6

Did the RNA World Exploit an Expanded Genetic Alphabet?

Steven A. Benner, Petra Burgstaller, Thomas R. Battersby, and Simona Jurczyk
Departments of Chemistry and Anatomy and Cell Biology
University of Florida, Gainesville, Florida 32611 and
Sulfonics, Inc., Alachua, Florida 32615

SINGLE BIOPOLYMER LIFE FORMS BASED ON RNA

In terms of its macromolecular chemistry, life on Earth can be classified as a "two-biopolymer" system. Nucleic acid is the encoding biopolymer, storing information within an organism and passing it to its descendants. Nucleic acids also direct the biosynthesis of the second biopolymer, proteins. Proteins generate most of the selectable traits in contemporary organisms, from structure to motion to catalysis.

The two-biopolymer strategy evidently works rather well. It has lasted on Earth for several billion years, adapting in this time to a remarkable range of environments, surviving formidable geobiological (and perhaps cosmic) events that threatened its extinction, and generating intelligence capable of exploring beyond Earth.

The terrestrial version of two-biopolymer life contains a well recognized paradox, however, one relating to its origins. It is difficult enough to envision a nonbiological mechanism that would allow either proteins or nucleic acids to emerge spontaneously from nonliving precursors. But it seems astronomically improbable that both biopolymers arose simultaneously and spontaneously, and even more improbable (if that can be imagined) that both biopolymers so arose with an encoder-encoded relationship.

Accordingly, a variety of "single-biopolymer" models have been proposed as forms of life that antedated the two-biopolymer system. These (presumably) could have emerged more easily than a two-biopolymer system. Such models postulate that a single biopolymer can perform the catalytic and information repository roles and undergo the Darwinian

evolution that defines life (Joyce 1994). For example, Rich (1962), Woese (1967), Orgel (1968), and Crick (1968) proposed that the first biopolymeric system that sustained Darwinian evolution on Earth was RNA. Usher and McHale (1976), White (1976), Visser and Kellogg (1978), and Benner et al. (1989) expanded on this proposal, recognizing that key elements of contemporary metabolism might be viewed as vestiges of an "RNA World" (Gilbert 1986), a time when the only encoded component of biological catalysis was RNA. The phenomenal discoveries by Cech, Altman, and their coworkers (Cech et al. 1981; Guerrier-Takada et al. 1983; Zaug and Cech 1986) showing that RNA performs catalytic functions in contemporary organisms has made the RNA World a part of the culture of contemporary molecular biology (Watson et al. 1987).

The notion that the RNA World was metabolically complex follows from the abundance of its vestiges in modern metabolism (Benner 1988; Benner et al. 1989). RNA fragments play roles in modern metabolism for which they are not intrinsically chemically suited, most notably in RNA cofactors such as ATP, coenzyme A, NADH, FAD, and *S*-adenosylmethionine. This suggests that these fragments originated during a time in natural history when RNA was the only available biopolymer, rather than by convergent evolution or recruitment in an environment where chemically better suited biomolecules could be encoded. If the RNA World developed ATP, coenzyme A, NADH, and *S*-adenosylmethionine, it follows that the RNA World needed these for some purpose, presumably for phosphorylations, Claisen condensations, oxidation-reduction reactions, and methyl transfers, respectively (White 1976; Visser and Kellogg 1978; Benner et al. 1989). This in turn implies complexity in the metabolism encoded by RNA-based life, implying in turn that RNA can catalyze a wide variety of chemical reactions. Conversely, the intellectual contribution of the RNA World model would be diminished were it not to embody a complex metabolism catalyzed by ribozymes, as there would then be no coherent explanation for the structures of contemporary RNA cofactors.

Accordingly, hopes were high when Szostak (1988), Joyce (1989a,b), Gold (Irvine et al. 1991), and their coworkers introduced in vitro selection as a combinatorial tool to identify RNA molecules within a pool that catalyze specific reactions. Elegantly conceived, the approach seemed likely to lead to the ultimate goal, the generation of an RNA (or DNA) molecule that would catalyze the template-directed polymerization of RNA (or DNA), a molecular system able to undergo Darwinian evolution. If selection procedures were appropriately designed, they should also produce RNA catalysts for almost any other reaction as well, at least if the RNA World model as elaborated above were a correct representation of natural history.

LIMITATIONS OF RNA AS A CATALYST

In contrast with these hopes (and only by this contrast), in vitro selection has been disappointing. RNA has proven to be an intrinsically poor matrix for obtaining catalysis, especially when compared with proteins. For example, to have a 50% chance of obtaining a single RNA molecule capable of catalyzing a template-directed ligation reaction by a modest (by protein standards) factor of 10,000, Bartel and Szostak estimated that one must sift through 2×10^{13} random RNA sequences 220 nucleotides in length (Bartel and Szostak 1993). Although many laboratories have tried, only a few have managed to extend the scope of RNA catalysis beyond the phosphate transesterification reactions where it was originally observed. For example, attempts to obtain an RNA catalyst for a Diels-Alder reaction using in vitro selection failed (Morris et al. 1994); the same reaction is readily catalyzed by protein antibodies (Gouverneur et al. 1993). Attempts to obtain RNA that catalyzes amide synthesis have succeeded, but with difficulty (Wiegand et al. 1997; Zhang and Cech 1997). The fact that such successes came only after many attempts is indicative of a relatively poor catalytic potential in oligonucleotides.

The comparison with peptides is instructive. For example, short (14 amino acids) peptides accelerate the rate-determining step for the amine-catalyzed decarboxylation of oxaloacetate by more than three orders of magnitude (Johnsson et al. 1990, 1993), not far below the acceleration observed for the first-generation ligases observed in the Bartel–Szostak selection beginning with 10^{13} random RNA sequences. Furthermore, the peptide is less than 10% the size of the RNA motif. Combinatorial experiments starting from this design (Perezpaya et al. 1996; L. Baltzer, pers. comm.) suggested that perhaps only 10^{7} random sequences must be searched to get a similar catalytic effectiveness as is observed in a library of 10^{13} RNA molecules. This suggests that peptides are intrinsically a million-fold fitter as catalysts than RNA.

The comparison is imperfect, of course, because it involves different reactions and different design strategies. This imperfection characterizes most of the comparisons that can be made at present. Not surprisingly, ribozymes are most frequently sought for reactions where oligonucleotides are most likely to be effective catalysts (for example, where oligonucleotides themselves are substrates), whereas peptide catalysts are most frequently sought for reactions suited for peptide catalysts (for example, those that make use of functional groups found on amino acid side chains). This makes the comparison nonquantitative, but useful nevertheless as an estimate of how well oligonucleotides and oligopeptides, respectively, perform when challenged by their favorite target reactions.

THE CHEMISTRY OF FUNCTIONAL CATALYSIS

The apparent superiority of proteins as catalysts compared with RNA reflects (at the very least) the availability to proteins of a wider range of building blocks and catalytic functionality than available in RNA. RNA lacks the imidazole, thiol, amino, carboxylate, and hydrophobic aromatic and aliphatic groups that feature so prominently in protein-based enzymes and has only hydroxyl, polar aromatic, and phosphate groups. An uncounted number of studies with natural enzymes and their models has illustrated the use of this functionality by protein catalysts (Dugas 1989).

Proteins also have advantages as catalysts over nucleic acids in their greater propensity to "fold." As is well known from the statistical mechanics of polymers, the repeating negative charge of the polynucleotide backbone causes the polymer to favor an extended structure (Flory 1953; Richert et al. 1996). Accordingly, the most prominent physical characteristics of nucleic acids are their solubilities in water, their ability to bind other oligonucleotides following simple rules, and their constancy of physical behavior over a wide range of sequences. In contrast, the most prominent physical characteristic of peptides is their propensity to fold, best known as a propensity to precipitate (which is, of course, a type of folding, in that peptide interacts with peptide rather than with water). A catalyst must fold if it is to surround a transition state and be effective, providing another reason that peptides might be intrinsically better catalysts than RNA (Benner 1989).

If it is necessary to generate trillions of long, random RNA sequences in order to have a 50% likelihood of finding one that catalyzes even modestly a simple ligation (a reaction that itself assumes the preexistence of long RNA molecules that act as templates and substrates), how many more random sequences must be generated to obtain a template-directed RNA polymerase? We cannot say, because such a ribozyme has not been generated. An optimistic guess is 10^{20}. This, the difficulty of obtaining plausible prebiotic syntheses of RNA molecules (but see Müller et al. 1990), and the observation that racemic mixtures of RNA do not effectively undergo abiological polymerization (see, e.g., Schmidt et al. 1997) have prompted many to question the RNA World as a viable model for generating the first life on Earth (Joyce et al. 1987; Miller 1997). The critique acknowledges the premise that the single-biopolymer system is more plausible as a first life form than the two-biopolymer system. It continues, however, by holding that the chemical properties of RNA are such that it could not have been the first living biopolymer, as it is too difficult to generate under abiotic conditions and provides too little catalytic power even if it could be generated.

EXPANDING THE STRUCTURAL REPERTOIRE OF NUCLEIC ACIDS

A decade ago, the intrinsic limitations of standard nucleic acids as a biopolymer for obtaining functional behavior under conditions of Darwinian selection were discussed, and several solutions to these limitations were proposed (Benner et al. 1987; Benner 1988, 1989; Switzer et al. 1989; Piccirilli et al. 1990). Each of these involved an expedient by which additional functionality was provided to the RNA.

One expedient was obvious. RNA might gain functionality using cofactors, much as contemporary proteins gain the functionality that they lack through vitamins.

A second solution was to append functionality to the standard nucleotides. Prompting this suggestion was the observation that contemporary tRNA and rRNA contain much of the functionality found in proteins but lacking in contemporary encoded RNA, including amino, carboxylate, and aliphatic hydrophobic groups (Fig. 1) (Limbach et al. 1994). These functional groups are introduced by posttranscriptional modification of encoded RNA. Some of these might even be placed by parsimony in the protogenome, the reconstructable genome at the trifurcation in the evolutionary tree joining the archaebacterial, eubacterial, and eukaryotic kingdoms (Benner et al. 1989; Limbach et al. 1994).

The third approach to expand the functional diversity of nucleic acids pursued the possibility of expanding the number of base pairs from the four found in standard oligonucleotides to include some of the

Figure 1 Transfer RNA contains a rich collection of functionalized standard nucleobases, created by posttranscriptional modification, that deliver functional groups (amino groups, carboxylic acid groups, aliphatic hydrophobic groups, in *green*) not found within unmodified RNA. Could these be vestiges of functionalized RNA originating in the RNA World?

nonstandard hydrogen-bonding patterns permitted by the geometry of the Watson-Crick base pair (Fig. 2) (Switzer et al. 1989; Piccirilli et al. 1990). Additional letters in the genetic alphabet could carry a richer diversity of functionality. Indeed, one might imagine a new type of biopolymer, one carrying functionalization like proteins but able to be copied like nucleic acids (Fig. 3) (Kodra and Benner 1997).

In a sense, the first approach had already been implemented in 1987. Most ribozymes require one or more metal ions to be effective catalysts. The metal ions are not encoded in the RNA sequence, provide a needed electrophilic center, and therefore compensate for the limited catalytic functionality of the biopolymer itself. Thus, metals can be considered to be "cofactors," and clearly improve the catalytic functionality of RNA. More recently, Breaker and his coworkers have expanded the approach to include organic molecules as second ligands in riboenzymes (Tang and Breaker 1997).

In contrast, the second and third approaches were far from implementation in 1987. Although standard bases carrying functionality were known to form stable base pairs and, in some cases, be accepted by polymerases (Prober et al. 1987), it was not clear that nonstandard bases would pair as expected, or whether polymerases would incorporate functionalized standard bases and nonstandard bases (Figs. 2 and 3) with sufficient speed and fidelity to be used in in vitro selection experiments. Furthermore, it was not known whether in vitro selection based on an expanded genetic alphabet might improve the binding and catalytic versatility of RNA.

Developing in vitro selection with an expanded genetic alphabet proved to be more difficult than developing in vitro selection with the standard nucleotides (A, T, G, and C), which was enabled by a rich collection of molecular biological tools. Nonstandard nucleobases needed to be synthesized (Switzer et al. 1989; Piccirilli et al. 1990; Vögel et al. 1993a; Vögel and Benner 1994). Their structures needed to be optimized for stability and pairing (Piccirilli et al. 1991a,b; Vögel et al. 1993b). New protecting group chemistry needed to be developed to permit automated synthesis of oligonucleotides containing them (Huang and Benner 1993; von Krosigk and Benner 1995). Polymerases were needed to catalyze their incorporation into oligonucleotides by the polymerase chain reaction (Horlacher et al. 1995; Lutz et al. 1996). These studies have been paralleled by work to append still more functionality onto standard nucleobases (Dewey et al. 1996; Kodra and Benner 1997). These experiments have established the chemistry of both functionalized standard and nonstandard nucleotides, and laid the ground for the first in vitro selection experiments using these.

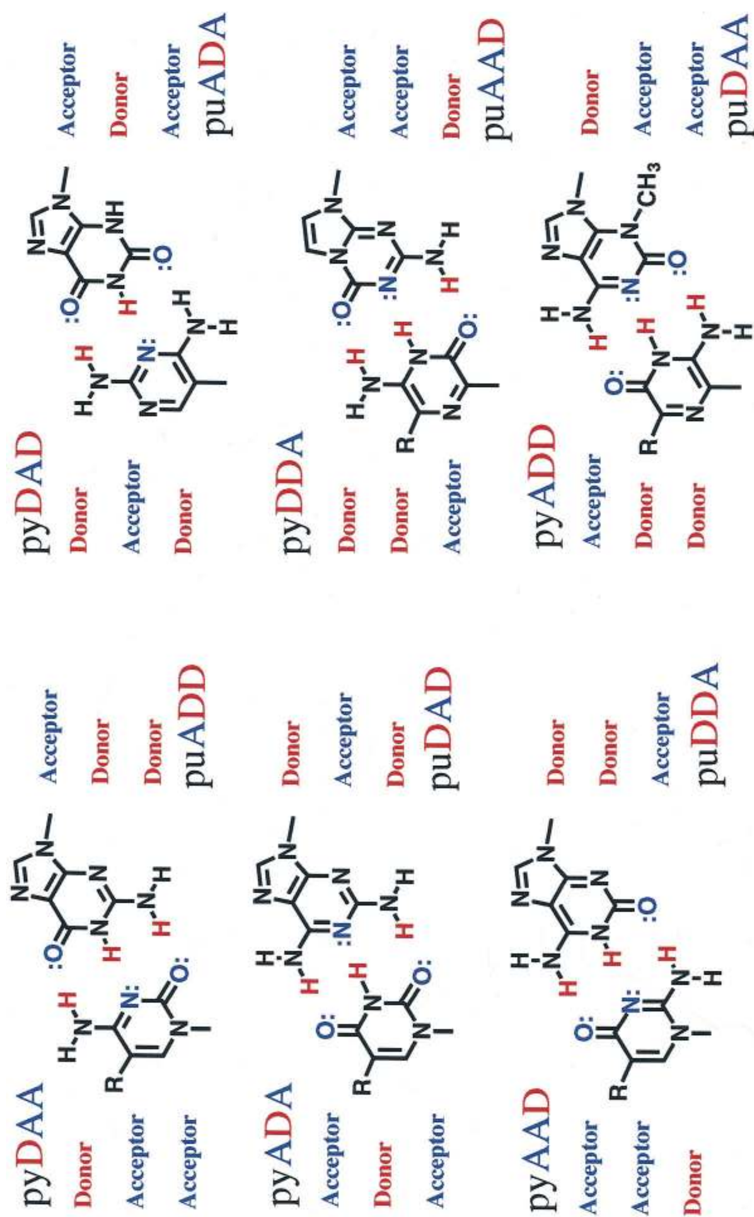

Figure 2 Twelve bases that are possible in a DNA- or RNA-based "alphabet" within the constraints of the Watson-Crick base-pair geometry. Pyrimidine base analogs are designated by "py," purine by "pu." The uppercase letters following the designation indicate the pattern of hydrogen bonding acceptor (A) (in *blue*) and donor (D) (in *red*) groups. Thus, cytosine is pyDAA, guanine is puADD, adenine is puDA- (diaminopurine, puDAD, completes the Watson-Crick base pair), and thymine is pyADA. The remainder of the base pairs are joined by nonstandard hydrogen-bonding schemes.

THE RNA WORLD HAD THE MOTIVE TO EXPLOIT MODIFIED NUCLEOTIDES

With these chemical developments, it has been possible recently to make a convincing, if not compelling, argument that the RNA World had both the motive and the opportunity to exploit nonstandard and functionalized nucleobases. Three results are central to this argument.

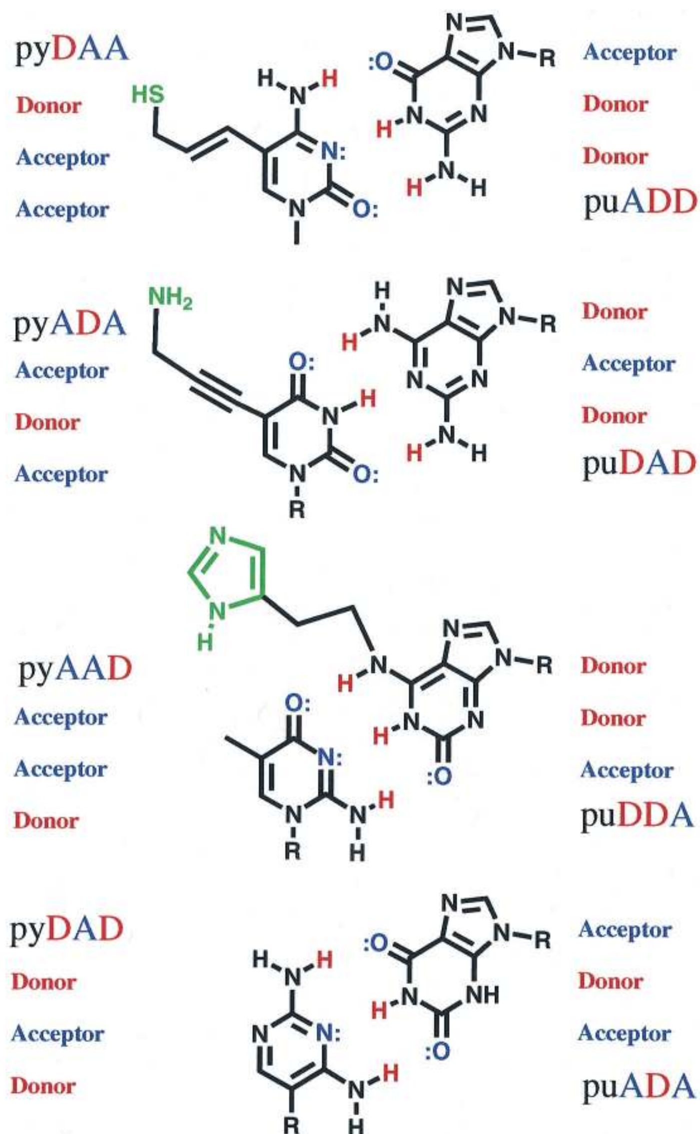

Figure 3 Nonstandard and standard nucleobases with functionality (in *green*). Note that the pyDAD nucleobase can be protonated below pH 7 (pKa = 7.4).

First, functionality has been incorporated into an RNA molecule that catalyzes a Diels-Alder reaction (Tarasow et al. 1997), starting from a functionalized standard pyADA nucleobase (Fig. 4, right). A selection starting with a library that did not contain functionalized nucleotides failed to yield a catalyst (Morris et al. 1994). The successful experiment with the functionalized pyADA base selected directly for a Diels-Alderase, however, whereas the experiment on the unfunctionalized library sought a Diels-Alderase by selecting for RNA molecules that bound to a transition-state analog for the reaction. The different selection strategies prevent us from saying conclusively that this particular functionalized nucleoside improves the intrinsic power of RNA as a catalyst for Diels-Alder reactions. Experiments that bear on this question will undoubtedly emerge soon.

Another functionalized selection experiment does support this conclusion. Burgstaller, Jurczyk, Battersby, and Benner prepared a different functionalized implementation of the pyADA nucleobase (trivially designated "J," Fig. 4) and incorporated it into an in vitro selection experiment seeking receptors for an adenosine derivative (P. Burgstaller et al., unpubl.). This experiment was done in strict parallel with experiments done by Huizenga and Szostak (1995) using a standard, unfunctionalized DNA library.

The functionalized library containing J yielded new motifs as receptors for ATP, including the following (the randomized region is underlined):

GGTCGTCTAGAGTATGCGGTAG<u>GAACGJCAGJGGGGGGAGCA</u>
<u>JAJGGJGJGAJACGCGA</u>CCGAAGAAGCJJGGCCCAJG

The motif prepared with unfunctionalized T replacing J does not bind ATP, suggesting that the ammonium functionality carried by J is essential for the binding properties of the new motif. A gel filtration experiment

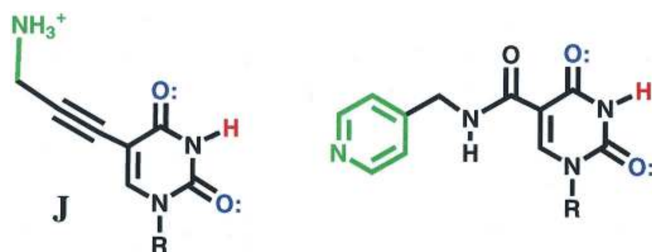

Figure 4 Functionalized standard bases that have been used in in vitro selections. Functional groups are shown in *green*, with the hydrogen-bonding acceptor and donor in *blue* and *red*, respectively.

was used to obtain an equilibrium binding constant (K_d) of 40 nM for affinity of this aptamer and ATP. This value is approximately 2 orders of magnitude greater than the reported K_d for the binding of RNA (Sassanfar and Szostak 1993) and DNA aptamers containing only standard bases to ATP (Huizenga and Szostak 1995). With the caveats that elution experiments permit only estimates of binding constants, and that further experiments with a wider range of ligands must be completed, these results suggest that introduction of a new functionality (an ammonium group bearing a positive charge) enhances the intrinsic value of a DNA library as a source of receptors by about two orders of magnitude.

These experiments make clear that functionalized oligonucleotides are superior to standard oligonucleotides as a matrix for generating receptors and catalysts. This would have given the RNA World a motivation to use functionalized oligonucleotides and an expanded genetic alphabet in its effort to generate diverse catalysts.

But did it? The third result comes from the field of "prebiotic chemistry," which seeks to discover ways by which the components of living systems might have emerged in the early Earth. Robertson and Miller (1995) showed how the intrinsic nucleophilicity of the 5-position of pyrimidines such as uracil might be exploited to generate functionalized uracil derivatives that carry positive charges at the 5-position under abiological conditions. Analogous chemistry can be used to generate other functionalized derivatives. The products resemble the amino group functionalized uracils found in some tRNA molecules (Fig. 1). This suggests that the RNA World may have had the opportunity to use some functionalized nucleosides when life first emerged on Earth.

Could nonstandard nucleobases (Fig. 2) also have been available during early episodes of life on Earth? The success of prebiotic chemists in generating organic species under prebiological conditions has expanded greatly the spectrum of molecules that might have been accessible to early life. Indeed, prebiotic chemistry might have been too successful, in that relatively simple prebiotic models can generate organic mixtures containing perhaps too many products (Khare et al. 1993). Contemporary prebiotic chemistry must become less an effort to show that a given moiety might be generated under prebiotic conditions, and more an effort to show how a useful moiety (such as a heterocycle or a ribose) arising under prebiotic conditions might be converted into one or more of its delicate derivatives (such as nucleosides) in the presence of organic gunk that emerges from a typical prebiotic experiment.

Notwithstanding these issues, several of the nonstandard nucleobases in Figure 2 do not appear to be less prebiotic than the standard nucle-

obases. The puADA nucleobase is, for example, a simple deamination product of the puADD base (also known as guanine). Thus, if guanine was generated on a prebiotic earth, puADA was a fortiori also generated on a prebiotic earth. Similar arguments can be made for the puDDA and pyAAD nucleobases. This suggests that if the RNA World had the opportunity to use the standard genetic alphabet, it may also have had the opportunity to use an expanded genetic alphabet.

CONTRADICTORY CHEMICAL REQUIREMENTS FOR CATALYSIS AND INFORMATION STORAGE

This evidence suggests that the RNA World had both access to a functionalized and/or expanded genetic alphabet and the motivation to use it. The case is made stronger by the functionalized nucleotides found in contemporary tRNA and rRNA (Fig. 1), presuming that these are vestiges of an RNA World.

Even assuming that further experimental work demonstrates the full catalytic potential of functionalized and expanded genetic alphabets, it is still not clear that they will support single-biopolymer systems of life, however. To support a self-sustaining chemical system capable of undergoing Darwinian evolution (Joyce 1994), a biopolymer must be able to search mutation-space independent of concern that it will lose properties essential for replication. We designate polymers that have this property as COSMIC-LOPER biopolymers (Capable Of Searching Mutation-space Independent of Concern over Loss Of Properties Essential for Replication), and comment briefly on the chemical constraints placed on biopolymers likely to have this property.

The need for the single biopolymer to be COSMIC-LOPER to support Darwinian evolution is nearly axiomatic. If a substantial fraction of the mutations possible within a genetic information system cause a biopolymer to precipitate, unfold, or otherwise no longer be recognizable by the catalyst responsible for replication, the biopolymer cannot evolve.

Curiously, catalysis on one hand and information storage on the other place competing and contradictory demands on molecular structure that make a single molecule that does both difficult to find. Specifically:

1. A biopolymer specialized to be a catalyst must have many building blocks, so that it can display a rich versatility of chemical reactivity. A biopolymer specialized to store information must have few building blocks, as a way of ensuring faithful replication (Szathmary 1992; Lutz et al. 1996).

2. A biopolymer specialized to be a catalyst must fold easily so that it can form an active site. A biopolymer specialized to store information should not fold easily, so that it can serve as a template.
3. A biopolymer specialized for catalysis must be able to change its physical properties rapidly with few changes in its sequence, enabling it to explore "function space" during divergent evolution. A biopolymer specialized to encode information must be COSMIC-LOPER, with its physical properties largely unchanged even after substantial change in its sequence, so that the polymer remains acceptable to the mechanisms by which it is replicated.

At the very least, a single biopolymer attempting to support Darwinian evolution must reflect some sort of structural compromise between these goals. No fundamental principle guarantees that a polymeric system will make this compromise in a satisfactory way, however. The demands for functional diversity, folding, and rapid search of function space might be so stringent, and the demands for few building blocks, templating ability, and COSMIC-LOPER ability so stringent, that no biopolymer structure achieves a suitable compromise.

Nor need a biopolymer exist that supports robust catalysis at the same time as it enables robust Darwinian evolution. If so, the single-biopolymer model for the origin of life would be unavailable as a solution to the "chicken-or-egg" paradox in the origin of two-biopolymer systems. Life would be scarce in the universe, and if a single-biopolymer system did arise, it would be poorly adaptable and easily extinguished by geobiological (and possibly cosmogenic) events. Conversely, if many polymeric systems exist that make an acceptable compromise between the demands of catalysis and the demands of information storage, life would have emerged rapidly via single-biopolymer forms and be abundant in the universe.

It is clear that proteins are not COSMIC polymers, even in cases where they can direct template-based replication (Lee et al. 1997). The physical properties of proteins (including their solubility) can change dramatically upon point mutation within the mutation space allowed by the 20 standard amino acids. Again, there are many examples of this phenomenon, but the peptides mentioned above that catalyze the decarboxylation of oxaloacetate are one. Altering their structure by a single acetyl group changes substantially their level of aggregation, and altering their internal sequence at a single residue changes substantially their helicity (Alleman 1989; Johnsson et al. 1990, 1993). If solubility and/or helicity are essential to the replicatability of a peptide template, a large range of plausible mutation would destroy it.

Natural oligonucleotides do not behave similarly. Indeed, molecular biologists rely on this fact. Every (or almost every) oligonucleotide will precipitate in ethanol. Every (or almost every, if we consider G-rich sequences [Wang and Patel 1994]) oligonucleotide will bind to its complement in a rule-based fashion. Every (or almost every) oligonucleotide will be a template for a polymerase. Every (or almost every) oligonucleotide will migrate as expected on an electrophoresis gel. This regularity is normal for oligonucleotides, but is exceptional for virtually every other class of organic molecule.

Even small steps taken from the natural backbone can destroy the COSMIC-LOPER properties of oligonucleotides. For example, work recently replaced the phosphate diester linkers in DNA and RNA by nonionic dimethylenesulfone linking units (Huang et al. 1991). The sulfone group is an "isosteric" and "isoelectronic" replacement for a phosphate. Nevertheless, these nonionic oligomers display some remarkable properties. First, they fold. For example, the octamer $A_{SO_2}U_{SO_2}G_{SO_2}G_{SO_2}U_{SO_2}$-$C_{SO_2}A_{SO_2}U$ folds in solution to give a folded form in water having a high melting temperature (~87°C) (Richert et al. 1996). Next, a synthetic intermediate leading to this oligosulfone was found to be a "catalyst" for a self-debenzoylation reaction (Richert et al. 1996). Still more remarkably, different oligosulfones evidently follow different strategies for folding and pairing. The dinucleotide analog $G_{SO_2}C$ in the crystal forms an antiparallel duplex approximately isomorphous with the analogous RNA (Roughton et al. 1995). In the crystal, the $A_{SO_2}T$ dinucleotide does not (Hyrup et al. 1995). The $U_{SO_2}C$ dinucleotide forms a complex featuring backbone-to-backbone and backbone-to-nucleobase hydrogen bonds (C. Richert, pers. comm.). Even within a relatively small search of sequence space, these nonionic oligonucleotide analogs retain no conformational or physical property that could be a ready basis for a common mechanism for replication. In this respect, oligosulfone analogs of DNA and RNA behave much the same as peptides and conventional small organic molecules, not the nucleic acids upon which they are modeled.

These results suggest that the need for a COSMIC-LOPER behavior is a strong constraint on what biopolymers might serve as the basis for single-biopolymer life. They also suggest that a polyelectrolyte (polyanion or polycation) structure is important for the COSMIC-LOPER behavior that we see in standard nucleic acids (Richert et al. 1996):

1. Phosphate groups force the interaction surface between strands as far distant from the backbone as possible, to the Watson-Crick "edge" of the nucleobases. Without interstrand phosphate-phosphate repulsion,

sugar-sugar interstrand interactions, sugar-backbone interstrand interactions, interactions between the sugar and backbone groups of one strand and the Hoogsteen edge of the nucleobases on the other, Hoogsteen-Hoogsteen interstrand interactions, and Watson Crick-Hoogsteen interstrand interactions all become important, and the recognition phenomenon ceases to be rule-based.

2. Phosphates discourage folding in an oligonucleotide molecule. The statistical mechanical theory of polymers suggests that the polyanionic backbone will cause natural oligonucleotides to adopt an extended structure (Flory 1953; Brant and Flory 1965). Nonionic oligonucleotide analogs should (and do) fold like peptides. By discouraging folding, the repeating polyanionic backbone helps oligonucleotides act as templates.
3. Electronic distribution in a molecule is described as an infinite series (monopole + dipole + quadrapole + . . .). The first nonvanishing term dominates. The repeating monopole (charge) in DNA makes dipolar interactions (hydrogen bonding) secondary to its properties, allowing the DNA molecule to mutate without changing greatly its physical behavior.

Returning to functionalized and expanded genetic alphabets (see figures), this discussion suggests that one must be careful when "decorating" oligonucleotides with functionality. At some level of functionalization, the COSMIC-LOPER properties that enable DNA and RNA to serve as an evolvable Darwinian system will be lost. Preliminary data suggest, for example, that extensive functionalization with hydrophobic side chains destroys these properties. It remains to be seen whether the level of functionality that must be introduced into DNA and RNA to enable it to support a complicated metabolism is greater than that required to destroy its COSMIC-LOPER properties.

CAN A SINGLE-BIOPOLYMER LIFE BE FOUND TODAY IN THE SOLAR SYSTEM?

Single-biopolymer models for Darwinian chemistry have relevance to the search for extraterrestrial life. For example, biologists have noted that the microfossils in the Allan Hills meteorite, which are as small as 20–100 nanometers across, are too small to be living cells (Kerr 1997). After all, the argument is made, the ribosome is 25 nm across, and ribosomes are a basic requirement for life.

This argument is, of course, narrowly formulated. Ribosomes are a basic requirement for life based on two biopolymers. If a single biopoly-

mer (such as RNA) can serve both genetic and catalytic functions, then ribosomes are not needed for life. Indeed, much of the metabolism of contemporary cells (aminoacyl tRNA synthetases and many amino acid biosynthesis enzymes, for example) comprising more than half of what is believed to be the core metabolism encoded by the protogenome (Benner et al. 1993) would also not be needed for life in an RNA World. A cell based on a single-biopolymer genetic system can be far smaller than one based on two biopolymers. This means that the fossils in the Martian meteorite structures are not too small to be remnants of a single-biopolymer form of life. Conversely, if the meteorite structures are indeed fossils, then they almost certainly are fossils of an organism that used only a single biopolymer as its molecular system capable of Darwinian evolution, and similar considerations should guide our search for nonterrean life.

The best place to search for single-biopolymer life may be here on Earth, however, assuming that terrestrial life originated here as a single-biopolymer Darwinian system. Whether such life remains on Earth depends on whether it was able to find a niche on the planet where it could compete with its descendants that developed two biopolymers. The superior power of proteins as catalysts provides presumptive arguments that a life form that did not exploit proteins as catalysts could not have competed with life that did. The biochemical innovation associated with translation almost certainly prompted an extinction more massive than the well-known extinctions at the end of the Cretaceous period.

A variety of ecological niches might provide single-biopolymer systems with an adaptive advantage over two-biopolymer systems, however, and may have provided ribo-organisms with the opportunity to survive on Earth even in the presence of two-biopolymer systems. For example, because cells containing single-biopolymer life can be much smaller than two-biopolymer cells, one-biopolymer life might have survived where small size offers a selective advantage. In subterranean matrices, for example, geological formations can have pore sizes that are too small to permit a two-biopolymer organism to live, but might permit a single-biopolymer cell to reside free from competition from its more adept protein-using cousins.

CONCLUSIONS

Experimental results suggest that the RNA World had both the opportunity and the motivation to use an expanded genetic alphabet. It remains to be seen how effectively functionalized oligonucleotides make a compromise between the structural demands for catalysis and the physical prop-

erties required for effective Darwinian evolution. Should experimental work show that they do so, we expect in vitro selections to provide effective new catalysts with the expanded genetic alphabet. In the most optimistic scenario, analogous single-biopolymer forms of life may be found elsewhere in the solar system, and perhaps in enclaves on planet Earth.

ACKNOWLEDGMENTS

We thank NASA, the Office of Naval Research, the National Institutes of Health, and the Swiss National Science Foundation for supporting some of the work described here. Petra Burgstaller acknowledges a fellowship from the Deutsche Akademische Austauschdienst.

REFERENCES

Allemann R.K. 1989. "Evolutionary guidance as a tool in organic chemistry." Ph.D. thesis no. 8804. Eidgenössische Technische Hochschule, Zürich, Switzerland.

Bartel, D.P. and Szostak J.W. 1993. Isolation of new ribozymes from a large pool or random sequences. *Science* **261:** 1411–1418.

Benner S.A. 1988. Reconstructing the evolution of proteins. In *Redesigning the molecules of life* (ed. S.A. Benner) pp. 115–175. Springer-Verlag, Heidelberg.

———. 1989. Enzyme kinetics and molecular evolution. *Chem. Rev* **89:** 789–806.

Benner S.A., Ellington A.D., and Tauer A. 1989. Modern metabolism as a palimpsest of the RNA World. *Proc. Natl. Acad. Sci.* **86:** 7054–7058.

Benner S.A., Cohen M.A., Gonnet G.H., Berkowitz D.B., and Johnsson K. 1993. Reading the palimpsest: Contemporary biochemical data and the RNA World. In *The RNA world* (ed. R. Gesteland and J. Atkins) pp. 27–70. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Benner S.A., Allemann R.K., Ellington A.D., Ge L., Glasfeld A., Leanz G.F., Krauch T., MacPherson L.F., Moroney S.E., Piccirilli J.A., and Weinhold E.G. 1987. Natural selection, protein engineering and the last riboorganism: Rational model building in biochemistry. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 53–63.

Brant D.A. and Flory P.J. 1965. The configuration of random polypeptide chains. Experimental results. *J. Am. Chem. Soc.* **87:** 2788–2791.

Cech T.R., Zaug A.J., and Grabowski P.J. 1981. In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*. Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27:** 487–496.

Crick F.H.C. 1968. The origin of the genetic code. *J. Mol. Biol.* **38:** 367–379.

Dewey T.M., Zyzniewski M.C., and Eaton B.E. 1996. The RNA world. Functional diversity in a nucleoside by carboxyamidation of uridine. *Nucleosides Nucleotides* **15:** 1611–1617.

Dugas H. 1989. *Bioorganic chemistry*, 2nd edition. Springer-Verlag, New York.

Flory P.J. 1953. *Principles of polymer chemistry*. Cornell University Press, Ithaca, New York.

Gilbert W. 1986. The RNA world. *Nature* **319:** 618.

Gouverneur V.E., Houk K.N., Depascualteresa B., Beno B., Janda K.D., and Lerner R.A. 1993. Control of the exo-pathway and the endo-pathway of the Diels-Alder reaction by antibody catalysis. *Science* **262:** 204–208.

Guerrier-Takada C., Gardiner K., Marsh T., Pace N., and Altman S. 1983. The RNA moiety of RNase P is the catalytic subunit of the enzyme. *Cell* **35:** 849–857.

Horlacher J., Hottiger M., Podust V.N., Hübscher U., and Benner S.A. 1995. Expanding the genetic alphabet: Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with non-standard hydrogen bonding patterns. *Proc. Natl. Acad. Sci.* **92:** 6329–6333.

Huang Z. and Benner S.A. 1993. Selective protection and deprotection procedures for thiol and hydroxyl groups. *Synlett* 83–84.

Huang Z., Schneider K.C., and Benner S.A. 1991. Building blocks for analogs of ribo- and deoxyribonucleotides with dimethylene sulfide, -sulfoxide and -sulfone groups replacing phosphodiester linkages. *J. Org. Chem.* **56:** 3869–3882.

Huizenga D.E. and Szostak J.W. 1995. A DNA aptamer that binds adenosine and ATP. *Biochemistry* **34:** 656–665.

Hyrup B., Richert C., Schulte-Herbrüggen T., Benner S.A., and Egli M. 1995. X-ray crystal structure of a dimethylene-sulfone bridged ribonucleotide dimer crystallized at elevated temperature. *Nucleic Acids Res.* **23:** 2427–2433.

Irvine D., Tuerk C., and Gold L. 1991. Selexion. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* **222:** 739–761.

Johnsson K., Allemann R.K., and Benner S.A. 1990. Designed enzymes. New peptides that fold in aqueous solution and catalyze reactions. In *Molecular mechanisms in bioorganic processes* (ed. C. Bleasdale and B.T. Golding), pp. 166–187. Royal Society of Chemistry, Cambridge, United Kingdom.

Johnsson K., Allemann R.K., Widmer H., and Benner S.A. 1993. Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides. *Nature* **365:** 530–532.

Joyce G.F. 1989a. Amplification, mutation and selection of catalytic RNA. *Gene* **82:** 83–87.

———. 1989b. Building the RNA world. Evolution of catalytic RNA in the laboratory. *UCLA Symp. Mol. Cell. Biol. New Ser.* **94:** 361–371.

———. 1994. Foreword. In *Origins of life: The central concepts* (ed. D.W. Deamer and G.R. Fleischaker). Jones and Bartlett, Boston, Massachusetts.

Joyce G.F., Schwartz A.W., Miller S.L., and Orgel L.E. 1987. The case for an ancestreal genetic system involving simple analogs of the nucleotides. *Proc. Natl. Acad. Sci.* **84:** 4398–4402.

Kerr R.A. 1997. Ancient life on Mars? Putative Martian microbes called microscopy artifacts. *Science* **278:** 1706–1707.

Khare B.N., Thompson W.R., Cheng L., Chyba C., Sagan C., Arakawa E.T., Meisse C., and Tuminello P.S. 1993. Production and optimal-constants of ice tholin from charged-particle irradiation of (1/6) C_2H_6/H_2O at 77 K. *Icarus* **103:** 290–300.

Kodra J. and Benner S.A. 1997. Synthesis of an N-alkyl derivative of 2′-deoxyisoguanosine. *Synlett* 939–940.

Lee D.H., Severin K., Yokobayashi Y., and Ghadiri M.R. 1997. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390:** 591–594.

Limbach P.A., Crain P.F., and McCloskey J.A. 1994. The modified nucleosides of RNA. Summary. *Nucleic Acids Res.* **22:** 2183–2196.

Lutz M.J., Held H.A., Hottiger M., Hübscher U., and Benner S.A. 1996. Differential discrimination of DNA polymerases for variants of the non-standard nucleobase pair between xanthosine and 2,4-diaminopyrimidine, two components of an expanded genetic alphabet. *Nucleic Acids Res*. **24:** 1308–1313.

Miller S.L. 1997. Peptide nucleic acids and prebiotic chemistry. *Nat. Struct. Biol.* **4:** 167–169.

Morris K.N., Tarasow T.M., Julin C.M., Simons S.L., Hilvert D., and Gold L. 1994. Enrichment for RNA molecules that bind a Diels-Alder transition-state analog. *Proc. Natl. Acad. Sci.* **91:** 13028–13032.

Müller D., Pitsch S., Kittaka A., Wagner E., Wintner C.E., and Eschenmoser A. 1990. Chemistry of alpha-aminonitriles: Aldomerisation of glycoaldehyde phosphate to rac-hexose 2,4,6-triphosphates and in presence of formaldehyde rac-pentose 2,4-diphosphates: Rac-allose 2,4,6-triphosphate and rac-ribose 2,4-diphosphate are the main reaction-products. *Helv. Chim. Acta* **73:** 1410–1468.

Orgel L.E. 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* **38:** 381–393.

Perezpaya E., Houghten R.A., and Blondelle S.E. 1996. Functionalized protein-like structures from conformationally defined synthetic combinatorial libraries. *J. Biol. Chem.* **271:** 4120–4126.

Piccirilli J.A., Moroney S.E., and Benner S.A. 1991a. A C-nucleotide base pair. Methylpseudouridine-directed incorporation of formycin triphosphate into RNA catalyzed by T7 RNA polymerase. *Biochemistry* **30:** 10350–10356.

Piccirilli J.A., Krauch T., MacPherson L.J., and Benner S.A. 1991b. A direct route to 3-ribofuranosyl-pyridine nucleosides. *Helv. Chim. Acta* **74:** 397–406.

Piccirilli J.A., Krauch T., Moroney S.E., and Benner S.A. 1990. Extending the genetic alphabet: Enzymatic incorporation of a new base pair into DNA and RNA. *Nature* **343:** 33–37.

Prober J.M., Trainor G.L., Dam R.J., Hobbs F.W., Robertson C.W., Zagursky R.J., Cocuzza A.J., Jensen M.A., and Baumeister K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238:** 336–341.

Rich A. 1962. On the problems of evolution and biochemical information transfer. In *Horizons in biochemistry* (ed. M. Kasha and B. Pullman), pp. 103–126. Academic Press, New York.

Richert C., Roughton A.L., and Benner S.A. 1996. Nonionic analogs of RNA with dimethylene sulfone bridges. *J. Am. Chem. Soc.* **118:** 4518–4531.

Robertson M.P., and Miller S.L. 1995. Prebiotic synthesis of 5-substituted uracils: A bridge between the RNA world and the DNA-protein world. *Science* **268:** 702–705.

Roughton A.L., Portmann S., Benner S.A., and Egli M. 1995. Crystal structure of a dimethylene sulfone linked ribodinucleotide analog. *J. Am. Chem. Soc.* **117:** 7249–7250.

Sassanfar M. and Szostak J.W. 1993. An RNA motif that binds ATP. *Nature* **364:** 550–553.

Schmidt J.G., Nielsen P.E., and Orgel L.E. 1997. Enantiomeric cross-inhibition in the synthesis of oligonucleotides on a nonchiral template. *J. Am. Chem. Soc.* **119:** 1494–1495.

Switzer C.Y., Moroney S.E., and Benner S.A. 1989. Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* **111:** 8322–8323.

Szathmary E. 1992. What is the optimum size for the genetic alphabet? *Proc. Natl. Acad. Sci.* **89:** 2614–2618.

Szostak J.W. 1988. Structure and activity of ribozymes. In *Redesigning the molecules of life* (ed. S.A. Benner) pp. 87–114. Springer-Verlag, Heidelberg.

Tang J. and Breaker R.R. 1997. Rational design of allosteric ribozymes. *Chem. Biol.* **4:** 453–459.

Tarasow T.M., Tarasow S.L., and Eaton B.E. 1997. RNA-catalyzed carbon-carbon bond formation. *Nature* **389:** 54–57.

Usher D.A. and McHale A.H. 1976. Hydrolytic stability of helical RNA: A selective advantage for the natural 3′,5′-bond. *Proc. Natl. Acad. Sci.* **73:** 1149–1153.

Visser C.M. and Kellogg R.M. 1978. Biotin. Its place in evolution. *J. Mol. Evol.* **11:** 171–178.

Vögel J.J. and Benner S.A. 1994. Non-standard hydrogen bonding in duplex oligonucleotides. The base pair between the acceptor-donor-donor pyrimidine analog and the donor-acceptor-acceptor purine analog. *J. Am. Chem. Soc.* **116:** 6929–6930.

Vögel J.J., Altorfer M.M., and Benner S.A. 1993a. The donor-acceptor-acceptor purine analog. Transformation of 5-aza-7-deaza-*iso*guanine to 2′-deoxy-5-aza-7-deaza-*iso*-guanosine using purine nucleoside phosphorylase. *Helv. Chim. Acta* **76:** 2061–2069.

Vögel J.J., von Krosigk U., and Benner S.A. 1993b. Synthesis and tautomeric equilibrium of 6-amino-5-benzyl-3-methylpyrazin-2-one. An acceptor-donor-donor nucleoside base analog. *J. Org. Chem.* **58:** 7542–7547.

von Krosigk U. and Benner S.A. 1995. pH-Independent triple helix formation by an oligonucleotide containing a pyrazine donor-donor-acceptor base. *J. Am. Chem. Soc.* **117:** 5361–5362.

Wang Y. and Patel D.J. 1994. Solution structure of the *Tetrahymena* telomeric repeat $d(T_2G_4)_4$ G-tetraplex. *Structure* **2:** 1141–1156.

Watson J.D., Hopkins N.H., Roberts J.W., Steitz J.A., and Weiner A.M. 1987. *Molecular biology of the gene,* 4th edition, p. 1115. Benjamin/Cummings, Menlo Park, California.

White H.B., III. 1976. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7:** 101–104.

Wiegand T.W., Janssen R.C., and Eaton B.E. 1997. Selection of RNA amide synthases. *Chem. Biol.* **4:** 675–683.

Woese C.R. 1967. *The genetic code. The molecular basis for genetic expression*. Harper & Row, New York.

Zaug A.J. and Cech T.R. 1986. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231:** 470–475.

Zhang B.L. and Cech T.R. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature* **390:** 96–100.