13

Building a Catalytic Active Site Using Only RNA

Thomas R. Cech and Barbara L. Golden
Howard Hughes Medical Institute
Department of Chemistry and Biochemistry
University of Colorado
Boulder, Colorado 80309-0215

WHERE DO RIBOZYMES FIT IN THE RNA WORLD HYPOTHESIS?

Two of the most fundamental requirements for life are information storage and catalytic function. Without storage, transfer, and replication of information, a system cannot learn from its past and improve its viability; that is, there can be no natural selection. Equally essential to life is catalytic (enzymatic) function. At the very least, there must be machinery to catalyze the copying of the informational molecules. This process must proceed with considerable fidelity, yet some frequency of errors is also necessary to provide the diversity that allows adaptation and evolution. Beyond replication of the genome, additional catalytic functions would be highly advantageous to provide basic metabolism for even a primitive self-reproducing system.

In contemporary organisms information is stored in the form of DNA. However, the persistence of RNA genomes in many viruses shows us that this sister nucleic acid is competent for information storage, at least for small genomes. Biocatalytic function in the modern world is mostly the domain of protein enzymes, although ribonucleoproteins (RNPs) still catalyze the essential cellular reactions of protein synthesis and RNA splicing. By what evolutionary pathway did this DNA-RNA-protein solution to the problem of life come about? The finding that RNA, an informational molecule, can by itself catalyze biochemical reactions has rekindled enthusiasm for the possibility that a key intermediate stage was an RNA World, with RNA providing both information and function, genotype and phenotype.

One version of this RNA World hypothesis is diagrammed in Figure 1. RNA is usually considered to be too complex a polymer to arise by
random chemistry, thus the proposal of a simpler progenitor (see Chapters 2 and 6). The advantages of replicating a more homogeneous polymer with a single type of building block may have driven the emergence of an RNA World, with RNA catalyzing its own replication (see Chapter 5). At this stage ribozymes, the subject of this chapter, would speed up and provide specificity to biochemical reactions. Even at the very earliest stages, ribozymes would probably work in concert with whatever randomly assembled peptides and other molecules were in their environment. But once a system stumbled upon a means to direct the synthesis of specific peptide reproducibly—that is, to translate information from RNA to protein (see Chapters 7 and 8)—ribozymes would start working as RNP enzymes. Thus, the RNA and RNP worlds would overlap extensively. Many of the ribozymes that function in contemporary cells continue to work as RNPs, as discussed in Chapters 14 and 18.

Will we ever know whether such an RNA World really existed? Perhaps not, but we can test the chemical plausibility of ribozymes performing diverse catalytic roles. Is RNA catalysis fast enough, specific enough, and versatile enough to support complex metabolism? How does catalysis by RNA compare with that attainable by proteins, and what is the structural basis of the similarities and differences? These are the questions we address here.

**RIBOZYME-CATALYZED REACTIONS**

**How Versatile is RNA Catalysis?**

Natural ribozymes (Table 1) are all involved in RNA processing reactions. RNase P cleaves the 5′ leader from primary transcripts of tRNAs. Chemically, this reaction is phosphodiester bond hydrolysis (Fig. 2a).
Group I and II introns mediate the splicing of the RNA in which they reside by a series of two concerted phosphodiester bond cleavage–ligation (transesterification) reactions (Fig. 2a). Group II RNAs also cleave and insert themselves into double-stranded DNA in a reaction that requires the cooperation of the ribozyme active site and a protein moiety (Zimmerly et al. 1995). Small ribozymes, such as the hammerhead, clip RNA replicative intermediates to make unit-size progeny molecules by a different transesterification mechanism (Fig. 2b) (see Chapter 11). All these reactions are quite similar, in that the substrates are themselves nucleic acids, and chemically they all involve attack of a nucleophile at a phosphate. This is only a very small segment of the constellation of reactions catalyzed by protein enzymes.

Ribozymologists have reengineered self-splicing and self-cleaving ribozymes to act as multiple-turnover catalysts by separating the portion

**Table 1 Natural ribozymes**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number sequenced</th>
<th>Biological sources</th>
<th>Reaction performed (reaction product)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-splicing RNAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>&gt;500</td>
<td>Eukaryotes (nuclear and organellar), prokaryotes, bacteriophage</td>
<td>transesterification (3’-OH)</td>
</tr>
<tr>
<td>Group II</td>
<td>&gt;100</td>
<td>Eukaryotes (organellar), prokaryotes</td>
<td></td>
</tr>
<tr>
<td>Self-cleaving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I-like</td>
<td>6</td>
<td>Didymium, Naeglia</td>
<td>Hydrolysis (3’-OH)</td>
</tr>
<tr>
<td>Small self-cleavers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hammerheads</td>
<td>11</td>
<td>Plant viroids and satellite RNAs, newt</td>
<td>transesterification (2’, 3’&gt;p)</td>
</tr>
<tr>
<td>hairpin</td>
<td>1</td>
<td>Satellite RNA of tobacco ringspot virus</td>
<td></td>
</tr>
<tr>
<td>HDV</td>
<td>2</td>
<td>Human hepatitis virus</td>
<td></td>
</tr>
<tr>
<td>VS</td>
<td>1</td>
<td>Neurospora mitochondria</td>
<td></td>
</tr>
<tr>
<td>RNase P RNAs</td>
<td>&gt;100</td>
<td>Eukaryotes (nuclear and organellar), prokaryotes</td>
<td>Hydrolysis (3’-OH)</td>
</tr>
</tbody>
</table>

of the molecule that contains the active site from the portion that undergoes reaction (Zaug and Cech 1986a; Zaug et al. 1986). This concept has been extended to so many catalytic RNA systems that it is already commonplace, yet it is critical for versatility: An “enzyme” restricted to a single round of reaction with a portion of itself would have extremely limited roles in an RNA World.

Ribozymologists have also coaxed natural ribozymes to catalyze alternative chemistries. Group I and group II ribozymes can cleave single-stranded DNA as well as RNA (Herschlag and Cech 1990a; Robertson and Joyce 1990; Mörl et al. 1992). The *Tetrahymena* ribozyme can trans-
fer a 3’-terminal phosphate from one RNA strand to another, and it can hydrolyze a 3’-terminal phosphate (Zaug and Cech 1986b). These reactions involve phosphate monoester substrates instead of the usual diester. These DNA-cleavage and phosphomonoester reactions do not extend the ribozyme repertoire greatly, as the substrates are still nucleic acid and the reactions still involve phosphorus centers. Although the *Tetrahymena* ribozyme has aminoacyl ester hydrolysis activity, a reaction that requires attack of water at a carbon center (Fig. 2d), the amount of catalysis is modest (Piccirilli et al. 1992).

The explosion of types of ribozyme-catalyzed reactions came with the advent of in vitro selection/evolution technology, developed in the laboratories of L. Gold, G. Joyce, and J. Szostak in 1990 (see Chapters 2, 5 and 6). In brief, a large combinatorial library of diverse RNA sequences is challenged to perform some “task,” like catalyzing a specific reaction, and the rare molecules that succeed are isolated and amplified to give a new population, enriched in competent molecules. The beauty of the method is that huge populations of sequences, such as $10^{15}$, can be sampled. In vitro evolution has led to the discovery of RNA and DNA molecules that utilize nucleoside triphosphate substrates (Fig. 2c) (Lorsch and Szostak 1994; Ekland and Bartel 1996); make and break amide bonds (Dai et al. 1995; Lohse and Szostak 1996; Wiegand et al. 1997), including an amide bond between two amino acids (Zhang and Cech 1997); alkylate a nucleoside or a thiophosphate (Wilson and Szostak 1995; Wecker et al. 1996), and add an amino acid to a nucleotide via an ester linkage (Illangasekare et al. 1995). In all of these cases, the substrates still have an essential nucleic acid component, but the reactions involve carbon centers as well as phosphorus centers. Thus, the spectrum of reactions that can be catalyzed by RNA is far greater than could be inferred from looking at natural ribozymes.

Two recent examples push the envelope even further. In vitro selection has led to the discovery of modified RNA that catalyzes a classic organic chemistry reaction, Diels-Alder cycloaddition, in the presence of Cu^{2+} ions (Fig. 2e) (Tarasow et al. 1997). In addition, both RNA and DNA molecules have been selected to catalyze the insertion of metal ions (Cu^{2+}, Zn^{2+}) into porphyrin rings (Fig. 2f), similar to a step in the biosynthesis of heme (Conn et al. 1996; Li and Sen 1996). The former reaction involves carbon–carbon bond formation, whereas the latter requires a conformational change in the substrate but no covalent bond formation. The other breakthrough in these studies was that the substrates were not directly attached to a nucleic acid component, so interactions other than substrate–ribozyme base-pairing are responsible for positioning the substrate within the ribozyme active site.
How Fast is RNA Catalysis?
Catalytic rate can be assessed in numerous ways, only two of which are discussed here. First, the turnover number of the enzyme ($k_{\text{cat}}$) is informative, because it gives the number of substrate molecules converted to product per minute by a single catalyst at saturating substrate concentration. Second, the rate constant for the chemical step of the catalyzed reaction can be compared to the rate constant for the same reaction uncatalyzed. In more technical terms, the first-order rate constant for the reaction of one enzyme-bound substrate (this is simply $k_{\text{cat}}$ in those cases where the chemical step is rate-limiting overall) is divided by the first-order rate constant for the uncatalyzed reaction at the same temperature. Both of these parameters are listed in Table 2 for a few select ribozymes and for some protein enzymes that catalyze related reactions.

One main conclusion from Table 2 is ribozymes that evolved in nature to catalyze reactions with nucleic acid substrates (e.g., group I introns) increase the intrinsic rate of a reaction by huge factors, within the range achieved by protein enzymes. Thus, RNA is not intrinsically a poor catalyst compared to protein. On the other hand, ribozymes selected in the laboratory to catalyze reactions with non-nucleic acid substrates have been measured to have more modest rate enhancements, no more than ~1000-fold. It remains to be seen if this represents a fundamental inferi-

<table>
<thead>
<tr>
<th>Ribozyme</th>
<th>Rate enhancement $k_{\text{cat}}/k_{\text{uncat}}$</th>
<th>Turnover number $k_{\text{cat}}$ (min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I, <em>Tetrahymena</em></td>
<td>10$^{11}$</td>
<td>0.1</td>
</tr>
<tr>
<td>Group I, <em>Anabaena</em></td>
<td>10$^{10}$</td>
<td>4</td>
</tr>
<tr>
<td>hammerhead</td>
<td>10$^{6}$</td>
<td>$\geq$1</td>
</tr>
<tr>
<td>RNA ligase</td>
<td>10$^{9}$</td>
<td>100</td>
</tr>
<tr>
<td>porphyrin metalation</td>
<td>460</td>
<td>0.9</td>
</tr>
<tr>
<td>Diels-Adler</td>
<td>800</td>
<td>—</td>
</tr>
<tr>
<td>T7 RNA polymerase</td>
<td>$3 \times 10^{11}$</td>
<td>14,000</td>
</tr>
<tr>
<td>DNA ligase, <em>E. coli</em></td>
<td>10$^{9}$</td>
<td>28</td>
</tr>
</tbody>
</table>


$^a_k_{\text{cat}}$ refers to the chemical step of the reaction (where known), and therefore may not be the same as $k_{\text{cat}}$ (turnover number). For the Diels-Adler reaction, rate enhancement is instead based on ratio of second-order rate constants.
Mechanisms of RNA Catalysis

ority of ribozymes for such reactions, or a limitation of the in vitro selec-
tion experiments.

Considering now the turnover numbers, even the ribozymes with
impressive catalytic rate enhancements are slow at processing substrates
under multiple turnover conditions. How is this possible? A reaction can
only proceed as fast as its slowest step. Ribozymes often bind nucleic acid
substrates by Watson-Crick base-pairing (secondary structure) plus addi-
tional “tertiary” interactions, which add up to give very tight binding. The
same interactions stabilize binding of the nucleic acid reaction product,
slowing its dissociation from the active site. Thus, rate-limiting product
dissociation makes some ribozymes such as the *Tetrahymena* ribozyme
very sluggish multiple-turnover catalysts (Herschlag and Cech 1990b;
Young et al. 1991). However, this should not be considered a deficiency,
because these catalytic introns have evolved as single turnover catalysts.
If these RNAs had a rapid turnover rate, they would be able to react with
other cellular RNAs once excised from their host RNAs.

**How Specific Is an RNA Catalyst?**

Protein enzymes are notable for their exquisite specificity, their ability to
distinguish between very closely related substrates. How do ribozymes
compare? Group I introns use guanosine (G) as a nucleophile to cleave
the 5’ splice site (Fig. 3), and other nucleosides have much reduced or
undetectable activity (Bass and Cech 1984). When Michel et al. (1989b)
located the G-binding site, they also identified specific hydrogen-bonding
interactions that explain the specificity (Fig. 4a). For the *Tetrahymena*
ribozyme, the specificity for G over 2-aminopurine ribonucleoside is
about 1000-fold (Legault et al. 1992), whereas single base-pair changes in
the G-binding site reverse this specificity. An even higher specificity for
G over deoxyG (Bass and Cech 1984) has been attributed to the favorable
interaction of a metal ion with the 2’-OH (Sjögren et al. 1997).

Another remarkable illustration of RNA’s potential to discriminate
between closely related molecules is an in-vitro-selected aptamer that
binds theophylline >10,000 times more tightly than caffeine, a molecule
that is identical except for one methyl group (Jenison et al. 1994). In this
case, NMR structural analysis supports the model that substitution of the
N7 proton with an N7 methyl group in caffeine would remove two H
bonds that the aptamer uses to bind theophylline and also disrupt stacking
interactions above and below the plane shown in Figure 4b (Zimmermann
et al. 1997).
On the other hand, ribozymes that rely on base-pairing to bind nucleic acid substrates often have their specificity limited by the fact that mismatches within nucleic acid helices are not very destabilizing (see Chapter 10). Even worse, the slow rates of helix dissociation can give enough time for cleavage to occur before a more weakly bound mismatched substrate can dissociate (Herschlag and Cech 1990c; Hertel et al. 1996). For example, the *Tetrahymena* ribozyme shows a 200-fold specificity for cleavage of the “matched” substrate GGCCCUCUAAAAA over the “mismatched” substrate GGCCCGCUAAAAA at low concentration of G nucleophile where the reaction is slow, but the specificity decreases to 4.5-fold at high G concentration (Herschlag and Cech 1990c). Mutant ribozymes that bind substrate less tightly do show increased sequence specificity (Young et al.)

---

**Figure 3** Group I intron structure and reactivity. (a) Expanded view of secondary structure of a minimal group I intron. P = paired region. Thick lines represent the RNA backbone. Thin lines show connectivity in this expanded view but in fact have zero length. Shaded regions = catalytic core. In the first step of self-splicing, exogenous guanosine or GTP binds in the G-site and cleaves the 5’ splice site (large arrowhead). The 3’-terminal (ω) guanosine of the intron (circled) then occupies the G-site for the second step of splicing. (b) Self-splicing. First step, 5’ splice-site cleavage by exogenous guanosine. Second step, 3’ splice-site cleavage by the 5’ exon at the bond following ωG (circled), resulting in exon ligation.
The hammerhead ribozyme also has some ability to distinguish mismatches in a ribozyme-substrate helix away from the catalytic core, but it has dramatic discrimination against mismatches that flank the core (Werner and Uhlenbeck 1995; Hertel et al. 1997). Presumably, these proximal mismatches destabilize the catalytically active structure. Thus, there are cases in which a ribozyme recognizes a nucleic acid substrate with more specificity than that intrinsic to base-pairing interactions. Such mismatches in a ribozyme-substrate helix away from the catalytic core, but it has dramatic discrimination against mismatches that flank the core (Werner and Uhlenbeck 1995; Hertel et al. 1997). Presumably, these proximal mismatches destabilize the catalytically active structure. Thus, there are cases in which a ribozyme recognizes a nucleic acid substrate with more specificity than that intrinsic to base-pairing interactions. Such
a feature seems critical to obtaining sufficient fidelity for RNA-catalyzed RNA replication in an RNA World.

Ribozymes discriminate easily between stereoisomers of the same substrate, as expected for any catalyst that provides a chiral three-dimensional active site. A dramatic example is provided by the *Tetrahymena* ribozyme, which cleaves the R_P and S_P diastereomers of an RNA substrate with a >1000-fold different rate, even though these differ only in the position of a subtle oxygen-to-sulfur substitution (Cech et al. 1992). As another example, an RNA selected for binding to D-tryptophan agarose bound >700 times more weakly to L-tryptophan agarose (Famulok and Szostak 1992). An exciting application of this stereospecificity is “mirror image” drug development: One can select a nucleic acid aptamer that binds to the mirror image (enantiomer) of a target protein, and then be assured that the enantiomer of the selected aptamer will be specific for the naturally occurring isomer of the target protein (Klussmann et al. 1996; Williams et al. 1997). The enantiomer of the aptamer, composed of unnatural L-ribose sugars, is not recognized by natural nucleases and therefore has the potential to be a long-lived drug in vivo.

**MECHANISMS OF RNA CATALYSIS**

How do ribozymes lower the activation energy and thereby speed up biochemical reactions? The much-studied group I ribozymes will be used to illustrate the catalytic strategies available to RNA, with other systems being compared and contrasted along the way (see also Narlikar and Herschlag 1997).

**Binding and Orienting Reactive Groups**

When enzymes bind substrates, they create a high local concentration and optimal orientation to increase the efficiency of reaction (Jencks 1969). Group I introns use this same strategy, binding guanosine in proximity to the 5’ splice site and converting what would normally be an intermolecular reaction into an effectively intramolecular reaction (Bass and Cech 1984). Guanosine binding is entropically driven, perhaps indicative of release of solvent from the RNA upon G-binding (McConnell and Cech 1995). Elements of the ribozyme that contribute to G-binding (Fig. 4a) (Michel et al. 1989b; Yarus et al. 1991) and to binding of the 5’-splice site have been identified. (The latter set of interactions is described below in Building a Catalytic Center Using Only RNA, because it provides a good general model for specific helix-packing interactions in large RNAs.)
Although we know that these two reactants are bound in a manner that allows them to react, there is no information about whether they are held in anything close to an optimal orientation.

Group II introns, found in certain mitochondria, chloroplasts, and bacteria, have conserved structure distinct from that of group I (Michel et al. 1989a). Group II introns also accomplish self-splicing by consecutive transesterification reactions, but with a twist: In the first step, the 2'-OH of a specific intronic adenosine (A) cleaves the 5' splice site to form a "lariat" intermediate, and in step two, the cleaved 5' exon attacks the 3'-splice site, ligating the surrounding exons (Peebles et al. 1986; Schmelzer and Schweyen 1986; van der Veen et al. 1986). In the laboratory, group II introns can catalyze trans reactions, such as binding a free 5'-exon analog and ligating it to the 3'-exon (Jacquier and Rosbash 1986). Both the cis and trans reaction rely on two exon-binding sites that base-pair with complementary 5'-exon sequences (intron binding sites or IBS, Fig. 5; Jacquier and Michel 1987). These long-range interactions are similar to the internal guide sequence-5'-exon pairing of group I introns (Davies et al. 1982; Michel et al. 1982). In addition, a portion of the conserved GUGYG sequence at the 5' end of group II introns pairs with a loop within domain I of the intron, further contributing to 5'-splice site recognition (Jacquier and Michel 1990). Presumably, additional, as-yet-unidentified tertiary interactions position the 2'-OH of the nucleophilic A in domain VI for attack.

Step 2 of group II intron self-splicing requires an alternate structural interaction (Chanfreau and Jacquier 1996; η/η' in Fig. 5). This is especially noteworthy, because although there is much evidence for conformational changes in ribozyme catalysis (see, e.g., Been and Perrotta 1991; Wang et al. 1993; Golden and Cech 1996), there are few examples where a specific structure has been identified that orchestrates a conformational switch. In the group II intron, the branch-site A for step 1 of splicing must be displaced from the active site to allow entry of the last intron nucleotide and the 3'-splice site (Steitz and Steitz 1993; Jacquier 1996). Branch formation or 5'-splice site cleavage may result in a displacement that allows the η/η' base-pairing to form, and its formation may in turn pull the branched A from the active site, allowing the last intron nucleotide to enter for step 2 of splicing.

**Metal Ion Catalysis**

Activity of most ribozymes requires, or is greatly stimulated by, divalent metal ions such as Mg++, Mn++, and in some cases Ca++. These could be
involved in folding the polyanionic RNA catalyst, in active-site chemistry, or in both. Plausible chemical roles have been much discussed previously (Pyle 1993; Yarus 1993) and include

1. electrostatic stabilization of an anionic attacking or leaving group
2. electrostatic stabilization of an electrophilic group (Lewis acid catalysis)
3. electrostatic destabilization of a substrate in the ground state
4. specific orientation of substrates by metal ion coordination
5. acid-base catalysis by metal-bound water or hydroxide
6. oxidation-reduction chemistry

In favorable cases, the structural and catalytic roles of metal ions have been separated. The *Tetrahymena* ribozyme requires divalent metal ions in order to attain its active three-dimensional folded structure (Latham and Cech 1989; Celander and Cech 1991), a requirement that can be met by a
variety of divalent cations and partially met by high concentrations of monovalent cation (Downs and Cech 1996). In addition, correct positioning of the 5’ splice site helix within its active site has a requirement for Mg++ that cannot be met by Ca++ (Wang and Cech 1994; McConnell et al. 1997).

Because numerous metal ions are required to fold the ribozyme, it was difficult to identify additional metal ions involved directly in active-site chemistry, especially since they are already bound at the 2 mM Mg++ concentration required for ribozyme folding (McConnell et al. 1997). This problem was addressed by site-specific substitution of atoms that would predictably perturb the binding and function of a coordinated metal ion. In the *Tetrahymena* ribozyme, substitution of bridging oxygen atoms by sulfur at the reactive bonds led to loss of activity with Mg++. Restoration of cleavage with thiophilic metal ions such as Mn++ and Cd++ in both the forward and reverse reactions led to the proposal of metal ions 1 and 2 in Figure 6. Metal ion 1 destabilizes the substrate in the ground state but stabilizes the developing negative charge as the O–P bond is broken in the transition state, thereby providing roughly 10⁶-fold catalysis (Piccirilli et al. 1993; Narlikar et al. 1995). Metal ion 2 acts to deprotonate the 3’-OH of the G nucleophile (Weinstein et al. 1997). These two metal ions may form the sort of “two metal ion center” proposed by Steitz and Steitz (1993), a catalytic strategy with clear connections to the world of protein enzymes (Beese and Steitz 1991; Kim and Wyckoff 1991). Substitution of the 2’-OH of the G nucleophile by an amino group also led to a metal specificity switch, leading to the proposal of metal ion 3 (Fig. 6) (Sjögren et al. 1997). Metal ion 3 could function to position or activate the nucleophile.

Divalent metal ions are required for the hammerhead ribozyme reaction under physiological conditions. There is evidence for their binding to the reaction-site phosphate, enhancing its ability to be attacked by the oxyanionic 2’ nucleophile, and also evidence for metal ion activation of the 2’-OH nucleophile (see Chapter 11). However, the hairpin ribozyme, which cleaves RNA by the same 2’, 3’-cyclic phosphate mechanism as the hammerhead, appears to use metal ions just for folding, not for chemistry. Cobalt hexamine, a metal complex inert to changes in its coordination sphere, supports full hairpin ribozyme cleavage, providing evidence against the importance of direct coordination of the metal to the RNA. Also unlike the hammerhead, phosphorothioate substitution at either non-bridging oxygen has small and metal-ion-independent effects on activity, evidence against any direct metal coordination (Hampel and Cowan 1997; Nesbitt et al. 1997; Young et al. 1997). This system serves as an impor-
tant reminder that RNA enzymes, like protein enzymes, will embrace diverse catalytic strategies.

**Covalent Catalysis**

Many protein enzymes use a nucleophilic amino acid (e.g., serine, lysine, cysteine, or histidine) to attack a substrate, forming a covalent intermediate with a portion of the substrate. In a second step, attack of the covalent intermediate by an external nucleophile (such as water) releases the product and restores the enzyme (e.g., serine proteases, or alkaline phosphatase in Fig. 7).

The *Tetrahymena* ribozyme performs reactions that are highly analogous to those carried out by *Escherichia coli* alkaline phosphatase (Fig. 7).
Because the biological function of this RNA involves self-splicing reactions at phosphodiester bonds, it was not obvious that it should have reactivity with phosphomonoester substrates, since the two have very different transition states (Zaug and Cech 1986b). Nevertheless, the ribozyme efficiently transfers a 3′-terminal phosphate from an oligonucleotide to its own 3′-terminal guanosine, forming a covalent E-P intermediate. The phosphate can then be transferred to a different oligonucleotide (one that can also bind in the active site) or it can be hydrolyzed, the latter reaction being enhanced at low pH. In either case, the free catalyst E is restored, ready for another catalytic cycle.

Why is covalent catalysis advantageous? It allows an active site to be occupied successively by two different substrates; in the example given above, the Tetrahymena ribozyme first binds the oligo CCCUCUp, stores the phosphate as an E-P complex, binds a second oligo UCU, and trans-
fers the phosphate to it to form UCUp (Zaug and Cech 1986b). More generally, the immobilization gained by forming a covalent intermediate may be advantageous for catalysis.

Contemporary protein enzymes use covalent catalysis to act as proteases, esterases, phosphoglucomutases, transphosphorylases, and ligases, often with a serine OH group as a nucleophile. In an RNA World, it would be feasible for ribozymes to perform similar reactions using a ribose 2'- or 3'-OH group as a nucleophile.

**General Acid–Base Catalysis**

Many chemical reactions require transfer of a proton to or from a reactant. The proximity of a proton donating or accepting group, that is, a general acid or base, can greatly speed such reactions. RNA does not have histidine, but functional groups on the RNA bases or backbone could perhaps serve as proton donors or acceptors (Pace and Marsh 1985). However, with the exception of the 5'-terminal phosphate and modified nucleotides such as 7-methylguanosine, the pKₐ values of these groups are normally far from pH 7. Could an RNA structure perturb a pKₐ in an active site to provide efficient general acid–base catalysis? In the absence of a confirmed example, it seems possible that ribozymes will in general use metal ions and metal-ion-bound solvent in lieu of general acid–base catalysis (Cech et al. 1992).

A DNAzyme that catalyzes the insertion of metal ions into mesoporphyrin IX works by enhancing the basicity of the bound mesoporphyrin substrate by 3–4 pH units, thereby enhancing its ease of metallation (Li and Sen 1998). This may occur by distortion of the planar mesoporphyrin structure by the DNAzyme, although involvement of a catalytic group such as a negatively charged phosphate is another possibility.

**Use of Binding Energy away from the Site of Reaction**

To decrease the energy barrier for a reaction, an enzyme must stabilize the reaction’s transition state more than it stabilizes the bound substrate(s) in the ground state. One strategy for doing so is to use binding energy away from the site of reaction to “force” an interaction at the reaction site, an interaction that induces the substrate to react (Jencks 1975). The enzyme can force entropic fixation of a substrate with respect to another substrate or with respect to active-site groups, or force electrostatic destabilization by juxtaposing a (partially) charged substrate atom with a like charge on the enzyme. In either case, if the strained interaction is relieved upon
approach to the transition state, catalysis will result. When enzymes use
this trick, the observed binding of the substrate(s) is weaker than it would
otherwise be, because part of the intrinsic binding energy is used to pay
the price of forcing the unfavorable ground-state interaction.

Recent studies with the *Tetrahymena* ribozyme have shown that this
sophisticated strategy is used by RNA as well as protein enzymes. The
*Tetrahymena* ribozyme binds its oligonucleotide substrate (CCCUCUpA)
and product (CCCUCU-OH) by base-pairing with the CCCUCU portion
plus additional “tertiary” interactions. The tertiary interactions are much
weaker with the substrate than with the reaction product, a difference
due to the phosphate rather than the 3'-A residue (Narlikar et al. 1995).
The authors suggest that the ribozyme uses binding interactions to force
the bridging oxygen atom, partially positive in the ground state, next to
the positively charged magnesium ion 1 (Fig. 6); this substrate destabi-
lization is relieved in the transition state as negative charge accumulates
on the oxygen atom. They estimate 280-fold rate enhancement from the
ground state destabilization and another 60-fold from the additional posi-
tive interactions in the transition state, for a combined catalytic effect that
accounts for $10^4$-fold of the total $10^{11}$-fold catalysis achieved by this
ribozyme (Narlikar et al. 1995).

**BUILDING A CATALYTIC CENTER USING ONLY RNA**

**Solvent-inaccessible Core**

Protein enzymes form a densely packed, hydrophobic core to support a
concave active site. Given the absence of hydrophobic side chains and the
polyanionic backbone of nucleic acids, how is it possible for them to form
a functionally equivalent structure? Single-stranded nucleic acids certainly
fold into more compact structures than their double-helical counterparts.
Transfer RNA and guanine quadruplexes provide well-studied examples,
but even the tRNA structure does not provide much solvent inaccessibili-
ty to its backbone; most of the backbone is still on the outside of the
molecule, bathed in solvent, very different from a folded protein.

Large ribozymes, such as the *Tetrahymena* group I intron, do form a
relatively solvent-inaccessible core as judged by protection from free-
radical bombardment (Latham and Cech 1989; Celander and Cech 1991;
Sclavi et al. 1997). This suggested that nucleic acid building blocks can
be packed together with the help of divalent cations to form something
roughly equivalent to the core of a globular protein. In the following sec-
tion, we explore the types of interactions that contribute to formation of
such a structure.
Interactions that Stabilize a Close-packed RNA Core

The crystal structure of one domain of the *Tetrahymena* ribozyme at 2.6 Å resolution provided the first atomic view of an RNA large enough to have a relatively solvent-inaccessible core (Cate et al. 1996). This 160-nucleotide domain, called P4-P6, comprises about half of the active site of the ribozyme. When synthesized as a separate RNA molecule, it assumes its native secondary structure as well as a higher order tertiary structure that appears to be a simple subset of its structure within the context of the whole, active ribozyme (Murphy and Cech 1993, 1994). Examination of the structure (Fig. 8a) reveals both the expected nucleic acid features and

![Figure 8](image)

*Figure 8* An RNA domain. (a) The P4-P6 domain of the *Tetrahymena* group I intron rendered as CPK atoms, showing the side-by-side arrangement of two helical subdomains. The two major sites of interaction between the two subdomains are highlighted: the junction of P4 and P6 (blue), the A-rich bulge (red), the GAAA tetraloop (green), and the tetraloop receptor (violet). (b) Coordination of two magnesium ions (violet) by the backbone of the A-rich bulge displays the nucleotide bases for tertiary interactions with the rest of the domain. (c) The GAAA tetraloop-receptor is stabilized both by stacking (highlighted in red) and specific hydrogen bonds between the two modules. For orientation purposes, the G of the GAAA tetraloop is drawn in blue.
also features that appear protein-like, at least superficially. The molecule has a sharp bend at the top, which allows two punctuated double-helical regions to be aligned side by side. The protein-like feature is the tightly packed core produced by sandwiching together the two halves of the molecule. Thus, the P4-P6 structure provides the first opportunity to evaluate the molecular basis for formation of such a higher order RNA structure, as discussed below.

Long-range Base Pairs and Triples

The two halves of the P4-P6 RNA domain are brought together mainly through two tertiary interactions: The A-rich bulge docks into the minor groove of P4 (Flor et al. 1989) and also into the P5abc three-helix junction (Fig. 8b), whereas the L5b tetraloop is bound by a receptor sequence (Costa and Michel 1995) located between P6a and P6b in the other half of the molecule (Fig. 8c). Each of these interactions involves base triples between a base pair within a duplex region and a third base that is far away in the secondary structure of the molecule. However, thinking of these as simply base triples oversimplifies the interactions, as they involve multiple H–bonds between 2'-hydroxyl groups and phosphate oxygens, and also extensive base-stacking interactions (Cate et al. 1996).

A similar lesson about base triples can be gleaned from an intermolecular tetraloop/minor groove interaction seen in the crystal structure of the hammerhead ribozyme (Pley et al. 1994). Again base triples provide the specificity of the interaction (often because substitution of a different base would cause a steric clash), but a larger number of long-range H–bonds involve ribose hydroxyls and phosphate oxygen atoms.

Comparative sequence analysis and site-specific mutagenesis studies have identified base triples and long-range base pairs in many other ribozymes whose detailed structures have not yet been solved. Examples in group I ribozymes include the triple-helical scaffold involving P4 and P6 (Michel and Westhof 1990; Michel et al. 1990), base-pairing between peripheral loops that stabilize the sunY and Tetrahymena introns (Michel et al. 1992; Jaeger et al. 1993; Lehnert et al. 1996), and a long-range triple between P4 and J8/7 that brings together the domains of the Tetrahymena and cyanobacterial introns (Tanner and Cech 1997). Examples in group II ribozymes so far include long-range base pairs between two hairpin loops or between a hairpin loop and an internal loop (Michel et al. 1989a) and a GAAA tetraloop-receptor interaction that joins domains I and V (Costa and Michel 1995), as summarized in Figure 5. RNase P RNAs are compacted by long-range pseudoknots, one of which involves a hairpin loop (James et al. 1988; Haas et al. 1991), and other interactions (see
Chapter 14). As in the interactions seen in crystal structures, it seems likely that each of these biochemically verified tertiary interactions will be buttressed by a network of additional H–bonds involving the RNA backbone.

The 2′-Hydroxyl Group

The 2′-OH of the ribose sugar, the group that distinguishes RNA from DNA, is exploited frequently for structural interactions in large RNAs. This should be expected, because once RNA engages in base-paired secondary structure, its backbone is its most available element, and the 2′-OH groups lining the minor groove are particularly available for higher-order interactions. As detailed above, H-bonds involving 2′-OH groups buttress base triple interactions. In addition, helical regions can be held together by ribose zippers—the ribose and attached base of one nucleotide donate and accept H-bonds from the same 2′-OH of a second nucleotide, and this interaction is repeated at the adjacent level in a “ladder” of interactions.

Ribose 2′-hydroxyls also have a key role in 5′-splice-site recognition in group I introns. This system is of general interest because it provides a model for specific recognition of base-paired helices within folded RNA molecules. The last few nucleotides of the 5′-exon are first recognized by base-pairing to an intronic internal guide sequence (IGS; Davies et al. 1982; Michel et al. 1982). The resulting duplex is called P1 (Fig. 3a). In a kinetically separable step, this P1 helix is then positioned within the ribozyme active site (Bevilacqua et al. 1992; Herschlag 1992). As shown in the model in Figure 9, certain 2′-OH groups on both strands of P1 are recognized by the intron core (see, e.g., Pyle et al. 1992; Strobel and Cech 1993; Strobel et al. 1998). Additionally, the exocyclic amino group of the G·U wobble base pair at the cleavage site is also important for splice-site recognition and specificity (see, e.g., Strobel and Cech 1995). This G·U wobble pair is conserved among most of the 500 known group I introns, and the two that have instead a G-C become more reactive when a G·U is substituted at this position (Hur and Waring 1995). The docking of the P1 helix within the group I active site illustrates how backbone and minor groove interactions can be used to recognize an RNA element that has its base sequence information “hidden” in a helix.

In group II self-splicing introns (Fig. 5), domain V is a phylogenetically conserved element that is essential for catalysis (Jarrell et al. 1988). As in the P1–group I intron case, specific 2′-OH groups on one side of the helix mediate binding of domain V to the remainder of the intron (Abramovitz et al. 1996). On the opposite (major groove) face of the domain V helix, specific 2′-OH groups, phosphates, and bases (including the G of a G-U
pair) participate in transition-state stabilization and therefore appear to form part of the active site for 5′ splice-site cleavage (Chanfreau and Jacquier 1994; Peebles et al. 1995; Abramovitz et al. 1996).

**Metal Ions**

The role of divalent metal ions in stabilizing sites of close backbone–backbone interaction has been studied in transfer RNA, both thermodynamically (Cole et al. 1972) and structurally (see Chapter 12). However, tRNA only provided details for the shorter-range interactions, not those which mediate helix packing on globular domains. The crystal structure of the P4-P6 ribozyme domain revealed metal-ion-mediated long-range interactions. The close packing of helices brings their phosphates into proximity; the resulting electrostatic repulsion is ameliorated

---

**Figure 9** Substrate recognition by a group I ribozyme. An oligonucleotide substrate (S) is first recognized by base-pairing to a short complementary sequence called the internal guide sequence (IGS), forming the helix P1. The cleavage site is defined by a G·U wobble pair at the top of the helix. The P1 helix then docks into the active site of the intron (including elements P3, P4, P5, P7, P8, J4/5, and J8/7). It is correctly positioned by interactions with 2′-OH groups (circled) and the exocyclic amine group of the G·U wobble at the cleavage site. (Figure template courtesy of Scott Strobel.)
by binding of hydrated magnesium ions. These outer-sphere magnesium-ion complexes bridge P5 and P5a as well as P5b and P6a (Cate et al. 1996). The crystal structure brought atomic detail to a picture that had been inferred from studies of ribozyme folding as a function of Mg\(^{++}\) concentration (Celander and Cech 1991; Banerjee et al. 1993; Jaeger et al. 1993; Murphy and Cech 1994; Szewczak and Cech 1997).

A striking role of divalent cations is seen in the A-rich bulge of P4-P6. The backbone wraps around two divalents, each of which makes an inner-sphere coordination complex with three phosphate oxygens (Cate et al. 1996). As a result, the backbone is on the inside of this local structure and the bases protrude to make tertiary interactions with distant portions of the molecule (Fig. 8b). This structure provides an answer to the paradox of Sigler (1975), who pointed out that proteins form sensible secondary structure, in that the side chains protrude to allow higher-order interactions, whereas nucleic acids form helices that tuck the side chains (bases) inside, not optimal for higher-order interactions. In this A-rich bulge, the metal ions turn a portion of the RNA “inside out,” constraining the backbone to be inside and displaying the adenosine “side chains” for tertiary interactions. Phosphorothioate substitution at the metal-coordinating sites interferes both with local folding and with folding of the P4-P6 domain, confirming the essentiality of this metal binding (Cate et al. 1997).

In going from the isolated P4-P6 domain to the whole ribozyme, much more surface area is buried. Therefore, metal ions are likely to be even more important. Consistent with this idea, a long string of functionally significant metal-binding sites in the ribozyme core has been mapped by Christian and Yarus (1993) using phosphorothioate interference and Mn\(^{++}\) rescue.

In summary, both inner-sphere and outer-sphere coordination of RNA functional groups by divalent cations contribute to the specificity and stability of RNA tertiary structure.

**Small Ribozymes Are Different**

At this point we must introduce a disclaimer. We have been developing a picture of a globular RNA with a concave active site that maintains a similar form with or without substrates bound. That this is a reasonable model for the *Tetrahymena* ribozyme is supported by the small amount of energetic coupling between the binding of the guanosine substrate and the P1 helix, indicative of a site that is only slightly reorganized upon binding either single substrate (Bevilacqua et al. 1993; McConnell et al. 1993). Similarly, chemical modification or cleavage studies have seen only
minor differences between the free ribozyme and substrate-bound versions. However, small ribozymes such as the hammerhead are much different. Here the binding of substrate is necessary for formation of the catalytic center (Hertel et al. 1997), and even after the substrate–ribozyme complex is formed, the catalytically active conformation may be formed only occasionally (see Chapter 11). Thus, the small ribozymes that were useful in the early stages of the RNA World were probably relatively floppy, and only after RNA self-replication became efficient enough to copy a larger RNA (>100 nucleotides) could ribozymes with permanent active sites come on the scene.

Ribozyme Dynamics

An atomic-resolution crystal structure or NMR structure of a macromolecular catalyst is the single most valuable set of structural information one can obtain. It provides a framework for planning and interpreting experiments that probe the relationships between structure and catalytic function. Yet it gives largely a static picture, and even if a completely rigid macromolecule could bind its substrates, it could never promote their reaction. Formation of a transition state requires movement. One type of movement—switches between distinct conformers—has been described above. We now consider another type of dynamics—thermal motions. These can be thought of as the excursions that a molecule undergoes from its most stable ground-state structure.

The magnitudes of thermal motions between two domains (the P1 substrate domain and the P3-P9 domain) of the *Tetrahymena* ribozyme have been investigated by disulfide cross-linking (Cohen and Cech 1997). Reactive groups were appended to the RNA backbone by an adaptation of the 2′-modification method pioneered by Sigurdsson et al. (1995). Upon collision, these groups can form a disulfide cross-link. Sites separated by 50 Å in the structure were found to be cross-linked at rates only 3 to 15 times slower than the rates for proximal sites, indicative of an unexpectedly high degree of flexibility. The motions were estimated to occur on the microsecond time scale at 30°C (Cohen and Cech 1997).

Presumably, there will be fewer degrees of freedom and therefore less flexibility within a single RNA domain than between two domains. Nevertheless, these measurements suggest that large, globular RNAs are much more dynamic than proteins. It remains to be tested whether one of the functions of RNA-binding proteins is to damp out these large-scale motions, thereby conferring a selective advantage to the RNP World over the RNA World (Fig. 1).
REFERENCES


Mechanisms of RNA Catalysis


