Chemistry & Biology, Vol. 10, 5–14, January, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S1074-5521(03)00003-6

In Search of an RNA Replicase Ribozyme

Review

Kathleen E. McGinness[1] and Gerald F. Joyce*
Departments of Chemistry and Molecular Biology
and The Skaggs Institute for Chemical Biology
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, California 92037

The theory that an RNA world played a pivotal role in life's evolutionary past has prompted investigations into the scope of RNA catalysis. These efforts have attempted to demonstrate the plausibility of an RNA-based genetic system, which would require RNA molecules that catalyze their own replication. The mechanistic features of modern protein polymerases have been used to guide the laboratory evolution of catalytic RNAs (ribozymes) that exhibit polymerase-like activity. Ribozymes have been developed that recognize a primer-template complex in a general way and catalyze the template-directed polymerization of mononucleotides. These experiments demonstrate that RNA replicase behavior is likely within the catalytic repertoire of RNA, although many obstacles remain to be overcome in order to demonstrate that RNA can catalyze its own replication in a manner that could have sustained a genetic system on the early Earth.

Introduction

The existence of an RNA world is a plausible step in the early evolutionary history of life on Earth. The RNA world hypothesis proposes that there was a time when RNA acted as both genotype and phenotype, roles that largely are fulfilled by DNA and proteins, respectively, in modern biological systems [1]. Although there is no direct evidence for the existence of an RNA world, it has been suggested that "molecular fossils" of RNA-based life are present in modern biology. For example, catalytic RNA molecules (ribozymes) operate in viral genome processing [2], mRNA processing [3], tRNA maturation [4, 5], and protein synthesis [6, 7]. Additionally, nucleotide-derived small molecules function as cofactors in a variety of protein-catalyzed reactions [8].

If there was an RNA world, then there must have been a pathway by which RNA-based life evolved into the DNA and protein-based system of modern biology. RNA continues to play a critical role in biological systems. RNA molecules, in the form of mRNA and tRNA, act as intermediaries between DNA-based genetic information and protein-based function. RNA functions as the genetic material in certain viruses, including retroviruses, that require copying of RNA-based information to DNA. Thus, mechanisms still exist to transfer RNA information into both protein and DNA, making it plausible that an RNA-based genetic system developed into one in which DNA and protein assumed most of the duties that once were performed by RNA.

For an RNA world to have existed, accurate transfer of genetic information would have been necessary. If RNA was the chief agent of catalytic function, then that process must have been carried out by one or more RNA molecules. The simplest system for the propagation of RNA-based information would involve two RNA molecules, a replicase ribozyme and its complement, with the replicase being responsible for producing copies of both itself and the complement. This activity could have evolved to allow copying of other RNA molecules whose functions were advantageous for survival. Although there is no way to be certain whether such a replicase ribozyme existed in evolutionary history, laboratory experiments can be used to determine whether RNA is capable of exhibiting the relevant catalytic behaviors. Inquiries into this issue have focused on the properties of modern protein polymerases, with the aim of developing ribozymes with polymerase-like activity. This review discusses the general features of catalysis by protein polymerases and various laboratory experiments that have shown how many of these features can be realized with RNA.

Generic Polymerase Model

In searching for RNA molecules that are nucleic acid polymerases, it is reasonable to look to modern protein polymerases for insight into the general mechanistic properties of this class of enzymes. There are striking similarities among known polymerases in terms of their structural topology, mode of substrate recognition, and proposed catalytic mechanism [9, 10]. The functional requirements of a nucleic acid polymerase include a generic mechanism for substrate recognition, catalysis of template-directed mononucleotide addition with high fidelity, and the ability to operate in a processive manner. For this discussion, the substrates of polymerization will be restricted to a primer-template complex and 5′-activated mononucleotides, typically in the form of mononucleoside 5′-triphosphates (NTPs), with RNA and DNA being treated as equivalent.

Protein polymerases recognize double-stranded nucleic acids in a largely sequence-independent manner. This is seen in the crystal structures of several polymerase enzymes, which reveal a primer-template complex bound in a positively charged channel of the protein with side chain contacts to the phosphate backbone of the nucleic acid duplex [11–19]. The DNA duplex bound in the 3′-exonuclease site of the Klenow fragment of *E. coli* polymerase I is recognized in a similar manner [20]. This mode of primer-template binding in which only general features of a nucleic acid duplex are recognized would be advantageous for any polymerase that is responsible for copying a large variety of nucleotide sequences.

General recognition of the incoming nucleotide also is important, while maintaining accurate information

*Correspondence: gjoyce@scripps.edu
[1]Current address: Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

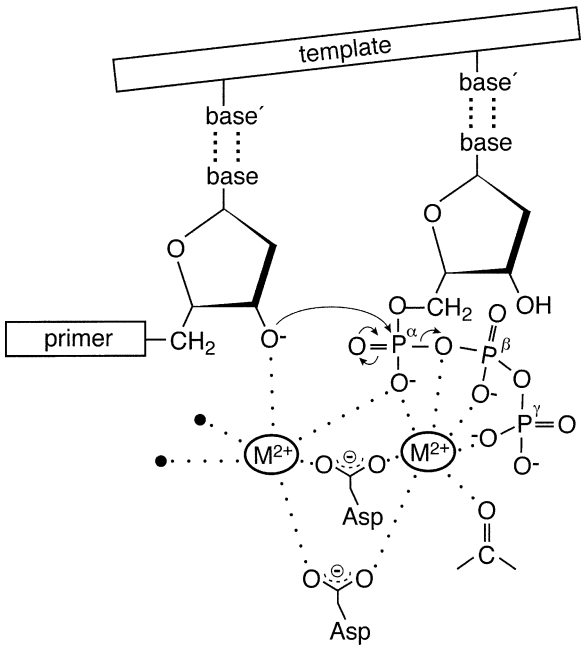

Figure 1. Proposed Chemical Mechanism of Primer Extension Catalyzed by Protein Polymerases

Addition of an NTP involves a two metal ion mechanism with no direct participation of the protein side chains. Electrostatic interactions involving divalent metal cations (M^{2+}) are indicated by dotted lines. Solid circles indicate metal-bound water molecules. (Adapted from [10].)

transfer through specific Watson-Crick pairing interactions with the templating nucleotide. The first NTP to be added onto the 3′ end of a primer has been seen in several crystal structures [12, 14–16, 18] and points to both geometric selection and induced fit as important mechanisms for binding the correct NTP [21]. The incoming nucleotide interacts largely with the primer-template complex. Hydrogen bonding between the incoming and templating bases has been observed, as well as base-stacking interactions between the incoming NTP and the nucleotide at the 3′ end of the primer. Some interactions have been observed between conserved protein residues and the minor groove side of the template nucleotide of the forming pair. The protein side chains also contact the ribose moiety and phosphate groups of the incoming NTP. Specific interactions in the nucleotide binding site are limited to those between the incoming nucleotide base and the templating base, allowing for a binding pocket that recognizes NTP substrates in a general way, while providing a framework for specific interactions that ensure accurate information transfer.

A conserved chemical mechanism has been proposed for the protein-catalyzed polymerization of NTPs (Figure 1) [10]. This mechanism involves two or three highly conserved acidic residues positioned within the active site of the enzyme. These residues interact with two divalent metal ions, which in turn contact the 3′-hydroxyl of the primer and the phosphate groups of the incoming nucleotide. One of the metals is implicated in binding and positioning the incoming nucleotide within the active site. The other is proposed to act as a Lewis acid to activate the 3′-hydroxyl group of the primer for attack on the α-phosphate of the incoming nucleotide as well as to stabilize the pentacoordinated phosphate of the transition state. According to this proposed mechanism, the protein side chains do not participate directly in catalysis; they only serve to orient the catalytic metal ions. This is of particular importance when considering a general mechanism for polymerase activity. RNA molecules are known to be capable of specific metal ion binding and precise orientation of metals for catalysis [22] and thus may be able to support polymerase-like reactions through a mechanism similar to that of protein polymerases.

Polymerases must operate with high fidelity to ensure accurate information transfer. Mechanisms that promote high fidelity can operate at the time of NTP addition or by proofreading following primer extension. For this discussion, only those mechanisms that occur at the time of NTP addition will be addressed. The error rates observed for protein polymerases that do not have a proofreading capability range from 10^{-3} to less than 10^{-6} per nucleotide [21]. This level of fidelity is achieved through several mechanisms in addition to Watson-Crick pairing between the template and incoming nucleotide.

There is a transition from B- to A-form DNA just upstream of the 3′ end of the primer that accommodates widening of the minor groove of the primer-template complex, allowing conserved protein residues to access this face of the nucleic acid duplex [11–16, 20]. Hydrogen bond acceptors in the minor groove, such as the O2 of pyrimidines and N3 of purines, are in roughly the same position for all four possible Watson-Crick base pairs, allowing polymerase proteins to interact with these groups in a generic manner to ensure high fidelity while maintaining a general mechanism for substrate recognition [21]. Interactions with these groups at the base pair upstream of the forming pair have been implicated in achieving high fidelity. If a mismatched pair is present at this upstream position, the polymerase is much less likely to continue primer extension [19]. Additionally, bending of the primer-template complex has been proposed to contribute to enhanced fidelity, allowing the polymerase to access a larger surface area when recognizing the forming base pair [11–16, 20]. A conformational change in the protein occurs upon binding the correct NTP within the enzyme active site, creating a binding pocket that allows for optimal geometric fit of Watson-Crick base pairs [12, 14, 16, 18]. RNA molecules, despite their reduced functional group diversity compared to proteins, are able to adopt complex structures involving a variety of hydrogen bonding modes, not limited to standard Watson-Crick pairing. Thus, RNA may be able to exhibit some of the same fidelity-enhancing mechanisms employed by protein polymerases at the time of NTP addition.

Processivity refers to the ability of a polymerase to catalyze hundreds to thousands of successive NTP additions without dissociating from the template. It is important for efficient polymerase activity, especially when there is a requirement to copy a long template. More generally, the ability of the polymerase active site to

translocate even one nucleotide along the template following the addition of an NTP can increase the rate at which a template is copied. Polymerases employ a variety of strategies to ensure that their catalytic domain remains associated with the primer-template complex, thereby enhancing processivity. The β subunit of *E. coli* DNA polymerase III holoenzyme acts as a sliding clamp that encircles the template-primer complex and serves to hold the polymerase domain close to the complex [23]. A similar mechanism is used by eukaryotic DNA polymerases δ and ϵ, which employ the proliferating cell nuclear antigen protein to achieve processivity [24]. Bacteriophage T7 DNA polymerase is able to "hijack" thioredoxin from the host cell to increase polymerase affinity for the primer-template complex [14], and there is some indication that other polymerases may use a similar mechanism [25]. Processivity factors also may be encoded within the polymerase itself, as has been proposed for low-processivity enzymes such as the Klenow fragment of *E. coli* DNA polymerase I [26] and DNA polymerase β [27, 28]. In general, processivity is conferred by an appendage to the polymerase, distinct from the substrate binding and catalytic domains, that allows the enzyme to remain closely associated with the template during copying. In an RNA world, such auxiliary domains may have been appended to a polymerase ribozyme to convert these enzymes from a dissociative to a processive mode of action.

In summary, nucleic acid polymerases are able to recognize a primer-template complex in a largely sequence-independent manner and catalyze the template-directed addition of NTPs through attack of the 3′ end of the primer on the α-phosphate of the incoming nucleotide, with concomitant loss of pyrophosphate. A two metal ion mechanism may be employed for catalysis, in which protein side chains position metals within the active site to interact with both the 3′-hydroxyl of the primer and the phosphate groups of the incoming nucleotide. High fidelity of information transfer is ensured largely through an induced fit mechanism, in which the polymerase folds around the forming base pair to generate a functional active site only if the nucleotides are correctly paired. Finally, processivity of the polymerase can be attributed to protein domains, either connected to or distinct from the catalytic domain, that help the polymerase maintain close association with the template that is being copied.

Template-Directed Oligomerization of Mononucleotides

Early investigations into the problem of information transfer in an RNA world focused on the nonenzymatic template-directed polymerization of activated mononucleotides. Mononucleotides were activated at the 5′-phosphate by either carbodiimide or imidazole [29], rather than the triphosphate activation of nucleotide substrates utilized by most protein polymerases. When suitably 5′-activated nucleotides are allowed to react in the presence of a complementary template, the oligomeric products contain a mixture of 2′,5′-, 3′,5′-, and 5′,5′-phosphodiester linkages [30]. Most protein polymerases synthesize oligonucleotides that contain 3′,5′ linkages, although the 2′,5′-oligoadenylate synthases catalyze formation of 2′,5′-linked RNAs [31]. While the latter enzymes do not operate in a template-dependent manner, they raise the possibility that 2′,5′-linked oligonucleotides may have existed in an RNA world.

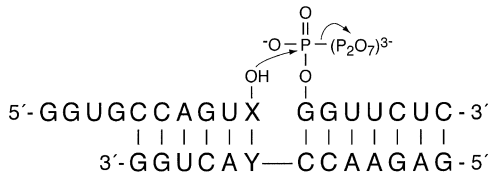

Figure 2. Uncatalyzed Template-Directed Ligation of Triphosphate-Activated Oligonucleotides

X and Y indicate nucleotides that were varied in order to study the effects of base pairing on the regiospecificity of ligation. (Adapted from [42].)

Efforts have been made to develop protocols for template-directed oligomerization that favor the formation of 3′,5′-phosphodiester linkages. The formation of 3′,5′-linked RNAs was favored with poly(C) templates when guanosine 5′-phosphorimidazolides were oligomerized in the presence of Zn^{2+} [32] or when guanosine 5′-phosphoro-2-methylimidazolide was employed [33]. All four activated mononucleotides could be oligomerized on poly(C,N) templates (N = A, G, or U) [34], but those reactions required a substantial excess of cytidine in the template [35, 36]. Similar template-copying reactions were demonstrated in systems that more closely resembled the primer extension reactions of protein polymerases [37–39]. Employing defined-sequence templates, accurate information transfer was demonstrated in the absence of proteins [40]. However, the varied efficiency of oligomerization of the four nucleoside 5′-phosphoro-2-methylimidazolides limited the number of different sequences that could be copied [29]. The detrimental effects of template intra- and intermolecular structure and the requirement for an excess of cytidine in the template also limited the scope of these reactions [29]. This suggests that generalized information transfer in an RNA world would have been difficult to achieve through the nonenzymatic template-directed polymerization of mononucleotides.

Other studies of uncatalyzed template-directed reactions involved the ligation of short RNA oligonucleotides that were activated with a 5′-triphosphate group, which may be considered mimics of NTPs (Figure 2) [41, 42]. Although the rates of reaction in these experiments were markedly reduced compared to those seen with 5′-imidazole activation, the ratio of 3′,5′- versus 2′,5′-linked oligonucleotide products was significantly increased, indicating that the leaving group at the 5′ end can influence the regiospecificity of the reaction [42]. The ligation reactions were found to be dependent on and catalyzed by the addition of divalent metal cations [41]. Additionally, an extended Watson-Crick duplex on both sides of the ligation junction was found to enhance significantly the 3′,5′ regiospecificity of the reaction [42]. Thus, employing substrates similar to those utilized by protein polymerases, the ligation of RNA oligonucleotides on an RNA template results in preferential formation of the type of linkages observed in most natural RNAs. How-

ever, the rate of these ligation reactions is very slow, typically $\sim 10^{-7}$ min^{-1}, indicating the need for catalysis to support replication in an RNA world.

In Vitro Evolution of Ligase Ribozymes

In vitro evolution has been employed to generate ribozymes with polymerase-like properties. Several ribozymes have been developed that catalyze ligation of a hydroxyl at the 3′ end of a template-bound primer to the triphosphate at the 5′ end of the ribozyme, with concomitant loss of pyrophosphate [43–47]. The same chemical reaction is involved in both oligonucleotide ligation and primer extension through mononucleotide addition. Thus, the generation of ligase ribozymes is an incremental step toward achieving RNA-catalyzed RNA polymerization. All of the known ligase ribozymes were generated employing the same general approach. A pool of RNA molecules was challenged to bind a primer molecule through Watson-Crick pairing and catalyze RNA ligation in order to be selectively amplified. Although a similar method was used to develop the various ligase ribozymes, each had a distinct lineage and exhibits unique biochemical properties. Thus, it is likely that there were many possible solutions to the problem of RNA-catalyzed information transfer in an RNA world.

The class I ligase was evolved from a starting pool of more than 10^{15} random-sequence RNAs (Figure 3A) [43, 44]. Ribozymes were selected for the ability to bind an RNA primer through Watson-Crick pairing to a template region and to catalyze ligation of the primer to their own 5′ end. In the initial pool of RNA molecules, the template region was located adjacent to an engineered hairpin loop at the 5′ end of the ribozyme. This was designed to hold the 5′ nucleotide of the ribozyme in an adjacent template-bound position to mimic the template-bound NTP in a primer extension reaction catalyzed by a polymerase protein.

The ligase ribozymes evolved from this pool were divided into three classes based on their predicted secondary structure and regiospecificity of ligation. The class II and class III ribozymes catalyzed formation of a 2′,5′ linkage. Ribozymes in both of these classes no longer employed the engineered hairpin at the 5′ end, and the class II ribozymes utilized a different template sequence than the one that had been intended. In contrast, the class I ligase ribozymes catalyzed formation of a 3′,5′ linkage. These ribozymes did take advantage of the engineered template region, although they too had disrupted the hairpin at the 5′ end. Interestingly, a single templating nucleotide for the 5′-terminal residue of the ribozyme was retained, allowing for a reaction akin to primer extension. Thus, a mechanism similar to that of protein polymerases may be proposed for substrate binding and orientation. In a slightly modified format, this ribozyme also catalyzes template-directed primer extension utilizing NTPs [48], as will be discussed below.

The L1 ligase ribozyme was developed in a similar manner, starting from a pool of random-sequence RNAs, each with an attached template region and 5′-terminal hairpin [45]. Again, the evolved ribozyme did not retain the 5′ hairpin and utilized a template region other than the one that was engineered. This ligase forms 3′,5′-phosphodiester linkages. It was found to operate in an allosteric manner such that catalytic activity was dependent on an oligonucleotide effector (Figure 3B) [45]. It was engineered to be activated by small molecule effectors [49] and has been evolved to respond to protein effectors [50]. The L1 ligase adopts a simple three-way junction secondary structural motif and catalyzes ligation of an oligonucleotide that is bound within one of the three stems. The nucleotides surrounding the ligation junction are not Watson-Crick paired, with the 3′-terminal uridylate of the oligonucleotide substrate engaged in a U•G wobble pair and the 5′ guanylate of the ribozyme positioned as a G•A mismatch. This is likely to result in orientation of the substrates within the ribozyme active site in a manner distinct from that used by the class I ligase and protein polymerases. The ability of the L1 ligase to be activated by external cofactors suggests a potential mechanism for phenotypic regulation in an RNA world. If this ligase could be converted into a polymerase, environmental factors might regulate the expression of a replication phenotype, allowing replication to occur only under favorable environmental conditions.

Another ligase that has been evolved from a random sequence library is the R3/R3C ribozyme [46]. In its initial form (R3), this ribozyme contained only three of the four nucleotides, lacking cytidine residues (Figure 3C). It then was evolved to contain all four nucleotides (R3C). Evolution was initiated from a pool of random-sequence RNAs that contained a fixed template region that was complementary to the 3′ end of the oligonucleotide substrate. The initially selected 155-nucleotide ligase ribozyme was reduced to an active structure that contained only 74 nucleotides arranged in a simple three-way junction motif with the substrate binding site at its 3′ end. The ligation reaction occurred in the context of a fully Watson-Crick paired substrate and an unpaired 5′-terminal nucleotide of the ribozyme, resulting in formation of a 3′,5′-phosphodiester linkage. Again, it can be surmised that proper orientation of the substrates is achieved by a mechanism distinct from that utilized by protein polymerases. Although a three-nucleotide genetic system would not be capable of supporting replication in an RNA world, the R3 ribozyme illustrates that catalysis can be realized with reduced chemical diversity, lending credence to the idea that simple genetic systems with fewer than four nucleotides may have existed, as has been suggested previously [51, 52].

The R3C ligase, which was evolved from the R3 ligase following the introduction of cytidine, contained one deleted nucleotide and eleven mutations relative to the R3 ligase. Seven of those mutations were added cytidine residues (Figure 3D). The structural motif of the R3C ribozyme is similar to that of the R3 ribozyme, with the newly introduced cytidine residues mainly serving to stabilize the stem regions. The catalytic rate of the R3C ribozyme was increased by 20-fold compared to the R3 ribozyme, but the ligation junction remained the same, with the 5′-terminal nucleotide of the ribozyme still being unpaired.

The fourth known example of an in vitro-evolved ligase ribozyme that forms a 3′,5′-phosphodiester linkage is the class hc ligase (Figure 3E) [47]. A somewhat different

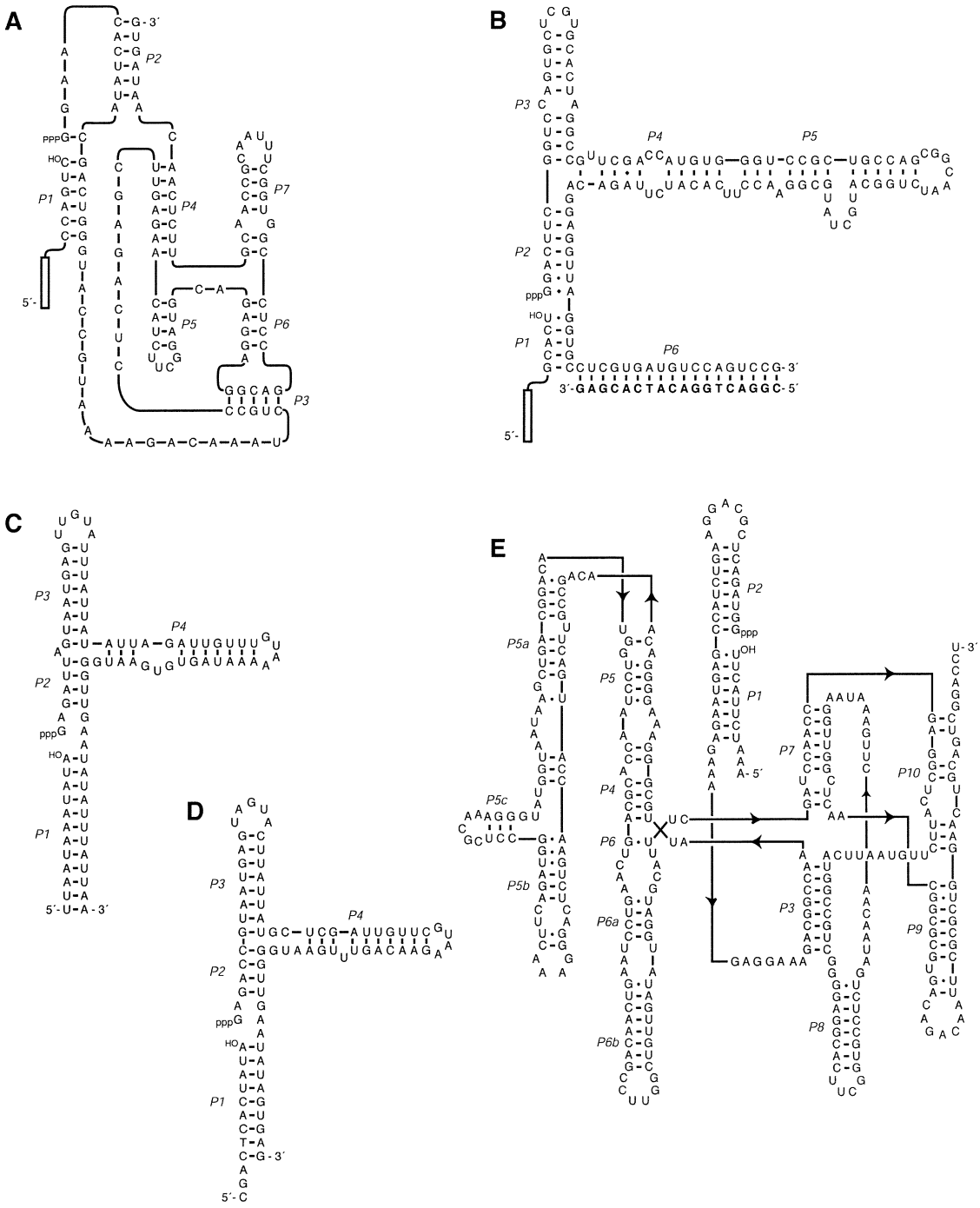

Figure 3. Ligase Ribozymes Obtained by In Vitro Evolution

The primary sequence and proposed secondary structure are shown for (A) the class I ligase [44]; (B) the L1 ligase, with the oligonucleotide effector shown in bold [45]; (C) the R3 ligase [46]; (D) the R3C ligase [46]; and (E) the class hc ligase [47]. Figures were adapted from the references indicated.

strategy was used to generate this ribozyme. The independently folding P4–P6 domain of the *Tetrahymena* group I ribozyme [53, 54] was employed as a structural scaffold upon which to build ligase activity. Three randomized regions totaling 85 nucleotides were appended to the structural scaffold. The ribozymes were evolved for the ability to catalyze ligation of the 3′ end of an oligonucleotide substrate to their own 5′ end, with ligation occurring in the context of a helix created by a constant stem-loop region at the 5′ end of the RNA. The final evolved ribozyme retained this stem-loop structure, allowing for a reaction in which the nucleotides both upstream and downstream of the ligation junction were fully paired. This provided a reaction format closely re-

sembling that of polymerase proteins. As mentioned above, the context of a complete duplex favors formation of 3′,5′ linkages in the uncatalyzed template-directed ligation of oligonucleotides [42]. In the case of the class hc ligase, there was a preference for substrates that contained a G•U wobble pair on the 5′ side of the ligation junction. However, the reaction also proceeded efficiently when a G-C pair was present at this position. The ribozyme was able to utilize duplex substrates with a variety of sequences upstream and downstream of the ligation junction, presumably employing a general mechanism for substrate recognition similar to polymerase recognition of a primer-template complex. Thus, the class hc ribozyme illustrates that a reaction format similar to primer extension and generalized substrate recognition is within the capabilities of RNA.

The various in vitro-evolved ligase ribozymes represent potential stepping stones toward the goal of developing a replicase ribozyme. They all catalyze the chemistry of polymerization, specifically the template-directed attack of a nucleotide 3′-hydroxyl on the α-phosphate of a nucleotide 5′-triphosphate. The class I [44, 55], R3/R3C [46], and class hc [47] ribozymes all have been shown to act as true enzymes, ligating two RNA substrates that are separate from the ribozyme. It is reasonable to surmise that the L1 ligase [45] also could be made to operate in an intermolecular reaction format. However, all of these RNA-catalyzed reactions require the substrates to form Watson-Crick pairs with the ribozyme and thus are not capable of sequence-independent substrate recognition as is exhibited by polymerase proteins. Furthermore, they all employ oligonucleotide 5′-triphosphate substrates rather than NTPs and catalyze only a single joining reaction rather than the polymerization of RNA substrates.

RNA-Catalyzed Primer Extension

Starting with the ligase ribozymes described above, the next step in realizing a replicase ribozyme is the development of RNAs that catalyze the addition of multiple substrates (e.g., NTPs) in a template-directed manner. It also is important to utilize a primer-template complex that is separate from the ribozyme and is recognized by the ribozyme in a sequence-independent manner. Both properties have been achieved for in vitro-evolved forms of the class I and class hc ligase ribozymes.

The class I ligase was converted to a format that supports reactions on an external template by dividing the ribozyme between the P1 and P3 stems (Figure 3A). The template and half of the P2 stem were provided as a separate RNA molecule that could bind an oligonucleotide primer (Figure 4A) [48]. The template strand still was recognized by the ribozyme through Watson-Crick pairing that recapitulated the P2 stem. This restructured ribozyme was able to catalyze the template-directed addition of three NTPs onto the 3′ end of the primer. Some mechanistic aspects of the reaction have been investigated, pointing to similarities with protein polymerases. It has been proposed, for example, that the ribozyme employs a two metal ion mechanism for catalysis, and that both binding and orientation of the NTP substrates are important for discrimination of Watson-Crick pairs [56].

All four NTPs could act as substrates for template-directed polymerization catalyzed by the class I ligase [48]. The average fidelity of Watson-Crick addition was 85% when all four NTPs were supplied at equimolar concentrations, with most errors occurring as a result of G•U wobble pairing. The fidelity was increased to 92% when the concentration of GTP was made 10-fold lower than that of the other three NTPs. If the ribozyme could copy a long template with similar fidelity, it would produce eight errors per 100 nucleotides, which is too many to support a stable replicative system.

An evolved form of the class I ligase was shown to support NTP addition reactions utilizing a portion of the ribozyme sequence as a template [57]. This ribozyme was obtained through continuous in vitro evolution, a method in which catalysis and selective amplification are made to occur within a single reaction vessel (Figure 5) [58]. In continuous evolution, the ribozymes bind and ligate a primer molecule that contains the sequence of the T7 RNA polymerase promoter element. All of the RNA molecules are reverse transcribed to form cDNAs, but only those that have ligated the primer onto their own 5′ end will give rise to cDNAs that contain a functional double-stranded promoter. These in turn are forward transcribed by T7 RNA polymerase to produce progeny RNA molecules that are eligible to enter another cycle of reaction and selective amplification.

In order to evolve ribozymes that are capable of catalyzing the template-directed addition of NTPs, the ligase ribozymes were challenged to react with primers that were shortened by one or two nucleotides at the 3′ end of the promoter sequence. In order to become eligible for amplification, the ribozymes were required to catalyze the addition of one or two NTPs, thereby completing the promoter, followed by ligation of the extended primer [57]. The E278-19 ribozyme that resulted from the continuous evolution process (Figure 4B) catalyzed the template-directed addition of two NTPs followed by RNA ligation. It was able to add NTPs in either a 5′→3′ direction or 3′→5′ direction. This ribozyme expands the notion of primer extension to include the possibility of NTP addition onto either end of a template-bound primer.

Although both of the ribozymes described above highlight the ability of RNA molecules to extend a primer by catalyzing the addition of NTPs, ribozyme-specific Watson-Crick pairing of the primer or template still was required. Thus, these ribozymes do not operate in a general manner that would allow copying of any template sequence. Sequence-independent recognition of a primer-template complex has been realized with an evolved form of the class hc ligase ribozyme [59]. The ribozymes were evolved to ligate an oligonucleotide substrate that was bound to a hairpin motif attached to the 5′ end of the ribozyme. During the evolution process, a pseudo-intramolecular reaction format was achieved by introducing a long poly(U) linker between the hairpin and ribozyme. The resulting 18-2 ribozyme was shown to catalyze ligation of two RNA substrates on an external RNA template (Figure 4C). Ligation took place either in the context of a hairpin molecule that acted as both template and triphosphate-containing oligonucleotide or in the context of two oligonucleotides that were bound

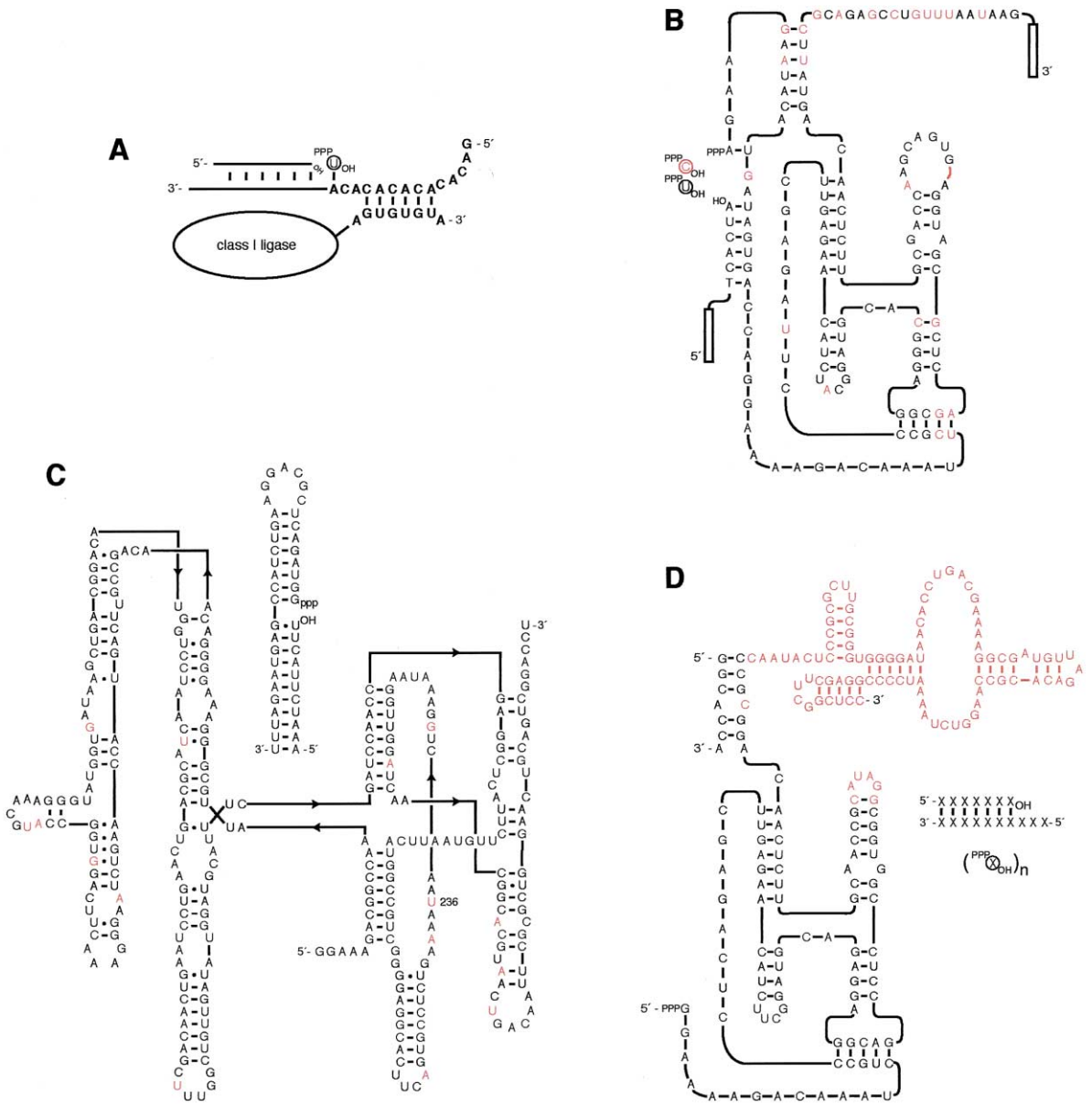

Figure 4. Ligase Ribozymes that Were Modified in Order to Exhibit Polymerase-like Activity

(A) The class I ligase ribozyme, restructured to support NTP addition reactions [48]; (B) the E278-19 ribozyme, which was evolved from the class I ligase [57]; (C) the class hc-derived 18-2 ribozyme [59]; and (D) the class I-derived polymerase ribozyme [60]. Mutations relative to the respective parent ribozyme molecules are shown in red. The NTP substrates are indicated by circled letters. Note that the 18-2 ribozyme contains two nucleotide changes compared to the previously published sequence due to correction of a typographical error: residues 236 and 237 should be U and A (rather than C and U), respectively. Figures were adapted from the references indicated.

to a separate RNA template. The reaction was shown to be general with respect to the sequence both upstream and downstream of the ligation junction, indicating that this ribozyme recognizes general features of an RNA duplex. The hc-derived 18-2 ligase also was shown to catalyze the template-directed addition of NTPs at a low level, but this occurred only in the context of a hairpin molecule and with varying fidelity, depending on the templating nucleotide. Thus, the sequence-independent recognition of a primer-template complex was realized but has not been extended to include the accurate template-directed polymerization of NTPs.

The furthest advance toward the development of an RNA replicase ribozyme involves an evolved form of the class I ligase that catalyzes the addition of multiple NTPs onto the 3′ end of a primer that is bound to an external RNA template (Figure 4D) [60]. In developing this ribozyme, the class I ligase was employed as a core catalytic domain, to which was appended a random-sequence domain of 76 nucleotides. During the in vitro-evolution process, a primer molecule was tethered to the 5′ end of the ribozyme, and the template RNA was provided as a separate molecule. Ribozymes were selected for the ability to catalyze extension of the primer by addition

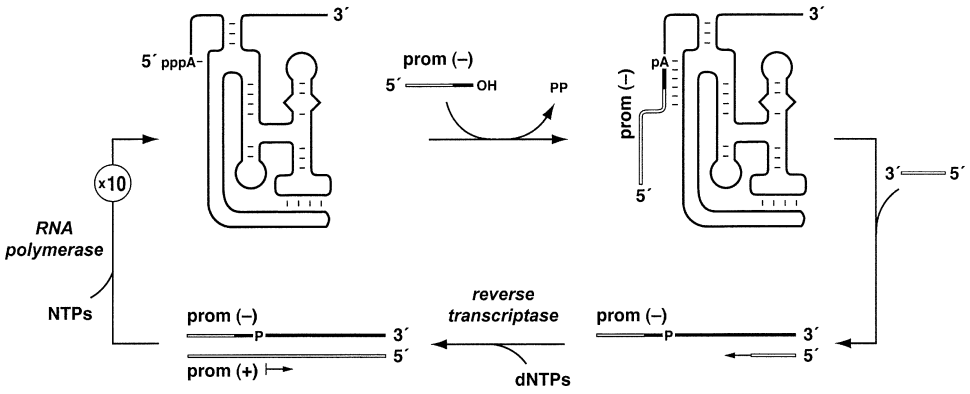

Figure 5. Scheme for Continuous In Vitro Evolution

A pool of ligase ribozymes is challenged to bind to a chimeric DNA-RNA substrate (DNA shown as open lines, RNA as solid lines) that contains the sequence of the T7 RNA polymerase promoter element [prom (–)]. The ribozymes are required to catalyze ligation of the 3′ end of the substrate to their own 5′ end. Complementary DNAs are generated by reverse transcriptase, extending a DNA primer that binds to the 3′ end of the ribozyme. Reacted RNAs give rise to a double-stranded promoter element and subsequently are transcribed by T7 RNA polymerase to yield progeny RNAs that are eligible to enter another cycle of reaction and selective amplification. (Adapted from [58].)

of NTPs that contained an attached biotin moiety. The evolved ribozyme was able to operate on a fully detached primer-template complex and in a sequence-general manner, extending a variety of primers by as many as 14 nucleotides. The ribozyme was able to extend template-bound primers that contained 10–60 nucleotides, indicating that it might be evolved to copy long templates.

The overall fidelity of NTP addition was found to be 96.7% and could be increased to 98.5% by biasing the concentrations of the four NTPs. As was seen with the restructured class I ligase [48], most errors were due to misincorporation of guanosine opposite templating uridine residues. Following a misincorporation event, the ribozyme was less likely to continue extending along the template. This indicates that abortive template copying may operate as a simple mechanism to ensure accurate production of full-length RNAs, similar to mechanisms employed by protein polymerases. The class I-derived polymerase ribozyme has true RNA-dependent RNA polymerase activity and exhibits many of the features of modern protein polymerases. However, the ribozyme would not be able to copy itself in a replicative manner; it only is able to extend a primer by 14 nucleotides, while the ribozyme itself contains 189 nucleotides. It has been suggested that the chief limit to further primer extension is the slow catalytic rate of the ribozyme, which might be improved by further in vitro evolution [60].

In summary, advances in recent years have led to the development of ribozymes with RNA polymerase activity. It has been shown that RNA is capable of catalyzing the chemistry of polymerization employing NTP substrates, recognizing an external primer-template complex in a sequence-independent manner, and operating with high fidelity in copying the template to a complementary strand. Significant obstacles still must be overcome, however, in order to generate an RNA replicase ribozyme. The fidelity of these reactions must be improved and the efficiency of primer extension must be increased in order to support copying of the entire ribozyme.

From RNA Polymerase to RNA Replicase

When considering the operation of an RNA replicase ribozyme in an RNA world, several issues in addition to the chemistry of polymerization must be addressed. These include the reaction conditions under which polymerization might have occurred, the minimum fidelity required for replication, the degree of processivity in the polymerization reaction, and the means by which the replicase discriminated between self and non-self RNA. The details of what might have occurred in an RNA world are unclear, but a general understanding of these issues can guide efforts to recapitulate an RNA replicase ribozyme in the laboratory.

The environmental conditions of the early Earth are not known with certainty [61]. Assuming that activated mononucleotides were available in sufficient concentration to support RNA replication, there is likely to have been an additional requirement for millimolar concentrations of magnesium ions or perhaps some other divalent metal cation. This is advantageous for the formation of well-defined secondary and tertiary structures of RNA and for the catalysis of RNA polymerization. Divalent metal ions are detrimental, however, to the longevity of RNA in solution because they accelerate the hydrolysis of RNA phosphodiesters. Similarly, it would be difficult to maintain RNA templates and replicase ribozymes under conditions of high temperature or alkaline pH. The class I-derived polymerase ribozyme preferably operates at pH 8.5 and 22°C in the presence of 200 mM $MgCl_2$, conditions that lead to substantial degradation of RNA. This is thought to be an important factor in limiting the extent of RNA-catalyzed polymerization [60]. It would be helpful to develop polymerase ribozymes that operate under conditions that are less likely to promote RNA degradation.

Another important issue concerning RNA replicase ribozymes is the degree of fidelity that would be needed to achieve sustainable replication [30, 62]. A critical factor is the length of RNA required to support replication, taking into account those residues that are more or less intolerant of substitution. It has been suggested that a copying fidelity of at least 98% would be needed to

sustain a system in which ribozyme molecules containing ~100 nucleotides generated an average of ten copies each (both plus and minus strands), assuming that 80 of the 100 nucleotides are intolerant of mutation [62]. This would allow the production of at least two RNAs, one ribozyme and its complement, that contained all of the nucleotides required for function. This level of fidelity has been achieved with the class I-derived polymerase in the presence of biased pools of NTP substrates [60]. However, the ribozyme contains almost 200 nucleotides, requiring an even higher fidelity of replication, assuming that each ribozyme molecule could generate an average of ten copies. Most protein polymerases exhibit a copying fidelity of greater than 99.99% [21], and a fidelity approaching this level may be required for an RNA replicase ribozyme. In any case, the issue of the fidelity required for a ribozyme to sustain its own replication cannot be addressed adequately until a ribozyme is obtained that can copy template sequences of comparable length to the ribozyme itself.

The degree to which processivity would be important for a replicase ribozyme is unclear. A processive mechanism, in which the ribozyme remains closely associated with the template, presumably would allow for more efficient copying of longer templates. This would allow shorter copying times to counterbalance the problem of RNA degradation discussed above. However, a rapid dissociative mechanism, in which the ribozyme repeatedly dissociates from and rebinds to the primer-template complex, may be easier to evolve. The extent to which the class I-derived polymerase is able to operate processively is not clear [60]. In any case, in vitro-evolution methods might be employed to develop or enhance the processivity of this ribozyme.

Another important issue is the ability of a replicase ribozyme to distinguish its own RNA from other RNAs present in the local environment. In order to ensure its own survival and facilitate evolutionary optimization of its catalytic activity, a replicase ribozyme must not be deflected from the task of copying its own genetic information. One possible solution to this problem is the "genomic tag" hypothesis, which suggests that the replicase might recognize a tRNA-like structure at the 3′ end of cognate template RNAs [63, 64]. The tag must be present in both the ribozyme and its complement and, upon recognition by the ribozyme, must not interfere with subsequent replication. Specific recognition of a tag runs counter to most laboratory experiments, which have attempted to achieve generalized polymerization. Thus, tag identification may require a recognition domain appended to a general-purpose catalytic domain, enabling copying of almost any template RNA following recognition of the tag.

Great strides have been made toward developing a ribozyme that is capable of RNA replicase activity. Several issues remain to be addressed, however, in order to demonstrate that such an RNA molecule could have sustained its own replication. Thus far, it has been fruitful to consider modern protein polymerases as a model for developing these ribozymes. Catalytic RNAs have been obtained that exhibit many of the properties of protein polymerases. However, using protein polymerases as the sole model to guide laboratory experiments may be too restrictive and perhaps may not take advantage of the unique properties of RNA. Many challenges remain in order to show that RNA alone could have supported an evolving genetic system and even more must be faced in showing that an RNA world could have existed under the conditions of the early Earth.

References

1. Gesteland, R.F., Cech, T.R., and Atkins, J.F. (1999). The RNA World (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
2. McKay, D.B., and Wedekind, J.E. (1999). Small ribozymes. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 265–286.
3. Lambowitz, A.M., and Caprara, M.G. (1999). Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In The RNA World, Gesteland, R.F., Cech, T.R., and Atkins J.F., eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 451–485.
4. Altman, S., and Kirsebom, L. (1999). Ribonuclease P. In The RNA World, Gesteland, R.F., Cech, T.R., and Atkins J.F., eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 351–380.
5. Kurz, J.C., and Fierke, C.A. (2000). Ribonuclease P: a ribonucleoprotein enzyme. Curr. Opin. Chem. Biol. *4*, 553–558.
6. Noller, H.F., Hoffarth, V., and Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. Science *256*, 1416–1419.
7. Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science *289*, 920–930.
8. White, H.B., III. (1976). Coenzymes as fossils of an earlier metabolic state. J. Mol. Evol. *7*, 101–104.
9. Arnold, E., Ding, J., Hughes, S.H., and Hostomsky, Z. (1995). Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. Curr. Opin. Struct. Biol. *5*, 27–38.
10. Brautigam, C.A., and Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Curr. Opin. Struct. Biol. *8*, 54–63.
11. Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Jr., Lu, X., Tantillo, C., Williams, R.L., Kramer, G., Ferris, A.L., Clark, P., et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA *90*, 6320–6324.
12. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. Science *264*, 1891–1903.
13. Eom, S.H., Wang, J., and Steitz, T.A. (1996). Structure of *Taq* polymerase with DNA at the polymerase active site. Nature *382*, 278–281.
14. Doublié, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature *391*, 251–258.
15. Li, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. *17*, 7514–7525.
16. Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669–1675.
17. Franklin, M.C., Wang, J., and Steitz, T.A. (2001). Structure of the replicating complex of a pol α family DNA polymerase. Cell *105*, 657–667.
18. Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J., and Pelletier, H. (1997). Crystal structures of human DNA polymerase β com-

plexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry *36*, 11205–11215.
19. Kiefer, J.R., Mao, C., Braman, J.C., and Beese, L.S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. Nature *391*, 304–307.
20. Beese, L.S., Derbyshire, V., and Steitz, T.A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science *260*, 352–355.
21. Kunkel, T.A., and Bebenek, K. (2000). DNA replication fidelity. Annu. Rev. Biochem. *69*, 497–529.
22. Fedor, M.J. (2002). The role of metal ions in RNA catalysis. Curr. Opin. Struct. Biol. *12*, 289–295.
23. Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. Cell *69*, 425–437.
24. Krishna, T.S.R., Kong, X.-P., Gary, S., Burgers, P.M., and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell *79*, 1233–1243.
25. Kelmann, Z., Hurwitz, J., and O'Donnell, M. (1998). Processivity of DNA polymerases: two mechanisms, one goal. Structure *6*, 121–125.
26. Minnick, D.T., Astatke, M., Joyce, C.M., and Kunkel, T.A. (1996). A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. J. Biol. Chem. *271*, 24954–24961.
27. Singhal, R.K., and Wilson, S.H. (1993). Short gap-filling synthesis by DNA polymerase β is processive. J. Biol. Chem. *268*, 15906–15911.
28. Pelletier, H., Sawaya, M.R., Wolfle, W., Wilson, S.H., and Kraut, J. (1996). Crystal structures of human DNA polymerase β complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. Biochemistry *35*, 12742–12761.
29. Joyce, G.F. (1987). Nonenzymatic template-directed synthesis of informational macromolecules. Cold Spring Harb. Symp. Quant. Biol. *52*, 41–51.
30. Joyce, G.F., and Orgel, L.E. (1999). Prospects for understanding the origin of the RNA world. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 49–77.
31. Justesen, J., Hartmann, R., and Kjeldgaard, N.O. (2000). Gene structure and function of the 2′-5′-oligoadenylate synthase family. Cell. Mol. Life Sci. *57*, 1593–1612.
32. Bridson, P.K., and Orgel, L.E. (1980). Catalysis of accurate poly(C)-directed synthesis of 3′-5′-linked oligoguanylates by Zn^{2+}. J. Mol. Biol. *144*, 567–577.
33. Inoue, T., and Orgel, L.E. (1981). Substituent control of the poly(C)-directed oligomerization of guanosine 5′-phophorimidazolide. J. Am. Chem. Soc. *103*, 7666–7667.
34. Inoue, T., and Orgel, L.E. (1982). Oligomerization of (guanosine 5′-phosphor)-2-methylimidazole on poly(C): An RNA polymerase model. J. Mol. Biol. *162*, 201–217.
35. Joyce, G.F., Inoue, T., and Orgel, L.E. (1984). Non-enzymatic template-directed synthesis on RNA random copolymers: poly(C,U) templates. J. Mol. Biol. *176*, 279–306.
36. Joyce, G.F., and Orgel, L.E. (1986). Non-enzymic template-directed synthesis on RNA random copolymers: poly(C,G) templates. J. Mol. Biol. *188*, 433–441.
37. Wu, T., and Orgel, L.E. (1992). Nonenzymatic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. J. Am. Chem. Soc. *114*, 317–322.
38. Wu, T., and Orgel, L.E. (1992). Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. J. Am. Chem. Soc. *114*, 5496–5501.
39. Wu, T., and Orgel, L.E. (1992). Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. J. Am. Chem. Soc. *114*, 7963–7969.
40. Inoue, T., Joyce, G.F., Grzeskowiak, K., and Orgel, L.E. (1984). Template-directed synthesis on the pentanucleotide CpCpGp CpC. J. Mol. Biol. *178*, 669–676.
41. Rohatgi, R., Bartel, D.P., and Szostak, J.W. (1996). Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. J. Am. Chem. Soc. *118*, 3332–3339.
42. Rohatgi, R., Bartel, D.P., and Szostak, J.W. (1996). Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3′-5′ phosphodiester bonds. J. Am. Chem. Soc. *118*, 3340–3344.
43. Bartel, D.P., and Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. Science *261*, 1411–1418.
44. Ekland, E.H., Szostak, J.W., and Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. Science *269*, 364–370.
45. Robertson, M.P., and Ellington, A.D. (1999). *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. Nat. Biotechnol. *17*, 62–66.
46. Rogers, J., and Joyce, G.F. (2001). The effect of cytidine on the structure and function of an RNA ligase ribozyme. RNA *7*, 395–404.
47. Jaeger, L., Wright, M.C., and Joyce, G.F. (1999). A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. Proc. Natl. Acad. Sci. USA *96*, 14712–14717.
48. Ekland, E.H., and Bartel, D.P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature *382*, 373–376.
49. Robertson, M.P., and Ellington, A.D. (2000). Design and optimization of effector-activated ribozyme ligases. Nucleic Acids Res. *28*, 1751–1759.
50. Robertson, M.P., and Ellington, A.D. (2001). *In vitro* selection of nucleoprotein enzymes. Nat. Biotechnol. *19*, 650–655.
51. Crick, F.H.C. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367–379.
52. Orgel, L.E. (1968). Evolution of the genetic apparatus. J. Mol. Biol. *38*, 381–393.
53. Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., and Woodson, S.A. (1998). RNA folding at millisecond intervals by synchotron hydroxyl radical footprinting. Science *279*, 1940–1943.
54. Treiber, D.K., Rook, M.S., Zarrinkar, P.P., and Williamson, J.R. (1998). Kinetic intermediates trapped by native interactions in RNA folding. Science *279*, 1943–1946.
55. Bergman, N.H., Johnston, W.K., and Bartel, D.P. (2000). Kinetic framework for ligation by an efficient RNA ligase ribozyme. Biochemistry *39*, 3115–3123.
56. Glasner, M.E., Yen, C.C., Ekland, E.H., and Bartel, D.P. (2000). Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. Biochemistry *39*, 15556–15562.
57. McGinness, K.E., Wright, M.C., and Joyce, G.F. (2002). Continuous *in vitro* evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. Chem. Biol. *9*, 585–596.
58. Wright, M.C., and Joyce, G.F. (1997). Continuous *in vitro* evolution of catalytic function. Science *276*, 614–617.
59. McGinness, K.E., and Joyce, G.F. (2002). RNA-catalyzed RNA ligation on an external RNA template. Chem. Biol. *9*, 297–307.
60. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science *292*, 1319–1325.
61. Mojzsis, S.J., Krishnamurthy, R., and Arrhenius, G. (1999). Before RNA and after: geophysical and geochemical constraints on molecular evolution. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 1–47.
62. Bartel, D.P. (1999). Re-creating an RNA replicase. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 143–162.
63. Weiner, A.M., and Maizels, N. (1987). tRNA-like structures tag the 3′ ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. Proc. Natl. Acad. Sci. USA *84*, 7383–7387.
64. Maizels, N., and Weiner, A.M. (1999). The genomic tag hypothesis: what molecular fossils tell us about the evolution of tRNA. In The RNA World, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 79–111.