In Vitro Selection of Functional Nucleic Acids

David S. Wilson and Jack W. Szostak

Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114-2696; e-mail: wilson@genetics.mgh.harvard.edu, szostak@frodo.mgh.harvard.edu

Key Words   aptamer, evolution, ribozyme, SELEX, combinatorial library

■ **Abstract**   In vitro selection allows rare functional RNA or DNA molecules to be isolated from pools of over 10^{15} different sequences. This approach has been used to identify RNA and DNA ligands for numerous small molecules, and recent three-dimensional structure solutions have revealed the basis for ligand recognition in several cases. By selecting high-affinity and -specificity nucleic acid ligands for proteins, promising new therapeutic and diagnostic reagents have been identified. Selection experiments have also been carried out to identify ribozymes that catalyze a variety of chemical transformations, including RNA cleavage, ligation, and synthesis, as well as alkylation and acyl-transfer reactions and N-glycosidic and peptide bond formation. The existence of such RNA enzymes supports the notion that ribozymes could have directed a primitive metabolism before the evolution of protein synthesis. New in vitro protein selection techniques should allow for a direct comparison of the frequency of ligand binding and catalytic structures in pools of random sequence polynucleotides versus polypeptides.

CONTENTS

611

INTRODUCTION

In vitro selection or SELEX (systematic evolution of ligands by exponential amplification) makes use of large populations of random RNA or DNA sequences as the raw material for the selection of rare functional molecules (1–3). These techniques have broadened our appreciation of the capabilities of nucleic acids (for recent reviews, see 4, 5), thereby increasing the plausibility of the RNA world hypothesis, as well as providing new reagents with potential therapeutic and diagnostic applications (6–9). This review highlights some of the recent advances in the in vitro selection of functional nucleic acids, with particular attention to some of the most challenging problems being addressed by these techniques. We emphasize selections that have demonstrated novel properties of nucleic acids, as opposed to those that have refined our understanding of natural protein-nucleic acid interactions (10, 11) or of mechanisms of ribozyme catalysis (12–14; for reviews, see 5, 15). Reviews on the early years of in vitro selection and its conceptual origins have appeared previously (1–3, 16–21). (For a recent review of the use of aptamers for diagnostic purposes and for the dissection of intracellular processes, see 11.)
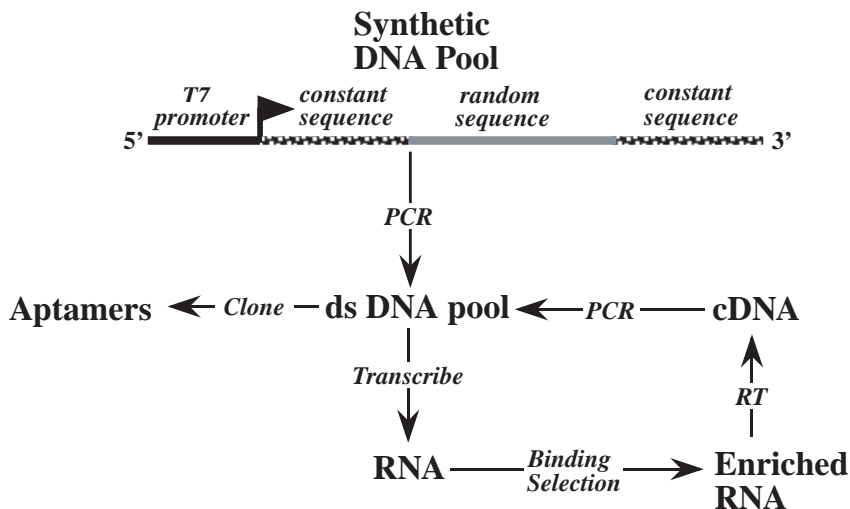
TECHNOLOGY AND ITS HISTORY

To apply genetic selections directly to populations of molecules, genotype and phenotype must be linked. In the case of a functional RNA, the same molecule can possess both genotype (represented by a sequence that can be copied by a polymerase) and phenotype (represented by some functional trait that varies according to sequence). The first to exploit this property of RNA was Sol Spiegelman, whose work in the 1960s with the RNA bacteriophage Qβ showed that Darwinian selection could operate in a cell-free system (reviewed in 22). This viral genome can be copied in vitro by the Qβ replicase protein. Serial dilution experiments permitted hundreds of generations of genome replication to be performed quickly, with the phenotype under selective pressure being replication speed. Because of the inherent mutation rate of this polymerase, genotypic variation is generated in vitro during exponential genome amplification. The result of applying selective pressure to a population of variants was evolution, where the Qβ genomes adapted by deleting sequences unnecessary for recognition by the polymerase, thus shortening replication time. Spiegelman also evolved RNA sequences that adapted to different "niches," such as the presence of ethidium bromide or an unbalanced composition of nucleotide triphosphates.

Though elegant and innovative, these initial in vitro selection experiments were restricted by the fact that replication speed was the only phenotype that could be selected for and also because the amount of genetic variation was limited by the error rate of the polymerase. Over 20 years later, the idea of imposing an additional phenotypic selection on populations of RNA molecules, that of

being able to bind a target compound, became practical. In principle, Spiegelman could have included such a step in his selections, but it is unlikely that the amount of variation present in his populations would have allowed for significant adaptation to such a selective pressure. By the 1990s, however, it was possible to generate immense sequence diversity by chemically synthesizing DNA chains in which some regions consisted of completely random sequence. The isolation of reverse transcriptase and the invention of polymerase chain reaction (PCR) made it easy to replicate almost any nucleic acid sequence in vitro (23) without the constraints of the Qβ system. These advances combined to enable the development of practical ways of using in vitro selection and directed evolution to search for sequence space for new functional RNA sequences (16, 24, 25).

The ligands that emerge from in vitro selections are called "aptamers." The general procedure for selection of an aptamer is as follows: a DNA pool is chemically synthesized with a region of random or mutagenized sequence flanked on each end by constant sequence and with a T7 RNA polymerase promoter at the 5′ end (Figure 1). This DNA is amplified by a few cycles of PCR and subsequently transcribed in vitro to make the RNA pool. The RNA molecules are then partitioned based on whether they bind to the chosen target compound, for example, by passing them through an affinity column derivatized with the target. The retained RNAs are eluted, reverse transcribed, amplified by PCR, transcribed, and then the entire cycle is repeated. All manipulations occur in vitro, which means that pool complexity is not limited by transformation efficiency. With successive rounds of selection, the ratio of active:inactive sequences increases. After 5–10 rounds, the pool becomes dominated by the once rare molecules that can bind the target ligand. The technique can also be simplified to find either double- (26) or single-stranded (27) DNA sequences that bind to targets, or it can be elaborated to allow for the selection of catalytic RNA or DNA species.

In most cases, selections start with a pool of sequences that are completely random except for the flanking constant regions. This allows for a maximally unbiased sampling of sequence space, so that a variety of different and independent adaptations to a given selection pressure can be recovered. In such experiments, one can sample over 10^{15} nucleic acid sequences, and after a few cycles of selection and amplification, one can recover the descendants of a single functional molecule from the initial population. Depending on the length (N) of the sequences being sampled, the coverage of sequence space can range from essentially complete (for $N = 25$) to vanishingly small (e.g. 10^{-117} for $N = 220$). For long sequences, the important parameter is not the fraction of sequence space that is sampled, but the probability of finding functional sequences. Functional sequences may be common enough to find experimentally by being of low informational complexity and thus statistically abundant, for example, a simple structural motif that can be embedded anywhere in a long sequence. Alternatively, many complex and thus individually rare motifs may collectively occupy a sufficient partial volume of sequence space as to be experimentally accessible (i.e. be common enough that there is a reasonable chance that at least one such sequence will be present in a pool of 10^{15} molecules).

**Synthetic
DNA Pool**

*T7
promoter*   *constant
sequence*   *random
sequence*   *constant
sequence*

5'   3'

PCR

Aptamers ← *Clone* — **ds DNA pool** ← *PCR* — **cDNA**

Transcribe   *RT*

RNA — *Binding
Selection* → **Enriched
RNA**

Figure 1   Example of typical aptamer pool structure and scheme for selection process.

APTAMER SELECTIONS

In vitro selection has been used to identify aptamers to targets covering a wide range of sizes, including simple ions (28), small molecules (24), peptides (29), single proteins (25), organelles (30), viruses (31), and even entire cells (32). In the case of complex targets like cells, the emerging aptamers will recognize the most abundant or easily recognizable molecular targets, which will not necessarily be the desired ones.

By determining the range of small molecules that RNA can specifically recognize, in vitro selection has helped to explore the implications of the RNA world hypothesis. By generating aptamers to proteins that do not normally interact with RNA, selections have been used to make reagents for interfering with the biological functions of specific proteins.

Aptamers for Small Molecules

If the RNA world hypothesis were correct, RNA would have to be capable of directing a primitive metabolism, leading to the production of compounds necessary to make, at the very least, RNA itself. To accomplish this, RNAs would require the ability to recognize small molecules, transition states, and cofactors with reasonably high affinity and specificity. Before the advent of in vitro selection, it was known that the group I intron could bind guanosine and arginine (33, 34), but it was not known how rare such RNAs were or with what other kinds of small molecules RNA was capable of interacting. This issue was addressed by

making a pool of ~10^{13} different RNA molecules with a random sequence of 100 nucleotides in length and then selecting those that were retained on columns derivatized with one of six different organic dyes (24). These dyes are polycyclic, aromatic planar molecules, with an overall negative charge and several potential hydrogen-bond (H-bond) donors and acceptors. Specific aptamers were identified to all of these dyes; it appeared that about one out of every 10^{10} RNA molecules could fold in such a way as to form a specific binding site for these rather arbitrarily chosen compounds. The capacity to generate so many diverse structures from polymers containing only four chemically similar subunits was surprising and raised questions about the minimal chemical requirements for the generation of stable structures. This difficult issue is only now being addressed by attempts to measure the frequency with which aptamers can be isolated from pools composed of only 2 or 3 different bases, as well as by attempts to expand the genetic alphabet and generate pools composed of 6 or 8 different bases (35–38). The latter approach also allows, in principle, for additional chemical functionalities to be exploited for the generation of functional nucleic acid structures.

The first RNA aptamer targeted to a small biomolecule was directed at ATP (39), one of the most important cofactors in contemporary metabolism. This aptamer was isolated by passing a random RNA pool over a column of immobilized ATP, washing away unbound sequences, and then eluting with free ATP. After six cycles of selection and amplification, ATP-binding sequences began to dominate the pool. Cloning and sequencing revealed a large number of highly divergent and thus independently derived sequences. Inspection of these sequences showed that they all possessed a common 11-nucleotide motif that always occurred within the same secondary structural context (a hairpin loop with an internal, asymmetric purine-rich bulge; see Figure 2a). Synthesis of this minimal structure confirmed that this domain was indeed responsible for the observed ATP binding, with a K_D of ~1 µM. Calculations based on cumulative enrichment and informational complexity suggest that these aptamers exist in random sequence at a frequency of about one in 10^{10}–10^{11}.

Not every aptamer selection reveals so many independent isolates of essentially the same structure. In some cases, for example, the selection for arginine or guanosine binders, many different independent sequences were isolated, but they turned out to represent several different structural classes (40, 41). In the case of the ATP aptamer, the observed structure must be much simpler (and therefore more common in sequence space) than the next more complex aptamer.

The large number of small-molecule aptamers that have been isolated over the past several years allows us to make some generalizations about both the selection process and the resulting aptamers. First, RNA is clearly capable of recognizing a large variety of small molecules. Successful selections have been performed on a diverse array of small targets, including other nucleotides and free nucleobases, amino acids, cofactors, basic antibiotics, and transition-state analogs (see Table 1). These include both planar and nonplanar compounds having overall negative or positive charges. Surprisingly, even molecules that are

a)
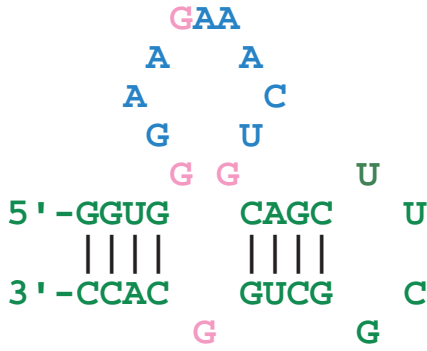
b)
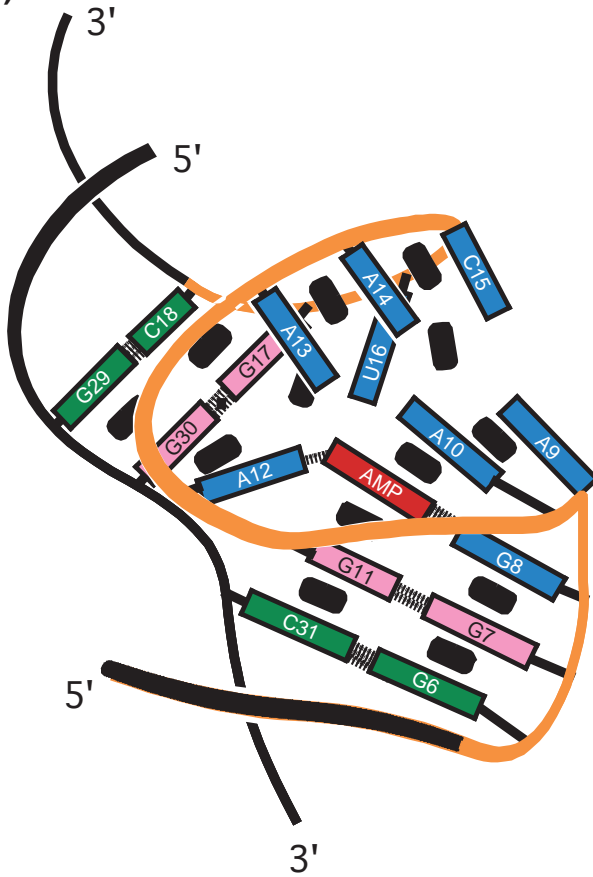
Figure 2   ATP aptamer structure. (*a*) Secondary structure diagram of the RNA sequence that binds ATP (39). (*b*) Three-dimensional structure of the ATP aptamer as determined by NMR spectroscopy (48–50). The color scheme is the same as in Figure 2*a*: the Watson-Crick stems are in *green*, the Gs that form the G-G base pairs are in *pink*, and the loop residues are in *blue*. The AMP ligand is in *red.* H-bonds are shown as *dotted lines* and stacking interactions as *solid rectangles*. Based on figure from Reference 62 (with kind permission from the authors and *Nature Structural Biology*).

largely hydrophobic, such as valine or tryptophan, can be recognized. Second, many of these RNA aptamers show impressive specificity. The theophylline aptamer, for example, discriminates against caffeine, which possesses one additional methyl group, by at least 10,000-fold (42). Third, it is also clear that different ligands present distinct levels of difficulty for recognition. This is reflected in the affinities of aptamers for their ligands (Table 1). Features that make ligands easy to recognize are planarity, presence of H-bond donors or acceptors, and positively charged groups. The highest affinities are observed for the aminoglycoside antibiotics, which possess multiple primary amines. Other high-affinity ligands are the (planar) nucleotides or nucleobases. The most difficult species to recognize appear to be nonplanar molecules with a neutral to negative overall charge or largely hydrophobic character.

Single-stranded DNA is also capable of recognizing a variety of small molecules, including organic dyes (27), ATP (43), porphyrins (44, 45), and arginine (46). As might be expected, most DNA aptamers do not function if they are converted into RNA, and vice versa. This is, presumably, due to the role of the 2′-OH in determining helical parameters and helix stability, and in contributing to tertiary interactions that stabilize aptamer structure and interact directly with the ligand. One exception is the G-quartet-based flavin aptamer, which, although selected from an RNA pool, binds with approximately equal affinity when composed of DNA (47).

What is the structural basis for ligand recognition by RNA and DNA aptamers? The simple aptamer structures tend to use purine-rich loops to recognize ligands. The bases of these loops engage in noncanonical base pairing interactions with each other to arrange the proper surfaces and H-bond donors and acceptors for ligand interaction. Often irregular chain topologies and across-helix stacking interactions stabilize the active conformation. These principles are illustrated by the following examples.

The solution structure of the ATP aptamer, as determined by NMR spectroscopy (48–50), is schematized in Figure 2*b*. It reveals how a small and apparently simple RNA secondary structure can form a precise binding pocket for its selected ligand. The purine-rich loop is highly ordered, and the trace of its backbone can be described by a Greek zeta (ζ–three consecutive turns). The two helices flanking the zeta are closed by mismatched G-G base pairs. This combination of irregular structural elements provides bases for H-bonding to and stacking against the adenine

TABLE 1   Aptamers for small molecules

Target	Estimated K_D (μm)	Reference
Nucleotides and nucleobases		
ATP/adenosine	1	39
ATP/adenosine (DNA)	6	43
Guanosine	32	41
Guanine/xanthine	1.8	184
7-Methyl-GTP	~0.5	185
Theophylline	0.11	42
Amino acids		
Arginine	0.33	61
Citrulline	62	58
Valine	12,000	186
Tryptophan	18	187
Cofactors		
Cyanocobalamin	0.09	188
N-methylmesoporphyrin IX	~14	44
N-methylmesoporphyrin IV (DNA)	~0.5	189
Flavin	0.5	190
NAD	2.5	47
RMP-botin	2	141

continued

portion of the ligand; other bases stack against the ribose moiety. About half of the ligand is buried within the binding cavity. This remarkable RNA conformation is stabilized by ligand binding; in the absence of ATP, this region is poorly ordered. Interestingly, the DNA aptamer to ATP (43) shows a very different sequence and secondary structure, but its NMR structure (51) reveals that many aspects of the actual binding site are similar. For example, in both cases the Watson-Crick face of the ATP H-bonds to the minor groove face of a G residue, and the ligand base stacks on top of a reversed Hoogsteen G-G base pair.

Another window into the surprisingly intricate nature of these aptamers comes from the NMR structure of the theophylline aptamer (52, 53). This RNA is a short hairpin with two small internal loops, one symmetric and one asymmetric. To form the binding pocket, one bulged strand forms an "S-turn" in which the direction of the chain reverses twice within a stretch of only five nucleotides (Figure 3). This results in several bases that make extensive stacking and H-bonding interactions with the theophylline, closing in around it to form a snug

TABLE 1   *continued*

Target	Estimated K_D (μm)	Reference
Antibiotics		
Tobramycin (aminoglycoside)	0.0008	55
Neomycin (aminoglycoside)	0.1	191
Lividomycin (aminoglycoside)	<0.2	192
Kanamycin (aminoglycoside)	<0.2	192
Streptomycin (aminocyclitol)	~1	193
Viomycin (basic peptide)	12	194
Chloramphenicol (small, neutral)	2.1	195
Transition state analogs		
Diels-Alder reaction	3,500	196
Bridged biphenyl isomerization	542	143
Other		
Dopamine	2.8	197
Peptide (substance P)	0.19	29
Divalent metals	~1	198

[a]Where multiple selections have been performed, this table lists only the highest-affinity case. All aptamers are RNA except where noted ("DNA"). RMP-biotin, ribose-monophospate-biotin, a carboxylate-phosphate anhydride of biotin and ribose-5-phosphate. K_D values were estimated by several different methods of variable accuracy and precision. Consult references for details.

binding pocket. The S-turn contortion is stabilized by two remarkable structures, a "1-3-2 stack," in which a base is stacked between the bases of the two preceding nucleotides, and a "base-zipper," in which bases from two different strands interdigitate so as to form a continuous stack.

Less dramatic structural arrangements can also result in high-affinity ligand-binding properties. A flavin-mononucleotide (FMN) aptamer structure, for instance, consists of a stem loop with an internal asymmetric purine-rich loop consisting of five nucleotides on one side and six on the other (54). This asymmetric loop is "zippered up" using base mismatches and one base triple without large backbone distortions. The planar tricyclic flavin moiety is stacked between a base triple and a G-G mismatch, and one edge of the chromophore H-bonds with the Hoogsteen face of an A. Another relatively simple structure is that of the tobramycin aptamer (55). The aminoglycoside antibiotic is bound by a hairpin loop with a single bulged A in the stem (56). The ligand resides in the major groove and is covered by a flap formed by a C residue from the loop (see also 57).

These structural studies have revealed that many of the nucleotides at the cores of aptamers play dual roles: they contact the ligand directly and, by participating in various mismatch, triplex, Watson-Crick, and other interactions with
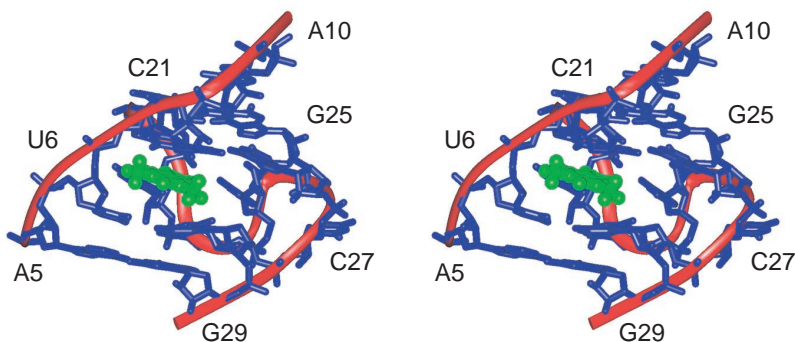
Figure 3   Stereo representation of the three-dimensional structure of the theo-phylline aptamer (42), as determined by NMR spectroscopy (52, 53). The RNA is shown in *blue*, with the backbone trace in *red*, respectively, and the theophylline is in *green*. Reproduced from Reference 52 with kind permission from the authors and *Chemistry & Biology*.

neighboring nucleotides in the aptamer, they determine the overall structure of the binding site. Such overlapping roles may make it difficult to generate altered specificities within the same structural context. This is to be contrasted with the case of protein-based receptors, in which the backbone architecture is mainly determined by the packing of the hydrophobic core; a largely distinct set of amino acid side chains, facing away from the core, establish the microanatomy of ligand binding sites. Nevertheless, an RNA originally selected to bind citrulline was later evolved to bind arginine, which differs by the replacement of a C=O in the urea group with a C=NH$_2$ to form the guanidino functionality (58). Three substitutions in the aptamer changed H-bond donors and acceptors so as to alter the specificity of the aptamer without causing major structural rearrangement. This aptamer appears to have an atypical structure, however, in that several of the bases that contact the ligand do not make extensive intramolecular interactions (59).

Structural models have also been reported for the citrulline and arginine RNA and DNA aptamers (59, 60). In addition, others aptamers apparently consist of pseudoknotted or G-quartet structures (19). It will be interesting to see whether more stringent selections for higher affinity ligand binding (61) will yield more complex aptamer structures.

For detailed reviews of aptamer structures, see elsewhere (62–64), and for a review of RNA structure in general, see Peter Moore's article in this volume.

Aptamers for Proteins

RNA or DNA must create a specific cavity to enclose and recognize a small mol-ecule. Proteins, however, possess extensive surfaces with ridges, grooves, pro-jections, and depressions, all decorated with numerous H-bond donors and

acceptors, suggesting that proteins should be excellent targets for aptamer selection. The first identification of a nucleic acid aptamer to a protein that does not normally interact with RNA or DNA was a single-stranded DNA aptamer to thrombin (65). The pool contained 60 random nucleotides flanked by constant regions, much like a typical RNA pool. A surprisingly high frequency (probably >1 in 10^5) of molecules bound to thrombin with nanomolar affinity. Analysis of the selected sequences uncovered a simple consensus, $GGNTGGN_{2-5}GGNTGG$, that was sufficient to inhibit thrombin activity and decrease the rate of blood clotting, with 50% inhibition at 25 nM aptamer (65, 66). The three-dimensional structure of the aptamer in solution has been determined by NMR (67, 68), and an aptamer-thrombin complex has been evaluated using X-ray crystallography (69). The DNA forms a G-quartet structure, but the details of the interaction with the protein are not completely clear (70).

Numerous selections of this type have now been reported, and it appears that it is possible to find high-affinity nucleic acid aptamers for most proteins. This has made in vitro selection of aptamers an effective method (71–74) for generating reagents that interfere with the activities of extracellular proteins (proteases, cell-surface receptors and their ligands, autoreactive antibodies) implicated in diverse pathological conditions. Over 30 "generic" proteins (i.e. those that are not believed to bind nucleic acids in vivo) have been used to generate RNA aptamers, with K_D values for their targets ranging from 50 pM to 50 nM (6, 19). This affinity is comparable to or greater than that typical of Fab fragments obtained through immunization. The frequency of these aptamers in random sequence is generally 10^{-10}–10^{-14}, depending on the target. This has raised the possibility that functional interactions between RNA and "generic" proteins may be more common in vivo than is presently appreciated (6, 75–77).

Aptamers that bind to and interfere with the action of proteins implicated in pathological conditions could potentially be used as pharmaceutical agents. One of the chief obstacles to this approach is the instability of RNA in blood due to ribonucleases. The rate of RNA degradation can be drastically reduced by replacing the 2′ hydroxyls of pyrimidines with amino or fluoro groups (78), and this modification is accepted by T7 RNA polymerase and avian myeloblastosis virus reverse transcriptase (79). Replacing all 2′-hydroxyls with other moieties, of course, tends to interfere with the properties of the aptamer, so the preferred method is so-called "front-loaded SELEX," in which the selection is performed using a pool already bearing all the modifications (19). Several aptamers selected from such libraries have been reported (79–82), and they display half-lives of up to 80 h in serum (79, 82). Such 2′-modified aptamers against basic fibroblast growth factor (83), immunoglobulin E (IgE) (84), interferon-γ (82), and keratinocyte growth factor (81) have been reported to inhibit cultured cells from undergoing normal physiological responses to these ligands. In addition, a nuclease-stabilized DNA aptamer to L-selectin was recently reported to inhibit L-selectin–dependent lymphocyte trafficking when injected intravenously into mice (85).

Other limitations of RNA as drugs, such as the high cost of synthesis and rapid renal clearance from the circulation, are also active areas of research (6, 86). Considerable stability to renal clearance has been accomplished by conjugating polyethylene glycol (PEG) to the RNA, or by covalently attaching dialkylglycerol to the RNA and then incorporating this conjugate into liposomes (87). A dialkylglycerol-conjugated, liposome-anchored vascular endothelial cell growth factor (VEGF)-binding RNA reduced VEGF-induced vascular permeability and angiogenesis in animals (87). The same aptamer shows antitumor efficacy in a human rhabdomyosarcoma nude mouse xenograft model (N Janjic & D Emerson, personal communication).

An alternative strategy for avoiding degradation is to invert all the chiral centers in the nucleic acid, thus creating a mirror-image structure that is resistant to blood and tissue-borne nucleases, which are enantiospecific. In practice, this has been accomplished by first making the enantiomer of the desired target and then selecting for aptamers from a pool of normal (*D*) RNA or DNA. Then a successful clone is synthesized as the L-RNA or DNA equivalent, which then recognizes the ligand with the natural chirality (88–90). Recently, Williams et al (88) used this approach to create an L-DNA aptamer for the cyclic L-peptide hormone vasopressin. This aptamer showed no signs of degradation in human serum after several days and interfered with vasopressin-mediated cyclic AMP synthesis in a cell culture assay.

Another interesting approach to using selection technology to inhibit physiological processes is referred to as "blended SELEX" (19). In this case, a relatively low affinity or specificity irreversible inhibitor of a target enzyme is linked to the random nucleic acid pool, and those sequences that facilitate the reaction between the inhibitor and a target enzyme are selected. This approach was used to enhance the reaction between a protease (neutrophil elastase) and a valyl phosphonate covalent inhibitor (91). The inhibitor was covalently linked to a DNA oligonucleotide that was in turn annealed to the constant region of a single-stranded DNA pool. This complex was then incubated with the protease, and ternary complexes of the pool molecules, DNA inhibitor, and protein were separated from protein-free pool DNA under conditions that denature proteins but not nucleic acids. The selection resulted in molecules that catalyzed the irreversible inactivation of the protease by ~100-fold over the starting pool plus inhibitor. One of these clones appears to inhibit rat neutrophil elastase in vivo, and consequently reduces neutrophil influx and lung injury in an acute inflammation model (92).

RIBOZYME SELECTIONS

A variation on the aptamer selection strategies can be used to isolate catalytic nucleic acid sequences. Such approaches have been used to change the function of known ribozymes, to create new ones, and to understand ribozyme structures, catalytic mechanisms, and even folding pathways (93–103). The excitement gen-

erated by the RNA world hypothesis has stimulated many experiments to determine the range of reactions that RNA can catalyze. There is also a growing recognition of the value of understanding the ways that a population of RNA sequences changes in response to selection pressures.

Because catalytic RNA structures are often significantly more complex than simple aptamers, it is less likely that an optimal catalyst will be isolated directly from a random pool. This has led to the now widespread use of mutagenic PCR amplification (96, 104) between rounds of selection to explore regions of sequence space surrounding those widely separated points sampled by the initial pool.

Catalysis of Reactions at Phosphate Centers

The majority of ribozyme selections have targeted phosphoester transfer reactions, mainly for technical reasons. In general, for catalytically active nucleic acids to be selected out of pools of largely inactive sequences, the reactions they catalyze must result in some kind of self-modification that results in the active molecules becoming selectable. This may seem paradoxical, as a true catalyst must accelerate the rate of a reaction without itself being consumed or otherwise modified. However, in practice, it is usually possible to convert a self-modifying RNA into a true catalyst by separating it into two distinct parts (which often interact by base pairing)—one part is then considered the catalyst and the other, the substrate (105). Multiple cycles of catalysis are then accomplished through dissociation of the product and subsequent binding of a new substrate.

One modification that provides an especially powerful selective handle is the addition of a particular nucleic acid sequence to the active pool members. This allows molecules that have catalyzed the reaction to be selectively retained on a column or to be copied in a selective reverse transcription or PCR reaction. This principle was validated by Robertson & Joyce (106), who made a pool of six variants of the *Tetrahymena* group I intron that could "tag" their 3′ ends by performing a transesterification reaction with a separate RNA fragment. This mixture was subjected to a new selective pressure in which a DNA substrate replaced the RNA substrate. After the introns were incubated with the substrate, reverse transcription was carried out using a primer that annealed to the tag sequence, thus ensuring that only those molecules that catalyzed the reaction could be copied into cDNA. This scheme allowed for the selective amplification of the most active variant in the pool.

A similar scheme was used in the first selection of ribozymes from completely random sequence (97). A pool of ~10^{15} different RNAs with 220 random nucleotides (N_{220}) was prepared by transcription from a DNA template, thus resulting in molecules with a triphosphate at the 5′ end (Figure 4). An attached substrate-docking domain aligned the 3′ end of a substrate RNA oligonucleotide with the 5′ end of the pool molecules. The 5′ triphosphate group served to activate the α-phosphate for nucleophilic attack by the 2′ or 3′ hydroxyl at the 3′ terminus of the substrate. The rare sequences that could catalyze this reaction

converted the interrupted hairpin loop into a continuous one, thus covalently tagging themselves. The tagged molecules were purified on a column derivatized with an oligo that was complementary to the tag sequence. Then, after reverse transcription, they were selectively amplified by PCR. This dual-selection procedure resulted in an extremely powerful selective force, so that only three rounds of selection and amplification were required before the population was composed mainly of active ligases. At that point, the pool consisted of ~100 different sequences, suggesting that about one in every 10^{13} random RNA molecules was an RNA ligase. In the three subsequent rounds of selection, mutagenic amplification was used to introduce new variation into the population, with the result that some species increased dramatically in abundance, while others declined to extinction. This pattern of new species becoming dominant while others were driven to extinction continued in subsequent cycles of selection, dramatically illustrating evolutionary forces in action at the level of molecules.

Analysis of the sequences surviving after 10 rounds of selection revealed three structurally distinct classes of ribozymes. The class I enzyme catalyzes the formation of 3′–5′ phosphodiester bonds, whereas the class II and III enzymes catalyze 2′–5′ bond formation (107; class I secondary structure shown in Figure 8). Further optimization of the class I ligase resulted in an RNA with an observed ligation rate constant of >1 per s—faster than the protein enzyme DNA ligase (107, 108). This corresponds to a rate acceleration over the templated background reaction of $>10^9$, which is comparable to that obtained by many protein enzymes.

Three main conclusions can be drawn from this selection. First, ribozymes can be derived from a relatively limited sampling of sequence space (only ~10^{15} sequences of the possible 10^{132} sequences in N_{220} space were present in the original pool). In other words, it may not be dramatically more difficult to make an enzyme than it is to make an aptamer. Second, there are multiple solutions to the same problem in that at least three apparently unrelated structures emerged from the selection. Third, two of these structures are simple enough that multiple independent examples were isolated from the same starting pool. In contrast, the class I ligase is so complex that the chance of finding it in a pool of only 10^{15} molecules was estimated to be 0.1% (108). One interpretation of this result is that many possible complex structures may exist, so that although the likelihood of accessing any particular one is very low, the likelihood of finding some complex structure is reasonably high.

In vitro selection has been used to isolate many other novel ribozymes that perform a variety of phosphoester transfer or hydrolysis reactions (Figure 5). The nucleophiles include water and 2′, 3′, and 5′ hydroxyls, as well as phosphate oxygens; the attacked phosphate can be part of a phosphorimidazolide, a phosphodiester, a pyrophosphate, or the α or γ phosphate of a triphosphate moiety. Most of these selections relied on the covalent addition of a nucleic acid sequence to active pool molecules, as discussed above. Two exceptions to this were the selection of a polynucleotide kinase and of a phosphatase. The polynucleotide kinase was selected by mixing ATP-γ-S with a random RNA pool containing a doped ATP
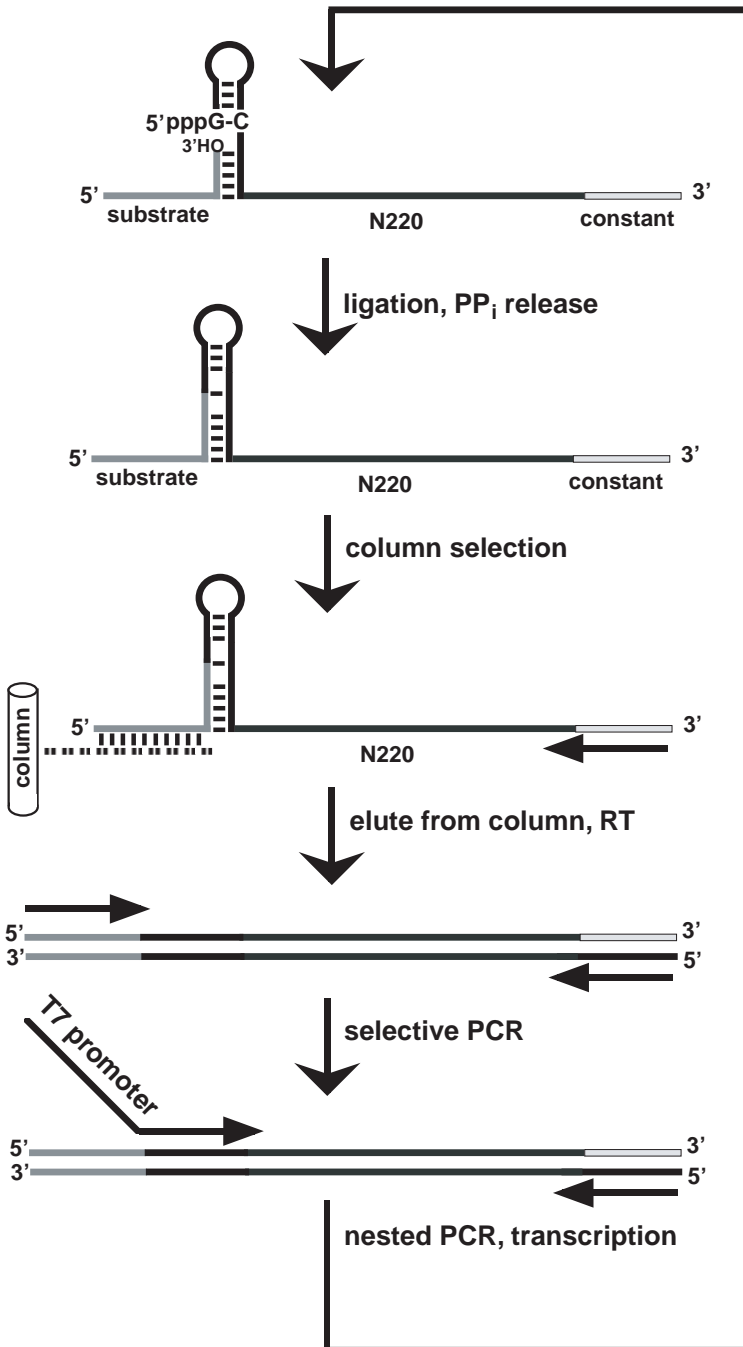
Figure 4   RNA ligase ribozyme selection scheme (97).
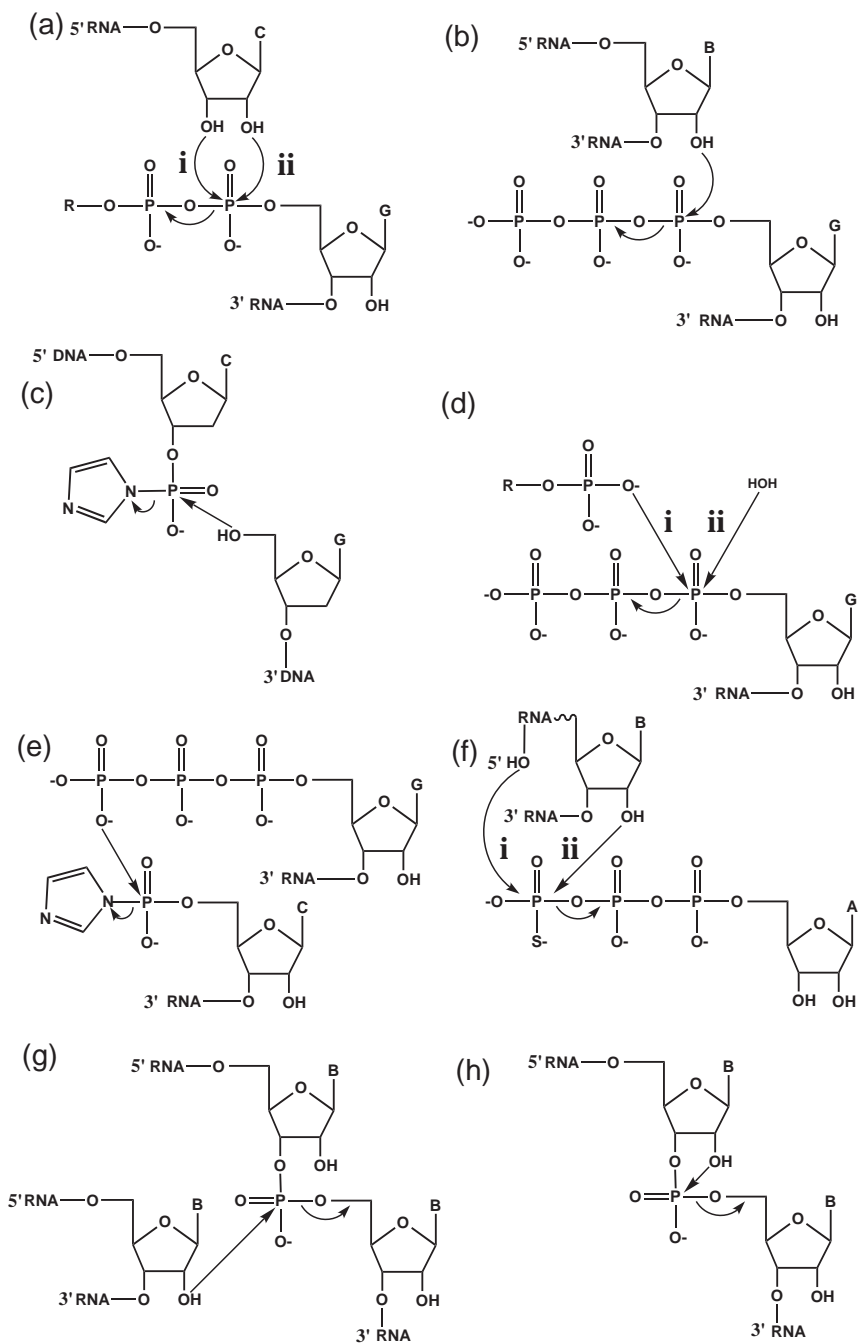
Figure 5   Phosphoester transfer and hydrolysis reactions catalyzed by in vitro-selected ribozymes and deoxyribozymes. (*a*) Either a (*i*) 3′ or (*ii*) 2′ hydroxyl at the 3′ end of an oligomer attacks the α phosphate of either a 5′ pppG (97, 107; R = phosphate) or 5′-5′–linked AppG (102; R = adenosine) motif, thereby forming a new phosphodiester bond and releasing pyrophosphate or AMP. (*b*) same as Figure 5*a*(*ii*) except that an internal 2′ hydroxyl is the nucleophile, so that a T structure is formed (103). (*c*) A ligation reaction catalyzed by a deoxyribozyme, in which a 5′ hydroxyl attacks a 3′ phosphorimidazolide, generating a 5′–3′ phosphodiester bond and releasing imidazole (128). (*d*) Either a (*i*) phosphorylated compound or (*ii*) water attacks the α phosphate of the pppG motif at the 5′ end of an RNA. R can be a variety of different groups, including orthophosphate, pyrophosphate, RNA, or various cofactors (111, 113). (*e*) The 5′ γ phosphate of one RNA attacks a phosphorimidazolide at the 5′ end of another RNA, to form a 5′-5′ tetraphosphate linkage (100). (*f*) Either the (*i*) 5′ hydroxyl or an (*ii*) internal 2′ hydroxyl of the ribozyme attacks the γ thiophosphate of ATP, thus thiophosphorylating itself (109, 110). (*g*) An RNA internal 2′ hydroxyl attacks an internal phosphodiester bond of another RNA, forming a 2′–3′ phosphodiester bond and releasing a 5′-OH RNA (103). (*h*) RNA cleavage mediated by attack of an internal 2′ hydroxyl on the adjacent 3′–5′ phosphodiester bond. This is the cleavage mechanism used by most of the selected RNA cleavers, as well as by most natural self-cleaving domains.

aptamer (109, 110). Pool molecules that could transfer the γ thiophosphate from ATP-γ -S onto one of their hydroxyl groups were then selectively retained on a thiophilic column. The selection yielded both 5′ and 2′ kinases (Figure 5*f*).

An RNA 5′ phosphatase was selected (111) by transcribing pool RNAs with T7 RNA polymerase, thus placing a triphosphate at the 5′ end. Those RNAs that were capable of removing the β and γ (but not α) phosphates then became substrates for T4 RNA ligase, which circularizes RNAs with 5′ monophosphates and 3′ hydroxyls. Circular RNA was then separated from linear by gel electrophoresis (112). The 5′ phosphatase ribozyme that emerged from this selection had the surprising quality that the selected nucleophile, water, could be replaced by phosphate or various phosphate-containing compounds, including some interesting metabolic cofactors (111, 113; Figure 5*d*).

Finally, numerous selections have identified RNA-cleaving ribozymes (Figure 5*h*). The first selection for this activity took advantage of the fortuitous self-cleaving acitivity of yeast tRNA[phe] in the presence of Pb[2+]. Pan & Uhlenbeck (112) created a pool of partially randomized tRNA[phe] sequences and circularized them with T4 RNA ligase. Those pool molecules that self-cleaved in the presence of Pb[2+] linearized themselves, thus allowing for their isolation from inactive, circular sequences on a polyacrylamide gel. Purified linear RNAs were then recircularized and amplified for the next round of selection. Some of the clones self-cleaved ~10-fold faster then tRNA[phe], and most cleaved at different sites than the parent molecule, suggesting that many distinct structures can accelerate this reaction. Several of these selections have aimed at characterizing or changing the activity

of natural RNA-cleaving ribozymes (for review, see 5), including the hammer-head RNase P (114, 115), the hairpin (116, 117), and the delta ribozyme (118). Novel RNA-cleaving ribozymes have also been selected (119), including some that function under the unusual conditions of low pH and the absence of divalent metals (120).

Selections have also been used to demonstrate that DNA can fold into cat-alytic structures (reviewed in 121). The first deoxyribozyme to be selected cleaved a single ribonucleotide phosphodiester bond embedded in a DNA mole-cule (122). Isolation of this catalyst was accomplished by making a single-stranded DNA pool with a 5′ biotin and a single internal ribonucleotide near the 5′ end; DNAs that could cut themselves off of a streptavidin column in the pres-ence of lead were selected. Later, similar techniques were used to find deoxyri-bozymes that perform the same reaction but in the presence of Mg^{2+} (123–125) or Ca^{2+} (126) or in the absence of any divalent metals (13, 125). Recently, deoxyribozymes that catalyze the same RNA cleavage reaction were selected that are divalent-metal-independent and use free histidine as a cofactor (127). The same type of selection scheme has recently lead to the development of a "generic" deoxyribozyme that appears capable of cleaving almost any RNA sequence, depending on the sequence of the substrate-binding arms of the enzyme (124). This deoxyribozyme consists of a 15-nucleotide loop in the center of the DNA enzyme-RNA substrate duplex and works with a k_{cat} of >1 min^{-1} in the presence of 50 mM Mg^{2+} ions. This surprisingly simple structure could potentially provide a general means of cutting almost any RNA sequence.

A DNA ligase deoxyribozyme has also been selected (128). In the presence of Cu^{2+} or Zn^{2+}, it facilitates the attack of a DNA 5′ hydroxyl on a phosphori-midazolide at the 3′ end of another DNA to make a 5′–3′ phosphodiester bond (Figure 5*c*).

Catalysis of Reactions at Non-phosphate Centers

The chemical reactions catalyzed by all known natural ribozymes are centered around phosphate groups (RNA cleavage and ligation, phosphotransfer, phos-phoester hydrolysis). The possibility that the chemical versatility of RNA might be limited to reactions involving phosphate was justified by invoking the general dependence of phosphotransfer and hydrolysis reactions, both in simple chemi-cal systems and protein-catalyzed reactions, on divalent metal ions, especially Mg^{2+}. Because the RNA backbone can be densely coated with Mg^{2+} ions, it seemed plausible that RNA would be especially suited for catalyzing phosphate chemistry. In vitro selections have shown, however, that nucleic acids can accel-erate several other types of reactions (Figures 6 and 7).

The first selection for a ribozyme that catalyzed a chemical reaction other than phosphate transfer or hydrolysis was for a self-alkylating RNA (129). To accom-plish this, a biotin-binding RNA aptamer was doped with random sequence, and this pool was mixed with an iodoacetyl derivative of biotin. Those molecules that could alkylate (and therefore biotinylate) themselves were then selected on immo-
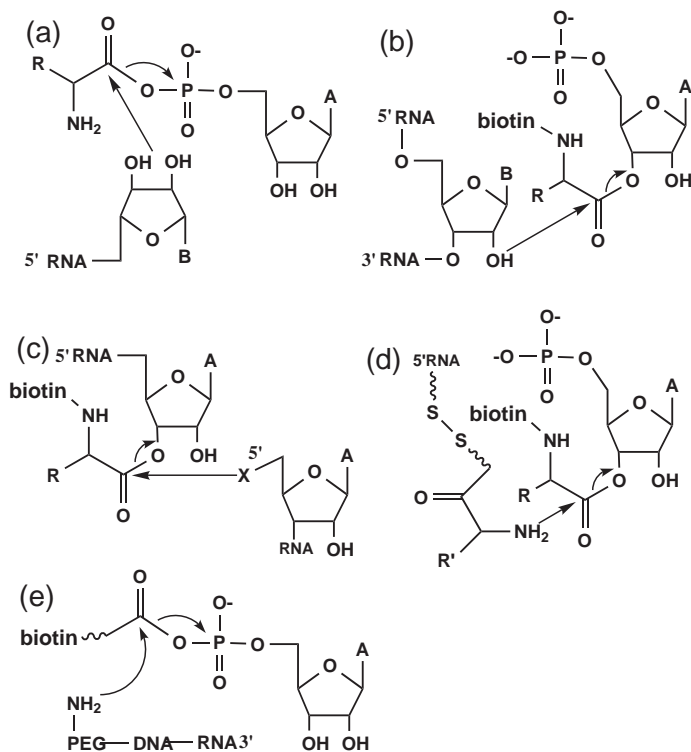
Figure 6 Acyl transfer reactions catalyzed by in vitro-selected ribozymes. (*a*) The 2′ (3′) hydroxyl of the ribozyme attacks the carbonyl carbon of an amino acid activated by AMP through a mixed carboxylate-phosphate anhydride linkage [130, 131; R = CH$_2$(C$_6$H$_5$)]. (*b*) An internal RNA 2′ hydroxyl attacks an aminoacyl ester of biotinylated phenylalanine and AMP, linked through the 2′ (3′) hydroxyl of AMP [199; R = CH$_2$(C$_6$H$_5$)]. (*c*) Aminoacyl transfer of an amino acid (N-biotinylated methionine) from the 2′ (3′) end of a donor RNA to the 5′ hydroxyl (X = OH) or amino (X = NH$_2$) group of the ribozyme (132, 133; R = CH$_2$CH$_2$SCH$_3$). When X = NH$_2$, an amide bond is formed. (*d*) Peptide bond formation (134). One amino acid is covalently tethered to the ribozyme. Its amino group attacks an aminoacyl adenylate to form a peptide bond between the two amino acids, releasing AMP [R = CH$_2$CH$_2$ S CH$_3$; R′ = CH$_2$(C$_6$H$_5$)]. (*e*) A primary amine at the end of a long PEG linker attacks the carbonyl carbon of a biotin-AMP carboxylate-phosphate mixed anhydride, forming an amide bond with biotin and releasing AMP (141).

bilized streptavidin (Figure 7*a*). The ribozyme that emerged from this selection catalyzed the alkylation of an N7 on an internal G residue with a k_{cat} of 0.05 min^{-1} and a K_M of ~1 mM, for a rate enhancement of ~3 x10^6 over the uncatalyzed rate.

Ribozymes that catalyze several other chemical transformations have since been identified through in vitro selection (Figures 6 and 7). The first aminoacyl
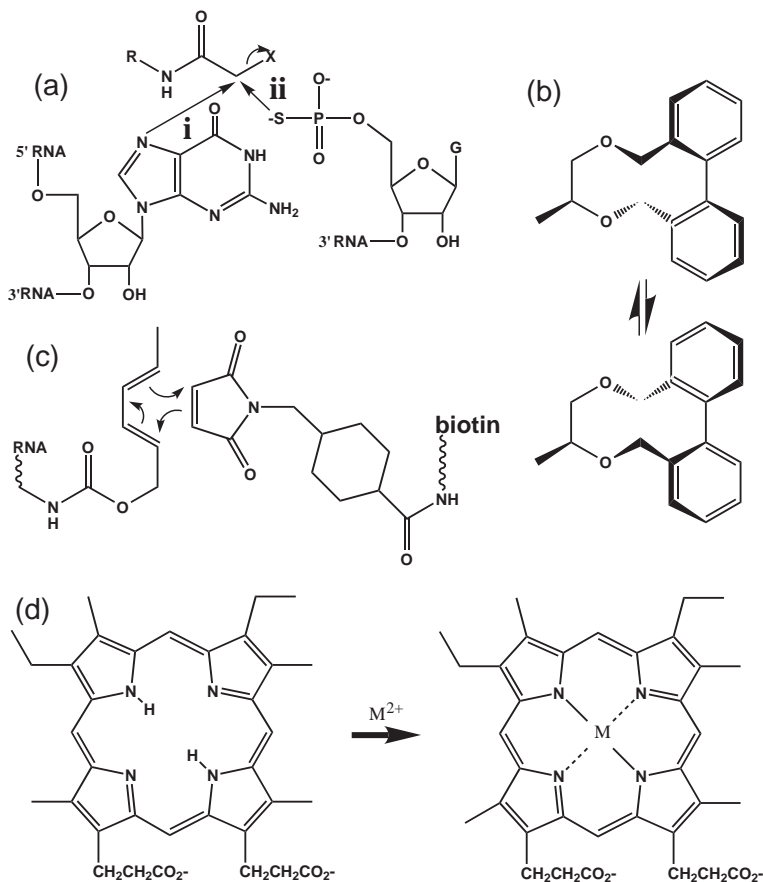
Figure 7   Additional chemical reactions accelerated by ribozymes or deoxyribozymes. (*a*) N- and S-alkylation: (*i*) The N7 of an internal G residue of the ribozyme attacks N-biotinoyl-N′-iodoacetyl-ethylenediamine, releasing I$^-$ (129); (*ii*) a thiophosphate at the 5′ end of the ribozyme attacks N-bromoacetyl-bradykinin, releasing Br$^-$ (200; bradykinin is a 9-amino-acid peptide). (*b*) Rotation of a sterically hindered C-C bond catalyzed by an RNA (143). (*c*) Diels-Alder reaction between an acyclic diene tethered to the 5′ end of the ribozyme and a biotinylated maleimide dienophile (140). (*d*) DNA- (45, 144) or RNA- (44) catalyzed insertion of Cu^{2+} (or Zn^{2+}) into mesoporphyrin IX. Certain RNA sequences have also been shown to be capable of cleaving amide bonds (165, 166) and aminoacyl esters (201). Finally, a deoxyribozyme that oxidatively cleaves DNA has also been reported (202, 203). See also Figure 10.

transferase was selected by incubating phenylalanyl-AMP (a mixed carboxylate-phosphate anhydride) with pool RNA (Figure 6*a*). RNA molecules that transferred the amino acid to one of their own hydroxyl groups could be distinguished by the presence of the α-amino group of phenylalanine; this amino group was made to

react with an N-hydroxysuccinimide ester of naphthoxyacetic acid, so that the aminoacylated RNAs were tagged with a hydrophobic group that caused retention on a reverse-phase column (130). The ribozyme thus selected transfers the amino acid from AMP to the 2′ (3′) hydroxyl at its own 3′ end, making this reaction analogous to the final step in the charging of tRNAs by aminoacyl-tRNA-synthetases. The enzyme recognizes only the AMP moiety of the substrate, and so it accepts other adenylated amino acids (131).

The key reaction in protein synthesis is peptide bond formation. Catalysis of amide bond formation was demonstrated for a ribozyme selected to transfer an N-biotinylated methionine from the 2′(3′) hydroxyl of a donor RNA oligonucleotide onto its own 5′ hydroxyl group (Figure 6c; 132, 133). Subsequent replacement of the 5′ hydroxyl nucleophile by an amine showed that the same enzyme could also form an amide bond. Catalysis of actual peptide bond formation between two amino acids was recently demonstrated by Zhang & Cech (134). To accomplish this, the carboxyl group of phenylalanine was covalently linked to the 5′ end of the pool RNA through a disulfide-containing linker, leaving the α-amino group free to react (Figure 6d). This H_2N-Phe-SS-RNA conjugate was then mixed with AMP-Met-Bio, a simple analog of initiator formyl Met-tRNA in which the carboxyl group of the N-biotinylated methionine is linked to the 2′ (3′) hydroxyl of AMP. Selection was for RNAs that catalyze the attack of the α-amino group of the tethered phenylalanine on the activated carbonyl of AMP-Met-Bio. Molecules that performed this reaction were isolated by binding to immobilized streptavidin. After washing away unreacted molecules, covalently bound RNAs were released by reducing the Phe-SS-RNA bond with dithiothreitol. The most successful molecule from this selection catalyzed peptide bond formation with a k_{cat} of 0.05 min^{-1} and a K_M for the AMP-Met-Bio of 190 μM. Ribozyme binding to the AMP-Met-Bio was mediated through the AMP portion, so other adenylated amino acids could substitute for Met. These results demonstrate that RNA alone can catalyze the chemical reaction performed by the ribosome; the question of whether the peptidyl transferase activity of the ribosome can be entirely RNA mediated is still under investigation (135–138).

This strategy for selecting ribozymes that catalyze the covalent joining of small molecules by tethering one of them to the pool RNA has also been described in three other reports (139–141). A Diels-Alder cycloaddition-catalyzing RNA was selected by covalently linking a dienophile to the RNA and incubating with a free diene (140), affording the first example of RNA-catalyzed carbon-carbon bond formation (Figure 7c). In another case an RNA pool with a primary amino group at the end of a long PEG tether was incubated with a mixed carboxylate-phosphate anhydride of biotin and AMP (141). RNAs that catalyzed attack of the amine on the activated carboxyl, to form an amide bond and release AMP, were isolated based on streptavidin binding (Figure 6e). The emerging ribozyme catalyzed amide bond formation but, unlike the amide-bond forming RNA described above (132), did not use base pairing to recognize either the nucleophile or the activated carboxylate. The generality of this technique should produce many interesting ribozymes in the future. Another similarity between

these two selections is that both were performed with RNA pools in which all of the uridines were modified at the 5 position to include additional functionalities [pyridylmethylcarboxamide (140) or imidazole (141)]. This way of broadening the chemical and therefore functional diversity of RNA has great potential for future applications, as discussed below.

RNA and DNA enzymes have also been discovered via a fundamentally different route that involves the isolation of aptamers to a transition state analog for a reaction, followed by screening of these aptamers for catalytic activity. This approach has been highly successful for producing catalytic antibodies (142), but fewer such achievements have been reported for nucleic acids, perhaps because few transition-state analogs have been designed with optimal binding to RNA in mind. The few selections that have succeeded have used transition-state analogs that are geometrically distorted from the ground-state molecules. By selecting an RNA aptamer for a planar transition-state analog for the rotation of a sterically hindered carbon-carbon bond, Prudent et al (143) found an RNA that catalyzed the rotation by ~100-fold (Figure 7*b*). More recently, RNA (44) and DNA (45, 144) enzymes that insert metals into porphyrins (Figure 7*d*) have been isolated after selection for aptamers to alkylated porphyrins, which are distorted from the usual porphyrin geometry so as to resemble the transition state. In at least one case, the rate enhancement appears to be accomplished by increasing the basicity of the bound porphyrin substrate, possibly by deforming it away from its normally planar conformation (145).

CURRENT GOALS AND FUTURE PROSPECTS

What fundamental questions can be most usefully addressed with in vitro selection technology? We discuss three broad areas. First, there are basic questions concerning the relationship between functionally selected RNAs and the populations from which they arise (either random sequence populations or mutagenized pools centered around RNAs with other functions). Second, there is the continuing evaluation of the plausibility of the RNA world hypothesis. Third, we would like to understand the advantages and disadvantages of RNAs versus proteins for building catalysts. Another important and very active area of research— the use of aptamers (and possibly ribozymes) as drugs, diagnostics, etc—has already been discussed above.

Evolutionary Pathways Leading to Functional Sequences

The majority of the selections performed to date have focused on the evolutionary outcome rather than the process itself. Ribozymes can arise either directly from pools of random sequence or through mutation of existing sequences. In both cases, there are important questions about the nature of the evolving populations that can be addressed by in vitro selection and directed evolution. Evolu-

tionary studies of evolving populations of molecules have the particular advantage that sequences from the entire history of the evolutionary trajectory are available for detailed retrospective analysis.

Random sequences are the greatest possible source of diversity, and random pools presumably model the earliest stages of RNA evolution in which random condensation of activated mononucleotides generated RNA populations. Random sequence pools allow us to ask what the probability is of finding a sequence with a given level of activity, and how this is affected by the length and composition of the random sequence (97, 108, 146). In the most basic sense, such selections allow us to determine the amount of information required to specify a given level of rate acceleration (or binding free energy for aptamers). Such information content cannot simply be evaluated by calculating the likelihood of a given successful ribozyme sequence arising from a random string of G, A, U, and C nucleotides because (*a*) many positions of a successful ribozyme are not important for activity (and therefore contain no relevant information); (*b*) there are many distinct sequences that form essentially the same structure, for example a particular helical segment within a ribozyme may be replaceable by many other helix-forming sequences; and (*c*) there may be many distinct structures that can satisfy the same selective pressure. In the case of the RNA ligases selected from an N_{220} random sequence pool (107), at least three apparently unrelated structural classes emerged.

How do ribozymes and aptamers mutate so as to change their specificity or the reactions they catalyze? How many mutations are required to accomplish a change of specificity? Will the functional changes be accommodated using the same structural context, or will extensive reconstruction be required? There is very little data relating to these questions. As discussed above, a citrulline aptamer was changed by three substitutions to an arginine aptamer, while maintaining essentially the same structure (58, 147). The *Tetrahymena* group I intron has been evolved to use a DNA (as opposed to the usual RNA) substrate, or to use Ca^{2+} (as opposed to Mg^{2+}) as a cofactor (96, 98). These latter two adaptations also arose from a small number of mutations, and in both cases, there appeared to be alternative, mutually exclusive substitutions that could satisfy the selection criteria.

In some cases, functional changes may require a new structural context. In RNA, a small number of mutations can give rise to a new secondary structure by destabilizing the parental fold and stabilizing one of the numerous competing ones. Any single RNA sequence probably has, in addition to its preferred structure, a "neighborhood" of secondary structures that become favorable upon 1–2 nucleotide substitutions (148). Also, because many sequences can form the same secondary structure, an RNA sequence can neutrally drift great distances through sequence space without significantly compromising its structure (and function), thus exploring new neighborhoods of secondary (and hence tertiary) structure space (149, 150). It will be instructive to conduct experiments designed to map the trajectories leading from one functional RNA to another.

In considering these trajectories, it is important not to ignore interactions between members of the population. In a recent selection for group I intron vari-

ants that could tag themselves with a DNA substrate, some of the most successful clones dispensed with an essential segment of the catalytic core, and survived by a kind of "parasitism" whereby they were acted upon in *trans* by intact, catalytically competent members of the population (151). Also, there may be important recombination events between pool molecules (e.g. occurring during PCR) that lead to improved fitness.

The fitness of molecules in an in vitro selection is a function not only of their catalytic activity but also of their relative efficiency of enzyme-mediated replication, as well as of the kinetic and thermodynamic parameters of folding. Depending on how the selection is executed, different steps, including folding, substrate binding, or catalysis, can become rate limiting and therefore the main targets of the selection. Treiber et al (152) recently designed a selection to isolate rapid folding variants of the *Tetrahymena* group I intron ribozyme. The P4–P6 domain of this intron folds rapidly (~60 min^{-1}), whereas the P3–P7 domain takes ~1 min to fold into the native conformation (153, 154). To isolate rapidly folding variants, a pool of mutagenized introns was briefly (5 s) exposed to folding conditions, after which unfolded molecules were selectively degraded. An additional step ensured that all the selected molecules were still catalytically active. After nine rounds of selection, clones that had three- to fivefold-faster folding rates were identified. Surprisingly, these clones all had mutations that destabilized the P4–P6 domain, implying that kinetically trapped, partially folded intermediates impede formation of the final native structure. Such a conclusion was supported by the fact that other mutations known to destabilize the P4–P6 domain also increased the overall folding rate, as did the addition of the denaturant urea.

Studies of the pathways leading to evolutionary adaptation may be facilitated by the "continuous in vitro evolution" strategy of Wright & Joyce (155). In this system (Figure 8), catalytic selection and amplification occur in the same reaction vessel, allowing many generations of selection to be conducted with maximal speed and minimal effort. To develop such a system, the substrate-binding site of the class I RNA ligase of Ekland et al (107) was modified so that it would accept an oligonucleotide substrate possessing the T7 RNA polymerase promoter sequence. This way, the most active ribozymes would tag themselves with a T7 promoter (Figure 8*a*). The reaction mixture also included reverse transcriptase (RT), all four deoxynucleoside triphosphates (dNTPs) and an RT primer that annealed to the 3' end of the pool molecules (Figure 8*b*). Therefore, after cDNA synthesis (Figure 8*c*), molecules that performed the ligation would possess a double-stranded T7 promoter, which would then direct another round of RNA synthesis by T7 RNA polymerase and nucleotide triphosphates (NTPs) (Figure 8*d*). In this way, all steps of the selection and amplification occurred simultaneously in the same vessel with no need for human intervention. At the point when the system became saturated with RNA, a small aliquot of this mixture was diluted into a fresh tube containing all the enzymes and reagents but lacking template. This process could be continued indefinitely. Although a typical ribozyme selection requires several days per round, Wright & Joyce performed 300 rounds
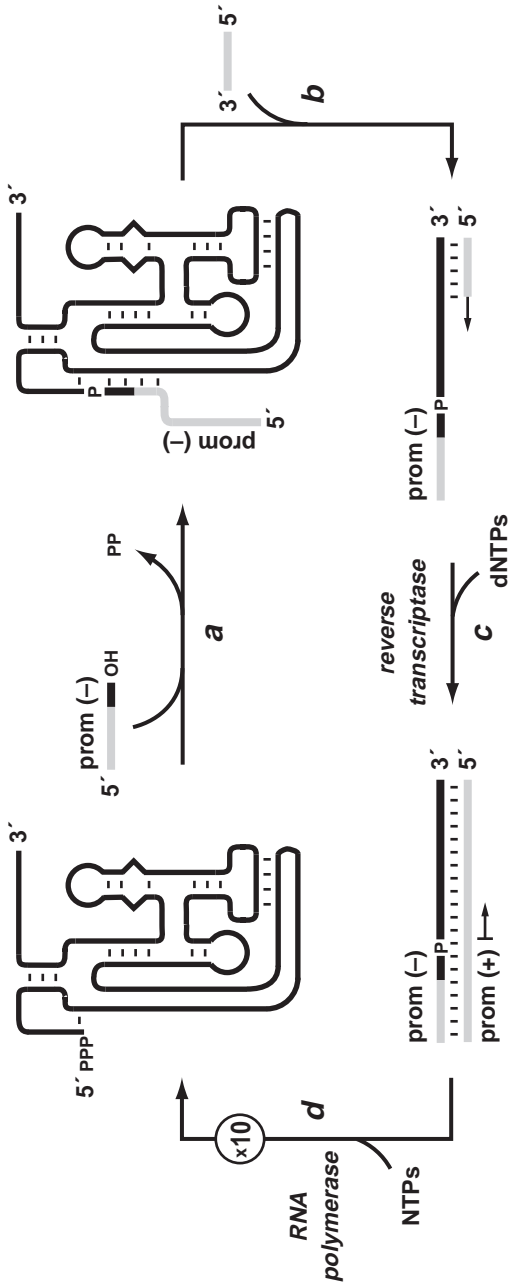
Figure 8 Continuous in vitro selection scheme of Wright & Joyce (155). The secondary structure of the class I ligase (107) is shown. The selection procedure is described in the text. Reprinted from Reference 155 with kind permission from the authors. Copyright 1997 American Association for the Advancement of Science.

in only 2 days by using this approach. With such a large number of selection rounds, even very small differences in ribozyme activity can give rise to dramatic changes in population distributions.

These "serial dilution" experiments are very analogous to the original selection experiments of Spiegelman and colleagues (see above), except that a catalytic property of the RNA molecules, rather than just their replication speed, is operated on by the selection. Spiegelman perhaps anticipated this when he wrote, almost 30 years ago: "Indeed, one might be able to persuade the RNA molecules themselves to acquire a weak catalytic function that would aid their replication" (22). The catalytic activities that can be selected for in a continuous in vitro evolution strategy are restricted to those that can in some way be linked to the replication cycle. It remains to be seen to what kinds of imaginative uses this system can be applied.

Exploring the RNA World Hypothesis

An experimental approach to history may seem paradoxical, but it has a long (and controversial) tradition in such diverse fields as archaeology, anthropology, geology, and evolutionary biology. In taking this approach, we are limited to testing the plausibility of scenarios for the origin and early evolution of life on Earth; nevertheless, being able to present even one plausible scenario would currently be a great advance.

There is now an impressive inventory of chemical activities catalyzed by nucleic acids (Figures 5–7), which is important given the variety of reaction types required for even the most rudimentary metabolism. However, there are two activities in particular that RNA must be shown to be capable of catalyzing if the RNA world hypothesis is to remain afloat. RNA must be able to catalyze self-replication and must be able to orchestrate the transition to the translation-based RNA-protein world. Progress in both of these arenas has been reviewed recently (156) but will be briefly updated here.

Early engineering and selection aimed at producing an RNA replicase ribozyme centered around the group I self-splicing intron (95, 157–161), but the most promising recent results come from work with the class I ligase selected from random sequence (107; Figure 4). In a dramatic advance, Ekland & Bartel (162) have reengineered the class I ribozyme to catalyze the templated addition of a single NTP onto the 3′ end of an RNA primer strand (Figures 9a and 9b). The reaction occurs with a k_{cat} of 0.3 min^{-1} and a K_M for GTP of 5 mM. Any nucleotide can be added, depending on the identity of the templating base, and the fidelity is reasonably high. Most interestingly, by inserting two additional nucleotides into the template strand (Figure 9c), between the region that binds the primer and that which binds the ribozyme, up to three successive templated nucleotides can be added (up to six can be added when using a dinucleotide repeat sequence, so as to allow Watson-Crick–based recognition in multiple registers). The rate of incorporation of the second and third nucleotides is severely
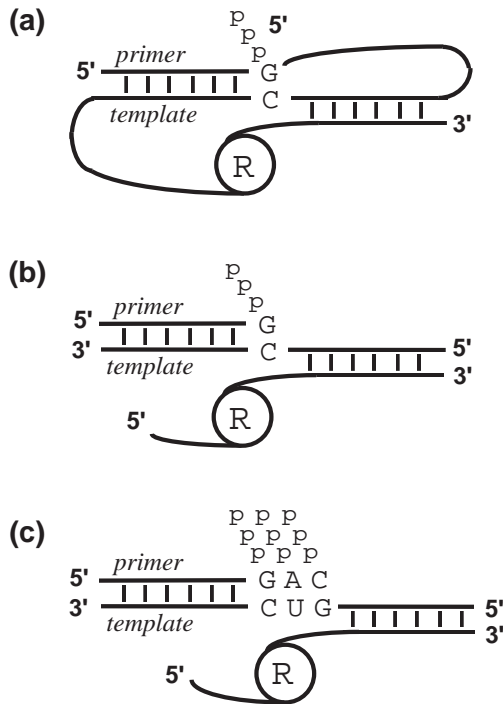
Figure 9   Changing the (*a*) class I ligase into a (*b*) terminal transferase or a (*c*) polymerase (162). "R" represents the ribozyme core structure.

reduced relative to the first, but this is nevertheless the first demonstration of a ribozyme that uses NTPs to carry out RNA synthesis.

The current version of the class I enzyme has one important limitation: the "template-binding region" of the enzyme recognizes the template by Watson-Crick base pairing. This rules out the possibility of translocation of the ribozyme active site along the template, which would be necessary for the addition of an unlimited number of successive nucleotides to a single RNA chain. To overcome this problem, the template-binding domain on the current enzyme must be replaced by a sequence-independent RNA binding domain, for example one that could recognize RNA backbone features. Whether it is possible for RNA to generate such a structure is unknown. The introduction of functional groups not normally present in RNA may be required.

Another important issue in RNA metabolism is the synthesis of the nucleotides themselves. Unrau & Bartel (163) recently selected a ribozyme that adds 4-thiouracil to phospho-ribosyl-1-pyrophosphate, to make 4-thiouridinemono phosphate (Figure 10). To achieve this, they tethered phosphoribosyl-1-pyrophosphate (via the 5′ phosphate) to the 3′ end of a random
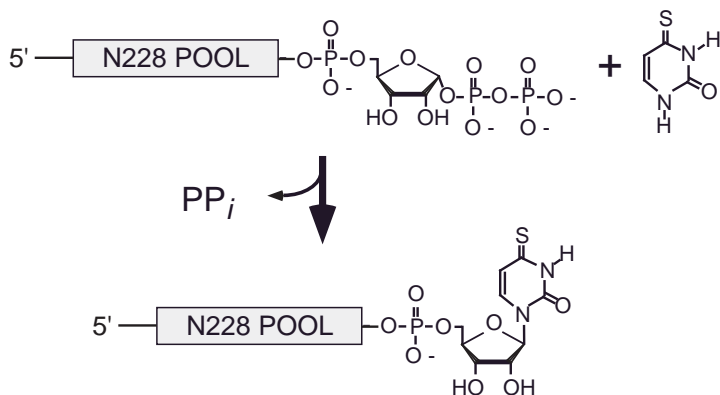
Figure 10 Reaction scheme for isolating UMP synthetase ribozymes (163). Reprinted with kind permission from the authors and from *Nature* (163), copyrighted 1998 Macmillan Magazines Ltd.

RNA pool and incubated this with 4-thiouracil. The pool molecules that facilitated the attack of thiouracil onto the pyrophosphate-activated C1 carbon of ribose thus labeled themselves with the thione group, allowing for efficient partitioning of active sequences by thiophilic reagents. The best ribozymes catalyzed this glyco-sidic bond formation with apparent second-order rate constants (k_{cat}/K_M) at least 10^7-fold above the background rate. This is the first example of a ribozyme that generates a building block for RNA synthesis.

Ribozymes that catalyze much of the chemistry of protein synthesis have been identified, as discussed above. RNA molecules have been selected that are capable of charging themselves by using activated amino acids, transferring these amino acids between RNA hydroxyl groups, and condensing amino acids together to form peptide bonds. What remains to be discovered are RNAs that perform multiple cycles of peptide bond formation, and ribozymes in which peptide synthesis is directed by an mRNA template. This will be a major technical challenge. Like the replicase, a translating ribozyme will need to engage the RNA template in multiple registers, so it too will require a sequence-independent RNA recognition domain. It is therefore extremely important to determine whether and under what circumstances such a domain can exist.

The most general question remaining is, what are the limits of ribozyme function? There are still numerous tasks that RNA has not been shown to be capable of performing, such as binding to small, non-planar compounds with an overall negative charge or hydrophobic character. The chemical structure of RNA suggests that it may be unable to perform at least some of these tasks. The capacity of structured RNAs may be broadened, however, by derivatization with additional chemical functionalities. It has been proposed that modified nucleobases may have existed at the time of the RNA world (164). Such bases would have maintained the structural features required for the formation of the Watson-Crick

base pairs but would be decorated with moieties that increased the chemical diversity of RNA. In particular, the C5-position of pyrimidine bases can be substituted with many different functional groups without compromising their ability to form helical base paired structures or to be copied by reverse transcriptases. Potentially useful functionalities would be hydrophobic groups, basic moieties, thiols, and imidazole. This approach was used to select the Diels-Alder cycloaddition-catalyzing ribozyme (140) and one of the amide bond-forming ribozymes (141) described above. For these selections, all uridines were substituted with either a 5-pyridylmethylcarboxamide or a 5-imidazole derivative, respectively. This idea could be extended by making a pool in which all the uridines were substituted with one side chain, and all the cytosines with another, for instance. In addition, one could link metabolic cofactors to the RNA either noncovalently (165, 166) or covalently (113, 167). Covalent attachment of cofactors to RNA could have been readily accomplished in the RNA world by ribozymes that specialize in such conjugation reactions. As mentioned above (113), a ribozyme has been isolated that catalyzes the covalent attachment of numerous phosphorylated cofactors, including thiamine phosphate, acetyl coenzyme A, NADP, ATP, and flavin mononucleotide (FMN), to the 5′ end of triphosphorylated RNA.

Another metabolically indispensable type of catalysis involves the coupling of a thermodynamically favorable reaction, such as ATP hydrolysis, to an unfavorable one, such as the aminoacylation of RNA from unactivated amino acids. To accomplish this, a ribozyme must couple the two reactions via a phosphorylated or adenylated intermediate. New selection strategies may be required to identify these types of ribozymes.

Comparison Between RNA and Proteins

Although RNA has been shown to be capable of performing many catalytic tasks, proteins clearly dominate in biology. Why is this? Our current understanding of the similarities and differences between RNA and protein catalysts is based on studies comparing natural, highly evolved protein enzymes with natural and artificial ribozymes (for an excellent review, see 168). In general, the ribozymes that have been selected do not live up to the standards of natural protein enzymes. Could this be due, at least in part, to the way in which selections are performed? The impressive properties of the natural group I intron and RNase P RNA support this idea. These RNAs are very complex and achieve a rate acceleration of $\sim10^{11}$-fold relative to the uncatalyzed reaction (169, 170). The group I intron also possesses some of the features of the most refined protein enzymes, such as the ability to perform two distinct reactions, separated by a conformational change (168). Its affinity for the exons changes over the course of the reaction to ensure that the intermediate is not lost and that the product dissociates rapidly to avoid reversal of the second chemical step (171, 172). A derivative of this intron also displays processivity in an endonuclease reaction (173).

The creation of such complex ribozymes in the laboratory may require selections that begin where previous ones finished, allowing us to build up

increasingly complex structures and, correspondingly, activities (4, 109, 129). Preselected modules have recently (174) been combined in a single RNA molecule and shown to interact with each other so as to create allosteric ribozymes (see also 175). Specifically, Tang & Breaker (174) fused a helix of the ATP aptamer with helix II of the hammerhead ribozyme. In the absence of ATP, this hammerhead cleaved its target at a rate of 0.054 min^{-1}. The addition of ATP reduced the catalytic rate by ~180-fold. In this construct, ATP binding locks the two helices of the aptamer into a defined structural relationship with each other such that one of them appears to enter into a steric conflict with helix I of the hammerhead, thus disfavoring the active ribozyme conformation (176).

There are several reasons for thinking that proteins may have intrinsically greater catalytic potential than do RNAs. The diversity of side chains in proteins provides a variety of functional groups not present in RNA for covalent catalysis (carboxylates, thiols, primary amines, imidazoles), for acid/base catalysis (imidazole), and for creating a low effective dielectric constant environment, which can dramatically enhance H-bonding and other electrostatic interactions between enzymes and substrates. The more rigid backbone of polypeptides as compared with that of RNA, and the ability to form tightly packed cores and solvent boundaries may also be important advantages. A protein active site may have ~8 amino-acid positions that directly contribute to substrate binding or catalysis; the number of possible combinations of side chains that evolution could sample would therefore approach 10^{10}, even though many of these would disrupt enzyme structure. A ribozyme active site would be much more restricted in the number of different combinations of side chains and functional groups that it could utilize.

To objectively compare the abilities of RNA versus protein to access catalytically active structures, we need to apply in vitro selection procedures to polypeptides. Unfortunately, this is not a trivial matter since selections require a linkage between phenotype and genotype. This linkage is lost, at least in vitro, as soon as translation ends—the ribosome releases its protein and mRNA to drift off in their separate directions. Two solutions to this problem are phage display (177, 178) and ribosome display (179–181), which maintain the linkage of phenotype and genotype through the noncovalent interactions that stabilize the structure of the phage and ribosome, respectively. These are powerful methods with unique advantages and disadvantages. The replication of phage libraries is fast and easy, but the requirement for an in vivo step limits the number of sequences that can be examined. Ribosome display is a fully in vitro procedure and can generate larger libraries but requires that a fragile molecular complex be maintained intact, which limits the procedures that can be used for selection.

An alternative approach has been to generate stable, covalent mRNA-protein fusions in a completely in vitro system by using puromycin (182). Puromycin is an antibiotic that interrupts translation by acting as a mimic of an aminoacylated tRNA acceptor. When puromycin accepts a nascent peptide chain, that chain becomes linked by a normal peptide bond to the O-methyl tyrosine moiety of puromycin. This mode of action was exploited by covalently linking
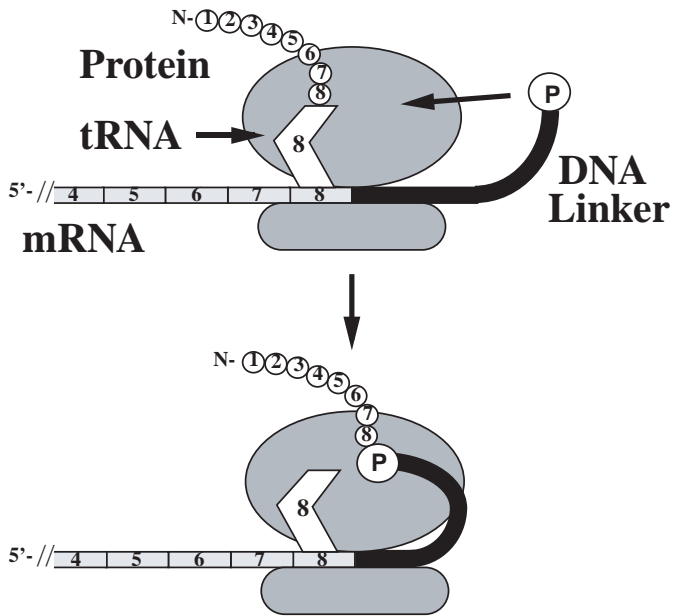
Figure 11 Using puromycin to generate stable mRNA-protein fusion molecules (182). For simplicity, an mRNA with eight codons is shown. (*top*) The eighth codon is recognized by the cognate tRNA to insert the corresponding amino acid. (*bottom*) The puromycin ("P") at the 3' end of the DNA linker enters the A site and accepts the peptide.

puromycin to the 3′ end of a synthetic mRNA through a DNA linker. A ribosome begins translation of such a template as usual, generating a polypeptide as it transits the open reading frame (Figure 11). When the ribosome arrives at the end of the open reading frame, it reaches the DNA linker and presumably stalls, allowing the nearby puromycin to enter the A site of the ribosome and accept the nascent peptide chain. The resulting RNA-peptide fusions can be efficiently formed from mRNAs encoding small peptides or large proteins. We have prepared libraries of fusions encoding ~10^{13} different random peptides (R Liu, A Keefe, T Heuer & JW Szostak, unpublished data) and are now selecting for aptamers and enzymes.

This technique should allow us to address the relative probability of finding an enzyme in collections of random protein versus RNA sequences. From our experience with RNA selections, we can estimate the fraction of RNA sequence space occupied by catalysts of certain chemical reactions; we can now determine the same number for protein sequence space. Closely related questions concern the relative ease of functional optimization and divergence (i.e. evolution of better functions and of novel functions) for protein versus RNA sequences. The answers to these questions have important implications for several issues related

to the origin and early evolution of life, such as the nature and evolutionary properties of the first biopolymers, and why protein largely replaced RNA as the polymer of choice for building robust and efficient catalytic machines.

CONCLUSIONS

In contrast to polypeptides, most random RNA sequences are highly structured (183). This results mainly from the favorable binding energy for Watson-Crick and G-U wobble base pairs and from the statistical likelihood that many such matches will form in a random string. The latter factor depends in turn on the small number of monomer types that make up a chain. Because of this simplicity, a random pool of RNA is a collection of relatively stable three-dimensional shapes. So immense is the number of different shapes in a complex RNA pool that at least one of them will likely bind a chosen target or accelerate a particular chemical reaction. Therefore, the art of selection consists not in synthesizing the functional nucleic acids but in contriving schemes for partitioning them away from the overwhelming majority of inactive molecules. Certain functions, however, might not exist in a random pool of 10^{15} unmodified RNA molecules, so, in these cases, we may need to resort to supplementing the RNA with cofactors or unnatural chemical functionalities or to biasing the region of sequence space being explored by adding previously selected domains. Using such techniques, it should be possible to estimate the range of activities available to nucleic acids. The new protein selection techniques should allow an objective comparison between the functional range and distribution of polynucleotides versus polypeptides.

LITERATURE CITED

1. Szostak JW, Ellington AD. 1993. See Ref. 204, pp. 511–33
2. Gold L, Allen P, Binkley J, Brown D, Schneider D, et al. 1993. See Ref. 204, pp. 497–509
3. Joyce GF. 1994. *Curr. Opin. Struct. Biol.* 4:331–36
4. Jaeger L. 1997. *Curr. Opin. Struct. Biol.* 7:324–35
5. Breaker RR. 1997. *Chem. Rev.* 97:371–90
6. Gold L, Brown D, He YY, Shtatland T, Singer BS, Wu Y. 1997. *Proc. Natl. Acad. Sci. USA* 94:59–64
7. Tian Y, Adya N, Wagner S, Giam CZ, Green MR, Ellington AD. 1995. *RNA* 1: 317–26
8. Davis KA, Abrams B, Lin Y, Jayasena SD. 1996. *Nucleic Acids Res.* 24:702–6

9. Drolet DW, Moon-McDermott L, Romig TS. 1996. *Nat. Biotechnol.* 14:1021–25

10. Bartel DP, Szostak JW. 1994. In *RNA-Protein Interactions*, ed. K Nagai, IW Mattaj, pp. 248–65. Oxford: IRL

11. Famulok M, Jenne A. 1998. *Curr. Opin. Chem. Biol.* 2:320–27

12. Frank DN, Pace NR. 1997. *Proc. Natl. Acad. Sci. USA* 94:14355–60

13. Geyer CR, Sen D. 1997. *Chem. Biol.* 4: 579–93

14. Costa M, Michel F. 1997. *EMBO J.* 16: 3289–302

15. Kumar PKR, Ellington AD. 1995. *FASEB J.* 9:1183–95

16. Szostak JW. 1992. *Trends Biochem. Sci.* 17:89–93

17. Chapman KB, Szostak JW. 1994. *Curr. Opin. Struct. Biol.* 4:618–22

18. Lorsch JR, Szostak JW. 1995. *Acc. Chem. Res.* 29:103–10

19. Gold L, Polisky B, Uhlenbeck O, Yarus M. 1995. *Annu. Rev. Biochem.* 64: 763–97

20. Eaton BE, Gold L, Zichi DA. 1995. *Chem. Biol.* 2:633–38

21. Hirao I, Ellington AD. 1996. *Curr. Biol.* 5:1017–22

22. Spiegelman S. 1971. *Q. Rev. Biophys.* 4: 213–53

23. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, et al. 1985. *Science* 230: 1350–54

24. Ellington AD, Szostak JW. 1990. *Nature* 346:818–22

25. Tuerk C, Gold L. 1990. *Science* 249: 505–10

26. Thiesen H-J, Bach C. 1990. *Nucleic Acids Res.* 18:3203–9

27. Ellington AD, Szostak JW. 1992. *Nature* 355:850–52

28. Ciesiolka J, Yarus M. 1996. *RNA* 2: 785–93

29. Nieuwlandt D, Wecker M, Gold L. 1995. *Biochemistry* 43:5651–59

30. Ringquist S, Jones T, Snyder EE, Gibson T, Boni I, Gold L. 1995. *Biochemistry* 34: 3640–48

31. Pan WH, Craven RC, Qiu Q, Wilson CB, Wills JW, et al. 1995. *Proc. Natl. Acad. Sci. USA* 92:11509–13

32. Morris KN, Jensen KB, Julin CM, Weil M, Gold L. 1998. *Proc. Natl. Acad. Sci. USA* 95:2902–7

33. Cech TR. 1993. See Ref. 204, pp. 239–69

34. Yarus M. 1993. See Ref. 204, pp. 205–17

35. Piccirilli JA, Krauch T, Moroney SE, Benner SA. 1990. *Nature* 343:33–37

36. Piccirilli JA, Moroney SE, Benner SA. 1991. *Biochemistry* 30:10350–56

37. Horlacher J, Hottiger M, Podust VN, Hubscher U, Benner SA. 1995. *Proc. Natl. Acad. Sci. USA* 92:6329–33

38. Lutz M, Held HA, Hottiger M, Hubscher U, Benner SA. 1996. *Nucleic Acids Res.* 24:1308–13

39. Sassanfar M, Szostak JW. 1993. *Nature* 364:550–53

40. Connell GJ, Illangesekare M, Yarus M. 1993. *Biochemistry* 32:5497–502

41. Connell G, Yarus M. 1994. *Science* 264: 1137–41

42. Jenison RD, Gill SC, Pardi A, Polisky B. 1994. *Science* 263:1425–29

43. Huizenga DE, Szostak JW. 1995. *Biochemistry* 34:656–65

44. Conn MM, Prudent JR, Schultz PG. 1996. *J. Am. Chem. Soc.* 118:7012–13

45. Li Y, Sen D. 1996. *Nat. Struct. Biol.* 3: 743–47

46. Harada K, Frankel AD. 1995. *EMBO J.* 14:5798–811

47. Lauhon CT, Szostak JW. 1995. *J. Am. Chem. Soc.* 117:1246–57

48. Jiang F, Kumar RA, Jones RA, Patel DJ. 1996. *Nature* 382:183–86

49. Dieckmann T, Suzuki E, Nakamura GK, Feigon J. 1996. *RNA* 2:628–40

50. Dieckmann T, Butcher SE, Sassanfar M, Szostak JW, Feigon J. 1997. *J. Mol. Biol.* 273:467–78

51. Lin CH, Patel DJ. 1997. *Chem. Biol.* 4: 817–32

52. Zimmermann GR, Jenison RD, Wick CL, Simorre J-P, Pardi A. 1997. *Nat. Struct. Biol.* 4:644–49

53. Zimmermann GR, Shields TP, Jenison RD, Wick CL, Pardi A. 1998. *Biochemistry* 37:9186–92

54. Fan P, Suri AK, Fiala R, Live D, Patel DJ. 1996. *J. Mol. Biol.* 258:480–500

55. Wang Y, Rando RR. 1995. *Chem. Biol.* 2:281–90

56. Jiang L, Suri AK, Fiala R, Patel DJ. 1997. *Chem. Biol.* 4:35–50

57. Jiang L, Patel DJ. 1998. *Nat. Struct. Biol.* 5:769–74

58. Famulok M. 1994. *J. Am. Chem Soc.* 116:1698–706

59. Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M. 1996. *Science* 272:1343–47

60. Lin CH, Patel DJ. 1996. *Nat. Struct. Biol.* 3:1046–50

61. Geiger A, Burgstaller P, von der Eltz H, Roeder A, Famulok M. 1996. *Nucleic Acids Res.* 24:1029–36

62. Feigon J, Dieckmann T, Smith FW. 1996. *Chem. Biol.* 3:611–17

63. Patel DJ, Suri AK, Jiang F, Jiang L, Fan P, et al. 1997. *J. Mol. Biol.* 272:645–64

64. Marshall KA, Robertson MP, Ellington AD. 1997. *Structure* 5:729–34

65. Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. 1992. *Nature* 355:564–66

66. Griffin LC, Tidmarsh GF, Bock LC, Toole JJ, Leung LLK. 1993. *Blood* 81:3271–76

67. Macaya RF, Schultz P, Smith FW, Roe JA, Feigon J. 1993. *Proc. Natl. Acad. Sci. USA* 90:3745–49

68. Wang KY, McCardy S, Shea RG, Swaminathan S, Bolton PH. 1993. *Biochemistry* 32:1899–904

69. Padmanabhan K, Padmanabhan KP, Ferrara JD, Sadler JE, Tulinsky A. 1993. *J. Biol. Chem.*

70. Kelly JA, Feigon J, Yeates TO. 1996. *J. Mol. Biol.* 256:417–22

71. Fitzwater T, Polisky B. 1996. *Methods Enzymol.* 267:275–301

72. Conrad RC, Giver L, Tian Y, Ellington AD. 1996. *Methods Enzymol.* 267:336–67

73. Davis JP, Janic N, Javornik BE, Zichi DA. 1996. *Methods Enzymol.* 267:302–14

74. Ciesiolka J, Illangasekare M, Majerfeld I, Nickles T, Welch M, et al. 1996. *Methods Enzymol.* 267:315–35

75. Keene JD. 1996. *Chem. Biol.* 3:505–13

76. Gold L, Singer B, He Y-Y, Brody E. 1997. *Curr. Opin. Genet. Dev.* 7:848–51

77. Singer BS, Shtatland T, Brown D, Gold L. 1997. *Nucleic Acids Res.* 25:781–86

78. Pieken WA, Olsen DB, Benseler F, Aurup H, Eckstein F. 1991. *Science* 253:314–17

79. Lin Y, Qiu Q, Gill SC, Jayasen SD. 1994. *Nucleic Acids Res.* 22:5229–34

80. Green LS, Jellinek D, Bell C, Beebe LA, Feistner B, et al. 1995. *Chem. Biol.* 2:683–95

81. Pagratis NC, Bell C, Chang YF, Jennings S, Fitzwater T, et al. 1997. *Nat. Biotechnol.* 15:68–73

82. Kubik MF, Bell C, Fitzwater T, Watson SR, Tasset DM. 1997. *J. Immunol.* 159:259–67

83. Jellinek D, Green LS, Bell C, Lynott CK, Gill N, et al. 1995. *Biochemistry* 34:11363–72

84. Wiegand TW, Williams PB, Dreskin SC, Jouvin MH, Kinet JP, Tasset D. 1996. *J. Immunol.* 157:221–30

85. Hicke BJ, Watson SR, Koenig A, Lynott CK, Bargatze RF, et al. 1996. *J. Clin. Invest.* 98:2688–92

86. Gold L. 1995. *J. Biol. Chem.* 270:13581–84

87. Willis MC, Collins B, Zhang T, Green LS, Sebesta D, et al. 1998. *Bioconjug. Chem.* 9:573–82

88. Williams KP, Liu X-H, Schumacher TNM, Lin HY, Ausiello DA, et al. 1997. *Proc. Natl. Acad. Sci. USA* 94:11285–90

89. Klussmann S, Nolte A, Bald R, Erdmann VA, Fürste JP. 1996. *Nat. Biotechnol.* 14:1112–15

90. Nolte A, Klussmann S, Bald R, Erdmann VA, Fürste JP. 1996. *Nat. Biotechnol.* 14:1116–19

91. Charlton J, Kirschenheuter GP, Smith D. 1997. *Biochemistry* 36:3018–26

92. Bless NM, Smith D, Charlton J, Czermak BJ, Schmal H, et al. 1997. *Curr. Biol.* 7:877–80

93. Michel F, Ellington AD, Couture S, Szostak JW. 1990. *Nature* 347:578–80

94. Green R, Ellington AD, Szostak JW. 1990. *Nature* 347:406–8

95. Green R, Szostak JW. 1992. *Science* 258:1910–15

96. Beaudry AA, Joyce GF. 1992. *Science* 257:635–41

97. Bartel DP, Szostak JW. 1993. *Science* 261:1411–18

98. Lehman N, Joyce GF. 1993. *Nature* 361:182–85

99. Tsang J, Joyce GF. 1994. *Biochemistry* 33: 5966–73

100. Chapman KB, Szostak JW. 1995. *Chem. Biol*. 2:325–33

101. Tsang J, Joyce GF. 1996. *J. Mol. Biol*. 262:31–42

102. Hager AJ, Szostak JW. 1997. *Chem. Biol*. 4:607–17

103. Tuschl T, Sharp PA, Bartel DP. 1998. *EMBO J*. 17:2637–50

104. Cadwell RC, Joyce GF. 1992. *PCR Methods Appl*. 2:28–33

105. Zaug AJ, Cech TR. 1986. *Science* 231:470–75

106. Robertson DL, Joyce GF. 1990. *Nature* 344:467–68

107. Ekland EH, Szostak JW, Bartel DP. 1995. *Science* 269:364–70

108. Ekland EH, Bartel DP. 1995. *Nucleic Acids Res*. 23:3231–38

109. Lorsch JR, Szostak JW. 1994. *Nature* 371:31–36

110. Lorsch JR, Szostak JW. 1995. *Biochemistry* 34:15315–27

111. Huang F, Yarus M. 1997. *Biochemistry* 36:6557–63

112. Pan T, Uhlenbeck OC. 1992. *Biochemistry* 31:3887–95

113. Huang F, Yarus M. 1997. *Proc. Natl. Acad. Sci. USA* 94:8965–69

114. Frank DN, Ellington AD, Pace NR. 1996. *RNA* 2:1179–88

115. Pan T. 1995. *Biochemistry* 34:8458–64

116. Joseph S, Berzal-Herranz A, Chowrira BM, Butcher SE, Burke JM. 1993. *Genes Dev*. 7:130–38

117. Siwkowski A, Humphrey M, De-Young MB, Hampel A. 1998. *Biotechniques* 24:278–84

118. Nishikawa F, Kawakami J, Chiba A, Shirai M, Kumar PKR, Nishikawa S. 1996. *Eur. J. Biochem.* 237:712–18

119. Williams KP, Ciafré S, Tocchini-Valentini GP. 1995. *EMBO J*. 14:4551–57

120. Jayasena VK, Gold L. 1997. *Proc. Natl. Acad. Sci. USA* 94:10612–17

121. Breaker RR. 1997. *Nat. Biotechnol*. 15:427–31

122. Breaker RR, Joyce GF. 1994. *Chem. Biol*. 1:223–29

123. Breaker RR, Joyce GF. 1995. *Chem. Biol*. 2:655–60

124. Santoro SW, Joyce GF. 1997. *Proc. Natl. Acad. Sci. USA* 94:4262–66

125. Faulhammer D, Famulok M. 1997. *J. Mol. Biol.* 269:188–202

126. Faulhammer D, Famulok M. 1996. *Angew. Chem. Int. Ed. Engl.* 35:2837–41

127. Roth A, Breaker RR. 1998. *Proc. Natl. Acad. Sci. USA* 95:6027–31

128. Cuenoud B, Szostak JW. 1995. *Nature* 375:611–14

129. Wilson C, Szostak JW. 1995. *Nature* 374:777–82

130. Illangasekare M, Sanchez G, Nickles T, Yarus M. 1995. *Science* 267:643–47

131. Illangasekare M, Yarus M. 1997. *J. Mol. Biol.* 268:631–39

132. Lohse PA, Szostak JW. 1996. *Nature* 381:442–44

133. Suga H, Szostak JW. 1998. *J. Am. Chem. Soc.* 120:1151–56

134. Zhang B, Cech TR. 1997. *Nature* 390:96–100

135. Noller HF, Hoffarth V, Zimniak L. 1992. *Science* 256:1420–24

136. Green R, Noller HF. 1997. *Annu. Rev. Biochem*. 66:679–716

137. Nitta I, Ueda T, Watanabe K. 1998. *RNA* 4:257–67

138. Nitta I, Kamada Y, Noda H, Ueda T, Watanabe K. 1998. *Science* 281:666–69

139. Hausch F, Jäschke A. 1997. *Bioconjug. Chem.* 8:885–90

140. Tarasow TM, Tarasow SL, Eaton BE. 1997. *Nature* 389:54–57
141. Wiegand TW, Janssen RC, Eaton BE. 1997. *Chem. Biol.* 4:675–83
142. Lerner RA, Benkovic SJ, Schultz PG. 1991. *Science* 252:659–67
143. Prudent JR, Uno T, Schultz PG. 1994. *Science* 264:1924–27
144. Li Y, Sen D. 1997. *Biochemistry* 36:5589–99
145. Li Y, Sen D. 1998. *Chem. Biol.* 5:1–12
146. Sabeti PC, Unrau PJ, Bartel DP. 1997. *Chem. Biol.* 4:767–74
147. Tang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M. 1996. *Science* 272:1343–47
148. Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. *Proc. R. Soc. London Ser. B* 255:279–84
149. Huynen MA, Stadler PF, Fontana W. 1996. *Proc. Natl. Acad. Sci. USA* 93:397–401
150. Huynen MA. 1996. *J. Mol. Evol.* 43:165–69
151. Hanczyc MM, Dorit RL. 1998. *RNA* 4:268–75
152. Treiber DK, Rook MS, Zarrinkar PP, Williamson JR. 1998. *Science* 279:1943–46
153. Zarrinkar PP, Williamson JR. 1994. *Science* 265:918–24
154. Sclavi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA. 1998. *Science* 279:1940–43
155. Wright MC, Joyce GF. 1997. *Science* 276:614–17
156. Hager AJ, Pollard JD, Szostak JW. 1996. *Chem. Biol.* 3:717–25
157. Doudna JA, Szostak JW. 1989. *Nature* 339:519–24
158. Doudna JA, Szostak JW. 1989. *Mol. Cell. Biol.* 9:5480–83
159. Doudna JA, Couture S, Szostak JW. 1991. *Science* 251:1605–8
160. Bartel DP, Doudna JA, Usman N, Szostak JW. 1991. *Mol. Cell. Biol.* 11:3390–94
161. Doudna JA, Usman N, Szostak JW. 1993. *Biochemistry* 32:2111–25
162. Ekland EH, Bartel DP. 1996. *Nature* 382:373–76
163. Unrau PJ, Bartel DP. 1998. Nature 395:260–63
164. Robertson MP, Miller SL. 1995. *Science* 268:702–5
165. Dai X, de Mesmaeker A, Joyce GF. 1995. *Science* 267:237–40
166. Dai X, de Mesmaeker A, Joyce GF. 1996. *Science* 272:18–19
167. Breaker RR, Joyce GF. 1995. *J. Mol. Evol.* 40:551–58
168. Narlikar GJ, Herschlag D. 1997. *Annu. Rev. Biochem.* 66:19–49
169. Herschlag D, Cech TR. 1990. *Biochemistry* 29:10159–71
170. Beebe JA, Fierke CA. 1994. *Biochemistry* 33:10294–304
171. Mei R, Herschlag D. 1996. *Biochemistry* 35:5796–809
172. Emerick VL, Pan J, Woodson SA. 1996. *Biochemistry* 35:13469–77
173. Herschlag D. 1992. *Biochemistry* 31:1386–99
174. Tang J, Breaker RR. 1997. *Chem. Biol.* 4:453–59
175. Peracchi A, Beigelman L, Usman N, Herschlag D. 1996. *Proc. Natl. Acad. Sci. USA* 93:11522–27
176. Tang J, Breaker RR. 1998. *Nucleic Acids Res.* 26:4222–29
177. Smith GP. 1985. *Science* 228:1315–17
178. Scott JK, Smith GP. 1990. *Science* 249:386–90
179. Mattheakis LC, Bhatt RR, Dower WJ. 1994. *Proc. Natl. Acad. Sci. USA* 91:9022–26
180. Hanes J, Pluckthun A. 1997. *Proc. Natl. Acad. Sci. USA* 94:4937–42
181. He M, Taussig MJ. 1997. *Nucleic Acids Res.* 25:5132–34
182. Roberts RW, Szostak JW. 1997. *Proc. Natl. Acad. Sci. USA* 94:12297–302
183. Fontana W, Konings DAM, Stadler PF, Schuster P. 1993. *Biopolymers* 33:1389–404
184. Kiga D, Futamura Y, Sakamoto K, Yokoyama S. 1998. *Nucleic Acids Res.* 26(7):1755–60

185. Haller AA, Sarnow P. 1997. *Proc. Natl. Acad. Sci. USA* 94:8521–26
186. Majerfeld I, Yarus M. 1994. *Nat. Struct. Biol.* 1:287–92
187. Famulok M, Szostak JW. 1992. *J. Am. Chem. Soc.* 114:3990–91
188. Lorsch JR, Szostak JW. 1994. *Biochemistry* 33:973–82
189. Li Y, Geyer R, Sen D. 1996. *Biochemistry* 35:6911–22
190. Burgstaller P, Famulok M. 1994. *Angew. Chem. Int. Ed. Engl.* 33:1084–87
191. Wallis MG, von Ahsen U, Schroeder R, Famulok M. 1995. *Chem. Biol.* 2:543–52
192. Lato SM, Boles AR, Ellington AD. 1995. *Chem. Biol.* 2:291-303
193. Wallace ST, Schroeder R. 1998. *RNA* 4:112–23
194. Wallis MG, Streicher B, Wank H, von Ahsen U, Clodi E, et al. 1997. *Chem. Biol.* 4:357–66
195. Burke DH, Hoffman DC, Brown A, Hansen M, Pardi A, Gold L. 1997. *Chem. Biol.* 4:833–43
196. Morris KN, Tarasow TM, Julin CM, Simons SL, Hilvert D, Gold L. 1994. *Proc. Natl. Acad. Sci. USA* 91:13028–32
197. Mannironi C, Di Nardo A, Fruscoloni P, Tocchini-Valentini GP. 1997. *Biochemistry* 36:9726–34
198. Hofmann H-P, Limmer S, Hornung V, Sprinzl M. 1997. *RNA* 3:1289–300
199. Jenne A, Famulok M. 1997. *Chem. Biol.* 5:23–34
200. Wecker M, Smith D, Gold L. 1996. *RNA* 2:982–94
201. Piccirilli JA, McConnell TS, Zaug AJ, Noller HF, Cech TR. 1992. *Science* 256:1420–24
202. Carmi N, Shultz LA, Breaker RR. 1996. *Chem. Biol.* 3:1039–46
203. Carmi N, Balkhi SR, Breaker RR. 1998. *Proc. Natl. Acad. Sci. USA* 95:2233–37
204. Gesteland RF, Atkins JF, eds. 1993. *The RNA World.* Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press