IN-55
371764

Quarterly Report
Contract NAS5-97210
March 10, 1998
Prepared by Steven A. Benner, Principal Investigator

Progress in the past three months has occurred in two areas:

I. Reconstruction of ancestral proteins:

Reconstructions of ancient forms of life using "maximum parsimony" and "maximuim likelihood methods offer a perspective on how contemporary life emerged, how its evolution is influenced by geological and cosmogenic events, and how new forms of life replace old forms of life by mass extinctions. Further, they allow one to approach the origin of life following a "present day backwards" strategy that is an alternative to prebiological experiments. One focus of this project is a "staged" reconstruction of ancient biomolecular structures, beginning with reconstructions that go only modestly back in time and continuing with recostructions that are more ancient. This strategy reflects the fact that more recent events are easier to reconstruct than more ancient events, as extinctions, rapid sequence evolution associated with functional adaptation, and non-Markovian processes (gene conversion, for example) cause the loss of information. Our strategy is to encounter and solve problems in this staging as they occur.

A second focus has been to demonstrate "practical" applications of exobiological research. In addition to the obvious pragmatic reasons for doing this (funds from sources other than NASA will be needed to carry this line of research to completion, and mission-oriented funding agencies always seem to have more money than basic science agencies), it remains a worthwhile test of *any* new research strategy to challenge it to solve real-world problems. This prevents the research programme from degenerating into a deep, scholarly study covering a narrow scope with little breadth in interest.

A. Ribonuclease (RNase).

RNase has been used as a system to examine very recent reconstructions, primarily because it is a protein with multiple paralogs (sister genes in the same organism) which can (and might have) talked to each other via gene conversion (a non-Markovian process that undermines the evolutionary model that underlies these reconstrutions), and which clearly has undergone rapid sequence evolution in some of its branches to acquire new biological functions. The protein family has now provided another "practical" applications of exobiological research; a reconstruction completed in the laboratory has permitted the correlation of in vitro behavior of the protein to physiological function. The approach is quite general, meaning that it should be possible to use analogous evolutionary reconstructions to identify physiologically relevant traits in many proteins. The exobiological research therefore addresses one of the classic, unsolved questions in experimental biochemistry: "Is the behavior that I am studying in vitro have any relevance physiologically?" A paper will be prepared when the final controls are completed, which we expect before the end of the contract period.

B. SH2 domains.

During the evolution of metazoa, tyrosine phosphorylation evolved as a crucial component of signal transduction, allowing cells within an organism to communicate with each other. Src homology 2 (SH2) domains are involved in recognizing phosphorylated tyrosine residues, allowing the signal generated from a binding event on a cell surface to be transduced to the nucleus where specific changes in gene expression result. These domains are small ~100 amino acid peptide motifs that have undergone major radiative divergence in metazoa.

While one yeast protein, SPT6, shows weak homology to metazoan SH2 domains [1], multiple gene duplication events postdating the yeast-metazoan divergence but predating the nematode-chordate divergence have resulted in a wide array of SH2 domains bearing a wide range of binding specificities and signaling capabilities [2,3]. Thus, it appears as if the SH2 domains originated their function at the time of the Cambrian explosion, one of the most important (for humans) events in the biosphere. The binding specificity of an SH2 domain for its substrate is dependent upon the three-five residues C-

terminal to the phosphotyrosine. We have reconstructed the evolutionary history of the SH2 domain and its signaling specificity to provide insight into the evolution of metazoan metabolism.

All domains that had been identified as SH2 domains in Swiss Prot Protein Database (October 31, 1997 release) were downloaded and submitted to Darwin (a bioinformatics tool developed in these laboratories) for a multiple alignment. Many of the SH2 domains showed weak similarity to human src that was not significant and were eliminated from the alignment. The multiple alignment obtained had many misplaced gaps and was realigned manually. These sequences were then placed in MacClade and a phylogenetic tree was generated using the principle of maximum parsimony, where a biological tree was utilized so long as it adhered to the maximum parsimony principle. Sequences resulting from Cambrian era gene duplications were branched solely based upon the principle of maximum parsimony.

The SH2 domain offers many of the classic problems associated with intermediate antiquity reconstructions. Many of the sequences from these gene duplication events had diverged significantly from each other, making reconstruction of the universal ancestral SH2 domains essentially impossible given available sequences. Therefore, in this case, the expressed to silent ratio analysis was discarded as a measure of adaptive evolution and the study focused in on one family of the tree. that containing the Grb2 domain.

Grb2 is an adapter protein containing both an SH2 domain and two SH3 domains. While homologues from *D. melanogaster* and *C. elegans* are functionally interchangeable at an organism viability level [4], the human, *D. melanogaster*, and *C. elegans* proteins show differences in binding specificity [2,3]. This binding specificity evolution was taken as a lead to study the evolution of binding specificity in this one subfamily dating back to the Cambrian explosion. This subfamily contained several vertebrate sequences- human, mouse, rat, Xenopus, chicken, and human GRAP (resulting from a gene duplication event that appears from its most parsimonious placement on the tree to be chordate-specific [5,6]), as well as the *D. melanogaster* and *C. elegans* sequences. Human nck was the most closely related outgroup in the reconstruction of the last common ancestor of grb2 homologues. The reconstruction from this tree contained several ambiguous residues. Attempts were made to resolve these using either DNA parsimony or one of several maximum likelihood techniques before deciding to clone the gene from additional invertebrate species to resolve the ambiguities.

Short primers were developed from the *C. elegans*, *D. melanogaster*, and chordate SH3 sequences independently and were used to clone the gene from various cDNA samples. The chordate primers were initially used to pull a band of the expected size (~350 base pairs) from *Ursus maritimus* cDNA (around the lab for another project) as a positive control. Subsequently, bands of the expected size were obtained from *Gekensia demissa* cDNA using the *D. melanogaster* primers and *Lytechinus variegatis* cDNA using the chordate primers. These bands will be sequenced once the gene has been cloned from additional invertebrate species. This should allow a faithful reconstruction of the Grb2 subfamily of SH2 domains, synthesis of the ancient protein in the laboratory, and analysis of its binding specificity to provide insight into the evolution of metazoan signal transduction and its role in the evolution of multicellularity.

1. MacLennan, A.J. and Shaw, G. 1993. "A Yeast SH2 Domain." *Trends in Biochemical Science* **18**, 464-465.
2. Songyang, Z. *et. al.* 1993. "SH2 Domains Recognize Specific Phosphopeptide Sequences." *Cell* **72**, 767-778.
3. Songyang, Z. *et. al.* 1994. "Specific Motifs Recognized by the SH2 Domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav." *Molecular and Cellular Biology* **14**, 2777-2785.
4. Stern, M.J. *et. al.* 1993. "The Human GRB2 and *Drosophila Drk* Genes Can Functionally Replace the *Caenorhabditis elegans* Cell Signaling Gene *sem-5*." *Molecular Biology of the Cell* **4**, 1175-1188.
5. Feng, G. *et. al.* 1996. "Grap is a Novel SH3-SH2-SH3 Adapter Protein That Couples Tyrosine Kinases to the Ras Pathway." *Journal of Biological Chemistry* **271**, 12129-12132.
6. Trub, T., Frantz, J.D., Miyazaki, M., Band, H., and Shoelson, S.E. 1997. "The Role of a Lymphoid-restricted Grb2-like SH3-SH2-SH3 Protein in T Cell Receptor Signaling." *Journal of Biological Chemistry* **272**, 894-902.

II. Nucleic acid sequences

Earlier work on nucleic acid analogs in these laboratories has discovered a resonance in the exobiological community. There are two reasons for this.

First, planetary exploration has almost arrived to the point where it is possible to retrieve samples of Martian or European (as in the moon) landscape in a form suitable for examination for chemical remnants of life or (if lucky), the chemistry of actual non-terrestrial life. This possibility has focused attention on the deficiencies in our knowledge of the chemical features of life that are likely to be universal regardless of genesis. The expansion of the genetic alphabet in these laboratories has shown that alternative chemical structures work, even within the context of DNA chemistry. Work with non-ionic analogs of nucleic acids has suggested specific limitations on this, however.

Second, an explosion of papers and manuscripts has shown that those attempting to create catalytic molecules from standard nucleic acids (seeking evidence for the RNA world) are constraining their search in a non-historical (and probably unproductive) way. In our laboratory (manuscripts enclosed), and at NexStar [7], it has now been shown that the catalytic potential of nucleic acids can be significantly expanded by adding functionality to the standard bases, while Miller and his coworkers have shown that functionalized nucleosides might have been available prebiotically [8]. These results show that it will not be long (months or years, but not a decade) before self-replicating, evolving systems based on an expanded genetic alphabet will be available in the laboratory. These will begin an entirely new phase of research into life, one that builds cells from the bottom up.

Two manuscripts are enclosed that address these two points. The first, prepared for a book on the RNA world being edited by Tom Cech, concerns universal chemical properties of genetic material capable of searching "mutation space" independent of concern of loss of properties essential for replication ("COSMIC-LOPER" behavior). The second reports the first in vitro selection experiment with a positively charged nucleotide analog.

7. Tarasow, T.M.; Tarasow, S.L.; Eaton, B.E. *Nature* **1997**, *389*, 54-57
8. Robertson, M.P.; Miller, S.L. *Science* **1995**, *268*, 702-705.

Submitted

Steven A. Benner
Principal Investigator
May 15, 1998

A Polymerase Chain Reaction and *In Vitro* Selection Incorporating a Cationic Nucleotide Analog

Petra Burgstaller,[§] Darwin Ang,[†] Thomas Battersby,[#] Simona Jurczyk,[#] Steven A. Benner[†#§*]

[§]Department of Chemistry and [†]Department of Anatomy and Cell Biology

University of Florida, Gainesville FL 32611

and

[#]Sulfonics, Inc. 12085 Research Drive, Alachua FL 32615

* To whom correspondence should be addressed at the Department of Chemistry, University of Florida

Present address: Hoechst Research and Technology Deutschland GmbH & Co.KG

65926 Frankfurt-am-Main, Germany

Abstract. 5-(3"-Aminopropynyl)-2'-deoxyuridine (J), a modified nucleoside with a side chain carrying a cationic functional group, was incorporated into an oligonucleotide library, which was successfully amplified using the Vent DNA polymerase in a polymerase chain reaction (PCR). When coupled to an *in vitro* selection procedure, the PCR amplification generated receptors (aptamers) that bound ATP, ADP, and AMP. This is the first example where a polymerase chain reaction (PCR) has been applied to an oligonucleotide library containing modified nucleotides carrying positively charged functional groups. The outcome of the *in vitro* selection incorporating the cationic nucleotide was compared with that of a standard selection using only natural nucleotides, a procedure documented by Huizenga and Szostak (*Biochemistry* **1995**, *34*, 656-665). The amplification of the functionalized library generated a motif containing J differing from the motif well known to arise in a standard selection experiment using only natural nucleotides. The motif containing J had an affinity for ATP ca. 2 orders of magnitude greater than that obtained from the standard selection, but no detectable affinity for ATP if thymidine replaces J.

Key words. In vitro selection, DNA, aptamer, polymerase chain reaction, Selex

In vitro selection is a combinatorial method that generates new receptors, ligands, and catalysts from nucleic acid libraries containing as many as 10^{15} different molecules.[1] It exploits (a) a "selection" procedure to separate RNA or DNA molecules with specific binding or catalytic properties from those lacking these properties, (b) the polymerase chain reaction (PCR), which amplifies single selected RNA or DNA molecules to give a large number of their descendants, and (c) mutation, which allows the descendent molecules to "evolve" to improve their binding or catalytic activities. Using this approach, a variety of ligands, receptors, and catalysts have been generated.[2]

These successes notwithstanding, quantitative analysis of *in vitro* selection experiments shows that the "intrinsic" potential of RNA as both a receptor and a catalyst is poor, especially when compared with that displayed by polypeptides.[3] For example, to have a 50% chance of containing a single RNA molecule capable of catalyzing a template-directed ligation reaction by a modest (by protein standards) factor of 10,000, a library of RNA molecules 220 nucleotides in length must have 2×10^{13} random sequences.[4] The limitation of nucleic acids appears to reside in the small number of encodable building blocks and little diversity in the functionality that they carry. Most notable in view of the polyanionic nature of RNA and DNA, standard nucleic acids have no encoded positive charge at neutral pH. This in turn limits the range of the binding and catalysis of products derived by *in vitro* selection.

While many non-standard[5] and functionalized derivatives of nucleic acids have been prepared,[6] and still more occur naturally,[7] it remains a challenge to develop *in vitro* selection protocols that incorporate them. Even modest discrimination by a polymerase against a functionalized nucleotide may lead to a selective loss of the functionality during PCR amplification. While polymerases are known to incorporate nucleotides carrying neutral functionality,[8] and two *in vitro* selection experiments have incorporated these,[9] it has proven especially difficult to successfully amplify by PCR oligonucleotides containing nucleotide analogs carrying positively charged functionality, the functionality most notably missing from standard DNA and RNA.

To surmount this obstacle, 5-(3"-aminopropynyl)-2'-deoxyuridine[10] (Figure 1), trivially designated as "J", was prepared by coupling N-propynyltrifluoroacetamide[11,12] with 2'-deoxy-5-

iodouridine (Sigma) with the aid of a palladium catalyst, and converted to the corresponding triphosphate following a procedure of Ludwig and Eckstein.[13] The side chain ammonium ion, with a pK_a of ca. 9.5, carries a positive charge at physiological pH. In preliminary experiments, JTP was examined as a substrate for a range of thermostable polymerases. As has been observed with previous experiments with non-standard bases,[14] different polymerases behaved quite differently with respect to the unnatural modification. Taq polymerase, for example, incorporates ATP opposite J in the template, but stops rather than incorporate JTP opposite A in a template. Tth and Tfl polymerases pause when encountering J both in a template and as a triphosphate. Vent polymerase, however, incorporates both JTP opposite A and ATP opposite J without pausing. Based on these experiments, Vent polymerase was chosen as a tool for incorporating J into a PCR experiment.

A series of experiments were then run to show that Vent was able to amplify oligonucleotides containing the positively charged nucleobase. Parallel PCR experiments were run with natural deoxynucleoside triphosphate mixes and mixes having TTP replaced by JTP (Figure 2). In both cases, amplification was seen, with the JTP supporting PCR amplification only slightly less well than TTP with the Vent polymerase.

To establish the value of the PCR amplification using JTP, an *in vitro* selection experiment was run incorporating it into a library following the procedure of Huizenga and Szostak,[15] who used *in vitro* selection experiments with natural triphosphates to obtain receptors for ATP. A parallel *in vitro* selection was run with JTP replacing TTP in the triphosphate mix. Nine rounds of selection were performed by passing successively enriched libraries through a column with immobilized ATP, ADP, and AMP, each separated by 10-14 cycles of amplification, were used.[16] Products were cloned and the clones sequenced.

These *in vitro* selection experiments using the J-containing DNA (J-DNA) library were benchmarked against parallel selection experiments using a library containing only standard DNA nucleotides. As these parallel standard DNA selection experiments followed the work of Huizenga and Szostak, they should generate a known motif that binds adenosine derivatives.[15] Accordingly, of 17

sequences isolated from the parallel selection with standard DNA, 11 showed a binding motif previously isolated by Huizenga and Szostak (motif A).

In the J-DNA selections, however, only 11 of the 39 sequences contained motif A, and several of these sequences originated from the same parent molecule. Another motif was found in 19 of the J-DNA sequences. This novel motif was built from a conserved 13-mer and a conserved 3-mer (motif B). Motif B was then prepared with unfunctionalized T replacing J. In this form, motif B does not bind ATP, suggesting that the cationic functionality carried by J is essential for the binding properties of the new motif.

These results show that a nucleotide derivative J carrying a functionalized side chain bearing a full positive charge at physiological pH can be accepted as a component of a polymerase chain reaction. The presence of J in defined motifs obtained from multiple clones shows that incorporation was at a defined sequence position.

These results also show that the PCR can be used as a part of an *in vitro* selection experiment to generate functionalized aptamers. It is intriguing to compare the results obtained with the functionalized library with those obtained by Huizenga and Szostak[15] and Sassanfar and Szostak[17] from unfunctionalized DNA and RNA libraries (respectively). Using the gel filtration method of Sassanfar and Szostak,[17] an equilibrium disassociation constant (K_d) of ca. 40 nM was measured for the binding of ATP to the functionalized aptamer whose sequence is shown in Figure 3. This value is approximately 2 orders of magnitude lower than the 6-8 μM disassociation constants reported for the binding of a representative example of the consensus motif A obtained from unfunctionalized in standard DNA.[15,17] This suggests that introduction of a cationic functionality (an ammonium group bearing a positive charge) enhances the "intrinsic" affinity of a receptor obtained from a DNA library by ca. two orders of magnitude. While more experiments must be done with more targets, it appears that the addition of the cationic functionality provided by an ammonium ion does indeed improve the quality (as measured in terms of affinity) of the aptamer product.

Figure 1. 5-(3"-Aminopropynyl)-2'-deoxyuridine ("J"), a modified nucleoside with a side chain carrying a cationic functional group.

Figure 2. Results of parallel polymerase chain reaction (PCR) amplifications of an oligonucleotide library incorporating the natural nucleoside triphosphates (ATP, CTP, GTP and TTP, bottom) or a functionalized mix of ATP, CTP, GTP, and the triphosphate of 5-(3"-aminopropynyl)-2'-deoxyuridine (J, top). Left and right hand lanes show molecular weight markers, while the middle lanes are for the cycle number indicated. An extended 5'-primer (51-mer, with the sequence CCGATTGAATCCTAGATCGCATGCTACTGATGACTGTGTAAGCTTGAGCAT) was used to create a gel mobility shift to distinguish the amplified template from the non-amplified template (note the double band in early cycles). The PCR conditions were as follows: 0.05 U/μL Vent polymerase, 260 nM each of dCTP, dATP, dGTP and TTP (bottom) or JTP (top), 4.7 μM primers, 1.6 nM template, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl, 2 mM $MgSO_4$, total volume 0.2 mL. Aliquots (10 μL) were removed after each cycle and loaded onto an agarose gel (4%), together with markers (50 base pairs, left and right lanes). The PCR cycles consisted of the following steps: 94°C (1 min), 55°C (2 min), 72°C (2 min).

Figure 3. Two motifs that bind ATP. Motif A is the wild type sequence reported by Huizenga and Szostak[15] as a typical ATP-binding motif obtained via *in vitro* selection from a standard, unfunctionalized DNA library. Motif B is a representative sequence obtained via a parallel *in vitro* selection experiment where the triphosphate of 5-(3"-aminopropynyl)-2'-deoxyuridine (J, Figure 1), a modified nucleoside with a side chain carrying a cationic functional group, is incorporated in the selection experiment instead of thymidine triphosphate. The motif is underlined, with the flanking regions included (not underlined).

References

1. (a) Szostak, J.W. In *Redesigning the Molecules of Life*. Benner, S.A., editor. Springer-Verlag, Heidelberg, 1988; pp. 87-114. (b) Joyce, G.F. *Gene* **1989**, *82*, 83-87. (c) Joyce, G.F. *UCLA Symp. Mol. Cell. Biol. New Ser.* **1989**, *94*, 361-371. (d) Irvine, D.; Tuerk, C.; Gold, L. *J. Mol. Biol.* **1991**, *222*, 739 -761.

2. For some recent examples and an introduction into the literature, see, (a) Wiegand, T.W., Janssen, R.C.; Eaton, B.E. *Chem. Biol.* **1997**, *4*, 675-683. (b) Zhang B.L.; Cech T.R. *Nature* **1997**, *390*, 96-100.

3. Johnsson, K.; Allemann, R.K.; Widmer, H.; Benner, S.A. *Nature* **1993**, *365*, 530-532.

4. Bartel, D.P.; Szostak, J.W. *Science* **1993**, *261*, 1411-1418.

5. (a) Switzer, C.Y.; Moroney, S.E.; Benner, S.A. *J. Am. Chem. Soc.* **1989**, *111*, 8322-8323. (b) Piccirilli, J.A., Krauch, T., Moroney, S.E.; Benner, S.A. *Nature* **1990**, *343*, 33-37. (c) Krosigk, U. von, Benner, S.A. *J. Am. Chem. Soc.* 1995, *117*, 5361-5362. (d) Vögel, J.J.; Benner, S.A. *J. Am. Chem. Soc.* **1994**, *116*, 6929-6930. (e) Vögel, J.J., Altorfer, M.M.; Benner, S.A. *Helv. Chim. Acta* **1993**,*76*, 2061-2069. (f) Vögel, J.J., von Krosigk, U.; Benner, S.A. *J. Org. Chem.* **1993**, *58*, 7542-7547.

6. (a) Robertson, M.P.; Miller, S.L. *Science* **1995**, *268*, 702-705. (b) Dewey, T.M.; Zyzniewski, M.C.; Eaton, B.E. *Nucleos. Nucleot.* **1996**, *15*, 1611-1617. (c) Kodra, J.; Benner, S.A. *Syn. Lett.* **1997**, 939-940.

7. Limbach, P.A.; Crain, P.F.; McCloskey, J.A. *Nucl. Acids Res.* **1994**, *22*, 2183-2196.

8. (a) Langer, P.R.; Waldrop, A.A.; Ward, D.C. *Proc. Nat. Acad. Sci. USA* **1981**, *78*, 6633-6637. (b) Prober J.M., Trainor G.L.; Dam R.J.; Hobbs F.W.; Robertson C.W.; Zagursky, R.J.; Cocuzza, A.J.; Jensen, M.A.; Baumeister, K. *Science* **1987**, *238*, 336-341.

9. (a) Latham J.A.; Johnson R.; Toole J.J. Nucl. Acids Res. **1994**, *22*, 2817-2822. (b) Tarasow, T.M.; Tarasow, S.L.; Eaton, B.E. *Nature* **1997**, *389*, 54-57.

10. (a) Hobbs, Jr., F.W. *J. Org. Chem.* **1989**, *54*, 3420-3422. (b) Casalnuovo, A. L.; Calabrese, J.C. *J. Am. Chem. Soc.* **1990**, *112*, 4324-4330.

11. Cruickshank, K.A.; Stockwell, D.L. *Tetrahedron Lett.* **1988**, *29*, 5221-5224.

12. Meyer, Jr., R. B. in *Protocols for Oligonucleotide Conjugates. Synthesis and Analytical Techniques*, Agrawal, S. editor. Totowa, New Jersey, Humana Press. pp 77-80. The protected nucleoside derivative was prepared as follows. **5-(3-Trifluoroacetamidopropyn-1-yl)-2'-deoxyuridine.** $C_{14}H_{14}F_3N_3O_6$ 5-Iodo-2'-deoxyuridine (Sigma, 1.412 mmol, 500 mg) was dissolved in dry DMF (12 mL). Ar was passed through this solution for 10 min. $(Ph_3P)_4Pd$ (0.1 equiv., 0.141 mmol, 163 mg) was added and Ar was passed through the solution for another 5 min. Triethylamine (2.0 equiv., 2.825 mmol, 285 mg, 0.393 mL) was added via syringe followed by addition of N-propynyltrifluoroacetamide[11] (2.5 equiv, 3.531 mmol, 533 mg) and CuI (0.2 equiv., 0.282 mmol, 53.7 mg). The mixture was stirred at 40°C for 5 h, the solvent evaporated, and the residue dissolved in MeOH/methylene chloride 1:1 (10 mL). Ion exchange resin Bio-Rad AG1 X8 (HCO_3^- form, 1.5 g, prepared from the chloride form by eluting through a column with the 16 fold volume of 1M NH_4HCO_3 solution followed by deionized water and finally with 0.5 M NH_4HCO_3 solution; no Cl^- was detected) was added to remove the $Et_3N{\bullet}HI$ by-product, and the mixture stirred at room temperature for 30 min. The mixture was filtered through Celite, the solid washed with MeOH/methylene chloride 1:1 (10 mL), and the combined filtrates evaporated. The residue was purified by column chromatography (chloroform/MeOH 8.25:1.75). R_F: 0.42 (chloroform/MeOH 8.25:1.75). Yield: >95% yellow foam (contains DMF). ^{1}H-NMR (DMSO-d_6): 2.13 (m, 2H, 2'), 3.60 (m, 2H, 5'), 3.81 (m, 1H, 4'), 4.24 (m, 3H, H-9, 3'), 5.12 (t, 1H, 5'-OH), 5.27 (d, 1H, 3'-OH), 6.11 (t, 1H, 1'), 8.22 (s, 1H, H-6), 10.09 (t, 1H, NH chain), 11.67 (s, 1H, NH cycl.). ^{13}C-NMR (DMSO-d6): 29.5 (C-9), 40.7 (2'), 61.0 (5'), 70.2 (3'), 75.4 (C-8), 84.8 (1'), 87.5, 87.7 (4', C-7), 97.7 (C-5), 113.9, 117.8 (q, CF_3, J = 287.95 Hz), 144.2 (C-6), 149.5 (C-2),

155.9, 156.4 (q, CQCF$_3$, J = 37.25 Hz), 161.7 (C-4). **5-(3-Trifluoroacetamidopropyn-1-yl)-5'-*O*-dimethoxytrityl-2'-deoxyuridine** $C_{35}H_{32}F_3N_3O_8$. The product from above (1.342 mmol, 505.9 mg) was coevaporated with pyridine, then dissolved in dry pyridine (10 ml) and cooled to 0°C. Et_3N (2 equiv, 2.684 mmol, 271.1 mg, 0.373 mL), DMAP (0.25 equiv, 0.3355 mol, 41 mg) and DMTCl (1.2 equiv, 1.61 mmol, 545.1 mg) were added and the mixture stirred at 0°C for 5 min. and at room temperature for 4 h. TLC (chloroform/10% MeOH, R_F = 0.47) did not show any starting material. MeOH (2 ml) was added and the mixture evaporated. The residue was extracted (ethyl acetate/aqueous $NaHCO_3$ solution), the combined organic layers washed with water, dried (Na_2SO_4) and the solvent evaporated. The residue was purified by column chromatography (chloroform/ 10% MeOH) to give 902 mg (99%) of product as a yellow foam. ^{1}H-NMR ($CDCl_3$): 2.44-2.61 (m, 2H, 2'), 3.34 (m, 2H, 5'), 3.73 (s, 6H, MeO), 3.89 (m, 1H, 4'), 4.14 (m, 2H, H-9), 4.59 (m, 1H, 3'), 6.34 (t, 1H, 1'), 6.80 (m, 4H, DMT), 7.14-7.33 , 7.61-7.70 (m, 9H, DMT), 8.21 (s, 1H, H-6). ^{1}H-NMR ref. [1] (DMSO-d6/D_2O): 2.28 (m, 2H, 2'), 3.08 (m, 1H, 5'), 3.28 (m, 1H, 5'), 3.91 (m, 1H, 4'), 4.06 (br s, 2H, H-9), 4.33 (m, 1H, 3'), 6.10 (t, 1H, 1'), 6.90 (m, 4H, DMT), 7.22-7.42 (m, 9H, DMT),7.94 (s, 1H, H-6).^{13}C-NMR ($CDCl_3$): 30.3 (C-9), 41.6 (2'), 55.2 (MeO), 63.5 (5'), 72.0 (3'), 75.3 (C-8), 86.0 (1'), 86.9 (Cq trityl), 87.0, (C-7), 87.3 (4'), 99.9 (C-5), 113.3 (DMT), 113.8, 117.6 (q, CF_3, J = 286.52 Hz), 126.9, 127.8, 128.0, 129.9, 135.4 (all DMT), 143.6 (C-6), 144.5 (DMT), 149.4 (C-2), 156.4, 156.9 (q, CQCF$_3$, J = 37.70 Hz), 158.5 (DMT), 162.6 (C-4). **3'-*O*-Acetyl-5-(3-trifluoroacetamidopropyn-1-yl)-5'-*O*-dimethoxytrityl-2'-deoxyuridine** $C_{37}H_{34}F_3N_3O_8$. The product from above (1.328 mmol, 902 mg) was coevaporated with pyridine and then dissolved in pyridine (10 mL) and DMAP (0.25 equiv, 0.332 mmol, 40.5 mg), Et_3N (2.5 equiv, 3.32 mmol, 335 mg, 0.462 ml) and Ac_2O (1.2 equiv, 1.594 mmol, 162.5 mg, 0.15 mL) were added. It was stirred at room temperature for 2.5 h (TLC: chloroform/10% MeOH, R_F = 0.75), then MeOH (5 ml) was

added to stop the reaction and it was evaporated to dryness. The residue was extracted (water/ethyl acetate), the organic layer dried (Na_2SO_4) and the solvent evaporated. The residue was purified by column chromatography (chloroform/ 10% MeOH) to give 937 mg (98%) of product as a yellow foam. ^{1}H-NMR ($CDCl_3$): 2.08 (s, 3H, Ac), 2.38-2.64 (m, 2H, 2'), 3.42 (m, 2H, 5'), 3.77 (s, 6H, MeO), 3.96 (m, 1H, 4'), 4.18 (m, 2H, H-9), 5.45 (m, 1H, 3'), 6.34 (t, 1H, 1'), 6.85 (m, 4H, DMT), 7.23-7.48 , 7.63-7.71 (m, 9H, DMT), 8.20 (s, 1H, H-6).^{13}C-NMR ($CDCl_3$): 20.8 (Ac), 30.2 (C-9), 38.7 (2'), 55.1 (MeO), 63.5 (5'), 74.9 (3'), 75.1 (C-8), 84.4 (1'), 85.3 (4'), 87.1 (Cq trityl), 87.2 (C-7), 99.4 (C-5), 113.3 (DMT), 113.6, 117.4 (q, CF_3, J = 287.42 Hz), 126.9, 127.7, 128.0, 129.9, 135.2 (all DMT), 143.1 (C-6), 144.3 (DMT), 149.4 (C-2), 156.3, 156.8 (q, $\underline{CO}CF_3$, J = 37.78 Hz), 158.6 (DMT), 162.0 (C-4), 170.3 (Ac).

13. Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631-635.

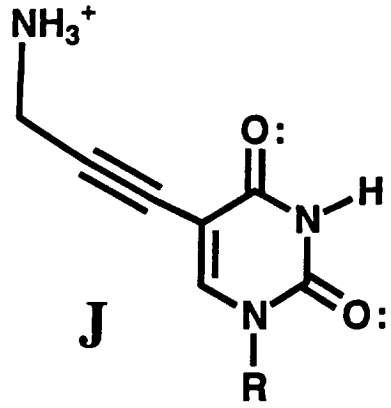
14. (a) Horlacher, J.; Hottiger, M.; Podust, V.N.; Hübscher, U.; Benner, S.A. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6329-6333. (b) Lutz, M.J.; Held, H.A.; Hottiger, M.; Hübscher, U.; Benner, S.A. *Nucl. Acids Res.* **1996**, *24*, 1308-1313.

15. Huizenga, D.E.; Szostak, J.W. *Biochemistry* **1995**, *34*, 656-665.

16. A library of DNA sequences was built to have 22-mer and 20-mer primer-binding regions sandwiching a random 70 base region containing the four natural bases in equal proportion. This was PCR amplified using JTP in place of TTP, with one primer tagged with biotin. The resulting double-stranded DNA library containing A, G, C, and J (J-DNA) was converted to the corresponding single stranded J-DNA library by loading the product on a streptavidin agarose column and eluting with 0.5 M NaOH. A preselection of the library was first performed against 6-aminohexanoic acid *N*-hydroxysuccinimide ester-Sepharose 4B quenched (0.1 M Tris, pH 8.0, 0.5 M NaCl) and then equilibrated in selection buffer (250 mM NaCl, 50 mM Tris, pH 7.6, 5 mM $MgCl_2$). The resulting library was then applied to the ATP-agarose column (N-8

linkage) and non-binding species were eluted with selection buffer. Binding components were then eluted with ATP, ADP or AMP (respectively) in selection buffer. The binding species were PCR amplified to complete the first round of selection. The second round of selection followed with the application of the amplified binding species to the ATP-agarose column and continuing the iterative process. The resulting DNA was cloned into the vector pCR 2.1-TOPO using the TOPO TA Cloning Kit (Invitrogen). For the standard DNA pools 10 clones were sequenced per selection experiment, and for the J-DNA pools 15 clones were sequenced. The randomized region of the J-containing sequences had been noticeably shortened during the selection, its average length after nine rounds of selection was 26 nucleotides compared to 70 nucleotides in the original library. Also, the content of the J nucleotide was reduced; instead of the expected 25%, it was only about 14%. This result suggests biases against longer oligonucleotide aptamers and J, although these results do not permit a conclusion as to whether these biases arise in the selection or the amplification steps.

17. Sassanfar, M. and Szostak, J. W. *Nature* **1993**, *364*, 550-553.



J

PCR Amplification with dJTP

PCR Cycles 0 1 2 3 4 5 6 7 8

PCR Amplification with TTP

PCR Cycles 0 1 2 3 4 5 6 7 8

Motif A

GTGCTTGGGGGAGTATTGCGGA
GGAAAGCGGCCCTGCTGAAG

Motif B

GGTCGTCTAGAGTATGCGGTAG
GAACGJCAGJGGGGGGAGCAJA
JGGJGJGAJACGCGA
CCGAAGAAGCJJGGCCCAJG

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 FL 32611

Sulfonics, Inc. 12085 Research Drive, Alachua FL

Running title: Expanded Genetic Alphabet in the RNA World

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 non-biological mechanism that would allow either proteins or nucleic acids to emerge spontaneously from non-living 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 (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; Zaug et al. 1996; Guerrier-Takada et al. 1983) 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 where 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 (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 & 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 (Zhang and Cech 1997; Wiegand et al. 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 1990; Johnsson et al. 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. Further, the peptide is less than 10% the size of the RNA motif. Combinatorial experiments starting from this design (Perezpaya et al. 1996; L. Baltzer, personal communication) 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, as 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), while 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 non-quantitative, 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 available than in RNA. RNA lacks the imidazole, thiol, amino, carboxylate, and hydrophobic aromatic and aliphatic groups that feature so prominently in protein-based enzymes. RNA has only hydroxyl groups, polar aromatic groups, 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 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 an effective, providing another reason why peptides might be intrinsically better catalysts than RNA (Benner 1989).

If one must 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 pre-existence 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, as 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, for example, 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 single biopolymer system is more plausible as a first life form than 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 proposed (Benner et al., 1987; Benner 1988; Benner 1989; Switzer at 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 (Figure 1) (Limbach et al. 1994). These functional groups are introduced by post-transcriptional 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 non-standard hydrogen bonding patterns permitted by the geometry of the Watson-Crick base pair (Figure 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 (Figure 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 & Breaker 1997).

In contrast, the second and third approaches were far from implementation in 1987. While 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 non-standard bases would pair as expected, or whether polymerases would incorporate functionalized standard bases and non-standard bases (Figures 2 and 3) with sufficient speed and fidelity to be used in *in vitro* selection experiments. Further, 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. Non-standard nucleobases needed to be synthesized (Switzer et al. 1989; Piccirilli et al. 1990; Vögel et al. 1993a; Vögel et al. 1994). Their structures needed to be optimized for stability and pairing (Piccirilli et al. 1991; 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 non-standard nucleotides, and laid the ground for the first *in vitro* selection experiments using these.

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 non-standard and functionalized nucleobases. Three results are central to this argument.

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 (Figure 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, while 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", Figure 4) and incorporated it into an *in vitro* selection experiment seeking receptors for adenosine derivative (unpublished). 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>GAACGJCAGJGGGGGGAGCAJAJGGJGJGAJACGCGA</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 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 & 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 ca. 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 have used 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 (Figure 1). This suggests that the RNA World may have had the opportunity to use some functionalized nucleosides when life first emerged on Earth.

Could non-standard nucleobases (Figure 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 derivative (such as nucleosides) in the presence of organic "gunk" that emerges from a typical prebiotic experiment.

Notwithstanding these issues, several of the non-standard nucleobases in Figure 2 do not appear to be less "prebiotic" than the standard nucleobases. 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, then it may also have had the opportunity to use an expanded genetic alphabet.

CONTRADICTING 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 (Figure 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 upon 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, then 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, while altering their internal sequence at a single residue changes substantially their helicity (Allemann 1989; Johnsson 1990; Johnsson et al. 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 non-ionic dimethylenesulfone linking units (Huang et al. 1991). The sulfone group is an "isosteric" and "isoelectronic" replacement for a phosphate. Nevertheless, these non-ionic oligomers display some remarkable properties. First, they fold. For example, the octamer $ASO_2USO_2GSO_2GSO_2USO_2CSO_2ASO_2U$ folds in solution to give a folded form in water having a high melting temperature (ca. 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 GSO_2C in the crystal forms an antiparallel duplex approximately isomorphous with the analogous RNA (Roughton et al. 1995). In the crystal, the ASO_2T dinucleotide does not (Hyrup et al. 1995). The USO_2C dinucleotide forms a complex

featuring backbone-to-backbone and backbone-to-nucleobase hydrogen bonds (C. Richert, personal communication). Even within a relatively small search of sequence space, these non-ionic 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 modelled.

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):

(a) 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.

(b) 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). Non-ionic oligonucleotide analogs should (and do) fold like peptides. By discouraging folding, the repeating polyanionic backbone helps oligonucleotides act as templates.

(c) Electronic distribution in a molecule is described as an infinite series (monopole + dipole + quadrapole + ...). The first non-vanishing 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 (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 biopolymer (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 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 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 non-terrean 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 properties 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.

ACKNOWLEDGEMENTS

We are indebted to the NASA, the Office of Naval Research, the NIH, 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

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

Allemann, R.K. 1989. *Evolutionary Guidance as a Tool in Organic Chemistry*. Ph.D. Dissertation No. 8804, Eidgenössische Technische Hochschule Zürich, Switzerland.

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

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

Benner, S.A. 1989. Enzyme kinetics and molecular evolution. *Chem. Rev.* **89**: 789-806.

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 Sym. Quant. Biol.* **52**: 53-63.

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* (eds. Gesteland, R., and Atkins, J.) pp. 27-70. Cold Spring Harbor Laboratory Press, N.Y.

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

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., Eaton, B.E. 1996. The RNA World. Functional diversity in a nucleoside by carboxyamidation of uridine. *Nucleos. Nucleot.* **15**: 1611-1617.

Flory, P.J. 1953. *Principles of Polymer Chemistry.* Cornell University Press, Ithaca, N.Y.

Gilbert, W. 1986. The RNA World. *Nature* **319**: 618.

Gouverneur, V.E., Houk, K.N., Depascualteresa, B., Beno, B., Janda, K.D., Lerner, R.A. 1993. Control of the exo-pathway and 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. USA* **92**: 6329-6333.

Huang, Z. and Benner, S.A. 1993. Selective protection and deprotection procedures for thiol and hydroxyl groups. *Syn Lett*: 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. *Nucl. 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. *Molecular Mechanisms in Bioorganic Processes* (eds. Bleasdale, C. and Golding, B.T.), pp. 166-187. Royal Society of Chemistry, Cambridge.

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.

Joyce, G.F. 1989b. Building the RNA World. Evolution of catalytic RNA in the laboratory. *UCLA Symp. Mol. Cell. Biol. New Ser.* **94**: 361-371.

Joyce, G.F. 1994. In *Origins of Life: The Central Concepts.* (eds. Deamer, D. W. and Fleischaker, G. R.) forward, Jones and Bartlett, Boston.

Joyce, G.F., Schwartz, A.W., Miller, S.L., and Orgel, L.E. 1987) The case for an ancestreal genetic system iinvolving simple analogs of the nucleotides. *Proc. Nat. Acad. Sci. USA* **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 optical-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. *Syn. Lett.,* 939-940.

Krosigk, U. von, 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.

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., McCloskey, J.A. 1994. The modified nucleosides of RNA. Summary. *Nucl. 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. *Nucl. Acids Res.* **24**: 1308-1313.

Miller, S.L. 1997. Peptide nucleic acids and prebiotic chemistry. *Nature Structural Biology* **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. Nat. Acad. Sci. USA* **91**: 13028-13032.

Müller, D., Pitsch, S., Kittaka, A., Wagner, E., Wintner, C.E., and Eschenmoser, A. 1990. Chemistry of alpha-aminonitriles: Aldomerisation of glycolaldehyde 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., 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.

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

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

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* (eds. Kasha, M. and Pullman, B.), pp. 103-126. Academic Press, N.Y.

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. What is the optimum size for the genetic alphabet ? 1992. *Proc. Nat. Acad. Sci. USA* **89**: 2614-2618.

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

Tang, J., and Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem. Bio.* **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. USA* **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.

Wang, Y. and Patel, D.J. 1994. Solution structure of the Tetrahymena telomeric repeat $d(T_2G_4)_4)_3$ G-tetraplex. *Structure* **2**: 1141-1156.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., and Weiner, A.M. *Molecular Biology of the Gene*. 4th ed. p 1115. Benjamin Cummings, Menlo Park.

White III, H.B. 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, N.Y.

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.

FIGURE LEGENDS

Figure 1. Transfer RNA contains a rich collection of functionalized standard nucleobases, created by post-transcriptional 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?

Figure 2. 12 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 upper case letters following the designation indicate the pattern ofhydrogen 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 *non-standard* hydrogen bonding schemes.

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

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.

Figure 1

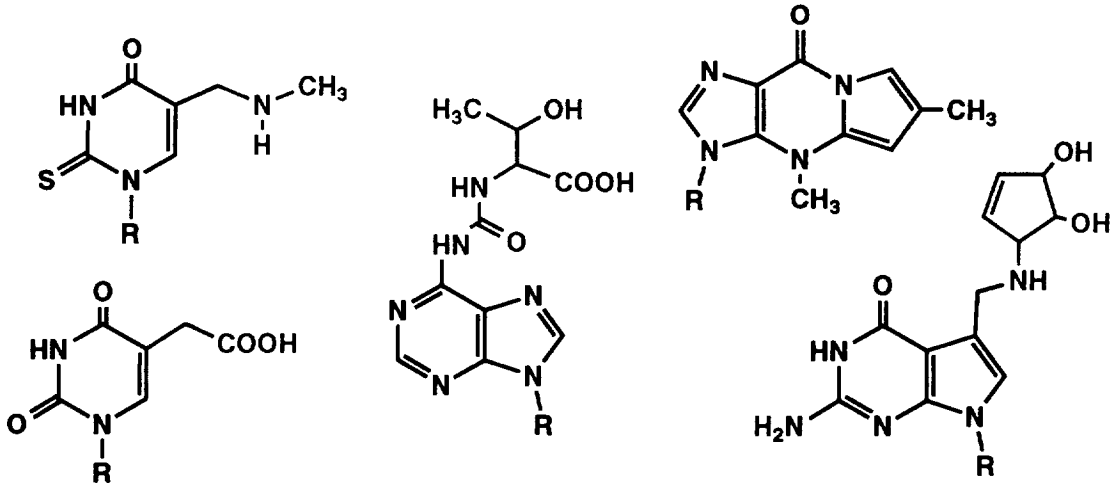

O
HN
N
CH3
H
S
N
R
H3C
OH
HN
COOH
HN
O
O
N
N
CH3
N
N
R
CH3
OH
OH
NH
O
HN
COOH
O
N
R
N
N
N
R
O
HN
H2N
N
N
R

Figure 2

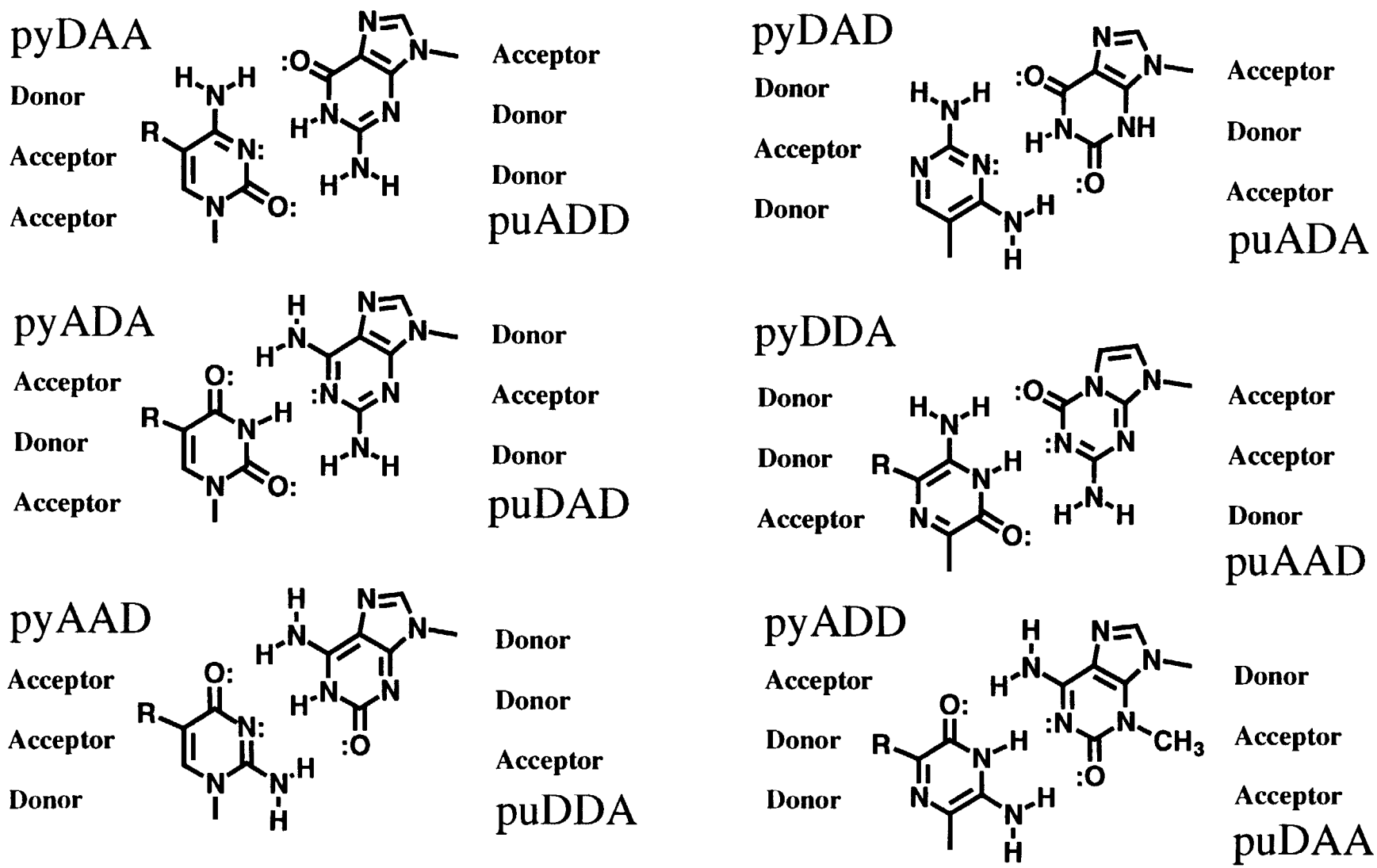

pyDAA
Donor
Acceptor
Acceptor
Acceptor
Donor
Donor
puADD
pyADA
Acceptor
Donor
Acceptor
Donor
Acceptor
Donor
puDAD
pyAAD
Acceptor
Acceptor
Donor
Donor
Donor
Acceptor
puDDA
pyDAD
Donor
Acceptor
Donor
Acceptor
Donor
Acceptor
puADA
pyDDA
Donor
Donor
Acceptor
Acceptor
Acceptor
Donor
puAAD
pyADD
Acceptor
Donor
Donor
Donor
Acceptor
Acceptor
puDAA

Figure 3

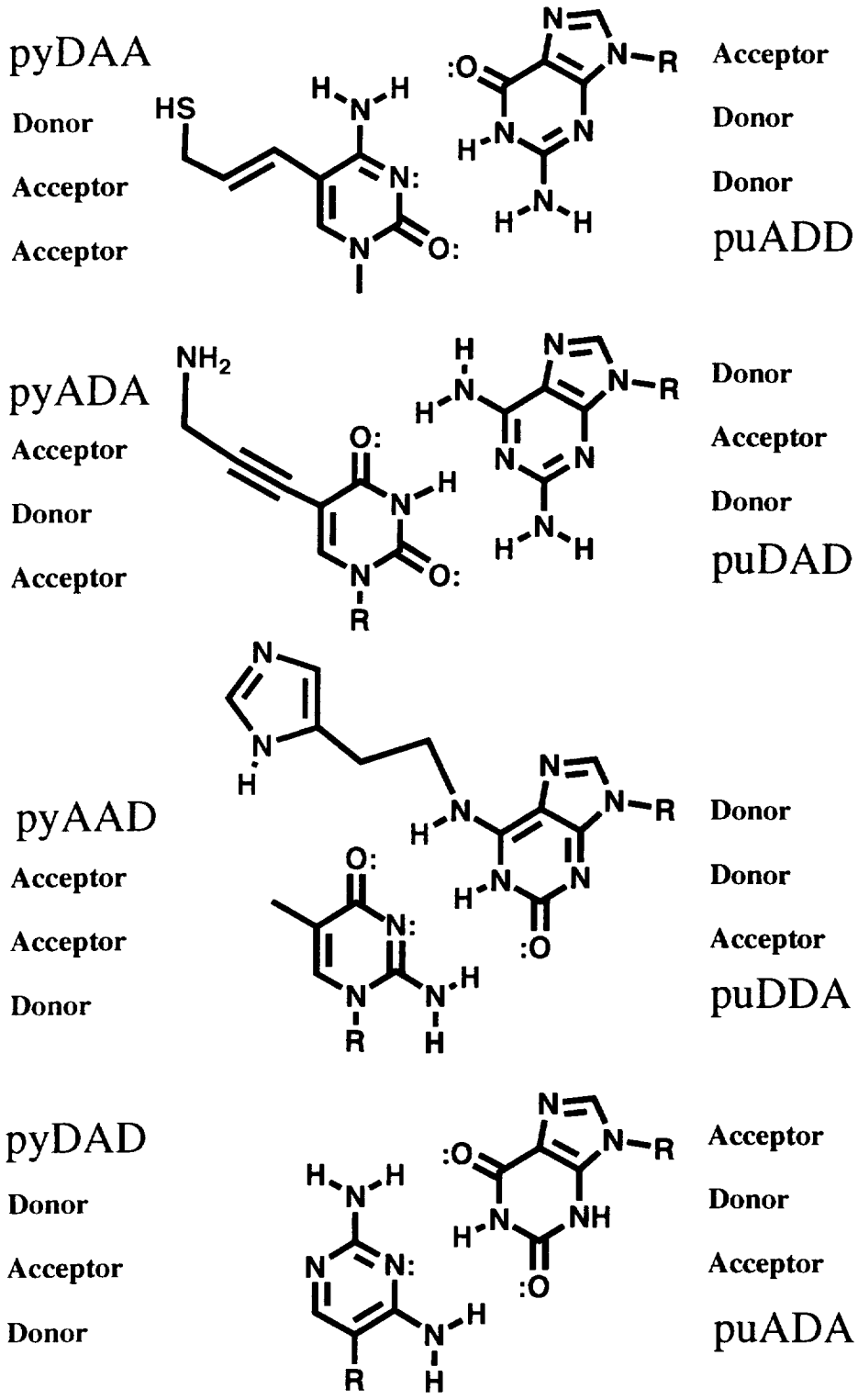
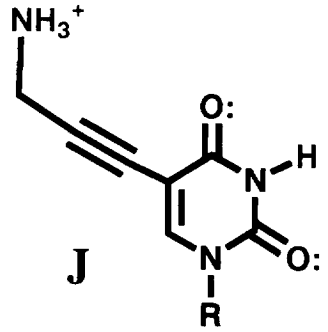
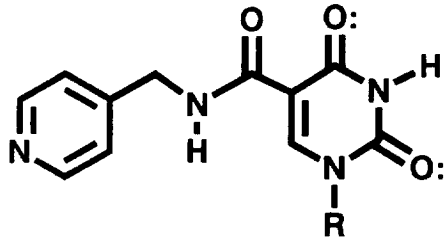

pyDAA
Donor
Acceptor
Acceptor
Acceptor
Donor
Donor
puADD
pyADA
Acceptor
Donor
Acceptor
Donor
Acceptor
Donor
puDAD
pyAAD
Acceptor
Acceptor
Donor
Donor
Donor
Acceptor
puDDA
pyDAD
Donor
Acceptor
Donor
Acceptor
Donor
Acceptor
puADA

Figure 4



J