Chance and Necessity in the Selection of Nucleic Acid Catalysts

JON R. LORSCH† AND JACK W. SZOSTAK*

Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Received August 2, 1995

> The equanimity of your average tosser of coins depends upon a law, or rather a tendency, or let us say a probability, or at any rate a mathematically calculable chance, which ensures that he will not upset himself by losing too much nor upset his opponent by winning too often. This made for a kind of harmony and a kind of confidence. It related the fortuitous and the ordained into a reassuring union which we recognize as nature.
>
> Guildenstern in *Rosencrantz and Guildenstern Are Dead*

In Tom Stoppard's famous play, the ill-fated heroes toss a coin 101 times. The first 100 times they do so the coin lands heads up. The chance of this happening is approximately 1 in 10^{30}, a sequence of events so rare that one might argue that it could only happen in such a delightful fiction. Similarly rare events, however, may underlie the origins of biological catalysis. What is the probability that an RNA, DNA, or protein molecule of a given random sequence will display a particular catalytic activity? The answer to this question determines whether a collection of such sequences, such as might result from prebiotic chemistry on the early earth,[1,2] is extremely likely or unlikely to contain catalytically active molecules, and hence whether the origin of life itself is a virtually inevitable consequence of chemical laws or merely a bizarre fluke. The fact that *a priori* estimates of this probability, given by otherwise informed chemists and biologists, ranged from 10^{-5} to 10^{-50}, inspired us to begin to address the question experimentally. As it turns out, the chance that a given random sequence RNA molecule will be able to catalyze an RNA polymerase-like phosphoryl transfer reaction is close to 1 in 10^{13}, rare enough, to be sure, but nevertheless in a range that is comfortably accessible by experiment. It is the purpose of this Account to describe the recent advances in combinatorial biochemistry that have made it possible for us to explore the abundance and diversity of catalysts existing in nucleic acid sequence space.

Jon Lorsch received his B.A. in chemistry from Swarthmore College in 1990. From there he returned to his native Cambridge, MA, with the intention of joining the laboratory of Prof. Jeremy Knowles. After having had Jon in his lab for only a short time, however, Prof. Knowles despaired of the future of enzymology and promptly gave up science. Jon was then granted asylum in Jack Szostak's lab, from which he was expelled in 1995 (shortly after completing his Ph.D. thesis). He is currently in exile as a postdoctoral fellow in Dan Herschlag's lab at Stanford.

Jack W. Szostak was born in England and raised in Canada, where he completed undergradauate studies at McGill University. He then moved to the United States for graduate and postdoctoral study with Prof. Ray Wu at Cornell University. He is currently professor of genetics at Harvard Medical School and Massachusetts General Hospital. Prof. Szostak's scientific interests include the study of the origins of life and biocatalysis and the exploration of sequence space in search of new ribozymes. Both Dr. Lorsch and Prof. Szostak greatly enjoy Tom Stoppard's plays.

In Vitro Selection

At the beginning of this decade, it was shown that RNAs with virtually any desired binding properties could be isolated from pools of random sequence molecules by iterative rounds of selection and amplification.[3,4] This process is called *in vitro* selection, or SELEX (systematic evolution of ligands by exponential enrichment;[3] Figure 1). The RNA receptors that result from such experiments are often referred to, at least by those who do not object to linguistic chimaeras, as aptamers (from the Latin, *aptus*, to fit, and the Greek suffix *-mer*). Aptamers that bind small molecules with good affinity and specificity are relatively abundant in sequence space: ~1 in 10^{10} random sequence RNA 100-mers can fold into structures capable of specifically binding certain organic dyes.[4] In the last few years a number of RNA and DNA aptamers for amino acids,[5−9] drugs,[10] and enzymatic cofactors[8,11−15] have been isolated. Typical association constants, measured in aqueous buffers, are in the range 10^5−10^7 M^{-1}. It is not possible, at present, to design RNA or DNA molecules with such properties.

Having shown that RNAs with specific binding properties could be isolated fairly easily, the next logical question to ask was whether RNAs with catalytic activity could also be isolated from pools of random sequences by *in vitro* selection. In principle, this can be viewed simply as an extension of the selection criterion from substrate binding to preferential binding of the transition state. While it had been known for some time that catalytic RNAs exist in nature,[16,17] it was not clear how easy or difficult it would be to find new ribozymes in collections of

† Present address: Department of Biochemistry, Beckman Center B400, Stanford University, Stanford, CA 94305.

(1) Joyce, G. F. *Nature* **1989**, *338*, 217−224.
(2) Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1993; pp 1−25.
(3) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505−510.
(4) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818−822.
(5) Famulok, M.; Szostak, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 3990−3991.
(6) Famulok, M. *J. Am . Chem. Soc.* **1994**, *116*, 1698−1706.
(7) Connell, G. J.; Illangesekare, M.; Yarus, M. *Biochemistry* **1993**, *32*, 5497−5502.
(8) Connell, G.; Yarus, M. *Science* **1994**, *264*, 1137−1141.
(9) Majerfeld, I.; Yarus, M. *Nature Struct. Biol.* **1994**, *1*, 287−292.
(10) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425−1429.
(11) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550−553.
(12) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973−982.
(13) Burgstaller, P.; Famulok, M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1084−1087.
(14) Lauhon, C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1246−1257.
(15) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656−665.
(16) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147−157.
(17) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849−857.

0001-4842/96/0129-0103$12.00/0

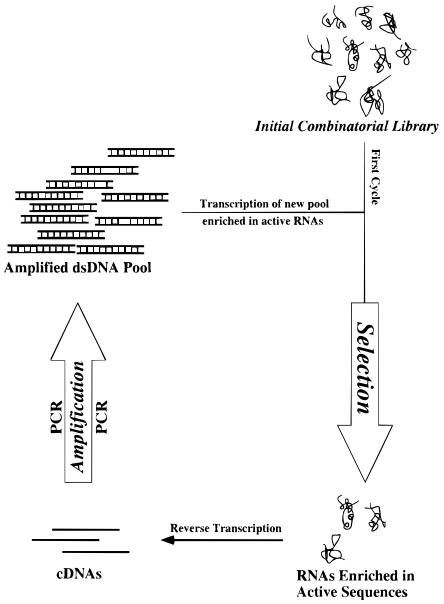

Figure 1. Schematic representation of *in vitro* selection and evolution. A combinatorial library of nucleic acid molecules is subjected to selection for some physical property (e.g., binding to a ligand, catalysis, etc.). Molecules possessing the desired trait are separated from inactive ones to yield a new pool enriched in functional sequences. The enriched pool is then amplified (e.g., by reverse transcription followed by the polymerase chain reaction) to obtain sufficient material to carry out another round of selection. The cycle is repeated until the pool is sufficiently enriched in active sequences. If the amplification procedure is carried out in an error-prone manner and the selection procedure is gradually changed or increased in stringency, the population of RNA or DNA molecules will undergo true Darwinian evolution.

random sequences. Furthermore, all naturally occurring ribozymes catalyze similar kinds of reactions. It seemed possible that this was because of fundamental limitations to the catalytic ability of RNA, which lacks the functional group diversity of proteins (4 similar nucleotides vs 20 diverse amino acid building blocks). We were interested in using *in vitro* selection to define the functional capabilities and limitations of RNA (and DNA). The study of new and different ribozymes generated by *in vitro* selection could make it easier to deduce general principles about how polynucleotides mediate catalysis. The promise of using ribozymes (and deoxyribozymes) as therapeutic agents is just beginning to be explored,[18] and *in vitro* selection may also aid in the development of these new therapeutic agents.

Indirect Selections: Transition State Analogues

The successful use of transition state analogues in the isolation of catalytic antibodies[19] immediately suggested that aptamers to transition state analogues might also exhibit catalytic activity. Aptamers that bind transition state analogues for ester and carbonate hydrolysis (phosphonates, alcohols)[20−23] and both inter- and intramolecular pericyclic reactions[22,24] have been isolated, but none have exhibited catalytic activity. In the one successful application of this approach, Prudent et al.[25] selected for aptamers to a planar phenanthrene derivative designed as a transition state analogue for the slow isomerization of two diastereomeric bridged biphenyls. One of three aptamers isolated was found to accelerate the isomerization reaction by ∼100-fold. Almost all of the binding energy of the RNA for the transition state analogue is realized in transition state stabilization. This modest success may reflect the simple requirements for catalysis in this case: pure geometric stabilization of the transition state, probably mediated by stacking of the RNA bases with aromatic rings, and no need for chemical catalysis, e.g., acid−base, nucleophilic, etc. A more general extension of this approach may require the synthesis of transition state analogues designed to interact strongly and specifically with RNA, perhaps less hydrophobic and with more hydrogen bond donors and acceptors than typically found on transition state analogues designed to interact wth proteins.

Direct Selections: Self-Modifying Ribozymes

In contrast to the difficulties encountered with the transition state analogue approach, direct selections for catalytic activity have led to the isolation of many new ribozymes (Table 1). Direct selections lead to the isolation of ribozymes that can accelerate a chemical reaction, but only for a single reaction cycle, because it is the self-modifying nature of the reaction that allows the active RNA molecules to be distinguished from inactive molecules during selection (Figure 2). However, in many cases it has turned out to be quite easy to engineer these single-turnover ribozymes into true multiple-turnover enzymes.

All known naturally occurring ribozymes catalyze cleavage/ligation reactions of phosphodiesters. Thus, the search for non-natural ribozymes began by hunting in sequence space for RNAs that could perform these kinds of reactions. Yeast $tRNA^{phe}$ undergoes hydrolytic cleavage at a specific site in its D loop in the presence of Pb^{2+}.[26−29] In an effort to isolate new RNAs that contain lead cleavage sites and to use the Pb^{2+} cleavage reaction to probe the sequence requirements of the tRNA fold, Pan and Uhlenbeck[30] selected, from a pool of partially randomized tRNAs, RNAs that self-cleave in the presence of lead ion (Figure 3). The positions that were randomized were known to form structurally important tertiary interactions near the Pb^{2+} binding site of the tRNA. In order to select for cleavage, Pan and Uhlenbeck developed a clever scheme in which the pool of mutagenized tRNAs was circularized using T4 RNA ligase. Molecules that cleaved themselves in the presence of Pb^{2+}, and thus

(18) Sullivan, S. M. *J. Invest. Dermatol. Suppl.* **1994**, *103*, 85S−89S.
(19) Lerner, R. A.; Benkovic, S. J.; Schultz, P. G. *Science* **1991**, *252*, 659−667.
(20) Ellington, A. D.; Szostak, J. W. Unpublished results.
(21) Famulok, M.; Szostak, J. W. Unpublished results.
(22) Lauhon, C. T.; Szostak, J. W. Unpublished results.
(23) Prudent, J.; Schultz, P. G. Personal communication.
(24) Morris, K. N.; Tarasow, T. M.; Julin, C. M.; Simons, S. L.; Hilvert, D.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 13028−13032.
(25) Prudent, J. R.; Uno, T.; Schultz, P. G. *Science* **1994**, *264*, 1924−1927.
(26) Werner, C.; Krebs, B.; Keith, G.; Dirheimer, G. *Biochim. Biophys. Acta* **1976**, *432*, 161−175.
(27) Brown, R.; Dewan, J.; Klug, A. *Biochemistry* **1985**, *24*, 4785−4801.
(28) Kryzosiak, W. J.; Marciniec, T.; Wiewiorowski, M.; Romby, P.; Ebel, J. P.; Giege, R. *Biochemistry* **1988**, *27*, 5771−5777.
(29) Behlen, L. S.; Sampson, J. R.; DiRenzo, A. B.; Uhlenbeck, O. C. *Biochemistry* **1990**, *29*, 2515−2523.
(30) Pan, T.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 3887−3895.

Table 1. Results of Successful Selections for Nucleic Acid Catalysts

target reaction	pool complexity/length	no. of ribozymes	reactions catalyzed	noteworthy requirements	ref
RNA					
RNA hydrolysis	$2.6 \times 10^5/1 \times 10^6$ 76-mers (9/10 random)	>50	RNA hydrolysis	Pb^{2+}	30
RNA ligase	10^{15} 274-mers (220 random)	~65	2′−5′ and 3′−5′ ligations		31
polynucleotide kinase	10^{16} 174-mers (100 random/22 mutagenized)	~10^4	2′ and 5′ (thio) phosphorylation	some require Mn^{2+}	43
self-alkylation	10^{14} 156-mers (24 random/93 mutagenized)	<10 (1 major seq)	N7 alkylation	biotin moiety required (not nucleophilic "hot spot")	42
rotamer isomerization	10^{15} 195-mers (128 random)	1	rotamer isomerization		25
self-acylation	10^{14} 95-mers (50 random)	~10	2′(3′) acylation (perhaps internal as well)	Ca^{2+}	45
3′ ligation of 5′-phosphorimidazolide	1×10^{15} 149-mers (90 random)	consensus	5′−5′ tetraphosphate ligation		36
DNA					
RNA hydrolysis	10^{14} 80-mers (50 random)	consensus	RNA hydrolysis (in otherwise DNA substrate)	Pb^{2+}	37
5′ ligation of 3′-phosphorimidazolide	10^{14} 158-mers (116 random)	consensus	5′ ligation of 3′-phosphorimidazolide	Zn^{2+} or Cu^{2+}	38

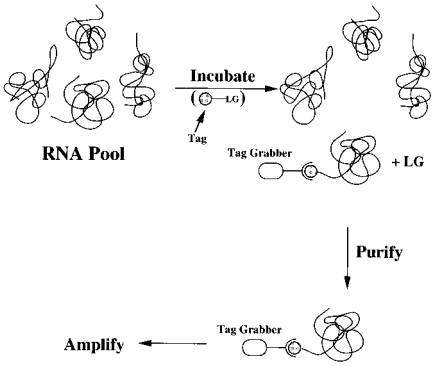

Figure 2. General method for the direct selection of catalytic nucleic acids. A pool of nucleic acid molecules is incubated with a tagged substrate. RNAs or DNAs that covalently label themselves with the tag are separated from inactive sequences using the tag as a handle. The tag could be, for example, a biotin group, and the "tag grabber" could be an immobilized streptavidin resin. The active sequences are amplified as in Figure 1, and the cycle can then be repeated. By derivatizing the RNA or DNA with a second substrate molecule, catalysts for reactions not involving nucleotides can be selected. While the RNA or DNA molecules as isolated accelerate only their own modification, it is often possible to engineer them into true enzymes that catalyze a multiple-turnover reaction. LG: leaving group.

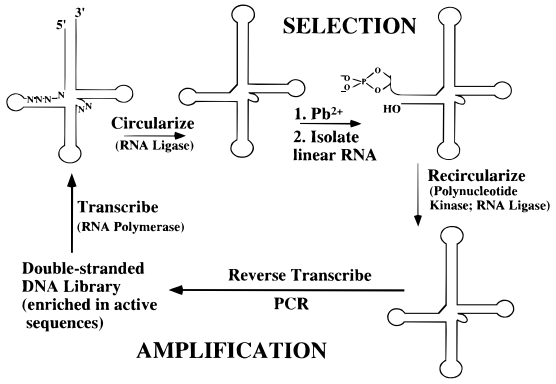

Figure 3. Selection scheme used by Pan and Uhlenbeck[30] to isolate tRNA variants that self-cleave in the presence of Pb^{2+}. A degenerate pool of tRNAs was circularized using T4 RNA ligase. These molecules were then allowed to self-cleave in the presence of Pb^{2+}. The resulting linear molecules were purified by polyacrylamide gel electrophoresis, recircularized, amplified, and transcribed to yield a new RNA pool enriched in active sequences. Modified from Figure 1 in ref 30.

became linear, were separated from circular ones by denaturing polyacrylamide gel electrophoresis. The partially purified molecules were recircularized with RNA ligase and amplified to allow for further rounds of purification. After six rounds of selection the pool of surviving RNAs self-cleaved in the presence of Pb^{2+} at rates comparable to that of wild-type $tRNA^{phe}$. Remarkably, only 3 out of 24 clones self-cleaved at the original site of hydrolysis. A number of the selected molecules self-cleaved more than 10 times faster than wild-type $tRNA^{phe}$. The fact that pools with only 10 or 11 random positions yielded so many apparently different catalytically active sequences suggests that RNA can find a great number of different ways to fold into structures capable of self-cleavage in the presence of Pb^{2+}.

Bartel and Szostak chose to search for RNAs that could catalyze ligation rather than cleavage.[31] Rather than bias their molecules toward any particular structure, they decided to select for RNA ligase ribozymes from pools of completely random sequence RNAs (Figure 4). The selection scheme involved isolating RNAs that could ligate an oligoribonucleotide substrate onto their own 5′-ends, with the incoming oligonucleotide displacing the pyrophosphate group from the 5′-triphosphate on the pool RNA to form a new phosphodiester linkage; the attached oligonucleotide acts to "tag" the catalytically active RNAs. Following the reaction, molecules that had performed the ligation reaction were separated from unreacted pool RNA on an oligonucleotide affinity column complementary to the tag sequence and were further enriched by selective (polymerase chain reaction (PCR)) amplification. Preparation and amplification of the original random sequence pool was rather labor intensive due to the large amounts of material involved (e.g., preparation of 16 copies of a pool of ~10^{15} sequences requires PCR on a scale close to 1 L); however, the great decrease in complexity that occurs after selection allows one to return to the more standard scale of Eppendorf tube PCR (<1 mL). After only three rounds of selection, catalytically competent

(31) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411−1418.

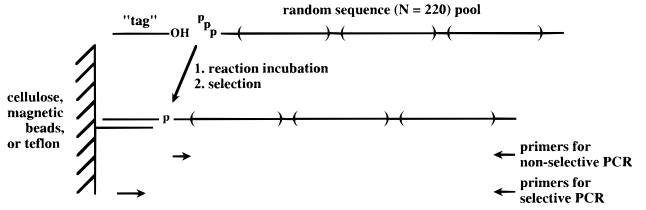

Figure 4. Selection scheme used by Bartel and Szostak[31] to isolate RNA ligases. A pool of DNA sequences containing 220 random bases was generated by ligating together three smaller random sequence pools (represented by parentheses). This DNA was transcribed to yield a random sequence RNA pool, with all molecules beginning with a 5′-triphosphate. Incubation with a substrate oligonucleotide resulted in rare catalytic sequences becoming ligated to the substrate *via* a new phosphodiester bond. The tagged sequences could then be purified by oligonucleotide affinity chromatography and further enriched by selective PCR using a primer complementary to the tag sequence.

RNAs had come to dominate the selected pool. Analysis of the selected pool RNA revealed that ~65 different RNAs, out of an initial pool of 10^{15} different RNA sequences, had been recovered. Thus, approximately 1 in 10^{13} random 220-mers is capable of catalyzing this ligation reaction at a rate sufficient to survive three rounds of selection.

For part of the selection (cycles 5, 6, and 7) the system was subjected to true Darwinian evolution: mutations were introduced via mutagenic PCR,[32] and selective pressure was applied by decreasing the time allowed for ligation. After 10 rounds of selection, the pool catalyzed the ligation reaction at a rate 7 million times faster than the uncatalyzed reaction, an improvement of almost 10 000-fold over the round 3 pool. The dramatic effect of *in vitro* evolution in this experiment is a consequence of the fact that the fraction of sequence space being sampled is extremely small (the number of possible 220-mers is 4^{220}, or $\approx 10^{132}$, compared to typical pool complexities of $\approx 10^{15}$ different sequences), and thus ribozymes isolated directly from these pools are unlikely to be optimal sequences. By allowing the pool to evolve one can search, in local sequence space, for mutants with improved activity. Indeed, analysis of DNA from each cycle of the selection showed that the system had behaved much like an evolving population of organisms: sequences arose and survived for a time, but were eventually out-competed as new, more efficient ribozymes were generated through mutagenesis.

The pool RNA from round 6 formed predominantly the 3′−5′ linkage, but recent experiments have shown that in later rounds, for reasons which are still mysterious, ribozymes that catalyzed the formation of 2′−5′-linked phosphodiesters came to predominate.[33] In the selection for RNA ligases,[31] it was expected that only ribozymes catalyzing the formation of 3′−5′ linkages would be isolated because of the reported inability of reverse transcriptase to read through 2′−5′ linkages.[34] However, we have since shown that, under the conditions used in the selection, 2′−5′-linked molecules are efficiently reverse-transcribed and amplified.[35] This is a good example of ribozymes emerging from *in vitro* selection but yielding unexpected reaction products, and it emphasizes the need to determine experimentally the specificity of every newly obtained catalyst.

Recently, Bartel and co-workers[33] were able to engineer and evolve a multiple-turnover version of one of the ribozymes that is over 3 orders of magnitude more active than the version of the ribozyme from the initial selection ($k_{cat} = \sim 100\ min^{-1}$), corresponding to a rate enhancement relative to the uncatalyzed but template-directed reaction approaching 10^9. The structure of the enzyme is quite complex, and the minimal catalytic domain appears to be on the order of 90 nucleotides in length, demonstrating that ribozymes similar in size and catalytic efficiency to those found in nature can be isolated by *in vitro* selection and evolution.

In an attempt to select for RNAs that can ligate oligonucleotides in a template-directed fashion onto their 3′-ends, Chapman and Szostak[36] isolated a small ribozyme that forms a 5′−5′ tetraphosphate linkage between its own 5′-triphosphate and a 5′-phosphorimidazole-activated oligonucleotide with a k_{cat} of approximately 0.1 min^{-1}. The formation of this linkage, while not the intended reaction, is of interest because the signaling molecule diadenosine tetraphosphate contains a 5′−5′ tetraphosphate, and a great many other biologically important molecules contain similar linkages (e.g., the 5′ cap structure of eukaryotic mRNAs contains a 5′−5′ triphosphate, and NAD and FAD contain 5′−5′ pyrophosphates).

These successes with RNA raised the possibility that DNA might have similar abilities. In the past year, two different deoxyribozymes have been isolated via *in vitro* selection, one catalyzing the cleavage of an otherwise entirely DNA substrate at a specific internal ribonucleotide in the presence of Pb^{2+},[37] the other catalyzing the Zn^{2+}- or Cu^{2+}-dependent ligation of two DNA molecules,[38] with rate enhancements on the order of 10^4−10^5.

It appears that both RNA and DNA are able to catalyze the formation and cleavage of a variety of different phosphodiester and phosphoanhydride bonds (e.g., 2′−5′ linkages, 3′−5′ linkages, pyrophosphate linkages, etc.). The number of different polynucleotide structures that can accelerate these sorts of reactions is almost staggering: Ekland et al.[33] have suggested, on the basis of a rather bold extrapolation, that there may be thousands or even millions of different RNA structures capable of catalyzing a simple RNA ligation reaction.

It should be noted that the recovery of only a single consensus structural motif[36−38] from a selection does not imply that there were no other functional structures within the pool. Rather, the observed consensus represents the simplest, and thus most common, solution to the particular chemical problem posed by the selection. The fact that these structures are often so small (<50 nucleotides in length) demonstrates just how easy it is for nucleic acids to fold into catalytically active structures.

As has been done for naturally occurring ribozymes, several *in vitro* selected nucleic acid catalysts have been engineered to act as true enzymes, that is, work

(32) Cadwell, R. C.; Joyce, G. F. *PCR Methods Appl.* **1992**, *2*, 28−33.
(33) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. *Science* **1995**, *269*, 364−370.
(34) Morl, M.; Niemer, I.; Schmeizer, C. *Cell* **1992**, *70*, 803−810.
(35) Lorsch, J. R.; Bartel, D. P.; Szostak, J. W. *Nucleic Acids Res.* **1995**, *23*, 2811−2814.
(36) Chapman, K.; Szostak, J. W. *Chem. Biol.* **1995**, *2*, 325−333.
(37) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223−229.
(38) Cuenoud, B.; Szostak, J. W. *Nature* **1995**, *375*, 611−614.

intermolecularly with multiple turnovers.[33,37,38] Thus, *in vitro* selected nucleic acids are not necessarily limited to performing self-modification reactions, even though it is self-modification that is usually the basis on which the molecules are selected.

Using Portable Binding Sites To Get a Grip on Cofactors

All of the ribozymes described above require a divalent cation for activity. Some of these metal ions almost certainly act as cofactors in the reactions (i.e., they are catalytic, as opposed to structural, metals). Yarus[39] has suggested that a ribozyme, in its simplest form, can be viewed as an RNA scaffold that holds metal ion cofactors in the appropriate places to perform the required chemistry. What other sorts of cofactors might RNA be capable of using? If RNA can fold into shapes that specifically bind enzymatic cofactors, it seemed possible that we could select for ribozymes that bind these cofactors and use them to perform a chemical reaction.

We thought we might have a better chance of finding such ribozymes by building a largely random pool around a core domain consisting of an aptamer for an enzymatic cofactor, i.e., an RNA pool that was biased toward structures that bind the cofactor of interest, and around which a variety of different ribozymes all using this cofactor (e.g., ATP) could evolve. These aptamer domains could be thought of as RNA versions of such recurrent protein motifs as the "Rossman fold", a nucleotide binding domain found in a great many protein enzymes.[40] This combinatorial assembly of distinct functional domains is largely the mechanism that Gilbert proposed for the evolution of functional ribozymes in the RNA world.[41] Two such selections have been successfully carried out in our laboratory. Wilson and Szostak used a two-part selection scheme to isolate RNAs that performed an autoalkylation reaction using an iodoacetyl derivative of the cofactor biotin.[42] An RNA receptor for biotin was first isolated from a random RNA pool. A pool consisting of the mutagenized aptamer sequence flanked on either side by 20 random nucleotides was then constructed, from which a ribozyme was isolated that alkylates itself with the biotin derivative. Greater than 50% of the molecules in the final pool were derived from a single ancestral sequence. In order to improve the activity of this ribozyme, yet another mutagenized pool was made, and yet another selection was performed. The final ribozyme alkylates itself at a specific guanosine N7 with a k_{cat} of approximately 0.05 min^{-1}, corresponding to a rate enhancement of more than 2×10^7 compared to the uncatalyzed reaction.

We have also isolated ribozymes that utilize the ubiquitous cofactor and substrate, ATP.[43] We constructed a pool consisting of an ATP binding site [11] surrounded by 100 bases of random sequence (Figure 5). Using the thiol group from the ATP analogue ATP-γS as a handle to pull out RNAs that could transfer a (thio)phosphate group covalently to themselves, we isolated approximately 10^4 different molecules capable of utilizing ATP($-\gamma$S). When we increased the stringency of the selection to isolate only the best catalysts, the number of active molecules was reduced to roughly 50. Thus, in a region of sequence space clustered around an ATP binding domain, there is a truly remarkable number of RNA sequences capable of utilizing the chemical energy in the phosphoanhydride bonds of ATP. Of the 50 ribozymes in the final pool, about half catalyze the transfer of a (thio)phosphate to their 5′-hydroxyls and half to internal 2′-hydroxyls. One of these 5′ kinase ribozymes has been engineered to (thio)phosphorylate the 5′-hydroxyl of an exogenous oligonucleotide substrate with multiple turnover, thus acting as a true enzyme.

One might ask whether it actually helped to bias the pool toward binding of the substrate in these selections. No one has yet repeated these selections using completely random pools to see how the number and catalytic efficiency of the solutions compare to those from the aptamer-based pools. Both the self-alkylating ribozymes and the kinases retain some of the conserved parts of the aptamer domains, but discard or greatly change others.[42,43] Moreover, it is not yet certain that the regions apparently conserved between these ribozymes and the respective aptamers are actually involved in substrate binding. The large number of functional sequences isolated in the kinase selection (~10 000) compared to the relatively few isolated in the selection for ligases from a completely random pool (~65)[31] suggests that the addition of the ATP binding site aided in finding functional sequences (the two reactions are of similar difficulty; the catalytic efficiencies of the early kinases are similar to those of the original ligases). These data suggest that it probably helped to use the aptamer sequences as the core of the pools, but we cannot yet rule out the possibility that many of the same solutions could have been found in a completely random pool as well.

Natural vs *in Vitro* Evolution: From Ribozymes to Monkeys

It has been hypothesized that the first life forms on earth were based on self-replicating RNA molecules.[41] The dual nature of RNA as an information carrier and enzyme makes it well suited to such a task. This view has been expanded to suggest that complex cells existed in which ribozymes performed many of the catalytic functions currently performed by proteins.[44] What have *in vitro* selection and evolution taught us about the likelihood of such an "RNA world"? First, the fact that RNA can catalyze the formation of 3′−5′ phosphodiester bonds using a pyrophosphate leaving group,[31] as well as other leaving groups (e.g., guanosine, imidazole), coupled with the fact that RNA can fold into structures that specifically bind nucleotide triphosphates,[8,11] immediately suggests that there may be RNA sequences that are capable of condensing nucleotide triphosphates in a template-directed manner. The selection of such ribozymes (RNA replicases) is currently a very exciting and rapidly progressing area of research.

In vitro selection has also shown that RNA catalysis is not limited to phosphate chemistry.[25,42,45] The first

(39) Yarus, M. *FASEB J.* **1993**, *7*, 31−39.
(40) Rossman, M. G.; Moras, D.; Olsen, K. W. *Nature* **1974**, *250*, 194−199.
(41) Gilbert, W. *Nature* **1986**, *319*, 618.
(42) Wilson, C.; Szostak, J. W. *Nature* **1995**, *374*, 777−782.
(43) Lorsch, J. R.; Szostak, J. W. *Nature* **1994**, *371*, 31−36.
(44) Benner, S.; Ellington, A. D.; Tauer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *86*, 7054−7058.

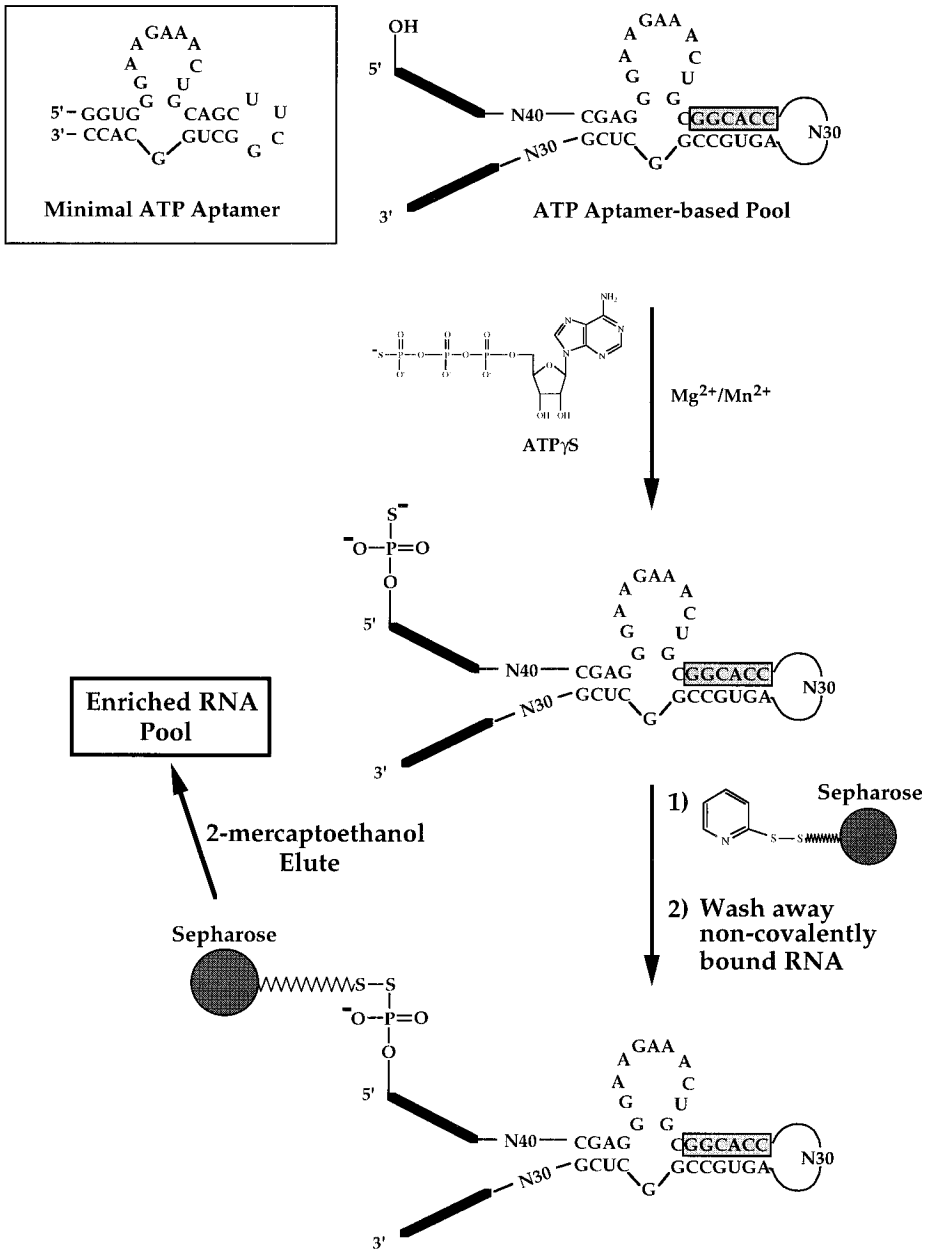

Figure 5. Selection scheme used by Lorsch and Szostak[43] to isolate polynucleotide kinase ribozymes. The boxed structure is a small RNA motif that binds ATP.[11] A partially random pool was generated by adding random sequences to the beginning, middle, and end of the ATP aptamer. Incubation of this pool with ATP-γS resulted in rare catalytic sequences tagging themselves by transfer of a thiophosphate group to their 5′-hydroxyl (or an internal 2′-hydroxyl). The active sequences were then purified by allowing the RNA to react with thiopyridine-activated thiopropyl Sepharose, followed by elution with 2-mercaptoethanol, and amplification. The cycle of selection and amplification was repeated until catalytic sequences dominated the pool. Reprinted with permission from ref 43. Copyright 1994 Nature.

demonstration that this was true was provided by Cech and co-workers,[46] who showed that the group I intron from *Tetrahymena* catalyzes (by about 10-fold) the hydrolysis of an ester on the 2′(3′)-hydroxyl of a small oligonucleotide complementary to the ribozyme's internal guide sequence (the oligonucleotide substrate binding site of the ribozyme). Recently, Yarus and colleagues have greatly expanded on these results and have used *in vitro* selection to isolate a ribozyme that can aminoacylate its own 2′(3′)-hydroxyl using as a substrate the phosphoanhydride of phenylalanine and AMP (Phe-AMP), at least 10^5 times faster than the background reaction.[45] The ribozyme is both Mg^{2+}- and Ca^{2+}-dependent. The authors raise the intriguing possibility that Ca^{2+}, which is known to be able to coordinate with sugar hydroxyls, may be used by the ribozyme to increase the nucleophilicity of the 2′- and/or 3′-hydroxyls. This ribozyme may be the specter of a long-vanished class of RNAs that charged their own 2′(3′)-hydroxyls with an amino acid. Such amino acid-charged RNAs could then react with other charged RNAs to form peptides. As we already know that RNAs can utilize the chemical energy in the phosphoanhydride bonds of ATP,[43] it does not seem too unreasonable to think that RNAs might exist that can synthesize the aminoacyl adenylates, as well.

RNA can clearly perform chemistry on the relatively

(45) Illangasekare, M.; Sanchez, G.; Nickles, T.; Yarus, M. *Science* **1995**, *267*, 643−647.
(46) Piccirilli, J. A.; McConnell, T. S.; Zaug, A. J.; Noller, H. F.; Cech, T. R. *Science* **1992**, *256*, 1420−1424.

labile ester group, but if an RNA world was to have evolved into a protein world, it is likely that RNA would have to have been able to catalyze reactions involving the considerably more stable amide group. Noller and co-workers[47] have provided evidence that the peptidyl transferase activity of the ribosome may be due, at least in part, to RNA rather than protein. More recently, Dai et al.[48] have shown that a version of the *Tetrahymena* group I intron that had previously been evolved to cleave DNA can accelerate the hydrolysis of an amide bond by over 3 orders of magnitude. The wild-type intron does not catalyze the reaction, and thus there is something about the active site of the ribozyme optimized for DNA cleavage that allows the acceleration of amide hydrolysis. We suspect that RNAs capable of catalyzing the formation of an amide bond will be described in the near future.

Guildenstern: Is that *it*, then? Is that all?

Rosencrantz: What?

Guildenstern: A new record? Is that as far as you are prepared to go?

Rosencrantz: Well...

Guildenstern: No questions? Not even a pause?

Challenges for *in Vitro* Selection

In vitro selection is in essence a technological development resulting from a combination of advances in DNA synthesis, separations technology, and sequence amplification technology, and as such it may be expected to benefit from further advances in these contributing areas. We will therefore first discuss a number of technical advances that are under active investigation and may be expected to greatly increase the power and widespread utility of this procedure. We will then discuss a few of the major biochemical and biomedical problems that *in vitro* selection may help to solve.

Perhaps the most important technical advance would be the development of methods for the direct selection of RNAs capable of multiple-turnover catalysis. All present selections are for self-modifying RNAs, since self-modification provides the basis for the separation of active from inactive RNA molecules. The attachment of a substrate to all RNAs in a pool by a long flexible tether could in principle provide a compromise situation in which active RNAs still become tagged by covalently attached product, while retaining activity against free (untethered) substrate. However, a more general approach would be to colocalize RNA and substrate by liposomal encapsulation, and to purify active RNAs by picking out vesicles containing accumulated product.

While direct selections for catalysis have so far proven to be more effective than indirect selections for binding to transition state analogues, improvements in the design of transition state analogues may also lead to more effective ways of isolating new polynucleotide catalysts. The transition state analogue approach has the virtue that, if catalysts can be found, they will usually catalyze the desired reaction with multiple turnover.

Refinements in the design of random and biased pools may also be expected to facilitate the isolation of interesting new catalysts. There has been considerable debate about the optimal length of the random sequence region in pools used to isolate new ribozymes. Longer pools provide a linear increase in the probability of finding any single sequence and a combinatorial increase in the probability of finding interacting sequence motifs, and they are of course necessary for the isolation of long complex structures such as the recently described class I ligase.[33] However, as pool length increases, "poisoning" of functional molecules by inhibitory sequences may also increase. In a number of RNAs selected in this laboratory, partial deletions often produce molecules with lowered activity, while larger deletions restore activity to the level of the full-length RNA. It is not yet clear what the optimal pool length is, or even whether there is an optimal length.

The presence of unwanted selective pressures during amplification is an unavoidable problem, the magnitude of which is still unclear. The amplification steps, because they use enzymes that themselves have preferences for given sequences or structures, bias the outcome of *in vitro* selection. For example, highly structured molecules may not be efficiently reverse-transcribed or PCR-amplified, and thus the most interesting molecules could be lost during repeated rounds of the selection. The development of new amplification procedures, using higher temperatures and more thermostable enzymes, added denaturants, single-strand binding protein,[49] etc., will be needed to circumvent these problems. Simpler, more robust and more efficient selection and amplification procedures could speed up, and ultimately lead to the automation of, the whole process of the *in vitro* evolution of new useful catalysts.

In vitro selection has already been used to answer questions about biological systems, largely by providing an efficient way of defining protein binding sites on RNA and DNA (for reviews, see refs 50 and 51). Although good evidence is lacking, RNA structures that bind small molecules may be involved in *in vivo* regulation of transcription or translation, raising the possibility that RNA motifs defined by *in vitro* selection experiments could be found in biological systems by screening the rapidly growing biological sequences database.[57] Selections for aptamers or catalysts from RNA pools transcribed from genomic or cDNA may lead to the identification of new RNA functions *in vivo*. The techniques of *in vitro* selection could also be applied *in vivo* to address questions of RNA localization and stability.

(47) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1420−1424.
(48) Dai, X.; Mesmaeker, A. D.; Joyce, G. F. *Science* **1995**, *267*, 237−240.
(49) Crameri, A.; Stemmer, W. P. C. *Nucleic Acids Res.* **1993**, *21*, 4410.
(50) Gold, L.; Allen, P.; Binkley, J.; Brown, D.; Schneider, D.; Eddy, S. R.; Tuerk, C.; Green, L.; MacDougal, S.; Tasset, D. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1993; pp 497−509.
(51) Bartel, D. P.; Szostak, J. W. In *RNA−Protein Interactions*; Nagai, K., Mattaj, I. W., Eds.; IRL Press: Oxford, 1994; pp 248−265.
(52) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Varmaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564−566.
(53) Tuerk, C.; MacDougal, S.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6988−6992.
(54) Pan, T.; Uhlenbeck, O. C. *Nature* **1992**, *358*, 560−563.
(55) Pan, T.; Dichtl, B.; Uhlenbeck, O. C. *Biochemistry* **1994**, *33*, 9561−9565.
(56) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1995**, *34*, 15315−15327.
(57) Richter-Dahlfors, A. A; Anderson, D. I. *Mol. Microbiol.* **1992**, *6*, 753−749.

One of the major motivations for work with *in vitro* selection is the hope that these techniques will generate therapeutically useful molecules. Aptamers that bind to and inhibit thrombin,[52] HIV reverse transcriptase,[53] and other physiologically important targets have been characterized. Ribozymes that cleave DNA could potentially recognize and destroy DNA viruses, or the integrated DNA copies of retroviruses. Breaker and Joyce[37] point out that catalytic DNA, being more stable than RNA, could be useful in gene therapy. Other stable nucleic acid analogues (e.g., phosphorothioates, 2′-*O*-methyl, 2′-fluoro or 2′-amino polynucleotides) could also be used to select for new aptamers or catalysts with therapautic goals in mind.

Quite apart from issues of application, the most fascinating opportunity raised by *in vitro* selection is the ability to study catalysis by biological macromolecules, not after four billion years of multifactorial optimization, but in their original state, newly emerged from the chaos of random sequences. Efforts to select for catalysts that may be similar to those that existed in the very earliest cells have already been discussed, but *in vitro* evolution also allows us to trace the various pathways by which primitive catalysts can evolve, step by step, into the remarkably efficient catalysts so common in current biology. With these advances many fundamental questions concerning catalysis by macromolecules have become experimentally accessible.

Why are enzymes (including ribozymes) generally so large? While it is obviously possible for small RNAs to fold into catalytically active structures (the hammerhead ribozyme, for example, is only ~40 nucleotides), it may be difficult for such small RNAs to rigidly position the number of functional groups required for large transition state stabilizations. Larger molecules, on the other hand, can recruit peripheral structural units to act as an exoskeleton to stabilize their catalytic cores, in essence paying the entropic price required to freeze the transition state stabilizing conformation with the enthalpy derived from intramolecular contacts. A comparison of ribozymes derived from short vs long random pools could shed light on this question.

Can substrate binding sites easily evolve into catalysts? In other words, how easy is it to start with an RNA that binds a substrate and mutate it into one that preferentially binds the transition state? It has not yet been definitively established that the incorporation of a binding domain into a pool increases the likelihood of finding efficient catalysts in a selection. In order to address this question it will be necessary to perform parallel *in vitro* selection experiments using a completely random pool and a pool containing a mutagenized aptamer domain embedded in random regions.

What catalytic mechanisms are easiest to implement in RNA (or DNA) hardware? There are many ways that ribozymes could, in principle, act to stabilize the transition state for a given reaction, ranging from simple substrate proximity and orientation to transition state specific hydrogen bonds and electrostatic interactions, to acid−base catalysis and metal-ion enhancement of nucleophilicity. It will be very interesting to see what kinds of mechanisms will be most commonly used by the primitive and poorly optimized catalysts resulting from *in vitro* selection experiments, and to see whether the *in vitro* optimization of these catalysts results simply from the refinement of the original mechanisms, or also from the recruitment of additional catalytic mechanisms. The few studies of the workings of non-natural ribozymes that have been carried out[33,54−56] suggest that there is much to be learned about how RNA and DNA structures act to catalyze reactions.

We are grateful to the members of our laboratory for helpful comments on this paper and to NASA, the NIH, the NSF, and Hoechst AG for supporting the research on which this review is based.

AR9501378