensions. Finally, the predictive value of the model
was tested and confirmed by photocrosslinking experiments.

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

To my wife Jennifer
Acknowledgements

First and foremost, I thank David Bartel for being my mentor and teacher for the past five years. His instruction and advice have had a profound influence on me, and I am incredibly grateful. The example he set as a scientist will always be part of the standards I set for myself, and for this I will always be in his debt.

I am deeply grateful to all the past and present members of the Bartel lab for their encouragement and advice. In particular, I thank Phil Zamore and Peter Unrau for teaching me some of the basics of enzyme kinetics, and Scott Baskerville for giving me helpful suggestions throughout the selection described in Chapter 3. I also thank these and all the other lab members for being very supportive at all stages of my graduate training. In many ways they have been more friends than coworkers.

I also acknowledge my collaborators, both in the Bartel lab and elsewhere. In particular, I thank Catherine Yen and Nelson Lau for working with me on the projects in Chapters 3 and 4, respectively. I also thank Eric Westhof and Valerie Lehnert for collaborating with me in the development of the 3-D ligase model, and James Berger for helping in many ways as I learned crystallographic techniques.

My parents have been an incredible support to me since I started graduate school, and I thank them for all their encouragement and help. My relationship with them has been one of the most important anchors in my life.

Finally, I thank my wife Jennifer for all that she has done to support me. She willingly interrupted her own graduate studies and moved to Boston to be with me, and since then has been a support to me in more ways than I can list. Her help has been the deciding factor in many of the difficult times in graduate school, and I owe her a debt of gratitude that can hardly be measured. Consequently, this thesis is dedicated to her.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Dedication</td>
<td>3</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>4</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>5</td>
</tr>
<tr>
<td>Introduction</td>
<td>6</td>
</tr>
<tr>
<td>Chapter I</td>
<td>32</td>
</tr>
<tr>
<td>A Kinetic Framework for Multiple-Turnover Ligation by the Class I Ligase Ribozyme</td>
<td></td>
</tr>
<tr>
<td>Chapter I has been published previously as:</td>
<td></td>
</tr>
<tr>
<td>Chapter II</td>
<td>74</td>
</tr>
<tr>
<td>A Kinetic Framework for Self-ligation by the Class I Ligase Ribozyme</td>
<td></td>
</tr>
<tr>
<td>Chapter III</td>
<td>104</td>
</tr>
<tr>
<td>Selection of an Improved Class I Ligase Ribozyme</td>
<td></td>
</tr>
<tr>
<td>Chapter IV</td>
<td>144</td>
</tr>
<tr>
<td>A Three-Dimensional Model of the Class I Ligase Ribozyme</td>
<td></td>
</tr>
<tr>
<td>Future Directions</td>
<td>178</td>
</tr>
<tr>
<td>Appendix</td>
<td>189</td>
</tr>
</tbody>
</table>

The Appendix has been published previously as:
Introduction

The RNA World and RNA self-replication

Early in the development of life on Earth, primitive organisms may have depended on RNA for both information storage and catalysis\textsuperscript{1-6}. The case for such an "RNA World" was first put forward because RNA plays essential parts in DNA replication and protein synthesis, and was strengthened by the later realizations that RNA also plays a vital role in other cellular processes like mRNA splicing, telomere synthesis, and even protein translocation across membranes\textsuperscript{7}. Further evidence supporting the idea that RNA lies at the root of modern biochemistry can be found in the wide variety of nucleotide-derived coenzymes, such as NAD, FAD, and Coenzyme A, that are essential for cellular metabolism\textsuperscript{8}. Their presence hints that RNA might have existed before any protein enzymes, and in this regard they are somewhat like molecular fossils\textsuperscript{9}. In a similar way, the fact that both dNTPs and histidine are synthesized from ribonucleotides suggests that RNA predates both of these biomolecules and, by extension, DNA and proteins\textsuperscript{7}.

Perhaps the most compelling support for the RNA World hypothesis came with the discovery that RNA could perform catalysis\textsuperscript{10,11}. Since it can perform both informational and functional roles, RNA neatly sidesteps the classic paradox that troubled origin-of-life research (i.e., that functional and informational molecules each require the other for replication\textsuperscript{12}). This finding brought the idea of RNA as the original biopolymer to a greater level of attention\textsuperscript{5,6}, and allowed a more extensive examination of the assumptions that are implicitly made in the RNA World hypothesis.
One of the most basic of these assumptions touches on the one of the defining features of an RNA World: RNA-catalyzed self-replication. This activity would have been one of the first requirements for the establishment of an RNA World, since evolution could not have taken place without accurate replication. Given this, the RNA World hypothesis relies heavily on the assumption that somewhere in sequence space there is an RNA molecule that can function as an RNA replicase, synthesizing copies of itself and other RNAs\textsuperscript{12}. In its simplest form, such an enzyme would have to make copies of itself faster than it is degraded and catalyze this reaction with a level of fidelity that ensures that daughter copies retain the parent activity\textsuperscript{12}. If these standards are met in the context of a primitive cell, Darwinian evolution could lead to better replicases and, eventually, to the emergence of other catalytic activities.

Testing the possibility of RNA self-replication has been difficult, because four billion years of evolution have erased any naturally occurring examples. Nevertheless, it has been possible to explore the capabilities of existing ribozymes, and from these studies draw conclusions as to whether RNA-catalyzed RNA synthesis is possible\textsuperscript{13}. Early work along these lines began with the \textit{Tetrahymena thermophila} Group I intron, when it was shown that derivatives of this ribozyme could perform a modified splicing reaction in which pentacytidylic acid (C\textsubscript{5}) was converted into both shorter and longer strings of C's (up to C\textsubscript{30})\textsuperscript{14}. Further, the ribozyme could catalyze the conversion of 5'-CpU and 5'-GpN to 5'-CpUpN and G in a reaction analogous to the second step of splicing (Figure 1A). Here 5'-CpU corresponds to the 5'-exon and 5'-GpN corresponds to the 3'-splice site (the 3'-terminal base in the intron and the 5'-terminal base in the 3'-exon)\textsuperscript{15}. 
From this point, the replicative capabilities of the Group I intron were extended in several ways. First, combining the two reactions described above resulted in a primer extension reaction in which up to 6 nucleotides were added sequentially to pentacytidylic acid (Figure 1B). While this reaction was an important first step toward RNA replication, it was not applicable to most RNA sequences because it was only partially template-dependent (the primer could bind in multiple registers, and there was no base that paired with the incoming nucleotide). An added restriction was that only three of the four nucleotides could be used—GpG was problematic because of competing cleavage reactions. Finally, the ribozyme did not use an external template, since the primer bound to the 5'-exon binding site within the ribozyme.
The first two obstacles, lack of template-direction and inability to use GpG, were circumvented by redesigning the primer, template, and ribozyme. Adding a discrete template base and adjusting the template sequence so that the primer could only bind in a single register allowed the examination of template influence. The problems associated with GpG were bypassed by switching to a mutant intron that recognized 2-aminopurine-activated nucleotides instead of guanosine-activated monomers\textsuperscript{17}. With these changes, template-directed extension by any of the four nucleotides was observed (Figure 1C). Expanding the reaction to include use of an external template, however, proved more difficult, as separation of the template from the ribozyme dropped activity by four orders of magnitude.

The difficulties associated with formation of an enzyme•template•substrate(s) complex were solved using a slightly different system in which the Group I intron spliced together oligonucleotide substrates rather than a primer and activated monomers (Figure 1D)\textsuperscript{18}. In these reactions, small oligonucleotides (6-10 nt long) were annealed to an external template RNA, and the ribozyme ligated the fragments into a single complementary strand. Further work showed that this ligation reaction could be used to synthesize the entire complementary strand of a shortened ribozyme (Figure 1E)\textsuperscript{19}. It should be noted, though, that this reaction has a limited relevance to general RNA synthesis because the large number of required substrates would be difficult to provide at appropriate concentrations.

Although all of these studies provided key demonstrations of RNA’s ability to catalyze RNA synthesis, they have several limitations with regard to general RNA replication. First, neither of the Group I intron-catalyzed reactions (primer extension or
oligonucleotide assembly) has the fidelity necessary to ensure accurate transmission of activity from one generation to the next. Oligonucleotide assembly reactions suffer from the fact that mispaired oligonucleotides can only be discriminated against if the competition between matching and mismatching substrates comes to an equilibrium. Because of the relatively slow dissociation rates seen with even short oligonucleotides\textsuperscript{20}, it is unlikely that such a population could reach equilibrium in the time frame of ribozyme catalysis. Furthermore, unless the mismatch was very near the ligation site, the ribozyme would likely be unable to discriminate between matching and mismatching substrates based on a difference in $k_{\text{cat}}$, since the two would probably be similarly oriented for catalysis. These problems have been partially overcome in a system in which the intron used trinucleotide substrates, but even then mismatches were frequent, especially in wobble pairs\textsuperscript{21}. One might imagine that primer extension reactions would allow the ribozyme the greatest chance for accurate RNA synthesis, since discrimination might be possible in the context of both $K_m$ and $k_{\text{cat}}$, but these reactions also showed a relatively high error rate (~35%)\textsuperscript{17}. A fidelity of >99% would probably be necessary to ensure that progeny molecules are functional\textsuperscript{22}, and it appears that none of the Group I intron-catalyzed reactions can achieve this level of accuracy.

The second problematic characteristic of Group I-catalyzed RNA synthesis is the leaving group. All the reactions described involve attack of a 3’-hydroxyl on an internal phosphate, with a mononucleotide leaving group. The reaction is then more properly a disproportionation, in which one RNA is extended at the expense of another. Although some of the RNAs in these reactions grow longer, the total number of phosphodiester
linkages per molecule remains the same. Thus, only a small fraction of the RNAs in
solution can reach a large size because of competing reactions.

In vitro selection

These issues limited the extent to which the Group I intron could be used as a
paradigm for RNA-catalyzed RNA replication. Billions of years of evolution have
optimized the ribozyme for a specific, single-turnover reaction, so it can hardly be
surprising that some of the aspects that make it a very efficient self-splicing intron also
limit its ability to perform general RNA synthesis. Given this finding, the most desirable
model system would be a ribozyme that has evolved to catalyze an accurate and efficient
RNA synthesis reaction without using disproportionation. Although the naturally
occurring RNA catalysts do not include a ribozyme that can perform this reaction\textsuperscript{23,24}, the
development of in vitro selection methods made the search for such an RNA possible.

In vitro selection of RNA begins with a large pool of random sequences, often
numbering $10^{15}$-$10^{16}$. These molecules are challenged to perform a task, such as binding
to a given target or performing a desired self-modification reaction, and those sequences
that succeed are isolated and amplified. The resulting DNA forms the basis for a new
pool, and a new cycle of selection and amplification begins. The process is carried out
iteratively, with each round enriching the pool further in sequences that are able to carry
out the desired function.

In principle, the activity that is sought in an in vitro selection experiment can be
any that facilitates the separation of active sequences from those that are inactive. This
inherent flexibility has made selection a useful tool for the study of the intrinsic
properties of RNA. In particular, the system has been valuable in isolating improved variants of naturally occurring ribozymes\textsuperscript{19,25,26}, in better defining RNA-protein interactions\textsuperscript{27,28}, and in identifying RNAs that are able to bind small organic ligands\textsuperscript{29}. Perhaps most impressive, though, is the capacity of this technique to isolate entirely new catalysts. In this way, the known repertoire of RNA catalysis has been expanded to include many reactions not catalyzed by naturally occurring ribozymes\textsuperscript{30}, including glycosidic bond synthesis\textsuperscript{31}, phosphorylation\textsuperscript{32}, alkylation\textsuperscript{33}, aminoacylation\textsuperscript{34-36}, and Diels-Alder cycloaddition\textsuperscript{37,38}.

\textit{The Class I Ligase Ribozyme}

The ability to select ribozymes \textit{de novo} provided a unique solution to the problem of RNA-catalyzed RNA replication. Instead of engineering a ribozyme that had already evolved to perform a different reaction, it was now possible to begin with random sequences and search for ribozymes that catalyze reactions resembling RNA replication. Given this possibility, Bartel and Szostak designed an \textit{in vitro} selection that would identify ribozymes able to catalyze a ligation reaction that approximated a single step in biological RNA polymerization: attack of a 3'-hydroxyl on the $\alpha$-phosphate of a 5'-triphosphate, forming a new 3'-5' phosphodiester linkage with concomitant displacement of pyrophosphate\textsuperscript{39}.

It is the use of pyrophosphate as a leaving group that most clearly distinguishes this reaction from those catalyzed by naturally occurring RNAs. Activation of the $\alpha$-phosphate by pyrophosphate is universal in nature, and avoids the disproportionation issue described earlier. Furthermore, pyrophosphate-activated monomers are less
susceptible to hydrolysis than mononucleotide-activated monomers because of the lack of an internal attacking group\textsuperscript{40}. Pyrophosphate activation is also thought by some to be plausible in potentially prebiotic conditions\textsuperscript{41}. Finding a ribozyme that can catalyze ligation using pyrophosphate as a leaving group would therefore be an important step toward demonstrating that RNA could catalyze a more general RNA polymerization reaction.

Selection for this activity resulted in the isolation of more than 65 different ribozymes\textsuperscript{39}. Most of these were structurally simple, and catalyzed the formation of 2'-5' phosphodiester bonds. One sequence, though, performed exactly the desired reaction—formation of a new 3'-5' linkage using pyrophosphate as a leaving group. This ribozyme, designated the Class I ligase, was mutagenized and subjected to further selection\textsuperscript{42}. The second selection yielded clones with more diverse sequences, and comparative sequence analysis was used to solve the secondary structure of the RNA (Figure 2; ref. 42). When complexed with substrate, the ribozyme forms a double-

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{Figure2.png}
\caption{The Class I ligase}
\end{figure}
pseudoknotted structure with seven helices and several critical joining regions. This structure was unexpectedly complex for an RNA that emerged from a sparse sampling of random sequences, and points to the likelihood that a large number of similarly complex catalysts exist in sequence space\textsuperscript{43}.

Perhaps even more remarkable than the complex structure was the efficiency with which the ribozyme catalyzed ligation. Initial characterization showed that at pH \(\geq 7.5\), self-ligation reactions (Figure 3A) were essentially complete within five seconds, making the rate of catalysis too fast to measure by manual pipetting\textsuperscript{43}. When the substrate-binding portion of the ribozyme was separated from the rest of the RNA multiple turnover ligation was observed (Figure 3B), and under these conditions ligation rates

\[\text{Figure 3. RNA-synthesis reactions catalyzed by the Class I ligase. Adapted from ref. 13.}\]
could more conveniently be measured. This reaction was also remarkably fast, with a $k_{cat}$
greater than 100 min$^{-1}$. This multiple-turnover rate exceeds those of other multiple-
turnover ribozyme reactions$^{43,44}$, and represents an enhancement of $10^9$ over the rate for
uncatalyzed ligation involving similar template-aligned oligonucleotides.

*Class I Ligation as a Model for RNA Self-Replication*

Because its reaction directly parallels that of biological RNA polymerases, further
experiments were done to test whether the Class I ligase could perform RNA
polymerization. Deletion of the first four nucleotides in the 5'-terminus, substitution of
the appropriate template base, and addition of free nucleoside triphosphates allowed the
ligase to be examined in a primer extension context (Figure 3C). In this format, the
ligase extended the primer by one nucleotide, with a high degree of fidelity$^{45}$. On
average, the Watson–Crick match was added 85% of the time when all four nucleotides
were at equal concentrations, and this accuracy was as high as 92% when the
concentration of GTP was dropped 10-fold. Both figures represent a substantial
improvement over the best previously observed fidelity in a Group I intron-catalyzed
primer extension reaction (65% correct).

As selected, the ligase contained only one template base, so primer extension was
limited to addition of one nucleotide. When the template region was extended to three
bases (separating helices P1 and P2 by three nucleotides instead of one) and the ribozyme
was given the appropriate nucleoside triphosphates, sequential, template-directed addition
of three nucleotides was observed (Figure 3D). This short RNA polymerization reaction
could be catalyzed with a variety of template sequences, suggesting that the ribozyme
does not make sequence-specific contacts with the synthesized RNA or its template. Instead, since the observed fidelity was much higher than that expected from a simple preference for Watson–Crick base-pairing\textsuperscript{46}, it appears that the ribozyme might make more generic contacts that promote accurate polymerization.

In short, the ligase ribozyme is able to catalyze accurate, template-directed RNA polymerization using nucleoside triphosphates, and has no requirements for sequence identity in the primer, the template, or the nucleotide being added. In these characteristics, the ribozyme is very similar to biological RNA polymerases. The ligation reaction as studied here, however, suffers from one major drawback: polymerization beyond three nucleotides is blocked because the ribozyme pairs with a nearby region of the template. To a limited extent, this obstacle could be overcome by redesigning ribozyme–template binding (helix P2) such that the two could pair in multiple registers. In this case, extension by up to six nucleotides was observed, though this format is of limited interest because only highly repetitive sequences can be replicated\textsuperscript{45}. Consequently, a more universal mode of ribozyme–template binding was necessary.

Recently, this objective has been attained using a combination of engineering and further \textit{in vitro} selection\textsuperscript{47}. In this case, a large random sequence was added to the 3' terminus of the ligase, and the ribozyme core was mutagenized. Ribozyme–template pairing was also prevented by addition of a competitor oligonucleotide. The ribozyme that emerged from this selection retained the entire ligase sequence, and had a new, well-structured 3'-domain. It was able to bind the primer–template pair in a sequence-independent way, and catalyze RNA polymerization of up to 14 nucleotides (Figure 3E). In addition, the accuracy of the reaction was increased to 96.7\%, so the polymerization
reaction not only proceeds through an entire turn of an RNA helix, but does so with remarkably few errors.

These results are arguably the most promising demonstration of RNA-catalyzed RNA polymerization to date. Significantly, although engineering and selection were both used in generating polymerase derivatives, the core ligase sequence remained constant, and presumably its structure is essentially unchanged in all the described variants. This finding makes the ligase itself appropriate for future studies pertaining to the mechanism and efficiency of its polymerase derivatives, and a point at which to begin further efforts toward improvement.

*The Class I Ligase and Biological RNA Polymerization*

In addition to being relevant to exploring the possibility of RNA self-replication, the Class I ligase offers a unique point of comparison to protein RNA polymerases. Although its mechanism is not yet well understood, initial work combined with insights from studies of protein-catalyzed RNA polymerization suggest that the same mechanism might be used by both the Class I ligase and biological polymerases.

The three-dimensional structures of a variety of polymerases have been solved, including DNA-dependent DNA polymerases\textsuperscript{48-54}, a DNA-dependent RNA polymerase\textsuperscript{55}, a reverse transcriptase\textsuperscript{56}, and RNA-dependent RNA polymerases\textsuperscript{57, 58}. Each of these structures contains an active site that is built around a set of highly conserved acidic amino acids. These carboxylate residues are always in the same position relative to substrates, and are critical for catalysis in that they bind and precisely position an
Figure 4. A proposed Two-Metal-Ion mechanism for general phosphoryl transfer. Adapted from ref. 63.

essential pair of divalent metal ions$^{59,60}$. Both structural and biochemical data suggest that these metals interact with substrates and play direct roles in catalysis$^{61,62}$.

Based on this data, and the finding that none of the protein side chains near the active site appear to participate directly in the chemistry of polymerization, a two-metal-ion-catalyzed mechanism was proposed$^{63}$. In this mechanism (Figure 4), the first metal (metal A) lowers the pKa of the attacking 3'-hydroxyl, and so promotes the attack of the 3'-O- on the α-phosphate. The second metal (metal B) facilitates the formation of the pentavalent transition state and stabilizes the growing negative charge on the pyrophosphate leaving group.

The applicability of this mechanism for general phosphoryl transfer is demonstrated by the fact that similar pairs of metal ions, separated by roughly the same distance (4 Å), have also been observed in the active sites of HIV RNase H$^{64}$, alkaline phosphatase$^{65}$, a phospholipase$^{66}$, and an exonuclease$^{67}$. Although all of these enzymes seem to use the same mechanism for phosphoryl transfer, they share no other apparent
structural homology. This fact implies that there may be many different ways to properly position the required metal ions, and that otherwise the structural context can be quite variable.

The ability of RNA to tightly and specifically bind metal ions has long been known. Early structures of tRNA showed several high-affinity Mg\(^{++}\)-binding sites that were involved in stabilizing loops and bends, and more recently solved RNA structures have shown similar metal-binding sites. Furthermore, biochemical studies have been able to localize Mg\(^{++}\) ions in the active site of several ribozymes. In view of these studies, it seems likely that RNA might be capable of positioning the required metal pair at a polymerase active site. If so, perhaps the two-metal-ion mechanism described above could be extended to RNA-catalyzed phosphoryl transfer.

Several parallels between biological DNA and RNA polymerases and the Class I ligase suggest that the same mechanism might be used in both cases. Recent work has shown that the ligase is almost completely dependent on Mg\(^{++}\) ions, with very low levels of activity in Mn\(^{++}\) and no activity in any other conditions (M. E. Glasner, N. H. B., D. P. B, manuscript in preparation). Similar preferences are seen for protein polymerases, where either Mg\(^{++}\) or Mn\(^{++}\) ions are absolutely required for catalysis. Furthermore, sulfur substitution at the non-bridging oxygens of the reactive phosphate results in the same stereospecific effect in both the ligase ribozyme and protein enzymes, consistent with the idea of a Mg\(^{++}\) bound in the same position relative to substrate in both cases.

If the ligase does in fact use the same mechanism as protein phosphoryl-transfer enzymes, it would likely prove to be a valuable model enzyme for the study of RNA polymerization. Perhaps the ribozyme would provide a glimpse into the minimal
requirements for a polymerase active site, or shed light on what advantages proteins might have over RNA in performing phosphoryl transfer reactions. However, if the mechanism of the ligase is unlike those previously seen, it might instead provide an interesting comparison to the wide range of enzymes which do utilize a two-metal-ion mechanism. In either case, the Class I ligase presents an unusual opportunity in which to study the biochemistry of RNA polymerization.

*The Class I Ligase and Continuous Evolution*

The self-ligation reaction promoted by the Class I ligase, together with its speed, has also made possible the development of a unique extension of *in vitro* selection, termed ‘continuous evolution’\(^8\). This approach differs from conventional ‘step-wise’ selection in that all the phases of *in vitro* selection (catalysis, reverse transcription, transcription) occur at the same time in the same tube. Specifically, when the sequence of the Class I substrate is changed to that of the T7 RNA polymerase promoter, ligation followed by reverse transcription results in a new transcription template. When transcribed, this template creates a new ribozyme that can then begin a new cycle of ligation, reverse transcription, and transcription. Since ligation occurs in competition with reverse transcription (reverse transcriptase inactivates the RNA by converting it to a double-stranded molecule and cleaving the RNA strand), each RNA is copied at a frequency that is at least in part dependent on its catalytic efficiency.

As all of the steps of selection take place simultaneously, and in the same tube, continuous evolution experiments require far fewer manipulations on the part of the researcher. Typically, the solution is simply diluted periodically into a solution that
contains fresh polymerases, NTP's, and oligonucleotide substrate and primer. Because of this advantage, selection can be performed much faster than in typical in vitro selection experiments, with up to 100 rounds taking place in the timeframe of a few days\textsuperscript{82}.

This speed, combined with the link between a given sequence's catalytic efficiency and its chances for replication, has made continuous evolution methods valuable in modeling evolutionary processes. Several groups have followed ribozyme populations for up to 300 generations, and have observed sequence changes that conferred both simple rate enhancements and more complex advantages such as a higher correlation between genotype and phenotype (less conformational heterogeneity)\textsuperscript{82,83}. Furthermore, the system has recently been adapted to the study of predator-prey dynamics, with continuous evolution performed in the presence of another ribozyme designed to cleave the ligase\textsuperscript{84}. In light of these studies, it seems that continuous evolution experiments may be useful in the future as a model system for the study of population biology and other aspects of complex living systems that are otherwise difficult to address in vitro.

\textit{Summary of this Study}

Since further study of the Class I ligase is likely to impact several areas of interest, we sought to characterize the ribozyme and its reaction and establish it as a model system. As noted earlier, initial characterization showed that the ligase was remarkably efficient; self-ligation reactions were too fast for manual measurements except at low pH\textsuperscript{43}. This finding was somewhat surprising, as the ribozyme was not selected under conditions that would be expected to specifically favor very fast variants\textsuperscript{42}. 

21
The unexpected speed with which ligation occurs makes the ultimate understanding of the ribozyme's reaction more interesting, but also somewhat more difficult.

The studies reported here began with an examination of the kinetics of multiple-turnover ligation. In this format, the core of the ribozyme was separated from a substrate helix containing both RNAs to be ligated. Under these circumstances each ribozyme could catalyze multiple ligation events prior to the first timepoint, so even very fast ligation rates could be measured. The rate constants were determined for each of the individual steps in multiple-turnover ligation, and these data were used to construct a minimal kinetic framework for the reaction. The rate of the chemical step \((k_c)\) showed a log-linear dependence on pH, and reached 300 min\(^{-1}\) at pH 8.0. This rate was surprisingly fast, and roughly comparable to rates for the fastest naturally occurring ribozymes. The overall rate of catalysis \((k_{ca})\) was dependent on the strength of product binding; weakening this interaction increased ligation rates. Together, these results form a foundation for further mechanistic studies of the Class I ligase, and allow for a more accurate comparison between the ligase and other ribozymes.

Following these experiments, kinetic analysis was extended to include the original self-ligation reaction. Rapid-quench flow techniques made it possible to follow single-turnover ligation over a large range of conditions, and a new Ca\(^{++}\)-prefolding method allowed the separation of ribozyme folding and catalysis. The rate for the chemical step \((k_c)\) was log-linear with pH in the range 5.7-8.5, and above this point the rates accelerated at a slower pace, reaching 800 min\(^{-1}\) at pH 9.0. In contrast, the rate of ribozyme folding was independent of pH, and remained constant at 60 min\(^{-1}\). These data, together with measurements of substrate association and dissociation, allowed the construction of a
kinetic framework for self-ligation. This framework broadened our understanding of Class I ligation, and suggested ways in which specific steps in the ribozyme reaction might be targeted for further optimization.

To test whether the rate of chemistry in Class I ligation could be accelerated beyond rates that have been previously reported for RNA catalysts, in vitro selection was performed in a way that was designed to specifically optimize the chemical step. After seven rounds of selection, most randomized positions had become conserved, suggesting that the ligase sequence was nearly optimal, at least within the explored sequence space. Although ligases that emerged from this selection had only a marginally improved chemical step, the Mg$^{++}$ dependence of these ribozymes had changed substantially, such that at low Mg$^{++}$ concentrations, the improved clones were up to 50-fold faster than the parent ribozyme. The improvements observed in selected variants could be localized to a few mutations, and this information provides a key link between the architecture of the ribozyme and its activity.

Finally, the structure of the Class I ligase was characterized using chemical probing and three-dimensional modeling. Hydroxyl radicals, which cleave an RNA wherever the backbone is accessible to solvent, were used to identify portions of the ligase that are internalized when the ribozyme is folded. These areas were used to monitor folding in a range of temperatures and Mg$^{++}$ concentrations, and were also used in conjunction with phylogenetic analyses of selection data to generate a three-dimensional model of the ligase (in collaboration with V. Lehnert and E. Westhof, IBMC-CNRS, Strasbourg, France). Nelson Lau performed UV-dependent crosslinking experiments that confirmed the predictive value of the model.
Collectively, these studies explore the remarkable efficiency of the Class I ligase, provide a working model for its structure, and begin to define links between the ribozyme’s structure and function. The kinetic frameworks lay the groundwork for more detailed examinations of the ribozyme’s mechanism, and provide a basis for interpretation of all future work on the ligase. Similarly, the structural model of the ligase provides a context in which to interpret further structural studies, as well as solution-based data that will be valuable in validating a high-resolution crystal structure. Based on this work further studies of Class I mechanism, structure, and folding are conceivable and can now be addressed experimentally.
References


A Kinetic Framework for Multiple-Turnover Ligation

by the Class I Ligase Ribozyme
The work presented in this chapter was a collaborative effort between myself and Wendy Johnston. Specifically, Wendy measured the Michaelis-Menten parameters for the 207t and 210t ribozymes, and I performed all of the other experiments.
ABSTRACT: The class I RNA ligase ribozyme, isolated previously from random sequences, performs an efficient RNA ligation reaction. It ligates two substrate RNAs, promoting the attack of the 3'-hydroxyl of one substrate upon the 5'-triphosphate of the other substrate with release of pyrophosphate. This ligation reaction has similarities to the reaction catalyzed by RNA polymerases. Using data from steady-state kinetic measurements and pulse-chase/pH jump experiments, we have constructed minimal kinetic frameworks for two versions of the class I ligase, named 207t and 210t. For both ligases, as well as for the self-ligating parent ribozyme, the rate constant for the chemical step ($k_c$) is log-linear with pH in the range 5.7-8.0. At physiological pH, the $k_c$ is 100 min$^{-1}$, a value similar to those reported for the fastest naturally occurring ribozymes. At higher pH, product release is limiting for both 207t and 210t. The 210t ribozyme, with its faster product release, attains multiple-turnover rates ($k_{cat} = 360$ min$^{-1}$, pH 9.0) exceeding those of 207t and other reported ribozyme reactions. The kinetic framework for the 210t ribozyme describes the limits of this catalysis and suggests how key steps can be targeted for improvement using design or combinatorial approaches.
†Abbreviations: PPₐ, inorganic pyrophosphate; S⁰OH, smaller of two substrates (5’-aaaCCAGUC, DNA bases lowercase); ppps, larger substrate (5’-pppGGAACACUAUACGACUGGCACCA or pppGGAACGAAAUACGACUGGCACCA); EDTA, Ethylenediaminetetraacetic acid; MES, 2-[N-Morpholino]ethanesulfonic acid; BES, N,N-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid; EPPS, N-[2-Hydroxyethyl]piperazine-N’-[3-propanesulfonic acid]; CHES, 2-[N-Cyclohexylamino]ethanesulfonic acid; RNase P, Ribonuclease P.
RNA ligase ribozymes were previously isolated from a large pool of random sequences based on their ability to join a substrate oligonucleotide to their own 5' terminus (1). These new ribozymes promote attack by a terminal hydroxyl of the substrate RNA upon the $\alpha$-phosphate of the ribozyme triphosphate, joining the two RNAs, with release of pyrophosphate (eq 1, where $S^{\text{OH}}$ is the substrate oligo, $\text{PPPR}$ is the ribozyme, P is the ligation product, and $\text{PP}_i$ is inorganic pyrophosphate). Seven of these ligases have been examined and grouped into three classes based on secondary structure and ligation regiospecificity (2).

Of the three classes, the class I ligase (Figure 1A) has been the most extensively studied. This ribozyme promotes a reaction analogous in three respects to a single step of templated RNA polymerization: an RNA 3'-hydroxyl attacks the $\alpha$-phosphate of a 5'-triphosphate, a new phosphodiester bond is formed with concomitant displacement of pyrophosphate, and the molecules that are joined are aligned by Watson-Crick pairing (Figure 1A). Indeed, variants of the class I ligase have been designed that use an RNA template and nucleoside triphosphates to extend an RNA primer by three to six nucleotides (3). The ability of class I variants to synthesize RNA using the same reaction that is employed by biological polymerases supports the idea of RNA self-replication during the early evolution of life (3, 4).

The reaction promoted by the class I ligase is also uniquely suited for continuous in vitro evolution experiments. During continuous evolution, ribozymes are propagated based on their ability to ligate themselves to a substrate RNA before they are inactivated by becoming a template for cDNA synthesis (5). Because selection can be maintained by
simple serial transfer, the approach allows rapid and convenient sequence re-optimization following modification of selection criteria by, for example, changes to reaction conditions or the nature of the substrate (5). This process is also an attractive in vitro simulation of natural evolution, which may provide insight into the evolution of biocatalysts.

The class I ligase is remarkably large for a ribozyme that emerged from random sequences. The secondary structure of the class I ligase complexed with its substrate RNA is a nested double pseudoknot with seven stems and two critical extended joining regions (Figure 1A; 6). Finding such a large and complex ligase in a limited sampling of sequence space suggests the existence of a very large number of distinct structures of equivalent complexity and activity (2).

Perhaps as a consequence of its large size, it has been possible to use a combination of in vitro evolution and engineering approaches to generate variants of the original class I isolate (isolate b1) that are very efficient in the self-ligation reaction (6). The self-ligation rate of one of these variants (construct b1-207, Figure 1A) is too fast to measure accurately by manual pipetting, with the reaction mostly completed within the first 5 seconds (2). Nevertheless, ligation rates are easily measured in a multiple-turnover context, where a complex of the two RNA substrates to be ligated is separated from the core of the ribozyme (Figure 1B, eq 2), allowing each ribozyme to ligate many molecules prior to the first timepoint. With the multiple-turnover format, a version of the ligase (variant b1-210t; Figure 1B) has a $k_{cat}$ exceeding 1 sec$^{-1}$—a value greater than those of other ribozymes and approaching those of comparable protein enzymes (2).
To understand the basis for this unusually high catalytic rate, as well as the current limits of this rate, we have constructed a minimal kinetic framework for the ligation reaction. Identifying the rate-limiting steps of the reaction and understanding the factors that change them provides a basis for comparing the ligase with other characterized ribozymes and for comparing the prototype ligase construct with new variants that emerge from continuous evolution or other selection experiments. These results also suggest strategies for further improving catalytic rates and form a foundation for future mechanistic studies of RNA-catalyzed RNA ligation and polymerization.
MATERIALS AND METHODS

Substrate RNAs and Ribozymes. The smaller of the two substrates (S\textsuperscript{OH}) was an RNA-DNA hybrid (5'-aaaCCAGUC, DNA bases lowercase; 6). It was synthesized by standard phosphoramidite chemistry (7), and purified by anion-exchange chromatography (Nucleopac 9x250 column, Dionex). The larger substrates (PPPS) for the multiple-turnover derivatives (5'-pppGGAACACUAUACGACUUGCACCA and pppGGAACGAAAUACGACUGGC-ACCA; Figure 1) were made by in vitro transcription of synthetic DNA by T7 RNA polymerase (8). Transcripts were purified on 15% polyacrylamide/8 M urea gels, taking care to exclude longer transcripts with one or more untemplated residues. The self-ligating ribozyme (b1-207, Figure 1a; Genbank #U26413) was transcribed in vitro from a plasmid template linearized with Earl. Both multiple-turnover derivatives (Fig 1b) were transcribed from PCR-amplified DNA that was generated using the b1-207 plasmid and primers that produce the desired base changes and truncations. Ribozymes were purified on 6% or 8% acrylamide/8M urea gels. RNA concentration was determined spectrophotometrically at 260 nm, assuming an extinction coefficient that was the sum of those for the individual nucleotides (9).

Radiolabeling of RNAs. S\textsuperscript{OH} was labeled using T4 polynucleotide kinase and 32P-\gamma-ATP. PPPS RNAs were labeled using 32P-\alpha-cordycepin triphosphate and yeast poly(A) polymerase (United States Biochemical) to add a radio-labeled 3'-deoxyadenosine nucleotide to a 22-nt version of each substrate which lacked the 3'-terminal A. Ribozymes were body-labeled by including 32P-\alpha-UTP in the transcription.
Kinetic Assays and Measurement of Michaelis-Menten Parameters. All ribozyme reactions were performed in 50 mM buffer, 60 mM MgCl₂, 200 mM KCl, and 600 μM EDTA at 22° C. Buffers were MES (pH 5.7, 6.0, 6.5), BES (pH 6.7, 7.0, 7.4), EPPS (pH 8.0, 8.5), or CHES (9.0, 9.5). In all cases, the ribozyme was heated (2 minutes at 80° C in H₂O) and then cooled (2 minutes at 22° C) just prior to initiation of the reaction. In multiple-turnover reactions, PPPS was added to the ribozyme RNA before heating. Both self-ligation and multiple-turnover ligation reactions were initiated by simultaneous addition of buffer, salts, and S⁰H to the ribozyme solution. Aliquots were taken at specified time points and added to 2 volumes of a stop solution (8 M urea, 100 mM EDTA). Product and substrate were separated on 20% (reactions with labeled S⁰H) or 15% (reactions with labeled PPPS) polyacrylamide gels, then quantified by phosphorimaging (Fuji BAS 2000).

In most single-turnover experiments S⁰H was radiolabeled and used in trace quantities (<100 nM) with saturating amounts (usually 1 μM) of ribozyme. (Pilot experiments showed that 1 nM labeled S⁰H was half-saturated at 40 nM ribozyme.) More than 99% of S⁰H was capable of being ligated. Rates of single-turnover reactions were calculated from equation 3,

\[
\text{Fraction reacted} = F_s(1 - e^{-kt})
\]

where \( t \) equals time, \( k \) equals the rate of catalysis, and \( F_s \) equals the fraction of substrate that reacts in an initial burst. The \( F_s \) (typically 0.7) was interpreted to represent the fraction of substrate-ribozyme in an active conformation. Experiments measuring the fraction of enzyme capable of reacting were performed similarly, except that the ribozyme (1 μM) was labeled and incubated with saturating unlabeled substrate (10 μM).
For standard multiple-turnover experiments, PPPS was radiolabeled and used with sufficient unlabeled S\textsuperscript{OH} to ensure that nearly all the PPPS was complexed. Control experiments showed that a 25 \mu M excess of S\textsuperscript{OH} did not affect reaction rates. (This use of radiolabeled PPPS with saturating S\textsuperscript{OH} allowed accurate examination of low concentrations of S\textsuperscript{OH}•PPPS complex. Reactions with high S\textsuperscript{OH}•PPPS concentrations could also be performed using labeled S\textsuperscript{OH} and equimolar PPPS; in this study, the only multiple-turnover reactions in which S\textsuperscript{OH} was labeled were the reactions of 210t reported in Figure 3, where both substrates were at 35 \mu M). Extended time courses showed that 85-90\% of the PPPS was capable of reacting to form product. Rates of multiple-turnover reactions were measured as initial rates (<20\% of the substrate converted to product). Values for $k_{cat}$ and $K_m$ were determined by measuring the reaction rate at least twice at each of at least four PPPS concentrations, then fitting the data to the Michaelis-Menten equation using a non-linear, least squares algorithm (DeltaGraph 3.5, Deltapoint).

*Preparation of ligated product and determination of $K_i$ for each product•ribozyme complex.* Standard ribozyme reactions were scaled up to generate ligated product for b1-207t and b1-210t. Products were gel-purified and added to the ribozyme reactions with PPPS, prior to the 80°C incubation. The $K_i$ for each was determined from the best fit to equation 4.

$$k_{obs} = \frac{[S]k_{cat}}{[S] + \left(K_m (1 + \frac{[I]}{K_i})\right)}$$  \hspace{1cm} (4)

*pH-jump partitioning experiments.* Ribozyme (1 \mu M) and labeled S\textsuperscript{OH}•PPPS complex were mixed in reaction buffer at pH 6.0. The low pH slowed catalysis such that only 24\% of the S\textsuperscript{OH}•PPPS reacted in the first 10 seconds. PPPS was radiolabeled and
included at trace concentration (<25 nM), and SOH was kept at 25 μM, well above saturating concentration. At these concentrations, the majority of the SOH•PPPS complex is bound, and substrate dissociation rates reflect the dissociation of the SOH•PPPS complex from the ribozyme. After 10 seconds at 22° C, the reaction was mixed with an equal volume of a solution that increased the pH to the desired value (6.0-8.0). Each experiment was done with and without a substrate chase that added unlabeled SOH•PPPS complex to a final concentration of 20 μM. Parallel experiments were done for each pH in which the unlabeled SOH•PPPS complex was mixed with the labeled SOH•PPPS complex before starting the reaction.

*Computer simulation of pH-jump partitioning experiments.* Data obtained from the partitioning experiments were compared to simulated experiments done using the program KINSIM (10). Reactions were simulated using equations 5-10, where E =

\[ \begin{align*}
E + S & \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P \\
E + I & \rightleftharpoons EI \rightleftharpoons ER \rightleftharpoons E + R \\
F + S & \rightleftharpoons FS \\
F + P & \rightleftharpoons FP \\
F + I & \rightleftharpoons FI \\
F + R & \rightleftharpoons FR
\end{align*} \]

(5) (6) (7) (8) (9) (10)

enzyme, S = labeled substrate, P = labeled product, I = inactive substrate and unlabeled chase substrate, R = unlabeled product, and F = inactive enzyme.

Equation 5 is essentially the mechanism shown in Scheme I (neglecting pyrophosphate release, which is very fast), and equations 6-10 duplicate the mechanism so that inactive and active versions of both ribozyme and substrate can be considered independently. The rate constants presented in Scheme I are those that fit the data using the simplest model, wherein the reverse reaction is negligible and binding of substrate and product is unaffected by whether the ribozyme is active (E) or inactive (F).
Simulations of partitioning experiments were done by calculating the concentrations of all reactant species after 10 sec at pH 6.0 with 1 μM E and 25 nM S and using these values as the starting point for a new simulation reflecting the changed concentrations, volume, and pH at the beginning of the partitioning.

Note that the interconversions described by equations 11-15 are absent from our model. This simplification is justified if there are no significant changes in the equilibrium between active and inactive enzyme conformations upon binding substrate or if the interconversions are slow relative to the duration of the experiment. The slow reaction rate observed for the second phase of the self-ligation reaction suggests that the interconversion does in fact take much longer (30-60 minutes) than the longest partitioning experiments reported here (2 minutes).
RESULTS

*The chemical step ($k_c$) is rate-limiting at low pH.* As with many protein-catalyzed reactions, ribozyme reactions are often limited by product dissociation or conformational changes rather than by $k_c$, the rate constant for the chemical transformation (11-16). For ribozyme reactions involving phosphoryl transfer, the attacking hydroxyl typically needs to be deprotonated during $k_c$. Therefore, examining the influence of pH on $k_{cat}$ is useful for investigating whether $k_c$, the step involving chemical transformation, might be rate-limiting (12, 15, 17-20).

The influence of pH on reaction rate was tested using the single-turnover version of the class I ligase. The analysis of these data was complicated by the fact that the ribozymes in a given preparation had non-uniform activity. Experiments using radiolabeled ribozyme and unlabeled substrate showed that there were three populations of self-ligase ribozymes (unpublished data). At pH 8.0, the major fraction (70%) reacted within a 5-second burst. A minor fraction (15%) was inactive, which can be explained at least in part by loss of the 5'-triphosphate during transcription. The remaining fraction (15%) reacted, but at a slow rate, reaching a plateau only after 30 to 60 minutes. We suggest that this very slow-reacting fraction is trapped in an inactive conformation, and the additional time needed reflects the slow interconversion between inactive and active conformations. Although this fraction does eventually form product, its reaction rate is sufficiently slow to be grouped with the inactive fraction; all rates reported in this study reflect those of the fast-reacting fraction.
At pH \leq 7.0, self-ligation rates were slow enough to be measured accurately by manual pipetting (Figure 2). In the pH range 5.7-7.0, self-ligation activity was log-linear with pH, slope = 1.0 (Figure 3, circles). The direct correspondence between the hydroxide ion concentration and the reaction rate indicated that a single deprotonation is rate-limiting and suggested a rate-limiting \( k_c \) at low pH.

The effects of pH on the reaction rate of two multiple-turnover derivatives, 207t and 210t (Figure 1B), were also examined. Both substrates were at concentrations > 20 \( \mu \)M, well above the \( K_d \) of the P1 helix, allowing the substrate pairs (\( S^{OH} \) and \( PP\text{PPS} \)) to be treated as a unimolecular substrate complex (\( S^{OH} \cdot PP\text{PPS} \) in eq. 2). As with self-ligation, multiple-turnover ligation at pH < 7.0 was log-linear with pH, slope = 1.0 (Figure 3), again suggesting rate-limiting \( k_c \) at low pH. Indeed, at pH < 7.0, the rates of both the 207t and 210t reactions closely matched the rate of self-ligation. Thus, the break in the RNA chain engineered to convert self-ligation to multiple-turnover (Figure 1) does not compromise the activity of the ligase.

The suggestion of a rate-limiting \( k_c \) at a phosphate diester can be corroborated by observing an elemental effect when one of the non-bridging oxygens is replaced with sulfur (21, 22). For the class I ligase, thio effects are most conveniently examined using the primer extension format (3) and the two \( \alpha \)-thio-GTP diastereomers. In this format, the thio effect matches (\( S_p \) substitution) or exceeds (\( R_p \) substitution) the expected elemental effect (M. Glasner, C. Yen, E. Ekland, D.P.B., in preparation). Thus, given the pH-rate profiles together with this apparent elemental effect for the class I ligase in the primer-extension format, it is reasonable to conclude that \( k_c \) is rate-limiting at low pH.
As with other ribozymes, the critical deprotonation during the chemical step is most likely that of the attacking hydroxyl of the substrate (14, 15, 17, 22).

*Product release is rate-limiting at high pH.* Above neutral pH, the 207t construct did not maintain log-linear behavior (Figure 3). The 207t and 210t enzyme-substrate pairs differ only at three base pairs within the P2 helix—the helix that plays a major role in substrate and product binding (Figure 1B). Because the 207t P2 helix is predicted to be somewhat more stable than the 210t helix ($\Delta\Delta G^{27}_{22} = -1.3$ kcal/mol in 1 M NaCl; 23), the 207t ribozyme would be expected to have a slower product dissociation rate constant. Therefore, the simplest explanation for the divergence of the pH profiles for these two constructs is that product dissociation becomes rate-limiting as $k_c$ increases. Since the rate constant for product dissociation is expected to be slower for 207t than for 210t, the break in the pH-rate profile should occur at a lower pH, and the pH-independent rate constant should be slower, as observed.

Varying pH between 6.0 and 8.0 generally has little influence on RNA structure. The titration of moieties within this range would require a dramatic, though not unprecedented (24) perturbation in $pK_a$ (nearest $pK_a$'s of nucleotides: cytidine N3, 4.5; guanosine N1, 10.0; uridine N3, 10.1; ref. 25). Therefore, it would be surprising if the rates of product dissociation or ribozyme conformational transition were influenced by pH changes within this range so as to confound interpretation of our results. However, it should be noted that the $\gamma$-phosphate of $PPPS$ is expected to deprotonate within this range ($pK_a$ of the $\gamma$-phosphate of ATP = 6.8). The $\gamma$-phosphate is in the vicinity of the catalytic site, and its deprotonation might influence catalytic rate. A transition in the pH-rate profile for both constructs near pH 6.8 would have suggested the importance of this
deprotonation. However, the 210t profile had no transition in the vicinity of pH 6.8 (Figure 3); if deprotonation of the γ-phosphate is truly relevant, its pKₐ would have to be perturbed by over 2 units. It is interesting that the 207t rates accelerated again above pH 8.0. This could be explained by increased product-dissociation rates as multiple moieties involved in product binding began to deprotonate (26).

*Steady-state parameters* (kₐₜ, Kₘ, Kᵥ). To better understand the influence of pH on the multiple-turnover ligation reactions, Michaelis-Menten parameters were determined for the two constructs at two extreme pH's, 6.0 and 9.0, as well as at pH 8.0, the pH of the inflection point of the 207t curve (Table 1). At low pH, the kₐₜ's reflected the rates plotted in Figure 3, as expected, given that the concentration of SOH•PPPS was at least 10 fold above the pH 6 Kᵥ's. However, the Kᵥ for both constructs significantly increased with increasing pH, such that the 210t Kᵥ at pH 9 (25 μM) approached the substrate concentration used in Figure 3 (35 μM). Therefore, part of the plateau observed for 210t rates at high pH (Figure 3) can be explained by increased substrate concentrations needed for effective ribozyme saturation.

There are several explanations for the remaining deviation of 210t rates from log-linear behavior. Complete deprotonation of the attacking 3' hydroxyl would produce this deviation, but this would require a pKₐ shift of ~3 units (from 12.4 to ≤ 9.0). Similarly, it is difficult to rule out deprotonation at other sites that interfere with the chemical step (26). Nevertheless, we favor the possibility that, as with the 207t ribozyme, kᵥ remains pH-dependent, and the plateau arises from a change to a different rate-limiting step, i.e. product dissociation.
To supplement the Michaelis-Menten constants listed in Table 1 and to more accurately assess product binding, the inhibition constant ($K_i$) for the product of each ribozyme was measured. Because in each case the inhibition was competitive, $K_i$ equals the $K_a$ of the product (0.23 $\mu$M for 207t, 13 $\mu$M for 210t) (Figure 4). Studies of model RNA helices predict that the two product•ribozyme helices would differ by 1.3 kcal/mol, corresponding to a 9-fold difference in affinity (23). The measured difference in product-binding affinities between 210t and 207t was 6-fold more than that expected, and it is intriguing to consider the possibility that in shuffling the 3 base pairs that distinguish the two enzyme•substrate complexes (Figure 1B), we disrupted a tertiary contact that favors enzyme•substrate/product binding.

Despite this difference between the two ribozymes, the overall picture of product affinity largely matches that predicted from the simple formation of the P2 helix. The predicted $\Delta G^\circ_{298}$ of helix formation is 8.7 kcal/mol and 7.4 kcal/mol for the 207t and 210t, respectively (in 1 M NaCl, including $-2.5$ kcal/mol as the contribution of the A:A mismatches at both ends of the helix (23, 27). The $K_i$'s correspond to $\Delta G^\circ$'s of 9.0 kcal/mol and 6.6 kcal/mol for 207t and 210t; both are within one kcal/mol of the predicted values. The P2 helix is somewhat removed from the site of chemistry; for catalysis, there must also be ribozyme contacts to the ligation junction. How could essentially all the product binding be explained by the formation of the P2 helix? One possibility is that additional contacts that favor substrate/product binding are offset by contacts that destabilize the substrate/product without destabilizing the transition state, a known catalytic strategy of the group I intron (28).
Minimal kinetic framework for 207t. The 207t ribozyme reaction was modeled using the steady state data at pH 6.0 and 8.0. This modeling built on the notion that \( k_c \) is rate-limiting at low pH and that the product-dissociation rate becomes limiting at higher pH. It initially assumed that pH changes from 6.0 to 8.0 did not significantly change the association or dissociation rate constants of substrates or products. Subsequent experiments confirmed the validity of this assumption. The elemental rate constants determined by this modeling and refined in subsequent experiments are shown (Scheme 1).

Pyrophosphate binding and pyrophosphorolysis. The reverse of ligation (cleavage with PP\(_i\), forming a triphosphate) has been detected using the primer-extension format (3), where the cleaved product can dissociate before re-ligating (M. Glasner, C. Yen, E. Ekland, D.P.B., in preparation). This pyrophosphorolysis reaction is very inefficient \((k_c\text{avg}/K_{m\text{PP}} = 0.013 \text{ M}^{-1}\text{min}^{-1})\). It is not saturable by soluble amounts of PP\(_i\), indicating that the \(K_{m\text{PP}}\) exceeds 1 mM. Assuming that the rate constant for pyrophosphate association \((k_{-3})\) is comparable to association constants of other small-molecule-enzyme interactions \((10^4-10^9 \text{ M}^{-1}\text{min}^{-1}, \text{ref. 29})\) and that \(K_{m\text{PP}} = K_{d\text{PP}}\), then the lower limit on the \(K_{m\text{PP}}\) can be used to calculate a lower limit for rate constant for pyrophosphate release \((k_3)\) of \(10^5\) \(\text{min}^{-1}\) \((10^8 \text{ M}^{-1}\text{min}^{-1} \times 1 \text{ mM})\). Thus, \(k_3\) is denoted as "fast" in our kinetic scheme. The inability to saturate PP\(_i\) binding precluded determination of the rate constant for the reverse of the chemical step \((k_{-2})\), although inefficiency of the reverse reaction allows pyrophosphorolysis to be neglected when determining the remaining elemental rate constants.
Rate constants for dissociation \( (k_4) \) and association of product \( (k_{-4}) \). Although the measured \( k_{\text{cat}} \) for the 207t ribozyme at pH 8.0 was only 16 min\(^{-1}\), the rates measured for the 210t ribozyme suggested that the chemical step \( (k_c) \) for these ligases was much faster, exceeding the \( k_{\text{cat(pH 8.0)}} \) of 210t (140 min\(^{-1}\); Table 1). With this model, product release was by far the limiting step under these conditions, so \( k_4 \) closely approaches the \( k_{\text{cat}} \) (16 min\(^{-1}\)). Because the observed product inhibition was competitive, it follows that \( K_i = K_d \), and thus \( k_4/K_i = \) rate constant for product association \( (k_{-4}) = 7 \times 10^7 \text{M}^{-1} \text{min}^{-1} \).

Rate constant for the chemical step \( (k_c) \). The rate constant for product release is assumed to be constant throughout the pH range 6.0 – 8.0, so the \( k_4 \) of 16 min\(^{-1}\) was used in the formula \( k_{\text{cat(pH 6.0)}} = (k_c(\text{pH 6.0} \times k_4)/(k_c(\text{pH 6.0}) + k_4) \) to calculate the rate constant for the chemical step at pH 6.0 (3.75 min\(^{-1}\)). Because the rate of chemistry is log-linear with pH in this range, the \( k_c(\text{pH 8.0}) \) can be extrapolated to 375 min\(^{-1}\). The extrapolation is reasonable because catalysis at a rate of 360 min\(^{-1}\) can be observed directly for the 210t ribozyme at pH 9.0 (Table 1), and thus there are not likely to be intervening conformational steps to consider.

Rate constant for substrate association \( (k_i) \) and an initial estimate of the rate constant for substrate dissociation \( (k_{-4}) \). Under conditions where [S] is well below the \( K_m \) and the chemical step is both essentially irreversible and much faster than substrate dissociation, \( k_{\text{cat}}/K_m \) is equal to the rate constant for substrate association. These conditions hold for 207t at pH 8.0. The rate constant for substrate association \( (k_i) \) is thus given by \( k_{\text{cat(pH 8.0)}}/K_m(\text{pH 8.0}) = 7 \times 10^7 \text{M}^{-1} \text{min}^{-1} \). The rate constant for substrate dissociation \( (k_{-4}) \) can be

\[
k_{-4} = K_m (k_i (k_c + k_4))/k_4 - k_c
\]

(16)
calculated by equation 16 (29), assuming kinetic scheme 1, and yields an initial estimate of 15 min⁻¹.

*Measurement of substrate dissociation rate constant (k₋₁) and the ribozyme specific activity.* In order to refine the initial estimate of k₋₁, we measured this rate constant with an experiment in which a trace amount of radiolabeled substrate-enzyme complex was subjected to a pH jump in the presence of a large excess of unlabeled substrate. This also allowed us to test our initial assumption regarding the pH-independence of substrate binding between pH 6.0 – 8.0. Such a test was important because interactions involving the P⁴P⁴S γ-phosphate (pKa = 6.8) might have made substrate dissociation pH-sensitive over this range.

With the pH jump, a burst of product was observed, reflecting the amount of product formed as the preformed E•S complexes partitioned into either E•P (at a rate equaling kₑ) or E + S (at a rate equaling k₋₁). Further radiolabeled product formation was much slower because the presence of the unlabeled substrate chase prevented most of the enzyme from binding free labeled substrate. The magnitude of the initial burst corresponds to the ratio of the rates at which the two outcomes occur (kₑ/k₋₁). By changing the pH (from 6.0 to 8.0), this ratio was measured at five different kₑ’s, allowing measurement of substrate dissociation under a wide range of catalytic efficiency (Figure 5).

The results of the partitioning experiments were compared to those predicted using the chemical kinetics simulation program KINSIM (10). We observed that the data did not match the simplest model, which assumed that all of the ribozyme was in an active conformation (Figure 5, dashed lines). Instead, they fit well to a more complicated scheme in which 30% of the ribozyme was inactive yet could still bind substrate (Figure
5, solid lines). The two models are best distinguished at high pH, where the actual burst height was only 70% of that predicted by the simplest model.

Examination of the control reaction, which had a pH jump but not a chase of unlabeled substrate, further supported the conclusion that only 70% of the enzyme existed in an active conformation. If all of the ribozymes were in an active conformation and the catalytic step was significantly faster than the rate of substrate dissociation, all of the substrate would be converted to product very quickly (within the first second). If, however, a significant fraction of inactive enzyme existed, then there would be a corresponding fraction of substrate that was not converted to product as quickly, reflecting either the time necessary for the substrate complex to dissociate from an inactive enzyme and bind an active molecule or the time for an inactive E•S complex to convert to an active conformation. The results showed that although most of the substrate reacted very quickly (< 1 second), a small but significant fraction needed more time to react (Figure 5, pH 7.0, 7.5, 8.0). These data are inconsistent with a model in which the ribozymes were uniformly active (Figure 5, dotted line), and better match a model in which 30% of the ribozymes existed in an alternate, inactive conformation (Figure 5, solid line).

Using the refined model in which 70% of the ribozymes are in an active conformation, we found that the rate of substrate dissociation ($k_{-1}$) was 10 min$^{-1}$ throughout the pH range tested. This supported our earlier assumption that binding of substrate (and product) was not significantly affected by pH changes within the range of pH 6 to 8. Using $k_{-1} = 10$ min$^{-1}$ and $k_{1} = 7 \times 10^{7}$ M$^{-1}$ min$^{-1}$, the calculated $K_{d}$ of the E•S complex is 0.14 µM.
It should be noted that our favored model assumes that a ribozyme in an inactive conformation is essentially unable to promote ligation, yet binds substrate with the same affinity as does a ribozyme in the active conformation. It also assumes that the equilibrium between inactive and active conformations is unperturbed by binding of substrate or product. We propose that this is the simplest explanation for the data observed in the pH jump-partitioning experiments, as they are difficult to explain in any other way without invoking a complex relationship between pH and substrate affinity.

The presence of an inactive enzyme fraction made necessary the minor adjustment of the other individual rate constants. For clarity, all rate constants reported in this study have been corrected for this lowered ribozyme specific activity.
DISCUSSION

The 207t kinetic framework (Scheme 1) suggests the following view of ligation (at pH 8.0): The ribozyme molecules fold into either active or inactive conformations upon addition of buffer, salts, and $S_{\text{OH}}$. Although there are as yet no measurements of the rate of this folding for the multiple-turnover constructs, we suggest that it takes place quickly (< 1 second) because there is no observed lag in product formation during self-ligation, even with time points as early as 5 seconds (Figure 2). The majority of the ribozyme molecules quickly finds an active conformation and reacts in a rapid burst with a $k_c$ of approximately 375 min$^{-1}$. After ligation, product is released at a rate of 16 min$^{-1}$. This final step appears to be the major obstacle to improving the overall catalytic rate of 207t. Product release limiting the overall catalytic rate has been observed in other ribozymes, and might be expected given that most known ribozymes (derived both naturally and artificially) were evolved to perform precise single-turnover reactions.

The 210t ribozyme, which differs from its parent (207t) by the rearrangement of 3 base pairs, catalyzes the same reaction with a much higher $k_{\text{cat(pH 8.0)}}$ (140 min$^{-1}$ vs. 16 min$^{-1}$). We can model a similar framework for the 210t reaction (Scheme 2) in order to more completely describe this ribozyme’s activity and the differences from its parent. Michaelis-Menten parameters and product $K_i$’s were measured as with 207t (Table 1), and the reaction was modeled as follows, using the insights from the 207t reaction.

Rate constant for the chemical step ($k_c$). The 210t $k_{\text{cat}}$ at pH 8.0 (140 min$^{-1}$) suggested a very fast rate of product release, so $k_c$ at pH 6.0 must be very close to the $k_{\text{cat}}$ (3 min$^{-1}$). This value for $k_c$ agreed well with the corresponding rate constant from the 207t reaction.
As discussed earlier, the changes made between the two ribozymes did not significantly affect the chemical step.

**Rate constants for dissociation** \( (k_4) \) **and association of product** \( (k_{-4}) \). Using the measured \( k_{ca} \) and calculated \( k_c \) for the 210t reaction at pH 8.0, the rate constant for product dissociation \( (k_4) \) was estimated to be 270 min\(^{-1}\) \( (k_{cat(pH\ 8.0)} = (k_{cl(pH\ 8.0)} \times k_4)/ (k_{cl(pH\ 8.0)} + k_4) \). The rate constant for product association \( (k_{-4}) \) can be estimated as \( k_4/K_i = 2 \times 10^7 \) M\(^{-1}\) min\(^{-1}\). It is difficult to say whether the 3-fold difference between \( k_{-4} \) values of the two enzymes is significant. Day-to-day variation was generally less than ±20% for any two experiments performed with the same reagents, but as much as 2-fold when performed with different preparations of reagents.

**Rate constant for substrate association** \( (k_1) \) **and an initial estimate of the rate constant for substrate dissociation** \( (k_1) \). In contrast with the parent ribozyme, where \( k_1 \) was simply \( k_{cat}/K_m \), we do not have an accurate measure of \( k_1 \) for 210t. If the \( k_1 \) for 210t matched the \( k_1 \) for 207t, it would be approximately \( (7 \times 10^7 \) M\(^{-1}\) min\(^{-1}\)). On the other hand, for the parent ribozyme, \( k_1 \) equaled \( k_{-4} \); if the same was true for 210t, then its \( k_1 \) may be closer to \( 2 \times 10^7 \) M\(^{-1}\) min\(^{-1}\). The range for \( k_1 \) shown in Scheme 2 \( (2-7 \times 10^7 \) M\(^{-1}\) min\(^{-1}\)) reflects these two possibilities.

The low affinity of the 210t ribozyme for its substrate also hampered measurements of the substrate dissociation rate constant. The partitioning approach used for 207t was not applicable here; the high \( K_m \) for the 210t ligase would have required both high enzyme concentrations to saturate the substrate pulse and unattainable substrate levels for an effective chase. Approaches using dilution rather than an unlabeled chase to partition the labeled E•S complex were likewise beyond the scope of these studies because they
would have required near-instantaneous (millisecond) mixing. The rate constant for substrate dissociation was therefore estimated as \((K_d)(k_i) = 85-290 \text{ min}^{-1}\), where \(K_d\) was estimated from \(K_m\) at pH 6.

Although the two ribozyme reactions presented here are alike in most respects, the multiple-turnover rates of the 210t ribozyme place it in a unique position with regard to RNA catalysis. The RNA component of RNase P and a multiple-turnover derivative of the *Tetrahymena* self-splicing intron are also capable of achieving fast \(k_c\)'s (> 100 min\(^{-1}\)), but like 207t, they are usually held to relatively low turnover rates by their slow product release (11, 14). To circumvent this limitation, product-binding mutants can be used, but these mutants typically have lowered affinity for the substrate, necessitating high substrate concentration to observe the high multiple-turnover rate. Therefore, this strategy is expected to falter at the point that the binding mutants begin to affect \(k_c\) or the high concentration of RNA substrate begins to inhibit the ribozyme. At low Mg\(^{2+}\) concentration, RNase P RNA has a \(k_{\text{cat}}\) of 37 min\(^{-1}\) (30, 31) and should be able to exceed 100 min\(^{-1}\) with lowered NH\(_4\)^+ concentration (J. Kurz and C. Fierke, personal communication). The 210t ribozyme's ability to catalyze its reaction with \(k_{\text{cat(pH9)}} = 360\) min\(^{-1}\) shows that ribozymes can perform multiple-turnover reactions on RNA substrates without severe limitation by product release. Two factors contribute to the fast catalytic turnover of the 210t ribozyme at pH 9. First, the \(k_c\) is very fast, probably exceeding 1000 min\(^{-1}\). Second, the ribozyme binds its substrate at least as tightly as it binds its product. The 210t appears to bind substrate slightly (3 fold) more tightly than it binds product (compare substrate \(K_d\) and product \(K_i\) in Scheme II), allowing high multiple-turnover rates to be observed at somewhat lower substrate concentrations. This possible
preference for binding substrate would be unusual for RNA-catalyzed ligation; if it could be enhanced, more rapid rates could be achieved.

The catalytic efficiency of the class I ligase \( k_{\text{cat}}(\text{pH 8.0})/K_m \) approaches the substrate-enzyme association rate. This is a feature typical of ribozymes that bind their substrate by forming a Watson-Crick helix (11, 13, 32, 33) and can be thought to satisfy the criteria for “catalytic perfection” (11), although the ribozyme catalytic efficiencies are considerably less than those of classical diffusion-controlled enzymes. For the class I ligase, the P2 helix must form during binding of the substrate complex. The substrate-enzyme association rate constant is within the range observed for the association rate constants of model RNA helices (range: \( 10^7 \text{-} 10^9 \text{ M}^{-1}\text{min}^{-1} \); refs. 34, 35, 36). It is reasonable to propose that formation of P2 limits the substrate association rate constant, and thus attempts to improve the catalytic efficiency \( (k_{\text{cat}}/K_m) \) of the class I ligase would focus on increasing the rate constant for P2 formation. Perhaps additional contacts to the backbone of the ribozyme arm of P2 could increase the association rate constant by presenting the arm more favorably for helix nucleation.

Comparison of the 210t \( k_{\text{cat}}(\text{pH 8}) \) (140 min\(^{-1}\)) with the uncatalyzed reaction in the equivalent buffer, salt, and temperature \( k_{\text{uncat}}(\text{pH 8.0}) = 1.2 \times 10^{-7} \text{ min}^{-1} \); ref. 2) indicates a rate enhancement of \( 1.2 \times 10^9 \). This refinement of our previously reported value \( (8 \times 10^8 \); ref. 2) reflects the correction for ribozyme specific activity. The rate enhancement can also be expressed in terms of \( k_c \). The enhancement of \( k_{\text{c,pH 8}} \) (375 min\(^{-1}\)) over \( k_{\text{uncat,pH 8}} \) is \( 3 \times 10^9 \). The *Tetrahymena* group I intron and RNase P RNA also have fast chemical steps, yet they promote somewhat more difficult uncatalyzed reactions, and so
the $k_c$ for derivatives of these ribozymes are $10^{11}$ to $10^{13}$ times faster than $k_{\text{uncat}}$ (11, 30, 31, 37).

The most interesting optimization of the class I ligase would be an increase in the rate of chemistry. At 22° C and physiological pH, the $k_c$ of the class I ligase and $k_c$'s of the ribozymes with the fastest known chemical steps (Tetrahymena group I intron and RNase P RNA) are all about the same (100 min$^{-1}$). Although natural ribozymes are not necessarily optimized for speed, it is noteworthy that the ligase has attained a comparable $k_c$ without the benefit of billions of years of evolution. Indeed, it is curious that our prototype ribozyme (construct b1-207) has a fast $k_c$; it was generated by combining the features of ribozyme variants selected under conditions (a 30-second incubation at pH 7.4) that would not have stringently rewarded quick chemistry (6). With the finding that lowering pH can isolate $k_c$, it should be feasible, using rapid-quench techniques, to select for variants with faster rates of chemistry. Such an approach would more directly explore, and perhaps extend, the limits of the class I ribozyme as well as those of RNA catalysis in general.
ACKNOWLEDGMENTS

We thank Carol Fierke and Timothy McConnell for helpful discussions and comments on the manuscript. We also thank Carl Frieden and Chuck Merryman for assistance with the simulation software, and Phil Zamore, Mike Lawrence and other members of the lab for helpful comments on this manuscript.
REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>b1-207t</th>
<th></th>
<th>b1-210t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat,(\text{min}^{-1})}$</td>
<td>$K_m, (\mu\text{M})$</td>
<td>$k_{cat,(\text{min}^{-1})}$</td>
<td>$K_m, (\mu\text{M})$</td>
<td></td>
</tr>
<tr>
<td>pH 6.0</td>
<td>3</td>
<td>0.22</td>
<td>3</td>
<td>4.2</td>
</tr>
<tr>
<td>pH 8.0</td>
<td>16</td>
<td>0.23</td>
<td>140</td>
<td>7.8</td>
</tr>
<tr>
<td>pH 9.0</td>
<td>35</td>
<td>1.5</td>
<td>360</td>
<td>25</td>
</tr>
</tbody>
</table>
Scheme 1

\[
\begin{align*}
E + S & \xrightleftharpoons[k_1^{-1}]{k_1} E \cdot S \\
E \cdot S & \xrightleftharpoons[k_6]{k_5} E \cdot S \\
E \cdot P & \xrightleftharpoons[k_8]{k_9} E \cdot P \\
E \cdot P & \xrightleftharpoons[k_4]{k_4^{-1}} E + P
\end{align*}
\]

- \( K_d = 0.14 \mu M \)
- \( 7 \times 10^7 \text{ M}^{-1} \text{ min}^{-1} \)
- \( K_i = 375 \text{ min}^{-1} \) (pH 8.0)
- \( 3.75 \text{ min}^{-1} \) (pH 6.0)
- \( 207 \text{ t enzyme} \)
- \( K_s = 0.22 \mu M \)
- \( 7 \times 10^7 \text{ M}^{-1} \text{ min}^{-1} \)
- \( 16 \text{ min}^{-1} \)
- fast
Scheme 2

210t enzyme
FIGURE LEGENDS

Figure 1: The class I ligase. (A) The class I ligase promotes the attack of the 3'-OH of the substrate RNA (S^{OH}, outline typeface) on its own 5'-α-phosphate, creating a new 3'-5' linkage with concomitant release of pyrophosphate. P1 and P2 indicate the first two paired regions of the ribozyme-substrate complex; lowercase type indicates DNA. (B) Multiple-turnover derivatives ligate a bimolecular substrate complex (S^{OH}P_{PPS}, outline typeface). The 210t ribozyme was engineered by changing the arrangement of the base pairs in the P2 stem as shown (box).

Figure 2: Class I self-ligation. Representative time courses are shown for self-ligation of the 207 construct at pH 5.7 (△), 6.0 (□), 6.5 (▽), 7.0 (◇), and 8.0 (○). Radiolabeled substrate (<100 nM) was incubated in reaction buffer with 1 μM unlabeled ribozyme. Each curve indicates the non-linear, least-squares best fit to equation 3 using an F_a of 0.7.

Figure 3: The pH dependence of ligation. The rates shown are those of 207(○), 207t (□), and 210t(△). 207 rate constants are for single-turnover, self-ligation reactions, as shown in Figure 2. Multiple-turnover rates under steady-state conditions were measured for 207t using 10 μM radiolabeled P_{PPS}, 12 μM S^{OH}, and 25 nM enzyme. Multiple-turnover rates for 210t were measured using 35 μM P_{PPS}, 35 μM radiolabeled S^{OH}, and 25 nM enzyme. Each rate is the average of at least two independent measurements. k_{obs} approximates k_{cat} in most cases, except for 210t above pH 7, where the increasing K_m made saturating concentrations of substrate difficult to attain. The line is log-linear with
a slope of 1.0. Control experiments showed that at representative pH values the identity and concentration of the buffer did not noticeably influence observed rates (data not shown).

Figure 4: Inhibition of ligase by product. (A) Product inhibition in the 207t reaction. Ribozyme (25 nM) was incubated at pH 6.0 with S^{OH} (4 μM) and the indicated concentration of P^PPS. Rates shown are those with no added product (□), 150 nM product (○), and 600 nM product (△). The line through the points measured with no added product is the best fit to the Michaelis-Menten equation, and indicates a $k_{cat}$ of 3 min$^{-1}$ and a $K_m$ of 220 nM. The other lines show the rates expected with these Michaelis-Menten values and a $K_i$ of 230 nM (eq 4). Error bars indicate the standard deviation of ≥ 2 independent measurements (≥ 4 for uninhibited data set). (B) Product inhibition in the 210t reaction. Ribozyme (25 nM) was incubated at pH 6.0 with S^{OH} (25 μM) and the indicated concentration of P^PPS. Rates shown are those with no added product (□) and 15 μM product (○). The line through the points measured with no added product is the best fit to the Michaelis-Menten equation, and indicates a $k_{cat}$ of 3 min$^{-1}$ and a $K_m$ of 4.2 μM. The other line is the best fit to equation 4 using this $k_{cat}$ and $K_m$, and indicates a $K_i$ of 12.8 μM. Competitive behavior was confirmed for 207t. Error bars indicate the standard deviation of ≥ 3 independent measurements.

Figure 5: Measurement of substrate dissociation rate constants and enzyme specific activity by pH-jump partitioning experiments. Each panel shows three separate time courses. The pH-jump partitioning experiments (△) were performed as follows:
Ribozyme (1 μM) and labeled SOH\textsubscript{3PPPS} complex were mixed in reaction buffer at pH 6.0. After 10 seconds, the reaction was mixed with an equal volume of a solution that increased the pH to the desired value (6.0-8.0) and added unlabeled SOH\textsubscript{3PPPS} complex to a final concentration of 20 μM. Control experiments with no substrate chase (○) were performed in the same way, except that the chase solution raised the pH without adding any unlabeled SOH\textsubscript{3PPPS} complex. For the control experiment with no pulse of labeled substrate (□), the unlabeled SOH\textsubscript{3PPPS} complex was mixed with the labeled SOH\textsubscript{3PPPS} complex before starting the reaction. The solid lines through each set of points show the expected accumulation of product in a experiment simulated using the rate constants shown in Scheme I ($k_{-1} = 10 \text{ min}^{-1}$) and assuming that 87% of S is active and 70% of E is active. The dashed line in each panel represents a partitioning experiment simulated assuming 87% active S and 100% active E, where rate constants were adjusted from those in Scheme 1 to account for additional active E ($k_1 = 4.7 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$, $k_{-1} = 10 \text{ min}^{-1}$, $k_2 = 2.3 \text{ min}^{-1}$ at pH 6.0, $k_2 = 230 \text{ min}^{-1}$ at pH 8.0, $k_{-2} = 0$, $k_4 = 11 \text{ min}^{-1}$, and $k_{-4} = 4.7 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$). The dotted line in each panel represents a experiment simulated without a substrate chase, assuming 87% active S, 100% active E, and using the same adjusted rate constants as for the dashed line.
Figure 1

A

b1-207

B

b1-2071

b1-210
Figure 2
Figure 3

![Graph showing the relationship between pH and observed rate constant ($k_{obs}$)](image)
Figure 5

A

$\text{pH 6.0}$

B

$\text{pH 6.5}$

C

$\text{pH 7.0}$

D

$\text{pH 7.5}$

E

$\text{pH 8.0}$

Fraction reacted vs. Time (minutes)
A Kinetic Framework for Self-ligation by the

Class I Ligase Ribozyme
ABSTRACT: The Class I ligase, a ribozyme previously isolated from random sequence, catalyzes a reaction analogous to a single step of RNA polymerization in which a new 3'-5'-phosphodiester bond is formed and pyrophosphate is released. We have examined the kinetics of this reaction using a combination of standard and rapid-quench flow techniques, and with these data we have modeled a minimal kinetic framework. The ribozyme binds substrate with an affinity close to that expected from simple base pairing, yet requires substrate for productive folding. After substrate has hybridized, folding occurs rapidly (60 min⁻¹), and this rate was independent of pH in the range tested. A Ca²⁺-prefolding technique was developed, making possible the separation of folding and chemistry, and permitted measurements of the chemical step of ligation throughout the pH range 5.7-9.0. The rate constant for this step (k_chem) increases log-linearly with pH in the range 5.7-8.5, and continues to climb above pH 8.5, reaching a rate of 800 min⁻¹ at pH 9.0. This rate is very fast relative to other ribozyme reactions, and the kinetic framework suggests strategies by which in vitro selection might be used to further optimize specific steps in Class I ligation.
Abbreviations: HPLC, High-performance liquid chromatography; EDTA, Ethylenediaminetetraacetic acid; EGTA, Ethylene glycol-bis[β-aminoethyl ether]-N,N,N′,N′'-tetraacetic acid; MES, 2-[N-Morpholino]ethanesulfonic acid; BES, N,N-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid; EPPS, N-[2-Hydroxyethyl]piperazine-N′-[3-propanesulfonic acid]; CHES, 2-[N-cyclohexylamino]ethanesulfonic acid; R, Ribozyme; S, Substrate; RNase P, Ribonuclease P.
The Class I ligase ribozyme (Figure 1), isolated previously from random sequences, catalyzes an RNA ligation reaction in which a small RNA substrate is ligated to the 5'-end of the ribozyme, displacing pyrophosphate and forming a new 3'-5' bond (1-3). Initial characterization of the ligase showed that its reaction is surprisingly efficient; self-igation carried out at pH ≥ 7.4 is essentially too fast to measure using manual pipeting (3). This speed was unexpected, as the ligase was isolated under conditions that would not have been expected to reward rapid chemistry.

Because the speed of the reaction made the original, single-turnover ligase difficult to assay accurately, the kinetics of ligation were initially examined using a multiple-turnover derivative (4). At pH 8.0, the overall rate (kcat) for the multiple-turnover reaction was greater than 140 min⁻¹, a rate exceeding those of other multiple-turnover ribozyme reactions. Two factors contributed to this efficiency: first, ligated product was bound with low affinity and released very quickly (270 min⁻¹). Second, and perhaps more significantly, the chemical step was very fast, and was indirectly measured at 300 min⁻¹, a rate that is equivalent to those reported for the fastest naturally-occurring ribozymes (5, 6).

In order to better explain this efficiency, and to lay the groundwork for future mechanistic studies, we have examined the kinetics of the ribozyme in its original single-turnover context using a combination of methods including rapid-quench flow analysis. Experiments showed that the ribozyme folds at a rate of 60 min⁻¹ (as measured by attainment of catalytic competence), and this rate was constant over the entire pH range tested. In contrast, the rate of the chemical step was log-linear with pH over the range 5.7-8.5. Above pH 8.5 it continued to climb, though more slowly, and reached a rate of
800 min$^{-1}$ at pH 9.0. This rate is the fastest yet reported for a ribozyme reaction, and the wide range of rates for the chemical step coupled with the constant rate of folding suggests ways in which specific rate constants might be targeted for optimization through \textit{in vitro} selection.
MATERIALS AND METHODS

Ribozyme and substrate RNAs. The Class I ligase ribozyme (Genbank no. U26413) was transcribed in vitro from a plasmid template linearized with EarI (4). Transcripts were purified on 6% polyacrylamide/8M urea gels and stored in water at -20°C. Radiolabeled ribozyme was prepared by including $[^32]P$-α-UTP in the transcription reaction. The substrate for the ribozyme reaction was purchased (Dharmacon Research, Inc.) as a synthetic RNA-DNA hybrid [5'-aaaCCAGUC, DNA bases lowercase]. It was purified by anion-exchange HPLC$^+$ and radiolabeled using T4 polynucleotide kinase and $[^32]P$-γ-ATP. RNA concentrations were measured spectrophotometrically at 260 nm, assuming an extinction coefficient that was the sum of those for the individual nucleotides.

Manual self-igation assays and measurement of Michaelis-Menten parameters. Standard ribozyme reactions were done at 22°C in reaction buffer containing 50 mM buffer, 60 mM MgCl$_2$, 200 mM KCl, and 600 μM EDTA. Buffers were MES (pH 6.0) or EPPS (8.0). In all cases, the ribozyme was heated (80°C, 2 min, in H$_2$O) and then cooled (22°C, 2 min) just prior to initiation of the reaction. Ligation reactions were initiated by addition of buffer, salts, and trace $^{32}$P-labeled substrate. Aliquots were taken at appropriate time points and added to 2 volumes stop solution containing 120 mM EDTA and 8 M urea. Product and substrate were separated in 20% polyacrylamide gels and quantified by phosphorimaging. Ligation rates were determined from equation 1,

$$\text{Fraction reacted} = F_s(1 - e^{-kt}) \tag{1}$$
where \( t \) is time, \( k \) is the rate of ligation, and \( F_a \) is the fraction of ribozyme•substrate complexes that are active. \( F_a \) was typically 0.7 for short timecourses (4).

Experiments measuring Michaelis-Menten parameters for ligation were performed by incubating trace radiolabeled substrate (<1 nM) with varying concentrations of ribozyme (1-300 nM) at pH 6.0 as above. Because the ribozyme tended to stick to tubes and thus give erratic results when at concentrations below 100 nM, tRNA (1 µM) was added to the H₂O prior to addition of ribozyme. Control experiments established that added tRNA did not affect reaction rates (data not shown). Reaction rates were measured at least three times at each of four or more ribozyme concentrations. Values for \( k_{cat} \) and \( K_m \) in the self-ligation reaction were determined from the best fit to equation 2.

\[
k_{obs} = k_{cat} \frac{[S]}{[S] + K_m}
\]  

(2)

Experiments measuring the rate constant for substrate association were done in a similar way with trace radiolabeled substrate and various concentrations of ribozyme, except that they were performed at pH 8.0.

**Pulse-chase experiments measuring substrate dissociation.** Ribozyme (1 µM) was mixed with trace radiolabeled substrate (≤ 10 nM) and incubated at pH 6.0. At this pH, catalysis is sufficiently slow to ensure that while ribozyme•substrate complexes form, only a small fraction (~24%) reacts to form product in the first 10 sec. After 10 seconds, the reaction was mixed with an equal volume of chase solution that added 25 µM unlabeled substrate while maintaining standard buffer conditions. Parallel reactions were performed in which the chase was omitted.
Rapid-quench flow assays. Ribozyme reactions requiring time points faster than 5 seconds were performed in an RQF-3 rapid-quench flow apparatus (KinTek corporation, Austin, TX). Standard rapid-quench flow reactions were done as follows: Ribozyme and water were first heated to 80 °C for 2 minutes, then incubated at 22 °C for 2 minutes. EDTA was added to a concentration of 1 mM to chelate any trace divalent metal contaminants of RNA, and radiolabeled substrate was then added. To start the reaction, ribozyme-substrate complex was mixed with an equal volume of reaction buffer. Reaction buffers were formulated such that final reaction conditions were 1 µM ribozyme, ≤ 16 nM *substrate, 60 mM MgCl₂, 200 mM KCl, 0.6 mM EDTA, and 50 mM buffer. Buffers were BES (pH 7.0, 7.4), EPPS (pH 8.0, 8.5), or CHES (pH 9.0). Reactions were quenched with 0.5 volume 500 mM EDTA, then collected and analyzed by gel electrophoresis and phosphorimaging.

Rapid-quench flow reactions in which CaCl₂ (3 mM) was used to pre-fold the ribozyme were performed in the same way, except that CaCl₂ was added to the ribozyme-substrate mix. In all but initial experiments, EGTA was also added to the reaction buffer to a concentration of 5 mM.

Rapid-quench flow reactions in which ligation was initiated by addition of substrate were also performed essentially as described for standard reactions, with the following changes: Ribozyme was radiolabeled, and after heating and cooling in water the RNA was mixed with buffer and salts (final concentrations 50 mM buffer, 60 mM MgCl₂, 200 mM KCl, and 0.6 mM EDTA). The reaction was begun by mixing with an equal volume of a solution containing unlabeled substrate (final concentration 300 µM), quenched with 500 mM EDTA, and analyzed as described.
RESULTS AND DISCUSSION

Initial studies of Class I self-ligation were carried out without a detailed understanding of substrate-binding, as rates could be measured accurately by simply ensuring that the unlabeled component of a reaction was at a concentration far above that expected to be required for saturation of a 7-bp helix (≥ 1 μM). Rapid-quench flow experiments, however, made a more comprehensive examination of these rate constants necessary. The relatively large volumes involved made a more rigorous definition of saturation advantageous in that it allowed conservation of material, and a measurement for substrate association was necessary to ensure that this step was not unintentionally rate-limiting. With this in mind, we began our characterization of self-ligation kinetics by examining substrate binding, first by measuring the Michaelis-Menten parameters ($k_{\text{cat}}$ and $K_m$) for the reaction, and then by determining rates for both substrate dissociation and association individually.

Measurement of Michaelis-Menten Parameters for Self-Ligation. Ligation rates were measured at pH 6.0 using radiolabeled substrate and various concentrations of ribozyme. From these data, the $k_{\text{cat}}$ and $K_m$ for the self-ligation reaction were determined (Figure 2). The $k_{\text{cat}}$ for the reaction was measured at 3.0 min$^{-1}$, which agrees well with multiple-turnover rates measured at this pH (4). The $K_m$ was found to be 40 nM.

Determination of the rate constant for substrate dissociation ($k_{\text{dissoc}}$). To establish a rate constant for substrate dissociation, pulse-chase partitioning experiments were performed (Figure 3). Ribozyme (1 μM) was incubated with a trace amount of radiolabeled substrate ([*S]/[R] ≤ 0.01) at pH 6.0. After 10 seconds, a chase mixture was
added to the reaction that brought with it a large excess of unlabeled substrate ([S]/[R] = 50). The excess unlabeled substrate ensured that if R•S complexes dissociated, the two components only rarely re-associated because of competition from unlabeled substrate molecules. A reference ligation reaction was performed under the same conditions, but without the chase.

After addition of the chase mix, a burst of product formation was observed, corresponding to the amount of product formed as the R•S complexes partitioned into either product (at a rate equal to $k_{\text{chem}}$) or free R and *S (at a rate equal to $k_{\text{dissoc}}$). The height of this burst, as compared to the reference reaction performed without the chase, is dependent on the ratio of the rate constant for the chemical step to that of the sum of the rates constants for chemistry and substrate dissociation ($k_{\text{chem}}/k_{\text{chem}} + k_{\text{dissoc}}$). In this case, roughly 1/3 of the complexes reacted to form product. The remaining 2/3 consisted of the complexes that were in an active conformation but dissociated before catalysis could occur, and those complexes that were in an inactive conformation. Since 30% of the R•S complexes are known to be inactive (4), the fraction that was active and dissociated is approximately equal to the fraction that was active and formed product. Thus, the rate constant for substrate dissociation ($k_{\text{dissoc}}$) is roughly equal to that of the chemical step ($k_{\text{chem}}$) at pH 6.0, and can be estimated at 3 min⁻¹.

Measurement of the rate constant for substrate association ($k_{\text{assoc}}$). Changes in pH between 6.0 and 8.0 generally have little effect on RNA structure, so the rate for substrate dissociation is likely to be relatively constant in this range. Given this, and previous data that showed that multiple-turnover ligation rates (and thus the rate for the chemical step) could exceed 1 sec⁻¹ at pH 8.0 (4), it was reasonable to assume that at pH 8.0 the rate for
the chemical step was much faster than that for substrate dissociation. Under these conditions, \( k_{\text{cat}}/K_m \) approaches the rate constant for substrate association (\( k_{\text{assoc}} \)). We therefore measured ligation rates in standard ribozyme reactions using trace (< 1 nM) radiolabeled substrate and various concentrations of ribozyme (1 nM-1\( \mu \)M). At high concentrations of ribozyme ([R] \( \geq \) 300 nM) the rate of ligation was too fast to measure by manual techniques. This obstacle prevented an accurate measurement of the \( K_m \) for self-ligation at pH 8.0, but the lack of a plateau in ligation rates measured at \( \leq 100 \) nM ribozyme suggests that the \( K_m \) is probably \( \geq 300 \) nM (data not shown). At very low ribozyme concentrations ([R] \( \leq 30 \) nM), the observed rates decreased linearly with ribozyme concentration (Figure 4). The slope of this linear dependence gave a measurement of \( k_{\text{cat}}/K_m \) (and thus \( k_{\text{assoc}} \)) of \( 2 \times 10^8 \) M\(^{-1}\)min\(^{-1}\). An association rate of \( 2 \times 10^8 \) M\(^{-1}\)min\(^{-1}\) is in the middle of the range expected based on simple helix formation between two oligonucleotides (\( 10^7-10^9 \) M\(^{-1}\)min\(^{-1}\); ref. (7-9)) and equivalent to the helix formation step in other ribozyme reactions (10-12).

From this value, and the rate determined for substrate dissociation above, an apparent dissociation constant (\( K_d \)) for the substrate-ribozyme helix can be calculated as \( k_{\text{dissoc}}/k_{\text{assoc}} = 15 \) nM. Studies modeling the stability of RNA helices predict that the dissociation constant for this substrate-ribozyme helix would be slightly higher (29 nM) (13). It is difficult to say that the two values are significantly different. Although day-to-day variation was typically less than \( \pm 25\% \), it occasionally ran as high as 2-fold in experiments done on different days with different preparations of RNA. Regardless, the two values differ by very little, though it is not possible to make conclusions regarding the probability of non-Watson–Crick contacts between ribozyme and substrate because of
the folding issues discussed later and because as in other ribozyme reactions, additional contacts might stabilize the transition state without contributing to substrate binding (14, 15).

*Measurement of the rate of ribozyme folding* \((k_{\text{folding}})\). Previous studies have shown that the rate of folding of the ligase was relatively fast (taking \(\leq 1\) sec, since lags were never observed in reaction times as short as 5 sec), but a more accurate measurement was unattainable with manual techniques. Likewise, the rates for the chemical step at \(\text{pH} > 7.0\) were prohibitively fast, and were only seen indirectly in multiple-turnover experiments (4). We therefore used a rapid-quench flow (RQF) apparatus to measure ligation rates at \(\text{pH} 7.0-9.0\), as this method allowed us to follow the course of ligation with time points as short as 2 milliseconds. Ligation rates, as measured by RQF analysis, matched previous data at \(\text{pH} 7.0\). At higher \(\text{pH}\), rates increased slightly and reached a plateau of 60 min\(^{-1}\) (Figure 5A).

Previous studies showed that the chemical step is rate-limiting at low \(\text{pH}\) (4), and the plateau observed at \(\text{pH} \geq 8.0\) suggests that a step prior to chemistry is rate-limiting at high \(\text{pH}\). The rate of this previously unseen step is \(\text{pH}\)-independent, as the rates at \(\text{pH} 8.0\) and 9.0 are identical. Because the reactions were begun by addition of \(\text{Mg}^{2+}\), and Fe-EDTA mapping experiments have shown that the ribozyme’s tertiary structure is \(\text{Mg}^{2+}\)-dependent (N. H. B., N. C. Lau, V. Lehnert, E. Westhof, and D. P. B., manuscript in preparation), we propose that this step is ribozyme folding. A rate of 60 min\(^{-1}\) is somewhat faster than rates reported for the folding of other, larger ribozymes (16, 17), and roughly comparable to rates reported for the 160-nt P4-P6 domain of the Group I
intron (17, 18). We note that although this rate is fast relative to most other RNAs, it is still ~100-fold slower than the folding of tRNAs (19), and thus may be far from optimal.

*Separation of folding and catalysis and direct measurement of $k_{\text{chem}}$ at high pH.*

Multiple-turnover data suggested that at pH $\geq 7.5$ the rate of the chemical step could be much faster than the plateau described above (4), but $k_{\text{chem}}$ had not been directly measured at high pH. We expected that this might be accomplished by simply prefolding the ribozymes in Mg$^{2+}$, then beginning the reaction by addition of substrate (at concentrations high enough to ensure that substrate–ribozyme association was not rate-limiting). This scheme was tested at pH 9.0 using rapid-quench flow techniques. We observed that the reaction begun by addition of substrate followed a course similar to reactions begun by addition of Mg$^{2+}$ (Figure 6), suggesting that substrate is required for productive ribozyme folding.

Because of this finding, separation of folding and chemistry required conditions in which ribozyme and substrate could be mixed and folded without ligation activity. In the course of examining the metal dependence of the ligase, it was noted that in the presence of Co(NH$_3$)$_6^{3+}$ or Ca$^{2+}$ ions the ribozyme assumes an apparently native structure, though there is no detectable activity in ribozyme reactions containing either metal alone (M. E. Glasner, N. H. B., and D. P. B., manuscript in preparation). We therefore investigated whether either salt could be used to pre-fold the ligase. Ribozymes were incubated with substrate in either Co(NH$_3$)$_6^{3+}$ or Ca$^{2+}$ for ~30 min, and ligation was begun by addition of Mg$^{2+}$. Reactions to which Co(NH$_3$)$_6^{3+}$ was added showed complete inhibition even after addition of a large excess of Mg$^{2+}$ ([Mg$^{2+}$]/[Co(NH$_3$)$_6$] $\geq$ 4000) (data not shown). When ribozymes were prefolded in Ca$^{2+}$, however, ligation was much faster (5 sec$^{-1}$ at pH 9.0)
than in standard RQF reactions, though still not as fast as expected if chemistry were rate-limiting.

Previous experiments showed that in addition to promoting an apparently native structure, Ca\textsuperscript{2+} ions inhibit the activity of the ribozyme. We therefore modified our reaction scheme such that Ca\textsuperscript{2+} was added for prefolding and then removed at the start of the reaction by simultaneous addition of Mg\textsuperscript{2+} and EGTA (EGTA binds Ca\textsuperscript{2+} ions 10\textsuperscript{5}-fold more tightly than it binds Mg\textsuperscript{2+} ions) (Figure 5B). In experiments using this method, ligation rates matched previous measurements at pH 7.0 and were log-linear with respect to pH in the range 7.0-8.5 (Figure 5A). Above pH 8.5 the rates continued to climb with increasing pH, though at a slower pace, presumably because deprotonation of individual bases began to denature the ribozyme. The ability of calcium-preincubation to extend the range in which ligation rates are log-linear with pH suggests that chemistry is rate-limiting at high pH for the prefolded ribozyme. Thus, measurements of the chemical step were possible in a previously inaccessible pH range.

*Construction of a Kinetic scheme for Class I self-ligation.* The data presented in this study allowed us to begin the construction of a kinetic framework that models the Class I self-ligation reaction (Figure 7). The proposed framework was built with several simplifications: First, folding is represented as a single step, which in our case represents attainment of catalytic competence. In reality, folding is almost certainly several distinct steps, as has been seen for other RNAs (16-19). Second, the chemical step (R\textsubscript{•}S → P + PP\textsubscript{i}) is represented as irreversible, since previous work has shown that the reverse reaction (pyrophosphorolysis with formation of a triphosphate) is slow enough to be ignored (4, 20).
In the proposed framework, substrate binding and ribozyme folding are represented as separate events, and both of the two potential paths from $R_r+S$ to $R_t+S$ are formally possible. Note however, that reactions using ribozymes that had been preincubated in Mg$^{++}$ were virtually identical to reactions done without the Mg$^{++}$ preincubation, suggesting that the ligase does not stably exist in a folded form without bound substrate. Because of these data, the equilibrium between $R_r+S$ and $R_t+S$ is assumed to lie far to the side of $R_r+S$, and thus measured ligation rates largely reflect reactions in which folding is preceded by substrate association.

Given this assumption, the measured rates for substrate association and ribozyme folding can be placed on the framework as in Figure 7. We note that the equilibrium between $R_r\cdot S$ and $R_t\cdot S$ is assumed to lie far toward the folded form, because Ca$^{++}$-preincubation experiments did not detect the presence of an unfolded population. If this is true, and folding is much more stable in ribozymes that have bound substrate, it would follow by thermodynamic symmetry that folded ribozymes would bind substrate much tighter than ribozymes that are unfolded. This implication seems reasonable, since folded ribozymes could facilitate substrate binding by forming contacts to substrate in addition to base-pairing, or by positioning helices such that the ribozyme$\cdot$substrate helix is stabilized by stacking interactions.

The placement of the measured rate for substrate dissociation on the kinetic framework is somewhat difficult. The pulse-chase experiments reported here showed that dissociation of substrate from active ribozyme$\cdot$substrate complexes occurs at 3 min$^{-1}$, but it is not clear what step(s) this rate represents. One possibility is that 3 min$^{-1}$ simply reflects a simple dissociation event in which substrate is released from active
ribozyme-substrate complexes, leaving a free, folded ribozyme ($R_fS \rightarrow R_f + S$). The proposed framework would then suggest that such folded ribozymes exist only transiently in a folded conformation before unfolding. Alternatively, if the affinity between substrate and folded ribozyme is in fact very high, direct dissociation might be quite slow and the measured rate of $3 \text{ min}^{-1}$ might reflect a more complicated dissociation process in which the folded complex first unfolds, then dissociates. Though the latter hypothesis seems somewhat more likely because it conforms with the idea that folding enhances the ability of the ribozyme to bind substrate, the data presented here do not distinguish between these possibilities. Furthermore, it is difficult to imagine experiments in which they could be easily tested because of the relative inaccessibility of the free, folded state of the ribozyme.

**Conclusions.** The kinetics of Class I self-ligation, as modeled in Figure 7, can be summarized as follows: In the presence of Mg$^{++}$, free ribozymes exist primarily in an inactive conformation. Once substrate binds, the ribozyme-substrate complex folds rapidly (at a rate of $60 \text{ min}^{-1}$) into an active conformation, and then reacts to form product at a rate that is log-linear with pH, ranging from $3 \text{ min}^{-1}$ at pH 6.0 to $800 \text{ min}^{-1}$ at pH 9.0.

The rates reported for the chemical step of ligation at high pH are extremely fast for a ribozyme reaction, and prompt comparisons to the fastest naturally occurring RNA catalysts. The *Tetrahymena* group I intron and RNase P RNA are also able to achieve $k_{\text{chem}}$’s in this range ($350 \text{ min}^{-1}$, (5, 6)). Although these ribozymes have not necessarily been optimized for fast chemistry, it is intriguing that a comparably efficient enzyme could arise from a selection that did not specifically prefer extremely fast chemistry(2). The finding that three of the fastest ribozymes, disparate in origin and reaction, all reach
roughly the same level of catalysis raises questions of whether these catalysts might be limited by a property inherent to RNA, or whether further selection could be used to extend the activity of the Class I ligase to previously unseen levels.

A strategy for pursuing further optimization is apparent in the combination of pH-dependent chemistry and pH-independent folding. At pH 6.0, chemistry is roughly 20-fold slower than folding, and thus rate-limiting. Further selection under these conditions, therefore, would isolate the chemical step for improvement, and might result in new ligase variants that catalyze faster chemistry. At pH 8.5, the positions are reversed, with folding approximately 10-fold slower than chemistry. Selection at high pH, therefore, might result in ribozymes that folded faster than the parent. Although both of these strategies would require very fast selection steps (<1 sec), the rapid-quench flow techniques developed in this study make this feasible, and it will be of interest to see if either approach is able to extend the limits of the Class I ligase.
ACKNOWLEDGMENTS

I thank Jamie Cate for helpful experimental suggestions, and Margy Glasner and Bruce Tidor for helpful comments and discussion in the preparation of this chapter.
REFERENCES


FIGURE LEGENDS

Figure 1. The Class I ligase ribozyme. The ligase catalyzes the attack of the 3'-OH of a small oligonucleotide substrate (red, with DNA bases in lowercase) on its own 5'-α-phosphate, releasing pyrophosphate and forming a new 3'-5' linkage.

Figure 2. Class I self-ligation. Ligation rates are shown for reactions in which trace radiolabeled substrate (< 1 nM) was incubated with various concentrations of ribozyme at pH 6.0. Each point represents the average of at least three different measurements, and error bars indicate the standard deviation. The line represents the nonlinear, least-squares best fit to the Michaelis-Menten equation (equation 2, Materials and Methods). It indicates a $K_m$ of 40 nM and a $k_{cat}$ of 3 min$^{-1}$.

Figure 3. Measurement of substrate dissociation by pulse-chase partitioning experiments. Ribozyme (1 μM) was incubated for 10 seconds at pH 6.0 with trace radiolabeled substrate ([*S]/[R] ≤ 0.01). After 10 seconds, a chase mix containing a large excess of unlabeled substrate (final [S]/[R] = 50) was added. Shown are representative time courses done with (red) and without (blue) the added chase mix. The dotted line reflects the amount of product formed prior to chase addition.

Figure 4. Measurement of substrate association. Trace radiolabeled substrate (<1 nM) was incubated with the indicated concentration of ribozyme at pH 8.0, and observed rates
are plotted against ribozyme concentration. The line indicates the linear dependence of ligation rate on ribozyme concentration, with a slope of $2 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$.

Figure 5. pH dependence of folding and chemistry in Class I self-ligation. (A) Measurements of self-ligation using manual techniques (blue) were performed by incubating trace radiolabeled substrate with ribozyme (1 μM) in reaction buffer at the appropriate pH. Experiments measuring ligation rates at pH ≥ 7.0 (green) were conducted in essentially the same way in a rapid-quench flow apparatus. Experiments in which Ca$^{2+}$ was used to prefold the ribozyme in order to measure the rate of the chemical step (red) were performed as diagrammed in (B). Each rate is the average of at least two independent measurements. (B) A Ca$^{2+}$-preincubation method used to separate folding and chemistry in Class I self-ligation. Ribozyme and substrate are allowed to fold in Ca$^{2+}$ (3 mM), and the reaction is begun by addition of Mg$^{2+}$ (60 mM) and EGTA (5 mM). Rates measured using this technique are shown in red in (A).

Figure 6. Folding of the Class I ligase requires both Substrate and Mg$^{2+}$. Shown are ligation reactions done at pH 9.0 in which the ribozyme was either preincubated with substrate and the reaction initiated by addition of Mg$^{2+}$ (■) or the ribozyme was preincubated with Mg$^{2+}$ and the reaction initiated by addition of Substrate (●).

Figure 7. A Kinetic Framework for Class I Self-Ligation. A minimal kinetic framework is shown for ligation at 60 mM Mg$^{2+}$. R, ribozyme; R$_u$, unfolded ribozyme; R$_f$, folded
ribozyme; S, oligonucleotide substrate; P, ligated product. Measured rate constants are indicated at the appropriate step.
Figure 1
Figure 2

Ribozyme concentration (mM)

\[ k_{obs} \] (min⁻¹)
Figure 3

Fraction reacted

Time (min)
**Figure 5**

Diagram showing a reaction pathway involving MgCl$^2$ ETA and the effect of pH on ligation rate.
Selection of an Improved Class I Ligase Ribozyme
The work reported in this chapter was a collaborative effort involving Catherine Yen and myself. Catherine did the bulk of the work in synthesizing the randomized ligase pool, and is continuing to characterize the isolates that emerged from the selection. Everything else reported here is my own work.
Abstract

Background: The Class I ligase ribozyme catalyzes a Mg^{2+}-dependent RNA ligation reaction in which a small RNA substrate is ligated to the 5'-terminus of the ribozyme, forming a new 3'-5'-linkage and releasing pyrophosphate. Studies of the kinetics of both single- and multiple-turnover ligation have shown that the ribozyme is very fast for an RNA catalyst ($k_c = 300 \text{ min}^{-1}$ at pH 8.0), and have suggested strategies by which specific rate constants might be targeted for further optimization using in vitro selection.

Results: Unpaired regions of the ligase were randomized in both length and sequence, and in vitro selection was used to search for variants of the ribozyme that could perform extremely fast ligation under conditions in which the chemical step is rate-limiting.

Ribozymes isolated after seven rounds of selection had a slightly faster chemical step and substantially improved Mg^{2+}-binding, such that at 0.5 mM MgCl₂ improved clones were 50-fold faster than the parent ligase. This effect was largely due to a changed sequence in the joining region J1/3, which became remarkably adenosine-rich in improved clones.

Conclusions: Data from this study suggest that the chemical step of ligation is nearing an optimum, and that further improvement will be difficult without secondary structure changes. The sequence conservation seen in selected clones suggests that the structure of the ligase includes several as yet undefined tertiary interactions, and the mutations isolated in this study suggest links between the structure and function of the ligase.
Abbreviations

PCR, polymerase chain reaction; EDTA, Ethylenediaminetetraacetic acid; EGTA, Ethylene glycol-bis[β-aminoethyl ether]-N,N,N′,N′-tetraacetic acid; MES, 2-[N-Morpholino]ethanesulfonic acid; BES, N,N-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid; EPPS, N-[2-Hydroxyethyl]piperazine-N′-[3-propanesulfonic acid]; CHES, 2-[N-cyclohexylamino]-ethanesulfonic acid; RNase P, Ribonuclease P.
Introduction

It has long been suggested that a key stage in the development of life was an “RNA World” in which RNA acted as both genome and catalyst [1-5]. One of the most basic assumptions of this hypothesis is that RNA has the potential to catalyze its own replication, and considerable effort has been directed toward the search for RNA sequences with this ability [6-12].

One of the more promising leads in this search came in the discovery of the Class I ligase (Figure 1A), which was isolated from random sequences based on its ability to ligate a small RNA substrate to its 5'-terminus [11]. Its reaction is similar to that catalyzed by cellular RNA polymerases in that a template-aligned nucleoside triphosphate is ligated by the attack of a 3'-hydroxyl on the 5'-α-phosphate, displacing pyrophosphate [13, 14]. The link between the ligase and RNA polymerization was strengthened by subsequent work showing that an engineered derivative of the ribozyme could catalyze short primer extension reactions [15]. More recently, variants of the ligase have been isolated that are able to extend an RNA primer by up 14 nucleotides in an accurate, template-directed polymerization reaction [16].

A detailed mechanistic comparison between the ribozyme and biological polymerases has not yet been done, but several lines of data suggest the two may share a common mechanism. Biochemical and structural studies of protein polymerases have identified a pair of bound Mg^{++} ions that are essential for activity, a necessity that is paralleled in the Mg^{++}-dependence of the Class I ligase [17-24, M. E. Glasner, N. H. B., and D. P. B,
manuscript in preparation]. In addition, both the ligase and biological polymerases display the same stereospecific response to sulfur-substituted substrates, suggesting that there may be a metal ion bound at the same position relative to substrate in both classes of catalyst [25, 26]. Indeed, the currently proposed mechanism for protein-catalyzed phosphotransfer requires only those amino acid side chains involved in binding metal ions, and so seems particularly transferrable to an RNA catalyst [27, 28].

Recently the kinetics of the Class I reaction were examined in detail, using ligase ribozymes in both single- and multiple-turnover formats [29, M. E. G., N. H. B., D. P. B., manuscript in preparation]. These studies showed that ribozyme folding is constant with respect to pH, and occurs at 60 min⁻¹. In contrast, the rate of the chemical step for ligation (k₉) is log-linear with pH in the range 5.7-8.5, and continues to climb above pH 8.5, reaching 800 min⁻¹ at pH 9.0. The rate-limiting step, then, changes with pH; at pH 6.0, the rate of the chemical step (3 min⁻¹) is roughly 20-fold slower than folding, whereas at pH 9.0, chemistry is more than 10-fold faster than folding. This finding makes possible the isolation of individual steps in ligation, and suggests that rate constants might be specifically targeted for improvement in selection experiments.

In order to test this idea, and to better define the limits of the especially rapid chemical step seen in Class I ligation, an in vitro selection was performed under conditions in which the chemical step of ligation is rate-limiting. We have isolated an improved ligase that shows a slightly improved rate for the chemical step of ligation, and an unexpectedly large improvement in Mg²⁺-binding. Because of the latter effect, the improvement in
ligation rates achieved by the selection ranged from roughly 1.3-fold at saturating Mg\(^{++}\) (≥ 60 mM) to 50-fold at 0.5 mM. Under optimal conditions (pH 9.0, 60 mM Mg\(^{++}\)), the improved ribozyme has a \(k_c\) exceeding 1300 min\(^{-1}\), a speed that exceeds known RNA-catalyzed reactions.
Results

Selection and isolation of improved ligase variants

Previous selection experiments indicated that many of the unpaired residues within the ligase were important for activity, and suggested that single-stranded regions may play a role in defining the tertiary structure and active site of the ribozyme [30]. Based on these data, a randomized ligase pool was designed (Figure 1B) that explored more fully the possible sequences and lengths in these joining regions and loops. Residues that were highly conserved among the previously selected variants were mutagenized at a level of 10%. Less conserved residues were completely randomized in both sequence and length. We reasoned that allowing the ligase to eliminate as many unnecessary nucleotides as possible might lead to improved folding and/or catalysis, and would perhaps make the ribozyme more amenable to future crystallographic study. The helices of the ribozyme were left unchanged from the parent sequence, with a few exceptions as follows: The pair G73-C84 was changed to an A-U pair because experiments examining Class I-catalyzed primer extension showed that an A-U was slightly better at this position (E. H. Ekland, D. P. B., unpublished data). Helix P5 was extended by one base-pair so L5 sequences could be analyzed in a more stable context. Finally, for convenience in selection, the sequence of the region of the substrate that binds to the ribozyme was changed from 5'-CCAGUC to 5'-CCAGUA to approximate the sequence of the T7 RNA polymerase promoter. Similarly, G13, the 5'-terminal nucleotide in the segment of the ribozyme that pairs to substrate, was changed to a U in order to properly pair with this substrate.
Randomized ligase molecules were subjected to successive rounds of selection performed as diagrammed in Figure 2A. The procedure was essentially the same as performed previously, with two important differences. First, in this selection the substrate used was similar to the promoter for the T7 RNA polymerase. This allowed a direct progression between reverse transcription, PCR, and forward transcription, without the ligation steps that were formerly necessary to install the promoter. Second, selection steps conducted in this study were performed in a rapid-quench flow apparatus, so much faster reactions were possible. After 7 rounds, the activity of the pool reached a plateau (Figure 2B), so the cDNA was cloned, and individual isolates were sequenced.

Among the ligase sequences isolated, there were 35 unique clones, and ribozyme RNAs were prepared from each. None of the ribozymes tested depended on the presence of the reverse transcription primer or its binding site; in fact, the activity generally increased slightly when this primer-binding site was removed (data not shown). Thus, all ribozymes were prepared with a 3'-end that was truncated to match the length of the parent ligase. Furthermore, truncating the substrate from a 16-nt RNA/DNA hybrid to a 7-nt RNA did not affect activity (data not shown). The shorter substrate was used for all experiments following the initial characterization of clones.

The 35 unique ribozymes were assayed for ligation activity under the conditions of the final round of selection (pH 6.0, 10 mM Mg^{2+}, 200 mM KCl; Table 1). Ligation rates for these isolates ranged from 0.15 min^{-1} to slightly more than 2 min^{-1}. The fastest seven
clones (Numbers 101, 22, 96, 23, 91, 80, and 141) were essentially indistinguishable in rate, and they catalyzed ligation about twice as fast as the parent ribozyme.

Because selection was performed at 10 mM Mg\(^{2+}\), a concentration known to be below saturating for the parent ligase (M. E. G., N. H. B., and D. P. B., manuscript in preparation), we sought to determine whether selection had altered the Mg\(^{2+}\)-dependence of the ligase. Three representative clones from among the fastest group were chosen, and the activity of each was assayed at pH 6.0 in the presence of 1 mM Mg\(^{2+}\) (Table 1, column B). The activity of the parent ribozyme drops approximately 40-fold when the concentration of Mg\(^{2+}\) is dropped from 10 to 1 mM, yet the activity of the selected clones dropped by only 7- to 10-fold. One ribozyme in particular, clone 23, was able to catalyze ligation at a rate of 0.36 min\(^{-1}\), a pace 15 times faster than that of the parent ribozyme.

**Comparative sequence analysis of improved ligase variants**

The sequences of all isolates were aligned and compared to those of the starting pool and the parent ligase (Figure 3). With few exceptions, randomized positions showed a high degree of conservation, implying that after 7 rounds of selection almost every position was improved for catalysis. Of the positions mutagenized at 10% (A3-A4, A29-U34, G44-C47, C70-A71, A76, G88, A103, C113, and A121), only A76 and A101 changed identity consistently. The rest of these positions retained the parental sequence. As the chances of this outcome occurring randomly are very low, it seems clear that the parent ligase contained the optimal sequence for all of these positions except A76 and A103. The first of these, A76, became a U in 17 of 35 sequences, while the second position,
A103, became a C in 21 of 35 instances. Note in regard to the latter mutation that although in many of the selected clones this mismatch was converted into a Watson–Crick pair, the pair was almost always a C-G; the reciprocal A-U pair (caused by G88 → U) was observed only twice.

In a similar way, segments of the joining regions that had been completely randomized showed conservation in both length and sequence. The randomized segment within the region joining helices P1 and P3, J1/3, was almost always 10 nucleotides long in selected clones, even though in the initial pool this sequence varied from 2-10 nt. In addition to the observed length conservation, the region was remarkably A-rich in every isolate (on average, 75% of the bases in this segment were adenosine). The joining segment J3/4 showed somewhat less sequence conservation but was almost always the same size as in the parent ribozyme (4 nt), despite being randomized at 1-4 nt in the original pool. Of these four bases, the 5'-position (C40) was almost always the parental C, the 3'-position (A43) was usually the parental A, and the two interior bases (U41 and C42) varied considerably. Finally, as in J1/3 and J3/4, the length of the short segment joining helices P5 and P6 did not change from that found in the parent ribozyme (2 nt), and in most cases the parent sequence (5'-CA) was maintained as well.

In contrast with the joining regions, the two terminal loops, L5 and L7, showed no conservation in either size or sequence, except that loops of the largest allowable size (6 and 8 nt, respectively) often acquired an additional base-pair that extended the paired stem and decreased the size of the loop by two nucleotides. This lack of conservation is
similar to what has been seen in other selections [16, 30]. Apparently, the ligase has
minimal or no requirements for these sequences.

**pH and Mg**²⁺-dependence of the improved Class I ligase

Because the clone that performed best under more stringent conditions (clone 23) did not
include all of the observed changes, these mutations were combined in the context of that
close. Adding mutations to the sequence of clone 23 had little if any effect on ligation
rates (Table 2). The two changes that conferred slightly higher activity were combined
with the sequence of clone 23 to make an composite ligase sequence (Figure 4A),
hereafter referred to as the improved ligase, and this construct was used for all
subsequent experiments.

In order to better define the nature of the improvements attained in the selection, the rate
of the chemical step in ligation (kₜ) was measured for both the parent and improved
ligases. This was done at a relatively high Mg⁺⁺ concentration (60 mM) so that
differences in Mg⁺⁺ binding between the parent and the improved construct did not
obscure the interpretation of results. For the parent ribozyme reaction, the chemical step
is rate-limiting at low pH (< 7.0), and thus kₜ is equal to k_c [29]. At pH ≥ 7.0, however,
ribozyme folding becomes rate-limiting. This folding step requires both substrate and
divalent cations, so changing the order of reactant addition does not provide a way to
measure chemistry directly. Nevertheless, since the ligase is folded in Ca⁺⁺ without any
measurable activity, the chemical step can be measured directly by prefolding the
ribozymes in Ca⁺⁺, then replacing these ions with Mg⁺⁺ by simultaneous addition of
MgCl₂ and EGTA (M. E. G., N. H. B., and D. P. B., manuscript in preparation). Initial experiments showed that the fraction correctly folded and the folding rate were indistinguishable for the improved and parent ribozymes (data not shown), so the \( k_c \) for the improved ligase was determined as described above, using the Ca⁺⁺-prefolding method for measurement at pH ≥ 7.0 (Figure 4B). Like the parent ligase (Figure 4B, squares), rates for the improved ligase reaction (Figure 4B, circles) were log-linear with pH, with a slope of 1.0. At each point measured, the chemical step of the improved construct was slightly (1.2- to 1.5-fold) faster than that of the parent ribozyme, suggesting that the selective advantage in the mutations observed in the improved ligase did not lie in a greatly enhanced active site. Even so, there was this small yet reproducible improvement in the ribozyme’s chemical step, and it is intriguing to think that one of the acquired mutations might lie near the active site.

Because selected clones tolerated low Mg⁺⁺ concentrations better than did the parent ribozyme, the Mg⁺⁺-dependence of the improved ligase was examined in more detail. Ligation rates for the improved ligase and the parent ribozyme were determined at pH 6.0 in Mg⁺⁺ concentrations ranging from 0.5 mM to 100 mM (Figure 4C). At high Mg⁺⁺ concentrations (≥ 60 mM), the improved ligase was approximately 1.3-fold faster than the parent ribozyme. At 10 mM Mg⁺⁺, this margin increased to 2-fold. Below 10 mM, the parent showed a much sharper dependence on Mg⁺⁺ than did the improved ligase, and the improved construct was 15-fold faster at 1 mM Mg⁺⁺ and 50-fold faster at 0.5 mM Mg⁺⁺. Apparently, one or more Mg⁺⁺-binding sites in the parent ligase is either more easily saturated or eliminated by the changes in the improved ribozyme.
The two conserved mutations found in the composite ribozyme (J1/3 → 5' AUAAAAAAAGACAAAU and A76 → U) have a very different character, and thus might be expected to have correspondingly different effects on the ligase. To examine the individual effects of these two mutations in more detail, hybrid ribozymes were constructed. For each mutation, a ribozyme was constructed that was the improved sequence in every respect except that the mutation in question was changed back to the parental sequence (Table 3). (Note that in changing the sequence of J1/3 back to parent, the 5'-A residue was kept constant in order to maintain the 7 bp between substrate and ribozyme) An additional ribozyme was constructed in which the engineered change in P6 (G73-C84 → A-U) was switched back to the parental G-C pair in order to better define the effects of this change. Ligation rates were measured for each of these ribozymes under the same range of Mg²⁺ concentrations as above, and compared to the rates measured for the parent and improved ribozymes (Figure 4C). When the parental J1/3 sequence (5'-AACCGUAAAGACAAAU) was substituted into the improved ligase, almost all of the improvement was lost, and the hybrid showed a Mg²⁺-dependence very similar to that of the parent ligase. In contrast, when either position 77 (J6/3) or the changed pair in P6 was switched back to the parent sequence, the hybrid ribozymes showed a Mg²⁺-dependence similar to that of the improved construct, with only a small drop in activity over the entire range of Mg²⁺ concentrations. Thus, the difference in Mg²⁺-dependence between the improved ligase and the parent sequence is largely due to the change in J1/3.
Discussion

The *in vitro* selection data presented here provide a clearer, if not entirely new, view of the Class I ligase. Of the positions that were mutagenized in the construction of the randomized pool, most reverted to the parental sequence, and at almost every position examined the ribozyme showed a distinct sequence preference. Indeed, conservation was observed even in the regions in which the ligase was allowed to vary in both length and content (with the exception of loops L5 and L7). Note that several of the changes from the parent sequence were highly conserved, despite the fact that later experiments clearly showed that they conferred only slight selective advantages. We therefore suggest that the ligase sequence as shown in Figure 4A is, at least in regard to nearby sequence space, optimal, and that improved ligase variants are unlikely to be found without a selection exploring changes in paired residues.

Although the chemical step only improved slightly as a result of selection (Figure 4B), the Class I ligase in its improved form is an impressive ribozyme. At pH 8.0, the rate of the chemical step ($k_c$) is roughly equal to those reported for the fastest naturally-occurring ribozymes [31, 32], and at higher pH the ligase reaches rates never before seen in a ribozyme catalyzed reaction (22 sec⁻¹, pH 9.0). Comparing these rates with those of analogous protein enzymes, however, highlights just how much more efficient a protein polymerase active site can be. Multi-subunit RNA polymerases are much faster, as the $k_{cat}$'s measured for these enzymes are > 20 sec⁻¹ at physiological pH[33, 34], and replicative DNA polymerases are able to catalyze reactions with $k_{cat}$'s of 1000 sec⁻¹ [35].
Thus, although current variants may be nearing the catalytic limits of the Class I ligase, they remain far from the limits of a polymerase active site.

The observed improvement in the Mg\textsuperscript{2+}-dependence of the selected variants was somewhat surprising, because selection was performed at 10 mM Mg\textsuperscript{2+}, a concentration that is only ~5-fold below the parent ribozyme’s $K_{\text{m}}$ for Mg\textsuperscript{2+} (M. E. G., N. H. B., and D. P. B., manuscript in preparation). These conditions are not especially stringent, and thus improved Mg\textsuperscript{2+}-binding was not expected to confer a large selective advantage. Apparently, in lieu of ribozyme variants with vastly improved chemistry, ribozymes able to use Mg\textsuperscript{2+} more efficiently arose. This unexpected improvement has several interesting ramifications. First, the large improvement in catalysis at low Mg\textsuperscript{2+} (10-50-fold) means that the improved ligase may be better suited to physiologically relevant Mg\textsuperscript{2+} conditions (1-2 mM Mg\textsuperscript{2+}, [36-38]). Apart from raising questions of how well the ligase might function \textit{in vivo}, this finding supports the idea that ribozymes (perhaps like the ligase) performed RNA replication reactions in early organisms [39]. Current derivatives of the ligase catalyze RNA polymerization of up to 14 nt, but they prefer much higher concentrations of Mg\textsuperscript{2+} (200 mM) [16]. It would be of interest to see to what extent the improvements generated in this study can be transferred to these derivatives, and to see how much further the Mg\textsuperscript{2+}-binding properties of the ligase (and polymerase derivatives) could be improved when they are more directly targeted for improvement.

Perhaps the most compelling question that arises in light of the present study, however, is how the observed improvements are achieved by the few mutations found in selected
clones. The single base-pair change in P6 seems to make only a small difference, as does the mutation A76 → U (Figure 4C). Intriguingly, the latter base is only 10 Å from the ligation junction in the current three-dimensional model of the ribozyme (N. H. B., N. C. Lau, V. Lehnert, E. Westhof, and D. P. B., manuscript in preparation). It is tempting to propose that the small improvement in the chemical step is due at least in part to a subtle rearrangement caused by this change, though without a crystal structure this hypothesis is nearly impossible to prove.

The mutations observed in the joining region J1/3 seem to make the largest difference, and account for almost all of the improved Mg\(^{++}\)-dependence in the composite ligase. The remarkable result of changing this small sequence suggests that the nucleotides in J1/3 go beyond the role of a simple link between two helices, and somehow contribute to the stability of the entire ribozyme. A relatively simple explanation for this effect is that J1/3 participates in at least one critical Mg\(^{++}\)-binding site, and that the affinity of the site in question is increased in the improved clone such that it is always saturated in the Mg\(^{++}\) range tested. Alternatively, the changes in J1/3 might perform the same role at a distance, somehow permitting a slightly altered fold that changes the affinity of a Mg\(^{++}\)-binding site elsewhere in the ribozyme.

Either of these hypotheses is credible, but an alternative also seems conceivable: the improved ribozyme might have eliminated ≥ 1 Mg\(^{++}\)-binding sites, and thus faster rates are seen at low Mg\(^{++}\) because the ribozyme does not need as many ions bound. The striking prevalence of adenosine residues in J1/3 suggests several ways in which this
might occur. First, adenosine residues are ideal for forming structures based on stacking
energies, as adenine bases stack well yet have fewer side groups that might form
unwanted interactions [40]. One might imagine that adenosine residues in J1/3 would
contribute a greater propensity for stacking, and in this way J1/3 might be changed from
somewhat unstructured in the parent ribozyme to stacked in the improved ligase. Indeed,
stretches of poly(A) roughly as long as the Class I ligase J1/3 sequence have been shown
to form rigid, stacked structures under relatively mild conditions (neutral pH, 50-100 mM
NaCl, 20° C, [41, 42]). Given these data, it seems reasonable to propose that in the parent
ribozyme a Mg²⁺ ion might be needed for a stable J1/3 structure, whereas in the improved
close this sequence is strongly stacked on itself, rendering the metal ion unnecessary.

An alternative to an independently structured J1/3 is a scenario in which the improved
J1/3 sequence forms new contacts with other RNA elements. This sort of interaction is
easily conceivable given recent studies that have identified a generalized RNA tertiary
interaction in which consecutive, unpaired adenosine residues dock into the minor groove
of a helix, forming tandem base triples [43-45]. This ‘A-minor’ motif has been found in
a wide variety of RNA and RNA-protein structures, and appears to be a general mode for
RNA packing [43, 45]. If an analogous interaction appeared in the Class I ligase because
of the changed J1/3 sequence, it might stabilize a conformation that otherwise required
Mg²⁺ ions. In support of this idea, a similar effect has been observed in the group I intron
P4-P6 domain, where mutational analysis has shown that when an adenosine involved in
an adenosine-minor groove interaction is changed to uridine, the Mg²⁺-dependence of
folding changes sharply [43].
In the future, chemical-probing experiments may help to distinguish between these hypotheses. For instance, dimethyl sulfate methylates the N1 position of solvent-accessible adenine bases, and so might be expected to detect those adenosines that are involved in tertiary interactions and thus protected from solvent. This method has been useful in the past in identifying potential tertiary interactions [40, 46, 47], and might be similarly valuable in this case.

Ultimately, however, it seems likely that a detailed understanding of the mutations isolated in this study will require a high-resolution structure of the ligase. Structural studies have shed light on the mechanisms of folding and catalysis in several ribozyme systems [48-50], and a structure of the Class I ligase may not only clarify the ribozyme’s Mg\(^{2+}\)-binding properties and overall architecture, but also explain in more detail the mechanism of its exceptionally efficient reaction.
Materials and Methods

Pool design and construction

The DNA encoding a partially randomized Class I ligase pool was synthesized as a single, variable length oligonucleotide (5'-acgactcactataggAAcactatacttg(N_{10})-ACAATctgcc(N_{4})gagttgagaacatcg(N_{4})cgtg(N_{1})gaggAggcaagctctcgGtgg(N_{6})-ccaAgttcctcaacatatgtATTatctctgtgc, Figure 1B). Positions in lowercase were constant, and positions in uppercase were mutagenized at 10%. Thus, at these positions 90% of the molecules had the parental sequence, and the remaining 10% had one of the other three nucleotides. Positions noted as (N_{xy}) were completely randomized, and varied in length within the range noted. Randomized residues were synthesized using phosphoramidite mixtures that compensated for the different efficiencies with which each monomer adds to a growing DNA chain (relative efficiency 0.70:0.73:0.85:1.0, dA:dC:dG:dT). Length variation was introduced at the points indicated using a split-and-pool protocol [51]. This procedure resulted in a template pool varying from 134-150 nt, made up of a randomized ligase domain ranging from 107-123 nt that was flanked by constant primer-binding sites on the 5'- and 3'- ends. The synthetic DNA was purified by denaturing gel electrophoresis and used as a template in a large-scale PCR\textsuperscript{f} reaction (150 ml). Six cycles of PCR were done, where each cycle is 4 min at 96°C, 5 min at 42°C, and 7 min at 72°C. Primers for PCR were 5'-AAAGCAACAGGAATATT and 5'-TTCTAATACGACTCACTATAGG. Sequencing of clones from this PCR confirmed that randomization and length variation in the pool were as designed. Amplified DNA was phenol-extracted, precipitated, and used as a template for an \textit{in vitro} transcription reaction that generated pool RNAs.
Selection and amplification of improved ligases

The first round of selection was performed as diagrammed in Figure 2A. Pool RNA (580 μg, 1 μM final concentration) was heated in H₂O (80°C, 10 min) with a primer for reverse transcription (5'-AAAGCAACAGGAATATT) that hybridized to the pool’s 3'-terminus. After the RNA was cooled slowly to room temperature, biotinylated substrate (Dharmacon Research, 5'-biotin-taatacgactCCAGUA, DNA bases lowercase) was added to a final concentration of 1.2 μM along with buffer and salts (final concentrations, 10 mM MgCl₂, 200 mM KCl, 50 mM EPPS pH 8.0). The pool and substrate were incubated at 22°C for 5 minutes, and the reaction was stopped by addition of EDTA to 15 mM. The RNA was then precipitated and purified in a 10% polyacrylamide gel, using radiolabeled RNA markers that were 138 and 160 nt long (ligated pool ranged from 141-157 nt). After elution and precipitation, ligated product was heated (65°C, 10 min) and cooled (22°C, 10 min), then incubated with streptavidin paramagnetic beads (Promega) for 10 min according to the manufacturer's recommendations. After washes with 0.1X SSC (3 times), water (once), and reverse transcription buffer (once), additional RT primer was added and the remaining RNA was reverse-transcribed using an RNaseH-deficient reverse transcriptase (Superscript II, Gibco-BRL) while rotating in an incubator kept at 48°C. Parallel reactions were done without reverse transcriptase to detect any contaminating DNA in subsequent PCR steps. After 1 hour, EDTA was added to the reverse transcription mixture (final concentration = 3 mM), and the beads were washed three times with 0.1X SSC. Following the final wash, both the RT primer and a 5'-PCR primer (5'-TTCTAATACGACTCCAGTAGG) were added to 10 μM final concentration,
and the cDNA was removed from beads by alkaline denaturation (100 mM KOH, 20 mM Tris-Base). After removal from the beads, the cDNA solution was heated at 90°C for 10 min to hydrolyze any remaining RNA. It was then adjusted to pH 8.5 with HCl and used as a template for PCR.

Amplification of cDNA was performed using the RT primer and 5'-PCR primer described above, using a standard hot start protocol. After amplification, pool DNA was purified by Qiaquick column (QIAGEN). Concentrated DNA was used as a template in a second PCR in which the 5'-primer was changed to 5'-TTCTAATA-CGACTCAGTAGG.

Again, the resulting DNA was purified by Qiaquick column, and used as a template in a final PCR reaction using a 5'-primer containing the T7 RNA polymerase promoter sequence (5'-TTCTAATACGACTCAGTTATAGG). The final DNA was extracted with phenol and chloroform, then precipitated and used as template for an in vitro transcription which generated the pool RNA for the next round.

Further rounds of selection were performed as in round one, with the following changes: Beginning in round 2, RNA was treated with DNase (DNase RQ1, Promega) to eliminate template DNA prior to the selective ligation step. Also, in rounds 2-4 the gel-purification following the ligation step was omitted in favor of a simple precipitation. In rounds 5-7, gel-purification was again used to eliminate molecules that may have remained on the streptavidin beads because they were able to bind tightly to substrate. Furthermore, in rounds 5-7 reverse transcription preceded streptavidin-biotin affinity purification so that the RNAs were complexed with their cDNAs and thus unable to form structures that
might have a high affinity for the beads. The stringency of the ligation step was increased in each successive round by decreasing pH or reaction time (round 1, 5 min at pH 8.0; round 2, 20 sec at pH 7.0; round 3, 1 sec at pH 7.0; round 4, 2 sec at pH 6.0, rounds 5-7, 0.2 sec at pH 6.0). Buffers were at a final concentration of 50 mM, and were EPPS (pH 8.0), BES (pH 7.0) or MES (pH 6.0). For ligation steps requiring incubations of ≤2 sec, a rapid-quench flow device (KinTek corporation, Austin TX) was used. Finally, we observed that after round 6 the pool activity was identical with and without added RT primer, and thereafter it was omitted from the ligation step. After 7 rounds the pool cDNA was cloned (TOPO-XL kit, Novagen) and individual isolates were sequenced.

Ribozyme and Substrate RNAs

The parent Class I ligase ribozyme (Genbank no. U26413) was transcribed in vitro from a plasmid template linearized with EarI [29]. Transcripts were purified on 6% polyacrylamide/8M urea gels and stored in water at -20°C. Individual clone RNAs were transcribed from PCR templates and were purified in the same way. PCR templates were generated by amplifying specific clone cDNAs using a 5'-primer containing a T7 RNA polymerase promoter and either the RT primer used throughout selection or a primer (5'-AAAGCAACAGGAATATT) that truncated ribozymes to the same length as the parent ligase. Ribozyme constructs in which individual changes were made to a specific clone were prepared in the same way, using PCR primers to generate transcription templates with the desired changes.
All substrates used in this study (5'-aaaCCAGUC (parent ligase), 5'-taatacgactCCAGUA (selection) and 5'-UCCAGUA (ribozyme clones) [DNA bases lowercase]) were purchased (Dharmacon Research) and purified by anion-exchange chromatography (Nucleopac 9 x 250 column, Dionex). Substrates were radiolabeled using T4 polynucleotide kinase and [\(^{32}\)P]\(^\gamma\)-ATP.

**Manual ribozyme assays**

The ribozyme (1 \(\mu\)M final concentration) was heated (80° C, 2 min, in \(H_2O\)) and then cooled (22° C, 2 min), and the reaction was initiated by simultaneous addition of 50 mM buffer, the indicated concentration of Mg\(^{++}\), and trace \(^{32}\)P-labeled substrate. Buffers were the same as described for selection. Potassium chloride (200 mM) was included in initial experiments, but omitted in experiments measuring the Mg\(^{++}\)-dependence of ligation, and in all reactions in which the concentration of Mg\(^{++}\) was < 10 mM because at low Mg\(^{++}\) concentrations it inhibits the parent ligase (M. E. G., N. H. B., D. P. B, manuscript in preparation). Omitting KCl did not affect rates of either parent or clones at \(\geq 10\) mM Mg\(^{++}\). Aliquots were taken at appropriate time points and were added to 2 volumes stop solution containing 120 mM EDTA and 8 M urea. Product and substrate were separated in 20% polyacrylamide gels and quantified by phosphorimaging. Ligation rates were calculated as described previously [29].

**Rapid-quench flow ribozyme assays**

Ribozyme assays requiring time points shorter than 5 seconds were performed in a KinTek RQF-3 rapid-quench flow apparatus (KinTek corporation, Austin, TX). Final
reaction concentrations were 1 µM ribozyme, trace (<15 nM) radiolabeled substrate, 50 mM buffer, 60 mM MgCl₂, 200 mM KCl, 1.5 mM CaCl₂, 2.5 mM EGTA, and 0.8 mM EDTA. Reactions were initiated by mixing equal volumes of two solutions, one containing prefolded ribozyme-substrate complexes, and the other containing buffer and salts. The solution of prefolded RNA was prepared by first mixing ribozyme and water, then heating and cooling as for manual reactions. EDTA (1 mM) was then added to remove any metal contaminants, followed by addition of substrate. Finally, CaCl₂ was added to fold the RNA. Buffer solutions contained MgCl₂, KCl, buffer, EDTA (0.6 mM), and EGTA, which chelated the Ca²⁺ ions upon initiation of the reaction. Buffers were BES (pH 7.0, 7.4), EPPS (8.0, 8.5), or CHES (pH 9.0). Reactions were quenched with 500 mM EDTA, then collected and analyzed by gel electrophoresis and phosphorimaging.
Acknowledgements

We thank Scott Baskerville and Margy Glasner for helpful discussions and comments on this manuscript.
References


Figure Legends

Figure 1. The Class I ligase and the design of a randomized ligase pool. (A) The secondary structure of the ribozyme with substrate (green). Lower case indicates DNA residues. The ligase catalyzes the attack of the substrate’s 3’-OH on the 5’-α-phosphate of the ribozyme, with formation of a new 3’-5’ linkage and displacement of pyrophosphate. Helices are labeled P1-7. (B) The pool sequence paired to the substrate used for selection (green). Positions in red were either mutagenized at 10% (A, C, G, or U), or completely randomized (N). Regions labeled Nₙ₋ₙ also varied in length from X to Y nucleotides. The numbering system shown here reflects that of the composite improved ligase (Figure 4A).

Figure 2. Selection of improved ligase ribozymes. (A) In vitro selection scheme for improved ligase ribozymes. Ligase variants were incubated with a biotinylated substrate (B), and those ribozymes that ligated themselves to substrate were separated using streptavidin-coated paramagnetic beads. Reverse transcription and PCR regenerated dsDNA, and further PCR steps (corrective PCR) changed the substrate sequence into a functional T7 RNA polymerase promoter. The resulting DNA was transcribed to generate the RNA pool for the next round of selection. The first round was performed as shown, and further rounds were similar, as specified in Materials and Methods. The stringency of the ligation step was increased with each successive round by decreasing pH and/or reaction time. (B) Activity of pool RNA following each round of selection. Rates were measured at pH 6.0, 10 mM Mg²⁺, 200 mM KCl.
Figure 3. Comparative sequence analysis of selected clones. The sequence at top is the initial pool, and residues in red are those intentionally mutagenized. The sequence of the composite clone (Table 2, Figure 4A) is shown immediately below that of the pool. Clones are listed ranked by activity (Table 1). Nucleotides shaded dark blue are those differing from the pool sequence, and include all residues in regions that were completely randomized. Numbering reflects that of the composite improved ribozyme (Figure 4A). P refers to paired regions, J indicates joining segments, and L marks loop sequences.

Figure 4. An improved Class I ligase. (A) The secondary structure of the improved ligase sequence. Randomized positions where the parent sequence was retained are shown in green. Randomized positions that were conserved and different from the parent ligase sequence are shown in red. Positions in blue are those that were unconserved in selected clones. (B) pH-dependence of ligation by improved and parent ribozymes. Rates shown are those of the parent ligase (■) and the improved ligase (●). Ribozyme rates were measured in 60 mM Mg++, 200 mM KCl, 0.6 mM EDTA, and 50 mM buffer. Ligation reactions at pH 6.0 and 6.5 were performed using manual techniques, and those at pH ≥ 7.0 were done using a rapid-quench flow apparatus and a Ca++-preincubation technique that allowed the direct measurement of the chemical step (k ). (C) Mg++-dependence of improved and parent ribozymes. Rates are shown for parent ribozyme (■), improved ligase (●), and the improved ligase with the following mutations: A73–U84 → G–C (●), U76 → A (▼), and J1/3 → 5′-AACGUAAAAGACAAAU (▲).
Ligation rates were measured at pH 6.0 in a range of Mg^{++} concentrations ranging from 0.5 mM to 100 mM. Error bars represent the standard deviation of at least three independent measurements.
<table>
<thead>
<tr>
<th>Ribozyme</th>
<th>10 mM Mg$^{++}$</th>
<th>1 mM Mg$^{++}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
<td>1.10</td>
<td>0.025</td>
</tr>
<tr>
<td>Clone 101</td>
<td>2.50</td>
<td>0.22</td>
</tr>
<tr>
<td>Clone 96</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>Clone 22</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>Clone 91</td>
<td>2.20</td>
<td></td>
</tr>
<tr>
<td>Clone 23</td>
<td>2.15</td>
<td>0.36</td>
</tr>
<tr>
<td>Clone 80</td>
<td>2.15</td>
<td>0.24</td>
</tr>
<tr>
<td>Clone 141</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>Clone 159</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>Clone 84</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>Clone 173</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>Clone 178</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>Clone 50</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>Clone 162</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>Clone 69</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>Clone 55</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>Clone 89</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>Clone 70</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>Clone 71</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>Clone 175</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td>Clone 68</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td>Clone 66</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>Clone 2</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>Clone 106</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>Clone 18</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Clone 172</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Clone 77</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Clone 180</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Clone 1</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Clone 186</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Clone 124</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Clone 153</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Clone 35</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Clone 61</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Clone 187</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Clone 158</td>
<td>0.15</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Ligation rates of clones isolated from round 7 cDNA. Rates are shown for ligation under the conditions in which the final rounds (rounds 5-7) were performed (pH 6.0, 10 mM Mg$^{++}$, 200 mM KCl), or at pH 6.0, 1 mM Mg$^{++}$. Reported rates reflect the average of all measured rates for a given clone, using both 16-nt and 7-nt substrates.
Table 2

<table>
<thead>
<tr>
<th>Parent Ribozyme</th>
<th>Change</th>
<th>Relative Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 23</td>
<td>—</td>
<td>1.00</td>
</tr>
<tr>
<td>Clone 23</td>
<td>L5 → CUUCGG</td>
<td>0.87</td>
</tr>
<tr>
<td>Clone 23</td>
<td>L5 and L7 → CUUCGG</td>
<td>0.77</td>
</tr>
<tr>
<td>Clone 23</td>
<td>J1/3 → ATAAAAAAAGACAAAAU</td>
<td>1.06</td>
</tr>
<tr>
<td>Clone 23</td>
<td>J1/3 → GAAAAAAAAAGACAAAAU</td>
<td>0.78</td>
</tr>
<tr>
<td>Clone 23</td>
<td>A103 → C</td>
<td>0.88</td>
</tr>
<tr>
<td>Clone 23</td>
<td>J3/4 → CGCAGAGC</td>
<td>1.09</td>
</tr>
<tr>
<td>Clone 23</td>
<td>J1/3 → ATAAAAAAAGACAAAAU, J3/4 → CGCAGAGC*</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Table 2. Construction of an improved Class I ligase. Relative rates are shown for ribozymes derived from clone 23 in which mutations were added or changed. Rates reported are the average of ≥ 2 independent measurements at pH 6.0, 1 mM Mg++. The two changes that were reproducibly beneficial to the ribozyme were used in combination with clone 23 to generate the improved ligase sequence shown in Figure 4A. The asterisk marks the composite construct hereafter referred to as the improved ligase.

Table 3

<table>
<thead>
<tr>
<th>Parent Ribozyme</th>
<th>Change</th>
<th>Relative Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved ligase</td>
<td>—</td>
<td>1.00</td>
</tr>
<tr>
<td>Improved ligase</td>
<td>A73–U84 → G–C</td>
<td>0.63</td>
</tr>
<tr>
<td>Improved ligase</td>
<td>J6/3 (U76) → A</td>
<td>0.44</td>
</tr>
<tr>
<td>Improved ligase</td>
<td>J1/3 → AACCGUAAAAAGACAAAAU</td>
<td>0.13</td>
</tr>
<tr>
<td>Parent ligase</td>
<td>—</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 3. The effects of individual changes found in the improved ligase. Relative rates are shown for hybrid ribozymes derived from the improved construct, in which selected changes are changed back to the sequence found in the parent ligase. Rates reported are the average of ≥ 3 independent measurements at pH 6.0, 1 mM Mg++.
Figure 2

A

Transcription → RT primer-binding site → Ligase domain → Ligation → Corrective PCR → PCR → Reverse transcription

B

Graph showing the rate (min⁻¹) vs. rounds of selection.
Figure 3
Figure 4

A

Improved ligase

B

Ligation rate (min⁻¹)

C

Ligation rate (min⁻¹)

Mg²⁺ concentration
The Three-Dimensional Architecture

of the Class I Ligase Ribozyme
This work was a collaborative effort involving several colleagues. Nelson Lau performed the crosslinking experiments described here, and Valerie Lehnert and Eric Westhof modeled the structure of the ligase. All of the other work is my own.
Abstract

The Class I ligase ribozyme catalyzes a Mg$^{++}$-dependent RNA ligation reaction that is chemically analogous to a single step of RNA polymerization. The ligase is extremely efficient in both single- and multiple-turnover contexts, and so is informative for the study of RNA catalysis and RNA self-replication. Here we report the first characterization of the three-dimensional architecture of the ligase. Fe(II)-EDTA-generated hydroxyl radicals were used to measure the solvent accessibility at each position in the RNA backbone of the ribozyme. When the ligase was folded, several segments of the ribozyme became protected from cleavage, suggesting that when folded the RNA adopts a compact structure. Ribozyme folding was Mg$^{++}$-dependent, with a $K_{1/2}$[Mg] $< 1$ mM, and was observed over a broad temperature range (20° C-50° C). Together with comparative sequence analyses, these data were used to generate a three-dimensional model of the ribozyme. The predictive value of the model was tested and confirmed by a photocrosslinking experiment.
**Abbreviations:**

4SU, 4-thio-Uridine; EDTA, Ethylenediaminetetraacetic acid; MES, 2-[N-Morpholino]ethanesulfonic acid; BES, N,N-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid.
Introduction

Catalytic RNAs, like their protein counterparts, fold into specific and complex tertiary structures (Ban et al. 2000; Cate et al. 1996; Cate et al. 1999; Ferre-D’Amare et al. 1998a; Pley et al. 1994; Rupert & Ferre-D’Amare 2001; Scott et al. 1996). As these structures are intrinsically linked with function, an understanding of an RNA-catalyzed reaction requires complimentary studies of both the structure and activity of a given ribozyme. Unfortunately, though the reactions of many catalytic RNAs have been well defined, their structures are largely uncharacterized.

One such ribozyme, the Class I ligase (Figure 1A), catalyzes a reaction similar to that of biological RNA polymerases: attack by the 3'-OH of a small substrate RNA on a 5'-triphosphate, forming a new 3'-5' linkage with concomitant release of pyrophosphate (Ekland et al. 1995). Comparisons between this reaction and RNA polymerization were extended by experiments showing that engineered derivatives of the ligase were able to perform short primer-extension reactions (Ekland & Bartel 1996). More recently, variants of the ribozyme have been shown to catalyze template-directed polymerization of up to 14 nt (Johnston et al. 2001), supporting the idea that early in the origin of life RNA might have catalyzed its own replication (Bartel 1999; Joyce & Orgel 1999).

Recent studies have begun to define the reaction kinetics of the ligase, using ribozymes in both multiple- and single-turnover formats. In a multiple-turnover format, the ribozyme catalyzes ligation with a $k_{cat}$ greater than 2 sec$^{-1}$ at pH 8.0, a rate exceeding those of other multiple-turnover ribozyme-catalyzed reactions (Bergman et al. 2000). This speed is due
in large part to a very fast chemical step \( (k_c) \), and studies examining the original single-
turnover ligase have shown that this step can reach rates of 20 sec\(^{-1}\) at pH 9.0 (N. H. B.,
C. C. Yen, D. P. B., manuscript in preparation). The same studies of self-ligation also
showed that the ligase folds accurately and quickly (1 sec\(^{-1}\), as measured by attainment of
an active structure; M. E. Glasner, N. H. B., D. P. B., manuscript in preparation). This
suggests that the alternative folding pathways (and accompanying misfolding) seen in
other ribozymes (Pan & Sosnick 1997; Russell & Herschlag 1999) are less prevalent in
the case of the ligase.

Like protein RNA polymerases, the ligase has a near-absolute requirement for Mg\(^{++}\) ions
(M. E. G., N. H. B., D. P. B.). Furthermore, these protein enzymes and the ligase show a
similar stereospecific response to sulfur-substitution at the reactive phosphate, which is
consistent with the idea that one of these essential metal ions may be bound in the same
position relative to substrate in both catalysts (Eckstein 1985; Glasner et al. 2000). This
finding leaves open the possibility that the ligase uses the same mechanism as that
proposed for general protein-catalyzed phosphoryl transfer, and prompts questions of
whether the ribozyme might also share structural features with analogous protein
enzymes.

In order to better understand the relationship between the structure and function of the
ligase, we have begun to characterize its tertiary structure. The solvent accessibility at
each position along the sugar-phosphate backbone of the ribozyme was measured using
hydroxyl radical probing (Celander & Cech 1990; Celander & Cech 1991). These
measurements identified portions of the RNA that are protected from cleavage when the ligase is folded. Together with comparative sequence analysis, these data were used to model the ribozyme in three dimensions. Finally, a photocrosslinking experiment showed that the model successfully predicted three-dimensional proximity between ribozyme segments that were far apart in the secondary structure. The model provides key biochemical constraints necessary for the validation of future structural studies, and suggests strategies for crystallization of the ligase.
Results and Discussion

Hydroxyl-radical probing of the Class I ligase ribozyme

Hydroxyl-radicals, generated by either chemical methods (King et al 1993; Tullius et al 1987) or by synchrotron radiation (Sclavi et al 1998), have been used successfully to examine the structure of catalytic RNA's, ribosomal RNA's, and protein-nucleic acid complexes (Celandier & Cech 1991; Joseph & Noller 2000; Latham & Cech 1989; Tullius & Dombroski 1986). The radicals, when produced in solution with RNA, attack the ribose moieties in the nucleic acid backbone, causing strand cleavage (Hertzberg & Dervan 1984; Wu et al 1983). This cleavage is independent of sequence and secondary structure, and is instead dependent on the solvent accessibility of each position in the RNA backbone (Celandier & Cech 1990). In beginning investigations of the tertiary structure of the ligase, hydroxyl-radical probing was used to define portions of the ribozyme that are internalized by tertiary structure and thus protected from cleavage.

At 22° C and 10 mM Mg++, a significant number of nucleotides (17 out of 109 tested) became protected from hydroxyl-radical cleavage (protection factor >1.5, Figure 1) implying that the ligase assumes a compact structure in the presence of Mg++. The most striking protections were seen for nucleotides G45-C48, which make up the most conserved part of a joining region connecting helix P3 with helix P4, and for nucleotides G70-G75, which comprise the 5'-side of helix P5 (Figure 1). Interestingly, the opposite side of this helix (C81-C83) also had high protection factors, implying that this short helix might be almost completely surrounded by other RNA elements in the ribozyme core.

151
In addition to these regions, several single nucleotides became protected in the presence of Mg\(^{2+}\). Nucleotides A14 and C31 were consistently protected, though at a much weaker level than positions G45-C48 or G70-G75. Nucleotide U104 was strongly protected, perhaps because of the structure inherent in the junction of helices P4, P5, P6, and P7. Finally, nucleotide C62 was protected in the presence and absence of Mg\(^{2+}\), and may be protected because of local structure in the UUCG tetraloop (Cheong et al 1990).

**Temperature and Mg\(^{2+}\) dependence of Class I ligase tertiary structure**

Using the protection data from these nucleotides, the conditions of Mg\(^{2+}\) and temperature under which the ribozyme is folded could be identified. Hydroxyl-radical cleavage experiments were performed in a range of temperatures and Mg\(^{2+}\) concentrations, and the amount of ribozyme folding was measured by averaging the protection factors from positions G45-C48 and G70-G75 (Figure 2).

In the presence of 10 mM Mg\(^{2+}\), the ligase showed some structure at 10\(^\circ\) C, and was well-structured in the range 20\(^\circ\) C-40\(^\circ\) C. The Group I intron shows significant misfolding at lower temperatures, and the relatively lower protection seen at 10\(^\circ\) C may reflect a similar situation (Russell & Herschlag 1999). Note, however, that the pattern of protection did not change at low temperature, so if a misfolded form is more populated, it is not compact enough to form a structure with a new protection pattern. At 50\(^\circ\) C protection factors were again slightly lower, and at 60\(^\circ\) C most of the protection from hydroxyl-radical cleavage had disappeared (Fig 2A). At 60\(^\circ\) C and in the absence of Mg\(^{2+}\), the
ribozyme showed essentially no tertiary structure. These data agree well with UV thermal denaturation experiments, which show that in the presence of 10 mM Mg\(^{++}\) the ribozyme begins to show signs of denaturation at approximately 45\(^\circ\) C (N. H. B. and D. P. B., unpublished data).

When assayed at 20\(^\circ\) C, the ligase was essentially completely folded at all tested concentrations of Mg\(^{++}\), as low as 1 mM (Fig 2B). Fe(II)-EDTA probing could not be used as an assay for folding under conditions of < 1 mM Mg\(^{++}\), because of trace amounts of free EDTA in probing reactions. These results indicate that the \(K_{t/2(folding)}\) for Mg\(^{++}\) is below 1 mM, which is in agreement with data from similar studies (M. E. Glasner, N. H. B., and D. P. B., manuscript in preparation). In comparison, the \(K_{t/2(catalysis)}\) for Mg\(^{++}\) for is much higher (40-50 mM), suggesting that a native structure can be achieved without Mg\(^{++}\) ions bound in every catalytically useful binding site.

Interestingly, removing Mg\(^{++}\) ions from the ribozyme solution by adding EDTA did not completely denature the ribozyme’s tertiary fold. Rather, the areas that were found to be protected from cleavage when the ribozyme was folded in the presence of Mg\(^{++}\) were also partially protected by simply dropping the temperature from 60\(^\circ\) C to 22\(^\circ\) C in the absence of Mg\(^{++}\) (Fig 2C). Note that the presence of Mg\(^{++}\) changed the magnitude of each protection, but not the overall pattern of protection, suggesting that similar structures was present under both conditions. These data seem to suggest that in the absence of Mg\(^{++}\) the ribozymes are folded into a metastable, near-native conformation, though they do not rule out a scenario in which a small fraction of the ribozymes are folded correctly and the
majority are unstructured. If the former is true, the ligase might be similar to proteins in forming a semi-structured “molten globule” folding intermediate, though future experiments addressing folding transitions more directly will be necessary to substantiate this idea.

**Three-Dimensional Modeling of the Class I Ligase**

Together with comparative sequence information from secondary structure analysis (Ekland & Bartel 1995), the hydroxyl-radical probing data above were used to model the tertiary structure of the ligase (Figure 3A). In this model, the ligase assumes a compact structure in which the seven helices are closely linked by the double pseudoknotted secondary structure. Four of the seven helices (P4-P5-P6-P7) form two coaxial stacks that are arranged side by side, with P4 stacked on P7 and P6 stacked on P5. In this arrangement, the two stems P5 and P7 are parallel, and project out from the center of the ribozyme. This positions the two terminal loops, L5 and L7, near each other, but well away from any other portions of the ribozyme. The three remaining helices (P1-3) are arranged at the opposite end of the ribozyme. P1 is parallel to P4, and lies alongside it. The ligation junction lies near the center of the ribozyme, and is surrounded by P2, P4, P6, and J3/4. P2 extends from P1 in a third coaxial stack, projecting from the core of the ribozyme in a direction opposite to that of L5 and L7. P3 forms a pseudoknot that caps P6, and also changes the direction of the RNA strand more than 90° between J1/3 and J3/4. These two joining segments, together with P3, form an extended RNA strand that circles roughly three-fourths of the ribozyme.
Most of the nucleotides that were protected from hydroxyl-radical cleavage are internalized in the modeled structure. Nucleotides A14, G45-C48, and G70-G75 are packed tightly together, and form the interface between the parallel helices P1, P4, and P6. Nucleotide U104 is modeled well inside the ribozyme, at the junction of P4, P5, P6, and P7. As noted earlier, nucleotide C62 is in the middle of the L5 tetraloop motif, and probably is protected by local RNA structure. It is less clear how C31 and C81-C83 are protected. Recent selection data have suggested that tandem adenine bases in the joining region J1/3 might make contacts with the minor groove of another RNA helix, and it is tempting to suggest P6 in this role. If this were true, this interaction might explain the protection seen in P6 (C81-C83). Further experiments will be necessary before this idea could be confirmed, however.

**Verification of the Tertiary Model by Photocrosslinking**

The accuracy of the model is heavily dependent on the arrangement of helices P4, P5, P6 and P7. The model predicts that stems P5 and P7 project outward from the core of the ribozyme in a closely parallel fashion, thus positioning L5 next to L7. To test this prediction, we investigated whether a single 4-thio-Uridine (4SU) placed in the sequence of one loop could produce a crosslink to the adjacent loop. This particular crosslinking agent was chosen because it has been used successfully in a variety of settings (Christian & Harris 1999; Dontsova et al. 1994; Sontheimer & Steitz 1993) and because it generates photo-dependent crosslinks without a long linker arm (for review see (Favre et al. 1998)). The most straightforward way to incorporate 4SU involved breaking the ribozyme into two strands of RNA at either of the loops, so ligase molecules with a break in either L5 or
L7 were constructed and tested for activity. Ribozymes containing a break in L5 or L7 were 3-fold and 6-fold slower, respectively, than the unbroken parent ribozyme. Crosslinking experiments were conducted using a ribozyme containing a break in L5, since a 3-fold drop in activity is not expected to reflect a significant change in RNA structure. This two-piece ribozyme was composed of two RNA segments, designated fragment A and fragment B (see Materials and Methods).

Full ribozymes were assembled by mixing equimolar fragment A and fragment B RNAs (Figure 4A), then heating to 80° C in water and cooling to 22° C. Ribozymes lacking 4SU were assembled in the same way. Reaction buffer (buffer, MgCl₂, KCl, and EDTA) was then added, and ribozyme solutions were irradiated at 302 nM for 1 hour on ice. To these irradiated mixtures as well as control ribozyme solutions that were never irradiated, we added the 5′,32P-radiolabeled oligonucleotide substrate. Note that this substrate was prepared with a final phosphorylation step using unlabeled ATP so that essentially none of the substrate molecules had a free 5′-OH. After allowing the ligation reaction to proceed for 1 hour at 22° C, reaction mixtures were fractionated on a 10% denaturing polyacrylamide gel (Figure 4B).

When the 4SU-containing ribozyme mixture was analyzed following irradiation, five bands were detected above the major band corresponding to the uncrosslinked, reacted ribozyme (lane 4, Fig. 4B). Quantitation of the five bands indicated that each band contained approximately the same amount of radioactivity, and each was less than 1% of counts corresponding to the uncrosslinked, ligated product. Two of the five bands were
observed in the control reaction in which UV irradiation was omitted (lane 2, Fig. 4B), and the counts from these bands did not increase with irradiation. No bands were seen in control lanes showing ribozymes constructed without a 4SU residue, so it appeared that the three slowest migrating bands contained UV- and 4SU-dependent crosslinks.

All five crosslinked RNAs were purified and used as substrates in a kinase reaction, this time using [$^{32}$P]-$\gamma$ATP instead of [$^{33}$P]-$\gamma$ATP. This second radiolabeling identified products which crosslinked to the downstream RNA strand (fragment B), because only this type of crosslink would contain a free 5'-hydroxyl for labeling. Only one of the five isolated RNAs was efficiently radiolabeled (lane iii, Fig. 4C) and this product was purified from a denaturing polyacrylamide gel. The crosslink contained in this RNA was mapped by partial alkaline hydrolysis (Fig. 4D) and found to be to three residues on L7 (Fig. 4D). When the 3-D model was adjusted to mirror the break in L5, the 4-thio-Uridine is in clear proximity to these three residues (Fig. 4E). Thus, the identified crosslinks are in agreement with the predicted positioning of L5 and L7 in the 3-D model.

In addition to confirming one of the central predictions of the model, the crosslinking experiments reported here have practical implications for future crystallographic work. Previous studies suggested that the two terminal loops, L5 and L7, might be amenable to modification (Ekland & Bartel 1995), and this idea is confirmed by the finding that the ribozyme tolerates breaks and insertions in these regions. Studies of other RNAs have shown that crystallization often requires considerable intervention on the part of the researcher in the form of engineered RNA-RNA or protein-RNA interactions (Ferre-D'Amare et al. 1998b; Golden et al. 1997), and all available evidence suggests that these
loops would be an ideal position for such engineering. Further, these loops may prove useful as positions at which to incorporate heavy-atom-containing nucleotides. These advantages should help pave the way toward a high-resolution structure of the ligase.

Conclusion
With the data presented here, we have been able to gain the first view of the tertiary structure of the Class I ligase. As modeled, the structure of the ligase is quite compact, with extensive coaxial stacking. In a general sense, the model is similar to the x-ray structures of several recently characterized RNAs, in that are all based on parallel arrangements of coaxially-stacked helices (Cate et al 1996; Ferre-D’Amare et al 1998a). In the case of the ligase, these stacks seem to be held close by the pseudoknotted secondary structure and conserved joining regions that wrap around much of the ribozyme.

Almost all of the nucleotides that were protected from hydroxyl-radical cleavage are internalized in the modeled tertiary structure. Indeed, a calculated solvent accessibility profile based on the model fits well with experimentally determined values (data not shown). In addition, crosslinking data places L5 and L7 very close together, confirming the model’s predicted arrangement for the four-helix junction formed by helices P4-P5-P6-P7. Taken together, these results suggest that our view of the architecture of the ligase is accurate, and it will be of interest to see how well the modeled structure reported here compares with future structural studies.
Materials and Methods

Ribozyme and Substrate RNAs

The Class I ligase ribozyme (Genbank no. U26413) was transcribed in vitro from a plasmid template linearized with EarI as described previously (Bergman et al 2000). Transcripts were purified on 6% polyacrylamide/8M urea gels and stored in water at -20°C. The RNA components of ligases containing breaks in either L5 or L7 were transcribed from synthetic DNA templates and purified in 10% polyacrylamide/8M urea gels. The ribozyme fragment with a single 4SU\(^\ddagger\) residue (fragment A) was prepared as follows: a synthetic DNA template for transcribing the 5'-terminal 52 nucleotides (fragment A1) of fragment A was synthesized by standard phosphoramidite chemistry, except that 2-O-Methyl phosphoramidites were used for the last two positions of the template in order to reduce the heterogeneity at the 3'-end of the RNA usually seen in in vitro transcription reactions (Kao et al 1999). Radiolabeled fragment A1 RNA was prepared by adding \([^3P]\alpha\)-UTP to the transcription reaction. The 3'-terminal 11 nucleotides of fragment A (fragment A2) was purchased as an RNA oligonucleotide (5'-GAACAUUCC-[4SU]-U, Dharmacon Research). It included a terminal U nucleotide because 4SU-Controlled Pore Glass support beads were not available.

Fragments A1 and A2 were ligated to form fragment A as follows: fragment A1 (12 \(\mu\)M), fragment A2 (32 \(\mu\)M), and a 20 nt long DNA oligonucleotide that spans the junction of fragments A1 and A2 (24 \(\mu\)M), were heated to 80 °C in 1 mM EDTA, and 10 mM Tris-HCl (pH 7.5) for two minutes and allowed to cool slowly to room temperature (Moore & Sharp 1992). T4 DNA ligase buffer, 60 \(\mu\)M ATP and 3.7 units/\(\mu\)L of T4 DNA
Ligase (USB) were added, and the ligation reaction was incubated overnight in the dark at room temperature. Fully ligated products were separated in 10% polyacrylamide/8M urea gels. Ligation efficiency was ~60%, about twice that seen when fragment A was transcribed from a DNA template that lacked the 2'-methoxy-substitution.

The substrate for the ribozyme reaction was a synthetic RNA-DNA hybrid (5'-aaaCCAGUC, DNA bases lowercase (Bergman et al 2000)). It was radiolabeled using T4 polynucleotide kinase and either \([^{32}P]\gamma\text{-ATP}\) or \([^{33}P]\gamma\text{-ATP}\) (NEN). When substrate was used for crosslinking experiments, an additional 15 minute “chase” phosphorylation reaction containing excess unlabeled ATP was performed to ensure that nearly all substrate molecules had a 5'-phosphate. RNA concentrations were measured spectrophotometrically at 260 nm, assuming an extinction coefficient that was the sum of those for the individual nucleotides.

**Ribozyme Assays**

Ribozyme reactions in which the parent ligase and derivatives were compared were performed in 50 mM MES (pH 6.0), 60 mM MgCl₂, 200 mM KCl, and 600 μM EDTA at 22° C. In all cases the ribozyme was heated (80° C, 2 min, in H₂O) and then cooled (22° C, 2 min) just prior to initiation of the reaction. Ligation reactions were initiated by addition of buffer, salts, and trace \(^{32}\text{P}\)-labeled substrate. Aliquots were taken at appropriate time points and added to 2 volumes stop solution containing 120 mM EDTA and 8 M urea. Product and substrate were separated using 20% polyacrylamide gels and
quantified by phosphorimaging. Ligation rates were calculated as described previously (Bergman et al 2000).

*Hydroxyl-Radical Probing*

We note that a minority subset of ligase ribozymes is not active upon initial folding (Bergman et al 2000; Schmitt & Lehman 1999). To avoid probing the fold of misfolded ribozymes, ligase molecules were incubated with radiolabeled substrate, so that those that had assumed the active fold acquired the radiolabel and became visible in our analysis. Ribozyme (1 μM final concentration) was incubated with trace radiolabeled substrate (<100 nM final concentration) in buffer containing 50 mM BES pH 7.0, 10 mM MgCl₂, 200 mM KCl, and 0.1 mM EDTA. The ribozyme was heated (80° C, 2 min, in H₂O) and then cooled (22° C, 2 min) just prior to initiation of the reaction. Ligation reactions were incubated for 10 minutes at 22° C, at which point the reaction was diluted 10-fold into 50 mM BES pH 7.0 buffer with MgCl₂ sufficient to bring the concentration of Mg²⁺ to that indicated. When a final concentration of 0 mM Mg²⁺ was desired, the RNA was diluted 10-fold into a solution containing 50 mM BES pH 7.0 and 5 mM EDTA.

Labeled ligation product was subjected to hydroxyl-radical cleavage by adding to the RNA solution 0.1 volumes of a solution containing 20 mM (NH₄)₂Fe(II)(SO₄)₂, 20 mM ascorbic acid, and 22 mM EDTA. Solutions of (NH₄)₂Fe(II)(SO₄)₂ and ascorbic acid were prepared fresh before each experiment. Cleavage were typically for 15 min at 22° C. For probing at other temperatures, the RNA was also allowed to equilibrate for 15 minutes at the desired temperature prior to addition of the Fe/Ascorbate/EDTA solution.
Temperature did not noticeably affect cleavage reactions in the range 10° C–60° C. After addition of 2 volumes of a solution containing 8 M urea and 25 mM EDTA, cleaved RNAs were separated in 10% polyacrylamide gels and quantified by phosphorimaging (BAS2000, Fuji). Because a single gel provided accurate data on only ~50 nucleotides, reactions were typically loaded several times and electrophoresed for times ranging from 45 minutes to 4 hours in order to access most portions of the ribozyme. Individual positions were identified by comparison to partial alkaline hydrolysis and partial RNase T1 digestion ladders. Cleavage at each position was normalized to allow for differences in gel loading and cleavage efficiency (although cleavage was always done at levels that ensured that < 1% of the ligase molecules were cleaved), and protection factors were calculated for nucleotides 7-115. A protection factor was defined as the amount of cleavage at position N under denaturing conditions (60° C, 0 mM Mg++) divided by the amount of cleavage at position N under folded or experimental conditions (Pan 1995). The level of protection usually varied less than 30% from day to day, and the same positions were protected in each experiment.

Modeling of the Class I ligase

Molecular modeling was performed as described previously (Massire & Westhof 1998; Westhof 1993), using the program MANIP. The model was refined with the restrained least-squares program NUCLIN-NUCLSQ. Figures were produced using the program DRAWNA (Massire et al 1994).
Isolation and Mapping of Active, 4-thio-Uridine-Crosslinked Ribozymes

Active, crosslinked ribozymes were isolated as follows: The downstream half of the split ribozyme (fragment B) was dephosphorylated with alkaline phosphatase to expose a free hydroxyl on the 5'-end of the RNA. Fragment B RNA was then mixed with an equimolar amount of fragment A RNA to a final ribozyme concentration of about 1.2 μM in 10 mM MgCl₂, 200 mM KCl, 50 mM MES (pH 7.0), and 600 μM EDTA. As with the parent ligase, split ribozymes were heated and cooled prior to adding buffer and cations.

Solutions containing split ribozymes were then placed in a microtiter plate that was cooled to 4°C, and irradiated with a UV transilluminator (UVP) set at 302 nM. A polystyrene petri dish was used to filter out wavelengths lower than 300 nM. After irradiation for 1 hour, ³²P-labeled substrate was added (200 nM final concentration) and ribozymes were allowed to react for 1 hour in the dark at room temperature. Two volumes of a stop solution containing 8 M Urea and 25 mM EDTA was added to the reactions, and the mixed RNAs were separated in a 12% polyacrylamide/8M urea gel. Control experiments were performed in parallel omitting irradiation and/or the 4-thio-uridine substitution. UV-dependent, crosslinked ribozymes were detected by phosphorimaging. Crosslinked products were excised from the gel, eluted, and precipitated in ethanol.

The purified, crosslinked RNAs were then phosphorylated using [³²P]γ-ATP and purified in a 6% polyacrylamide/8 M urea gel. Inactive, crosslinked molecules should migrate differently in gels of different acrylamide percentages, so changing the gel percentage
between the first and second purifications allowed the active crosslinked ribozymes to be better separated from inactive molecules. RNAs that acquired a $^{32}$P label were detected by phosphorimaging and purified. The relabeled product was then subjected to limited base hydrolysis, and this reaction was run in a 10% polyacrylamide/8 M urea gel. The point at which crosslinking occurred was mapped by comparison to partial alkaline hydrolysis and partial RNase T1 ladders generated from uncrosslinked fragment B labeled with $[^{32}$P]γ-ATP.
Acknowledgements

We thank Chuck Merryman and James Berger for valuable discussions, and Peter Unrau, Craig Peebles, and Wendy Johnston for advice regarding the crosslinking experiments.
References


167


Figure Legends

Figure 1. Hydroxyl-radical cleavage of the Class I ligase. (A) The Class I ligase. The ligase catalyzes the attack of the 3'-OH of the substrate (shaded) on its own 5'-%alpha%-phosphate, forming a new 3'-5' linkage with release of pyrophosphate. P1-P7 are paired regions, and L5 and L7 are the loops extending from helices P5 and P7. Residues determined to be protected from hydroxyl-radical cleavage are colored red, and reflect those highlighted in (C). Residues colored blue are those for which solvent accessibility was not measured because they were too near the end of the RNA. (B) Representative gel showing hydroxyl-radical cleavage under different Mg++ concentrations. The gel shown was run for 90 min (see Materials and Methods). Data was collected for other positions by running gels for shorter or longer times. Nucleotides were identified by comparison to a ladder generated by partial digest of radiolabeled product using alkaline hydrolysis (lane designated OH) and RNase T1 (not shown). Control reactions in which Fe(II) or Mg++ were omitted are shown, along with a lane showing cleavage under denaturing conditions (60° C, 0 mM Mg++). (C) The histogram indicates the measured protection factors for nucleotides 7-115. Protection factors are defined as the ratio of cleavage under denaturing conditions to cleavage under experimental conditions. Positions determined to be protected when the ribozyme is folded are labeled and colored red.

Figure 2. Temperature- and Mg++-dependence of solvent accessibility in the Class I ligase. (A) Protection factors were averaged for G45-C48 and G70-G75, and the average is plotted for experiments in which hydroxyl-radical cleavage was performed at varying
temperatures in the presence of 10 mM Mg\textsuperscript{2+}. (B) Averaged protection factors are shown for the positions noted in (A) in experiments probing the ligase at different Mg\textsuperscript{2+} concentrations (at 22° C). (C) Radioactivity profiles of lanes from the gel in Figure 1B. Here the ribozyme was subjected to hydroxyl-radical cleavage in denaturing conditions (top), fully folded conditions (bottom), and conditions in which the secondary structure is formed but tertiary folding is not complete (middle). Regions in which cleavage protection was seen are shaded.

Figure 3. A three-dimensional model for the Class I ligase. (A) A ribbon representation of the 3-D model for the ligase. White spheres indicate those bases protected from hydroxyl-radical cleavage, and are those shaded white in (B). (B) A secondary structure diagram of the ligase is shown here that reflects the arrangement of helices proposed by the tertiary model. No bases or pairs have been changed from the original structure (Figure 1A). The color scheme reflects that of the model in (A).

Figure 4. Crosslinking analyses of the Class I ligase. (A) The secondary structure of the ligase ribozyme, indicating crosslinking results. The ribozyme is shown here with a broken L5, which facilitated insertion of a single 4SU residue. The black sequence is fragment A, and the red sequence is fragment B. Stars indicate the sites of radiolabeling on the two different strands comprising the ribozyme (black $^{32}$P, red $^{31}$P). Arrows mark the crosslinking sites as determined from (D). (B) Formation and separation of crosslinked ribozymes. In lanes 1 and 3, the two-piece ribozyme lacked 4SU; in lanes 2 and 4 it contained 4SU. In lanes 1 and 2 the RNA was not irradiated; in lanes 3 and 4
ribozymes were irradiated for 1 hour before radiolabeled substrate was added. Lanes 3 and 4 contain about twice as much sample as lanes 1 and 2. The arrows with Roman numerals point to products that were excised and eluted from gel slices for subsequent relabeling. (C) Relabeling of crosslinked RNAs. The lanes marked with Roman numerals correspond to the bands in panel (A). Only the RNA in lane iii was appreciably relabeled. (D) Mapping of the crosslinks within an active ribozyme. Lane 1 is unmodified, 5'-labeled 78.36 RNA. Lanes 2 and 3 are digests of labeled 78.36 RNA by RNase T1, and partial alkaline hydrolysis, respectively. Lane 4 is the partial alkaline hydrolysis of the relabeled crosslinked RNA shown in Figure 6b. Arrows mark the residues that form crosslinks within the the sequence of fragment B. (E) Magnified view of L5 and L7 as proposed by the three-dimensional model. The 4SU residue is colored red, and area on L7 forming crosslinks with 4SU is colored yellow.
Figure 2

A

![Graph A](image)

B

![Graph B](image)

C

![Graph C](image)
Future Directions

Taken together, the studies presented here place the Class I ligase in a somewhat unique position. The remarkable efficiency of the ribozyme reaction, the similarity of this reaction to those of biological RNA polymerases, and the possibility that Class I derivatives might someday be able to catalyze RNA self-replication all point to the fact that further study of the ligase will likely have an impact on a variety of topics. In this work we have defined the ribozyme as a model system, and established frameworks for both the reaction kinetics and the three-dimensional architecture of the ligase. Given these foundations, it should now be possible to examine both the structure and the mechanism of the ribozyme in more detail.

Although there are many different directions that future work could take, it seems that most of the readily imaginable experiments address one of three issues: the structure of the ligase, the mechanism of its reaction, and the process by which it folds. For any of these, an answer would be interesting not just for the sake of a better understanding of the Class I ligase, but also from the perspective of RNA catalysis, since the number of ribozymes whose structure, mechanism, or folding behavior is understood is still quite small. With this in mind, a direct method of seeking the answer to each of the above three questions is proposed below.

Structure of the Class I ligase

X-ray crystallographic analysis seems to be the logical choice for an examination of the structure of the Class I ligase. At even a low level of resolution (~5 Å) the electron-dense phosphate backbone of an RNA can be traced¹, and this information
would be valuable in confirming and/or correcting the current three-dimensional model.

At a higher resolution, a crystal structure would define the architecture of the ribozyme in terms of individual interactions\textsuperscript{2-5}. It would be of interest to see what non-Watson–Crick interactions exist inside the ligase, and to compare these data to the results of selections that repeatedly highlighted conserved residues in single-stranded joining regions. In particular, the results of the selection reported in this study could be much better understood, and the function of the J1/3 sequence and its remarkable preference for adenosine residues could be defined. Finally, it is interesting to note that fundamental RNA-RNA interactions are still being discovered\textsuperscript{6}, and a structure of this size might contain motifs that are informative in a very general sense.

Apart from defining the architecture of the ribozyme, a high-resolution structure would be interesting from the standpoint of the reaction mechanism. It may be possible to visualize metal ions that are bound at or near the active site, and deduce to some degree what their roles might be. The same would likely be true of RNA moieties in or near the active site. It might also be possible to identify the contacts the ribozyme makes with substrates (including both the RNA primer and the ribozyme’s 5’-oligonucleotide triphosphate). These contacts could be compared to analogous contacts made by protein RNA polymerases, and might help explain both the efficiency and the fidelity of the Class I reaction.

Historically, RNA crystallography has been quite difficult, largely because RNA is unstable and difficult to coax into well-ordered, highly diffracting crystals\textsuperscript{7}. Nevertheless, we have begun crystallization studies of the ribozyme, and have obtained promising results. Several constructs were screened for structural homogeneity, and one
was chosen for further study based on the finding that it ran as a single band in non-denaturing gels (N. H. B. and D. P. B., unpublished data). A large-scale purification protocol was then developed in which the ligase was produced by in vitro transcription, purified by denaturing gel, allowed to ligate substrate, and finally purified in a second denaturing gel. Initial crystallization conditions were screened using commercially available sparse matrix kits, and we found several conditions that resulted in crystallization (both microcrystals and long, thin needles). These conditions were further optimized, and several crystal forms were seen. One grew to a size that was large enough for preliminary x-ray analysis, and diffracted to 12-15 Å (N. H. B. and D. P. B., unpublished data).

Although this resolution is still a good distance from the range where useful structural information could be gained, several encouraging factors suggest that the structure of the ligase ribozyme can ultimately be solved. First, crystallization experiments so far have been conducted using only one ligase sequence, and no attempts have been made to optimize the sequence for crystallization. In the cases of some other ribozymes, several sequences had to be tested before one was found that diffracted well. Second, we have not yet tried to build crystal contacts using known RNA-RNA or protein-RNA interactions. In several cases the addition of known RNA-RNA interactions has made a drastic difference in how easily an RNA could be crystallized, and the same has been true when a well-characterized protein-binding site was added and the protein was included in the crystallization solution. Finally, preliminary diffraction was observed using a rotating anode x-ray source at room temperature. RNA crystals are often very radiation sensitive, and cryotechniques are usually crucial for data collection.
In light of this, it is conceivable that cooling alone could significantly improve
diffraction, especially if data collection can then be done using a synchrotron x-ray
source. Accordingly, the prospects for an eventual crystal structure of the Class I ligase
seem good.

*Reaction mechanism of the Class I Ligase*

Initial experiments have shown that the Class I ligase is Mg\(^{2+}\)-dependent, and
suggested that it may use the same two-metal-ion mechanism used by protein RNA
polymerases. Because of this, future experiments aimed at elucidating the reaction
mechanism of the ligase should begin by localizing and defining ribozyme-Mg\(^{2+}\) ion
interactions. Metal ions can sometimes be observed in RNA crystal structures, though
quite often they are hard to distinguish from bulk solvent. Mg\(^{2+}\) ions are especially
problematic in this regard, because they have an electron density similar to H\(_2\)O and
hydrated Na\(^+\) ions\(^{12}\). As a result, the most direct way to identify Mg\(^{2+}\)-binding sites in an
RNA is by biochemical methods.

Currently, the most effective biochemical technique for identifying Mg\(^{2+}\)-binding
sites in RNA involves the selective replacement of oxygen atoms by sulfur\(^{13}\). This
approach takes advantage of the fact that Mg\(^{2+}\) is an extremely hard Lewis acid, and so
does not bind well to sulfur ligands. Because of this, when an oxygen that normally
coordinates a Mg\(^{2+}\) ion is replaced with sulfur, metal-binding at that site is weakened, if
not eliminated. If the Mg\(^{2+}\)-binding site is essential for activity or folding, this effect is
easily detected, and potential ligands can be identified. To positively identify
substitutions that disrupt metal-binding, the experiment is typically repeated in the
presence of a thiophilic metal (e.g., Mn$^{++}$ or Cd$^{++}$) that might interact with the sulfur and replace Mg$^{++}$ at that position. Successful 'rescue' of activity or folding by the thiophilic metal is taken as evidence of inner-sphere metal-binding at that position.

Recent advances in chemical synthesis have made sulfur-substitution possible at many positions in RNA, including both pro-$R_p$ and pro-$S_p$ nonbridging phosphate oxygens and the 5', 3', and 2'-hydroxyls$^{13}$. Given these options, it should be possible to test some of the predictions that are made by the two-metal-ion hypothesis, and see whether the results are compatible with this mechanism as the basis for Class I ligation. Specifically, this model predicts that metal ions would contact the attacking 3'-hydroxyl and the pro-$R_p$ nonbridging oxygen of the $\alpha$-phosphate, as well as oxygens in the $\beta$- and $\gamma$-phosphates of the triphosphate. Already the two non-bridging $\alpha$-phosphate oxygens have been tested. When the pro-$R_p$-oxygen of the $\alpha$-phosphate was replaced with sulfur, there was a dramatic reduction in the rate of catalysis, while replacement of the pro-$S_p$-oxygen had a more modest effect$^{14}$. These results match the predictions of the model, and are in fact precisely the same as seen for protein polymerases$^{15}$. The next logical step is to test in a similar way the other relevant oxygens around the ligation junction, and see if the results continue to support, or eventually disprove, the idea that the ligase uses the same mechanism as protein polymerases.

It should be noted that the interpretation of sulfur-substitution experiments is sometimes difficult because sulfur can perturb RNA structure in ways other than by disrupting metal binding$^{13,16}$. For this reason, sulfur-substitution is most powerful in the context of a well-characterized structure, and it is tempting to imagine the experiments that might be possible if the structure of the ligase were solved. Such a combined
approach has been particularly successful in defining the metal-binding sites in the active site of a protein exonuclease and in the metal-ion core of a ribozyme domain\textsuperscript{16, 17}. With a defined structure in hand, one could also take a more general approach to sulfur-substitution. Several groups have used enzymatically incorporated phosphorothioates to examine potential metal-binding at each pro-\(R_p\) nonbridging phosphate oxygen throughout an RNA\textsuperscript{18-22}. These results, though difficult to interpret without a crystal structure, have been a valuable addition to structural studies\textsuperscript{16, 17}.

\textit{Folding of the Class I ligase}

In the past studies of RNA folding have shown that even relatively small RNAs follow complex folding pathways, often with significant kinetic traps\textsuperscript{4, 23, 24}. Large ribozymes in particular illustrate this, as both RNase P and the group I intron follow multiple folding pathways in the transition from denatured to fully folded, and a substantial fraction of the RNAs typically remain trapped in an incorrect fold\textsuperscript{25, 26}. In this regard, the Class I ligase is particularly well behaved. As measured by attainment of catalytic competence, Class I folding is quite fast (1 sec\textsuperscript{−1}), and roughly 70\% of the ribozymes in solution find the correct fold, suggesting that there are no major misfolding pathways. Some of this 'good behavior' can be attributed to the small size of the ligase relative to some of its peers, but even so it seems counterintuitive that a ribozyme that emerged from random sequences would fold faster or more accurately than some naturally occurring RNAs.

In light of this, an examination of the folding pathway for the Class I ligase seems in order. One of the most direct ways to do this is with time-resolved hydroxyl-radical
probing\textsuperscript{27}. In this technique, hydroxyl radicals are generated on a millisecond timescale by high-energy synchrotron radiation, and they cleave the RNA backbone at positions that are solvent-accessible. When the synchrotron is coupled with a stopped-flow device, probing can be carried out on RNA solutions as early as 20 ms after initiation of the folding process. This method has been successfully used to define ordered steps within group I intron folding and to measure the rate constants for transition between the steps\textsuperscript{27,28}.

The application of a time-resolved approach to the study of ligase folding might address several questions. As a first step, the time-dependence of ribozyme folding could be measured, and from this analysis it might be possible to identify specific folding steps. These data would likely reveal the general character of the folding pathway of the ligase, and show to what extent it follows a hierarchical pathway\textsuperscript{29}, as opposed to one based on a central collapse followed by smaller rearrangements\textsuperscript{30,31}. Finally, it would be interesting to see how direct measurements of folding compare with those measured by attainment of catalytic competence.

In performing the experiments described, it would be of interest to distinguish the folding pathway of ligated product from that of unreacted ribozyme. Further, folding of the latter could be observed when initiated by addition of either Mg\textsuperscript{2+} or substrate, and compared to the results obtained through kinetics experiments. It would also be of interest to explore further the conditions under which folding can occur, and here synchrotron-based methods provide a considerable advantage. Because the probing is not based on the addition of chemical reagents, a wider range of salt conditions can be examined, and this would be useful in measuring the Mg\textsuperscript{2+}-dependence of folding at low
Mg$^{++}$ concentrations, as well as defining the conditions under which the partially folded state can be found.

Overall, the Class I ligase could provide an engaging counterpoint to folding studies of naturally occurring RNAs. The folding pathways of some of the naturally occurring ribozymes are just now being worked out, and it seems that the general area of RNA folding would benefit from having another model system that arose in a completely different way (in vitro selection vs. natural evolution). Comparing the folding pathways of the ligase with those of other natural RNAs could show how evolution shapes the folding of an RNA molecule, and help explain some of the general aspects of RNA folding.
References


The Structural Basis for Terminator Recognition by the Rho Transcription Termination Factor

Cynthia E. Bogden,* Deborah Fass,† Nick Bergman,* Matthew D. Nichols,‡ and James M. Berger*†
*Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, Massachusetts 02142
†Department of Biology
Amherst College
Amherst, Massachusetts 01002
‡Department of Molecular and Cell Biology
University of California
Berkeley, California 94720

Summary

The E. coli Rho protein disengages newly transcribed RNA from its DNA template, helping terminate certain transcripts. We have determined the X-ray crystal structure of the RNA-binding domain of Rho complexed to an RNA ligand. Filters that screen both ligand size and chemical functionality line the primary nucleic acid-binding site, imparting sequence specificity to a generic single-stranded nucleic acid-binding fold and explaining the preference of Rho for cytosine-rich RNA. The crystal packing reveals two Rho domain protomers bound to a single RNA with a single base spacer, suggesting that the strong RNA-binding sites of Rho may arise from pairing of RNA-binding modules. Dimerization of symmetric subunits on an asymmetric ligand is developed as a model for allosteric control in the action of the intact Rho hexamer.

Introduction

The Escherichia coli transcription termination factor Rho is a hexameric helicase that can unwind RNA/DNA heteroduplexes upon ATP binding and hydrolysis (Brennan et al., 1987). The Rho protein binds to the nascent messenger RNA at specific Rho termination sites and may actively translocate along the RNA until it reaches the transcription complex (Geiselmarn et al., 1993; Platt, 1994; Richardson, 1996). There, it presumably facilitates termination by stripping the RNA from the template DNA. Although the Rho assembly is composed of six identical subunits, the enzyme appears to function as a trimer of dimers (Geiselmarn et al., 1993). For example, high-affinity ATP-binding sites exist at only three of the six potential sites (Stitt, 1988; Geiselmarn and von Hippel, 1992). Furthermore, three strong and three weak nucleic acid-binding sites have been detected (Geiselmarn et al., 1992; Wang and von Hippel, 1993b). These sites are apparently distinct, as the strong (primary) nucleic acid-binding sites are pyrimidine-prefering and DNA/RNA indiscriminant, while the weak (secondary) sites utilize C-rich RNA (Richardson, 1982; Bear et al., 1985; McSwiggen et al., 1988; Dolan et al., 1990; Wang and von Hippel, 1993b). Changes in ATP affinity and hydrolysis are observed upon adition of RNA, indicating that the nucleic acid-binding and ATPase functions are allosterically coupled (Richardson, 1982; Wang and von Hippel, 1993a).

The Rho ATP-binding region, which is homologous to the F, ATPase, is located in the C-terminal two thirds of the enzyme (Dombroski and Platt, 1988). The N-terminal third of Rho can be liberated by proteolysis from the rest of the protein and contains the strong, single-stranded nucleic acid-binding activity (Bear et al., 1985; Dolan et al., 1990). The structure of this domain in the absence of ligand has recently been determined and consists of an α-helical subdomain atop a 5-stranded β barrel (Allison et al., 1998; Briercheck et al., 1998). This β barrel belongs to the OB-fold class of single-stranded nucleic acid-binding motifs, which includes type II aminocyl tRNA synthetases and cold shock proteins (Cavarelli et al., 1993; Newkirk et al., 1994; Schindelin et al., 1994).

Structural homology to other OB-fold proteins, NMR chemical shift data, and mutagenesis has been used to assign the primary nucleotide-binding site to a cleft in the β barrel (Martinez et al., 1996a, 1996b; Allison et al., 1998; Briercheck et al., 1998).

Results and Discussion

To understand the molecular mechanism of RNA binding by Rho, we have determined the structure of the 13 kDa N-terminal RNA-binding domain, alone and in a complex with a 9-base, poly-cytidylc RNA oligonucleotide (Figure 1). We will refer to the protein domain as Rho13N and the RNA component as RC. The Rho13N apo structure was solved using multiple isomorphous replacement (MIR) methods and refined to a resolution of 2.0 Å with a free R factor of 26.1% and a working R factor of 19.3% (Table 1). This structure is comparable to that recently determined independently by X-ray crystallography and NMR (Allison et al., 1998; Briercheck et al., 1998). A polypeptide version of this structure was used as the search model for molecular replacement (MR) phasing of the Rho13N·RC crystals; the complex was refined to 2.4 Å with a free R factor of 28.7% and a working R factor of 24.8%.

Direct phasing of the data collected from Rho13N·RC crystals with the MR solution revealed strong density in the putative RNA-binding grooves of all three protein molecules in the asymmetric unit (Figure 1). Three-fold averaging and solvent flattening yielded maps into which two nucleotides per Rho13N domain could be readily built. An additional ribose/phosphate group for a third nucleotide'3' to the other two was also visible in each case, although the base for this group was invariably disordered. Overall, the size of the interaction is comparable with that of the aspartyl tRNA synthetase, which recognizes three nucleotides but contains a broader binding cleft (Cavarelli et al., 1993). No contacts

5 To whom correspondence should be addressed (e-mail: jnberger@ uclink4.barsley.edu).
† Present address: Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel.

"To whom correspondence should be addressed (e-mail: jnberger@ uclink4.barsley.edu)."
between Rho13N and the 2'-hydroxyl of the bases are observed, consistent with the observation that this domain binds either oligoribo- or oligodeoxyribonucleotides. The RNA oligorun 5'–3' from left to right across

![Figure 1. Structure of Rho13N Bound to RNA](image)

(a) Experimental electron density map superposed on a ball-and-stick model of the refined RNA. The map was generated using phases from the 3-folds-averaged poly-serine molecular replacement solution and the observed structure factors. Gold contours are at 1.0 σ and cyan at 2.5 σ.

(b) Refined 2F – F model-phased map of the same region. Gold contours are at 1.3 σ.

(c) Front view of the secondary structure of Rho13N monomer (green) shown with bound oligoribocytidine (ball-and-stick) in the OB-fold cleft. The cleft is formed on the surface of strands β2 and β3, with the β1–β2 and β4–β5 loops forming parts of the lower and upper walls, respectively. Helices and strands are labeled.

(d) View as in (c), rotated 90° about the vertical axis (a) and (b) generated by BOBSCRIPT and RASTER3D (Merritt and Murphy, 1994; Esnouf, 1997), (c) and (d) generated by RIBBONS (Craig, 1999).

Table 1. Multiple Isomorphous Replacement and Refinement Statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>Native</th>
<th>K, PτCl</th>
<th>Ter-Pt</th>
<th>Se-met</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20-2.0</td>
<td>20-3.1</td>
<td>20-3.1</td>
<td>20-2.6</td>
<td>14-2.4</td>
</tr>
</tbody>
</table>
| R
| 2 (last shell) (%) | 6.8 (16.8) | 6.3 (12.8) | 6.3 (14.9) | 6.8 (12.8) | 6.3 (14.9) |
| Completeness (last shell) (%) | 96.4 (94.8) | 90.9 (94.6) | 96.4 (94.6) | 96.2 (95.0) | 95.7 (96.8) |
| R
| 2 (last shell) (%) | — | 16.2 (17.0) | 16.1 (17.0) | 10.0 (12.7) | — |
| Number of sites | — | 2 | 3 | 6 | — |
| Phasing power* (last shell) | — | 3.0 (1.7) | 2.1 (1.7) | 4.3 (3.3) | — |
| R
| 2 (last shell) | — | 0.58 (0.78) | 0.75 (0.78) | 0.37 (0.45) | — |
| Figure of merit (20-2.6 Å) | 0.55 (0.30) |

<table>
<thead>
<tr>
<th>Refinement</th>
<th>Apo/Bound</th>
<th>Refinement</th>
<th>Apo/Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20-2.0/14-2.4</td>
<td>Number of non-H atoms</td>
<td>1902/3050</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>14,432/14,393</td>
<td>Protein</td>
<td>1812/2787</td>
</tr>
<tr>
<td>Working</td>
<td>13,276/13,685</td>
<td>Water</td>
<td>86/81</td>
</tr>
<tr>
<td>Free (8/5% of total)</td>
<td>1154/728</td>
<td>Ions</td>
<td>2/0</td>
</tr>
</tbody>
</table>
| R
| 2 (%) | 19.3/24.6 | rmsd bond lengths (Å) | 0.015/0.007 |
| R
| 2 (%) | 26.3/20.7 | rmsd bond angles (%) | 2.1/1.27 |

R = Σ |Fo| – |Fc| / Σ|Fo|,

where |Fo| is the intensity measurement for reflection j and |<Fo>| is the mean intensity for multiply-recorded reflections.

As the native set was obtained from three crystals, these values reflect Rref rather than Rfree.

R = Σ |Fo| – |Fc| / Σ|Fo|, where Fobs and Fcal are the derivative and native structure factors, respectively.

Phasing power = <Fobs> / |Fobs|, where <Fobs> is the root-mean-square heavy-atom structure factor and E is the residual lack of closure error.

R = Σ |Fo| – |Fc| / Σ|Fo|, where Fobs is the calculated heavy-atom structure factor.

Rfree = Σ |Fo| – |Fc| / Σ|Fo|, where the working and free R factors are calculated using the working and free reflection sets, respectively.

The free reflections were held aside throughout refinement.
minal subdomain between the apo and bound forms, although certain side chains have moved either to accommodate the RNA or due to lattice interactions.

An examination of the Rho13N-binding cleft shows how a nonspecific single-stranded DNA/RNA-binding domain can be adapted to select for pyrimidines, preferably cytidine (Figure 2, Table 2). The combination of Phe-62 and Leu-58 generates a hydrophobic shelf for the sugar moieties of the oligonucleotide, similar to that found in other OB-fold proteins such as replication protein A (RPA) and the aspartyl tRNA-synthetase (Cavarelli et al., 1993; Bochkarev et al., 1997). A distinctive feature of Rho13N, however, is that the β4/β5 loop extends directly over the β3 strand. Two amino acids, Glu-108 and Tyr-110, drape down from the sides of the loop, while Tyr-80 protrudes up from the β3 strand, to generate a three-wall enclosure that provides a snug fit for the base of the first cytidine. Hydrogen bonding to the Watson-Crick donor/acceptor groups of this base occurs exclusively from the main chain of the β4/β5 loop and seals off the top of this enclosure. While Rho could probably accommodate other pyrimidines at this site, binding of purines would require a large rearrangement of the loop. The NMR structure of the apo form suggests
Table 2. List of Protein Residues ≤4.0 Å from the Bound RNA in All Three Molecules

<table>
<thead>
<tr>
<th>RNA</th>
<th>Side Chain</th>
<th>Main Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H bond</td>
<td>VDW</td>
</tr>
<tr>
<td>Cyt7</td>
<td>Y80, E108, R109, Y110</td>
<td>E108, R109, Y110</td>
</tr>
<tr>
<td>Ribose</td>
<td>F62, Y80</td>
<td></td>
</tr>
<tr>
<td>Phosphate</td>
<td>Y80</td>
<td></td>
</tr>
<tr>
<td>Cyt2</td>
<td>R66, D78, A74, F64, Y110</td>
<td>G75</td>
</tr>
<tr>
<td>Ribose</td>
<td>F62</td>
<td></td>
</tr>
<tr>
<td>Phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

that this loop exhibits some mobility but is often in the conformation observed in the RNA-bound and both apo structures (Briecher et al., 1998).

The second nucleotide is more exposed than the first, although it appears more specifically coordinated by Rho13N (Figure 2). The contacts to the second base are entirely side chain mediated: the cytosine base stacks on the aromatic ring of Phe-64, and the guanidinium group of Arg-66 lies in line with the neighboring Asp-78 side chain such that the side chain donor/acceptor groups are coplanar and hydrogen bonded to the cytosine N4, N3, and O2 groups. This arrangement is strikingly reminiscent of hydrophobic and Watson/Crick interactions observed in duplex nucleic acid structures. Similar Watson/Crick mimicry and stacking of nucleotide bases on aromatic protein side chains occurs in other nucleotide-binding proteins, such as 7-methylguanosine mRNA cap-binding proteins (Hodel et al., 1997; Marrotigiano et al., 1997; Hakansson and Wigley, 1998).

In contrast, nonspecific OB folds, such as RPA, do not coordinate the base donor/acceptor groups so precisely (Bochkarev et al., 1997). Interestingly, Arg-66 of Rho13N is ion paired with Glu-56, which in conjunction with Asp-78 appears to lock the side chain of the arginine into the conformation observed in both the apo and bound forms. Thus, the extended side chains sterically select for pyrimidines at this site, while Watson/Crick-like pairing chemically selects cytosine over uracil. One may speculate that the weaker binding observed for this domain to uracil could potentially be accommodated through a shift in the Arg/Asp hydrogen bonding of the base, perhaps in a manner analogous to "G-U wobble."

In the lattice packing of the Rho13N::rC3 crystals, two Rho13N protomers share a single RNA chain separated by only one nucleotide, such that the two Rho13N domains bind in the space of only five cytidines (Figure 3). In the intact Rho hexamer, adjacent N-terminal domain protomers may also bind closely spaced sites on the RNA chain; biochemical studies support this proposal. For example, the length of RNA proposed to be bound by Rho in its three strong RNA-binding sites (~9-13 nt each) (McSwiggen et al., 1988; Geiselmann et al., 1992) is too long for an interaction between one Rho13N protomer and its target RNA. This length could be accommodated, however, if two Rho13N domains were to bind closely spaced RNA sites separated by a short spacer to constitute the strong RNA-binding interaction. Furthermore, increasing the length of short, single-stranded oligo-rCs from 6 to 8 nucleotides, or the order of the

Figure 3. Binding of a Single RNA by Two Rho13N Protomers
(a) View of two Rho13N molecules complexed to rC3 as packed in the crystal lattice. The two molecules are separated by a single base spacer. (b) Hypothetical reorientation of the two protomers seen in (a), generated by rotating one protomer and bound ligand about the free phosphodiesters bond of the spacer nucleotide.
(c) View as in (b), rotated 90° about the horizontal axis.
minimal length needed to span two Rho13N protomers, decreases the $K_a$ for ATP without affecting $v_{max}$ (Richardson, 1982). If, however, the primary binding sites are first saturated with poly(dA-rac-tC-tA) acid (poly-dC), then both $K_a$ and $v_{max}$ become independent of RNA-oligo length above six nucleotides (Richardson, 1982). These results can be explained if a minimal length of single-stranded nucleic acid were required to bridge the primary nucleic acid-binding sites of adjacent N-terminal domains and allosterically improve the efficiency with which the enzyme uses ATP (decreasing $K_a$). Once the N-terminal domains are bound, shorter cytosine-rich RNAs bound to the secondary site can stimulate ATP turnover (increasing $v_{max}$) (Richardson, 1982). It should be noted that cooperative binding of asymmetric ligands by homooligomeric proteins has been observed in a number of systems, often with profound regulatory and functional consequences (Cunningham et al., 1991; Hatsuda et al., 1995).

Rho forms a hexameric ring, similar to the F$_2$ ATP synthetase and to other hexameric helicases (Gogol et al., 1991; Abrahams et al., 1994; Yu et al., 1996b). The known structure of the F$_2$ ATPase thus provides a framework for incorporating the Rho13N$\cdot$rC$_5$ structure into a global model for Rho function. It has been suggested that the Rho13N domains lie horizontally in a head-to-tail arrangement about the ring, based on structural alignment of the Rho13N five-stranded OB fold with the N-terminal, six-stranded $\beta$ barrel in F$_2$ (Allison et al., 1998). However, the topology of the superposed regions of the $\beta$ barrels from the N-terminal domains of F$_2$ and Rho are different, suggesting that they are not derived from a common precursor and are therefore not constrained to an orientation that maximizes their structural alignment. Furthermore, the required shift in the radial position of Rho13N to accommodate the extra bulk of the N-terminal subdomain abrogates direct superposition. Finally, this arrangement orients the nucleic acid-binding grooves outward from, and parallel to, the 6-fold axis, requiring that RNA zigzag between clefts to retain the appropriate binding polarity.

Alternatively, Rho13N may occupy the same space in the Rho hexamer as the N-terminal domain of F$_2$ does relative to its ATPase domain, but oriented such that the nucleic acid-binding clefts are perpendicular to the 6-fold axis and tangential to the ring of Rho13N domains. In this model, the primary binding sites would be aligned such that nucleic acid would feed directly from one cleft to the next, with the polarity observed in the Rho13N$\cdot$rC$_5$ crystals (Figure 3). This proposal also suggests how Rho, despite being a homohexamer, executes certain functions as a trimer of dimers. Pairwise binding of adjacent Rho13N domains to closely spaced sites on the RNA could convert Rho from a symmetric a configuration to an (A)$_3$ state (Figure 4), similar to structural transitions that have been observed in DNA-dependent hexameric helicases (San Martin et al., 1995; Yu et al., 1996b). While the fundamental mechanism of Rho may therefore be analogous to those of its hexameric helicase counterparts (Egelman, 1996; West, 1996), the Rho helicase has been augmented with a set of nucleic acid "sensors" that home the protein to a particular site on the messenger RNA. Binding to this site in turn regulates the ATPase, duplex-unwinding, and translocation activities necessary for transcription termination by Rho.

**Experimental Procedures**

The Rho13N coding sequence (spanning residues 1-118 of the wild-type E. coli Rho protein) was amplified by PCR from a plasmid containing the intact Rho gene (C. E. B. and J. M. B., unpublished data) and cloned into pET28b (Novagen). Two residues, Gly-1 and His-6, were inserted at the N terminus during cloning. Protein expression was performed in BL21(DE3)pLYS5 cells by inducing with an $A_{600}$ of 0.3 with 1 mM IPTG for 2.5 hr at 37°C. Rho13N containing selenomethionine was prepared as described (Van Dyne et al., 1993), except with an induction time of 3 hr. Cells were lysed, and the extract was purified by ammonium sulfate precipitation (50%-70%), followed by Q-Sepharose and phosphoelulobase column chromatography in buffer A (50 mM Tris-HCl [pH 7.5], 10% glycerol, 1 mM EDTA, 1 mM EGTA, 1 mM pepstatin-A, 1 mM leupeptin, 1 mM PMSF + 100 mM KCl). The flowthrough containing Rho13N was concentrated and applied to a Pharmacia Sepharose S-300 gel filtration column equilibrated in buffer A + 200 mM KCl. Peak fractions were concentrated (Centricon-3, Amicon) to 25-30 mg/ml (Eidelberg, 1967), dialyzed against 10 mM Hepes-KOH (pH 7.5) + 50 mM NaCl, and diluted to 12 mg/ml. Purification of selenomethionine protein was done as for native protein, but with 2 mM DTT. Incorporation of selenium was verified by mass spectrometry (Voyager DE Bioworks Workstation, PerSeptive Biosystems; data not shown). rC$_5$ RNA was synthesized and deprotected (Tuschl et al., 1998), purified by anion-exchange HPLC (Biorad NucleoPac), typified, and resuspended in water to a concentration of 10 mg/ml. Rho13N was crystallized by the hanging drop vapor diffusion
molecular cell 492

method. Rho13N and well solution (10%–20% MPD, 50–100 mM ammonium acetate, 100 mM Tris (pH 8.5), 4 mM sucrose chloride, 2 mM eritum acetate) were mixed 1:1, and the well solution was diluted with 1 vol water before sealing. Thick plates (0.4 × 0.2 × 0.1 mm) grew in 1–2 weeks at 30°C and were equilibrated to 20°C for 24 hr before harvesting. Diffraction data were collected at 20°C on crystals harvested with well solution and mounted in capillaries, using a Rigaku X crystal on a Rigaku RU-300 rotating copper anode at 1.54 Å. Derivatives other than the selenomethionine were prepared by adding 0.2 µl of a 100 mM heavy atom stock directly to drops containing Rho13N crystals and soaking overnight. For the RNA-bound form, Rho13N was mixed with CsCl at a 1:0:1 molar ratio, then mixed 1:1 with a well solution of 600 mM ammonium acetate + 50 mM sodium acetate, before sealing over well solution. Thin plates (0.4 × 0.1 × <0.05 mm) grew in 1–2 weeks. These crystals were equilibrated with cryoprotectant by stepwise transfer into well solution plus increasing concentrations of ethylene glycol (5%, 10 min; 15%, 10 min; 20%, 5 min) before flash freezing in a gas-jet stream at 100 K for data collection. Data for the complex were collected at Brookhaven National Laboratory Beamline X4A, at a wavelength of 0.9779 Å. Apo-Rho13N crystallizes in the space group P1, with unit cell dimensions a = 30.42 Å, β = 36.18 Å, c = 54.25 Å, α = 90.06°, β = 102.84°, and γ = 96.43°, and two molecules per asymmetric unit. The RNA-bound complex crystallizes in the space group C2, with unit cell dimensions a = 132.8 Å, β = 31.3 Å, c = 155.8 Å, α = γ = 90°, β = 120°. The crystals contain these proteins molecules and two RNA oligonucleotides per asymmetric unit.

For data on all crystals were indexed and reduced with DENZO/SCALEPACK (Otwinski and Minor, 1997). The CCP4 set of programs was used for truncating and scaling native and derivative data sets (Collaborative Computational Project Number 4, 1994). Phase calculation and density modification on the unliganded Rho13N data were carried out with SHARP (Forteole and Brignone, 1997) and SOLOMON (Abrahams and Leslie, 1996), respectively. Model building for both structures was done with O (Jones and Kjeldgaard, 1997). Model refinement of the apo structure was carried out using XPLOR until the free R factor reached 36% (Brünger, 1993), followed by a single round of rebuilding into simulated-annealed omit maps (Hodg et al., 1992). The rebuilt model was subsequently refined using a Refmac5/ARP procedure (Lamzin and Wilson, 1993; Marzouk, et al., 1997) that incorporated a bulk-solution correction with automated water placement until refinement converged, yielding a free R factor of 25.1% and a working R factor of 18.3%. A poly-serine version of this model was used in molecular replacement (AMORE (Navaza, 1994) to solve the structure of Rho13N/RNA complex; the RNA was solvent flattened and 3-fold averaged using DM (Cowtan, 1994). This structure was refined using X-PLOR with NCS restraints, manually rebuilt into simulated-annealed omit maps (Jones et al., 1990; Ricci and Brünger, 1994), and then refined with CNS (Brünger et al., 1998), until the structure converged with a free R factor of 28.7% and a working R factor of 24.6%. The completed structures include residues 1–68 and 76–118 for the apo model and 1–118 for the complex.

Acknowledgments

The authors would like to thank David Bartel and members of his lab for technical assistance with RNA preparation, Richard Calden for critical reading of the manuscript, Craig Ogata and the Brookhaven X4A beamline staff for assistance collecting data from the Rho13N/X4A, and the Harold Y. and Leila G. Mathers Charitable Foundation, Whitehead Institute, and Berkeley MCB department for financial support.

Received November 24, 1998; revised January 25, 1999.

References


Fortelle, E.D.L., and Brignone, G. (1997). Maximum-likelihood...


Protein Data Bank Accession Numbers

The accession numbers for the coordinates of the Rho13N apo and complex structures are 1A8V and 2A8V, respectively.