19

Chance and Necessity in Biomolecular Chemistry

Is Life as We Know It Universal?

S.A. BENNER[1] and C.Y. SWITZER[2]

[1]Departments of Chemistry and Anatomy and Cell Biology, University of Florida, Gainesville, FL 32611, U.S.A.
[2]Department of Chemistry, University of California, Riverside, CA 92502, U.S.A.

Keywords: Darwinian evolution, biopolymer models, indicators of life

ABSTRACT

Life is defined as a self-sustaining chemical system capable of undergoing Darwinian evolution. The structures of biopolymers that might support life so defined are discussed in this chapter with an eye towards identifying features of these biopolymers that might be found in any life form, regardless of its genesis, and inaccessible (or improbable) as the product of any nonliving chemical system. Also explored is whether the structure of contemporary life is more or less complex than it needs to be, and whether systems with greater complexity existed earlier in the history of the planet. We then consider why some theories are successful at guiding the design and manipulation of complex organic molecules, while others are not.

INTRODUCTION

Unless it is intelligent enough to speak to us directly, extraterrestrial life will ultimately be recognized by some distinctive character of its chemistry. Chemistry may be even more useful than physiology (or fossil remnants of physiology) for this purpose, as the recent discussion of the "fossils" from Martian meteorites has shown (see

Simplicity and Complexity in Proteins and Nucleic Acids
Edited by H. Frauenfelder, J. Deisenhofer, and P.G. Wolynes

review in Kerr 1997). This discussion has been largely uninformed by relevant information from contemporary bio-organic chemistry, which must ultimately be brought to bear on the search for evidence of life on Mars. Nevertheless, it is clear from the published literature that the microstructures found in the Allan Hills 84001 meteorite might have been formed by nonliving processes as well as living processes. Thus, such microfossils cannot be taken as reliable indicators of life.

The same can be said, of course, for many chemicals that might be viewed as signatures of life. Fifty years ago, the discovery of adenine or alanine or ribose — three chemical building blocks of nucleic acids, proteins, and carbohydrates, respectively — in extraterrestrial samples might have been viewed as evidence for extraterrestrial life. No longer. "Prebiotic chemistry," which seeks to create abiotically in the laboratory the chemical components of terrestrial living systems, has produced adenine, alanine, and ribose derivatives, not only without the intervention of enzymes and cells, but also under conditions that (with a little imagination) are conceivable at least somewhere in the galaxy. While successful prebiotic experiments increase the possibility of spontaneous generation of life, every successful prebiotic experiment also diminishes the number of molecules that can be used as reliable signatures of life.

The problem is even more complicated, however. Samples where extraterrestrial life perhaps *used* to exist but exists no more (samples from the surface of Mars, for example) will undoubtedly be obtained by the NASA planetary exploration program before samples where extraterrestrial life perhaps *presently* exists (the oceans of Europa or the subsurface of Mars). Detecting the chemical and structural signatures of expired life is more difficult than identifying those of extant life, as degradation of carbon-containing molecules from life yields products that resemble even more the products of nonlife. For example, "polycyclic aromatic hydrocarbons" (PAHs) are molecular fragments of graphite. Such fragments are found in cigarette smoke and coal (indisputably derived from organic matter) and in interstellar space, presumably not arising from living matter (Li and Greenberg 1997). They will undoubtedly be found on Mars, provided that we look hard enough. From a chemical point of view, it will be difficult to tell whether these arose from carbon in living systems or carbon in nonliving systems. Likewise, in the evidently oxidizing top surface of Mars (Zent and McKay 1994), organic matter is likely to be degraded to carbon dioxide, also available from nonbiological sources.

Thus we are left with a conundrum. If structures that look like fossils of cells can be generated by nonlife, and if molecules that look like biochemicals (or their degradation products) can be generated by nonlife, how are we to identify life (or its remnants) on Mars, Europa, or elsewhere?

This conundrum can be solved if we can come to understand what chemical features are *universal* in living systems (that is, found in life regardless of its genesis) and *unique* to living systems (that is, not generated by any nonlife processes). To begin, we adopt the working definition of "life" proposed by a committee that addressed this problem for NASA in 1994 (Joyce et al. 1994), modified as suggested by Luisi (1998) to state explicitly certain aspects of the definition that NASA implied but did not state:

Life is a chemical system that is self-sustaining via the internal synthesis of components from external energy and nutrients, and capable of undergoing Darwinian evolution (reproduction, mutation, and natural selection).

The definition makes it evident that biology is not a discipline distinct from chemistry (a point perhaps obvious in 1999, when it is difficult to find a biologist who is not attempting to put a molecular face on the biological phenomena under study). The definition also captures the fundamentals of our understanding of how life works. Very few mechanisms have been proposed for how molecular systems could emerge with the high information content needed to be self-sustaining in a chemical sense:

1. Creationism, generation of heritable information in a living system by another system carrying still more information.
2. Lamarkianism, feedback of information from the environment to be encoded in a heritable form.
3. Darwinianism, the generation of heritable information by a process of random variation followed by natural selection.

Of these, Darwinianism is the only mechanism known to function in the natural world. Biology can therefore be defined as the subdiscipline of chemistry that studies molecular systems which undergo Darwinian evolution, that is, can reproduce, can suffer mutation, and evolve under the pressures of natural selection.

THE TWO-BIOPOLYMER MODEL

What features are universal in molecular systems that can undergo Darwinian evolution? One must begin with the only example of molecular systems capable of Darwinian evolution known to us: life on Earth. While one has little choice but to start here, we recognize the prejudice created by our knowledge of terrestrial life. A rather comprehensive knowledge of chemistry is required to go even one step beyond these prejudices (and even then, it is very difficult). Sociological problems are also confounding. "Chemistry for biologists" is frequently one of the worst courses in the academic curriculum. Conversely, most organic chemists do not involve themselves in issues central to defining organic reactivity that is general to life; many consider this type of research "flaky." Thus, it is difficult to match the problem with the expertise needed to solve it.

The one self-sustaining chemical system indisputably known to undergo Darwinian evolution is the "two-biopolymer" system built from nucleic acids and their encoded proteins of contemporary terrestrial life.[1] Proteins and DNA might be sought as

[1] Terrestrial life is actually a three-biopolymer system: DNA, RNA, and proteins. Molecular biologists often treat RNA and DNA as "the same," but they are not, and need not have been as similar as they actually are, given what we know about how small perturbations in backbone structure change the properties of oligonucleotides.

extraterrestrial markers of life. Indeed, if they were to be found together on Mars or Europa (and if the possibility of recent terrestrial contamination could be excluded), they would be strong indicators of extraterrestrial life in those locales.

However, before planetary expeditions to Mars or Europa are designed from this optimistic scenario, we must recognize that this starting point carries with it the prejudices noted above. We know that a two-biopolymer system works on Earth, but is the system universal? In particular:

1. Is it universal in its details, (i.e., will all life on all planets use proteins and nucleic acids with the same twenty amino acids and the same four heterocyclic nucleobases, which we will call *standard* biopolymers)?
2. Is it universal without its details, (i.e., will all life use proteins and nucleic acids, with the amide linkages and phosphodiester linkages as the universals, but with different amino acid side chains and the nucleobases)
3. Is the two-biopolymer system universal only in its broad outlines, (i.e., will all life use two biopolymers, one as an information repository, the other for catalysis, with the first encoding the other) but based on biopolymers other than those joined by peptide (amide) or phosphodiester bonds?

Considerable experimental work now shows that Pt. 2 is chemically possible. Nucleic acids with twelve distinct bases joined as base pairs via six different patterns of hydrogen bonding are competent as components of a genetic system (Switzer et al. 1989; Piccirilli et al. 1990). These can be used to expand the number of amino acids in encoded peptides (Bain et al. 1992), and technology has been developed by Hecht, Schultz, Chamberlin, and others that synthesize polypeptides with nonstandard amino acids by recruiting the stop codon (for a review, see Benner 1994). These experimental results rule out Pt. 1 as a defensible assumption when seeking signs of extraterrestrial life. Even if a two-biopolymer model for life is assumed to be universal, the scope of the search must extend past the chemistry of terrestrial biopolymers to at least Pt. 2.

THE ONE-BIOPOLYMER MODEL

Even the terrestrial version of the "two-biopolymer" model contains a well-recognized paradox, one that relates to "origins." On Earth, the two biopolymers (proteins and nucleic acids) are specialized to perform two functions (catalysis and information repository). The sequence of one is related to the sequence of the other by an encoded–encoder relationship. It is highly improbable, however, that both biopolymers arose spontaneously and simultaneously in an environment lacking life, and astronomically improbable that they arose in this way with an encoded–encoder relationship.

A variety of "single-biopolymer" hypotheses have been proposed to address this paradox. Such hypotheses hold that life originated as a single biopolymer that performed both catalytic and repository roles. A variety of these are based on "genetic takeover" models, where one biopolymer presumed to be easy to form under prebiotic conditions eventually templates the standard biopolymers known in contemporary terrestrial life.

The most popular model today is the hypothesis that the first (or perhaps an early) molecular system capable of Darwinian evolution was based on RNA, a hypothesis that dates back to seminal papers by Alex Rich (1962), Woese (1967), Orgel (1968), and Crick (1968), Usher and McHale (1976), White (1976), and Visser and Kellogg (1978). The phenomenal discoveries by Cech and Altman (Cech et al. 1981; Zaug and Cech 1986), that RNA performs catalytic functions in contemporary organisms, has made the "RNA world" a part of the culture in contemporary molecular biology (Watson et al. 1987).

Even simple "single-biopolymer" models have relevance to the search for extraterrestrial life. For example, biologists have taken pains to point out 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.

Actually, 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 amino acid biosynthesis enzymes, for example, which comprise more than half of what is believed to be the "core metabolism" [Benner et al. 1993]) is also not needed for life. 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.[2] A similar analysis can be applied to any structures that might be encountered in planetary exploration missions.

Because they do not require the invention of translation, we might naively view single-biopolymer life forms to be the most likely form of life to be encountered extraterrestrially. In fact, this view requires that three hypotheses be true:

1. Single biopolymers capable of self-sustenance and Darwinian evolution must exist (the "existence" hypothesis).

[2] If one accepts the notion that life on Earth began with an organism that used a single biopolymer such as RNA, one might ask whether there might be environments where a single-biopolymer organism would remain even after the invention of translation. Normally, the superior power of proteins as catalysts provides presumptive arguments that the older form of life was extinguished by the derived form, and none remains on Earth. The fact that cells containing single-biopolymer life can be much smaller than two-biopolymer cells suggests, however, that the one-biopolymer life might survive where small size offers a selective advantage. In subterranean matrices, e.g., 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.

2. The fitness of such biopolymers must be reasonable (the "fitness" hypothesis).
3. The "breakthrough" (Benner, Ellington et al. 1989) to translation must be chemically challenging.[3]

If these hypotheses are true, then we expect to encounter single-biopolymer systems in the galaxy with higher probability than two-biopolymer systems. If intrinsic chemical contradictions of a single biopolymer attempting to both store information and catalyze reactions (see below) render them intrinsically unfit, then single-biopolymer systems that fail to invent translation rapidly are expected to be easily extinguished, and the ratio of two-biopolymer living systems to single-biopolymer systems will be proportionately higher (at the same time as implying that the total number of living systems in the galaxy will be fewer).

IN SEARCH OF UNIVERSAL FEATURES OF A SINGLE-BIOPOLYMER LIFE

What then are expected to be molecular features both universal in and unique to single-biopolymeric systems capable of undergoing Darwinian evolution? What structures bring together replicatability, mutability, and adaptive function[4] (e.g., catalytic activity) in one biopolymer that can evolve in response to natural selection? As with any research, one cannot ask such questions in a vacuum. "Structure space" is far too large; there are simply too many polymers that might be made to undertake a research project to make them one at a time and inspect their properties, although this approach has certainly been tried (Freier and Altmann 1997). Instead, the search must be rationally constrained. Several constraints being used by many workers are briefly summarized below.

Constraint 1: Attempt to limit the range of possibilities using prebiotic chemistry

One of the successes of exobiology has been the generation of organic molecules characteristic of living systems under prebiotic conditions. If only a few organic molecules could have emerged spontaneously under conditions inferred for primitive Earth, prebiotic chemistry would be a powerful guide in our search for universal structures in life (assuming that wherever life emerged, the same prebiotic chemistry prevailed). It would force us to consider these (and only these) few molecules as the starting point to obtain Darwinian systems. Conversely, if many organic molecules could have been generated under conditions inferred for primitive Earth (or if many conditions can be imagined, each leading to different sets of prebiotic chemicals), then prebiotic chemistry is a less useful guide for research in Darwinian chemistry.

Unfortunately, prebiotic studies have been so successful that the latter is more the case than the former. An example is found in the "tholins" of the Cornell group, which have the deep red-brown color known to those who have cleaned out a chemical stockroom (Khare et al. 1993). Even when generated from simple organic compounds such as ethane (and more so if ammonia is added, and still more so if additional compounds are added to the mix), tholins contain so many compounds that it is difficult to exclude any specific compound in particular from the prebiotic environment, at least in small amounts. Further, the conditions under which these mixtures are formed (e.g., ethane in water with irradiation), appear to be quite accessible on both early and modern planets.

Indeed, an entirely new problem has emerged from prebiotic research: the problem of complexity. For example, template-directed synthesis of chiral compounds can be inhibited by building blocks of the opposite chirality (Orgel 1992; Schmidt et al. 1997). This is one example of a well-known problem relating to the chemistry of complex mixtures: one component of a prebiotic mixture can inhibit as well as facilitate a desirable reaction by another component of the same mixture. Chemists have long known that it helps to isolate compounds in relatively pure form before trying to crystallize them. Combinatorial chemistry, a new approach in medicinal chemistry derived by analog to evolution in natural systems (Thompson and Ellman 1996), is troubled by similar problems.

One task for the next phase of prebiotic/combinatorial chemistry is to define how complex mixtures might be resolved to obtain mixtures with useful complexity (but not greater). Indeed, at this stage in the development of prebiotic chemistry as a field, it would be more useful to ask how mixtures of organic chemicals might be fractionated on a primitive planet. Life may have been guided not by what organic molecules were synthesized in large quantities in a prebiotic setting, but by what compounds were efficiently concentrated by prebiotic chemical and mechanical processes.[5]

The second reason limiting prebiotic chemistry as a guide for Darwinian chemistry is the converse of the first. Even while many compounds are now known to be generated abiotically, there is no guarantee that a compound *not* presently known to be a product of abiotic synthesis will not become "prebiotic" by some future abiotic experiment. This has led to a sociological problem in the field, the propensity of prebiotic

[3] None of these points are free of controversy. For example, dozens of laboratories have attempted to identify an inherent chemical relationship between the DNA/RNA codon and the encoded amino acid ever since the genetic code was established. Such efforts are ongoing in many laboratories (we cannot cite all of these papers; two recent papers with leading references are DiGiulio [1997] and Alberti [1997]). If an intrinsic link exists between the chemistry of proteins and nucleic acids, then inventing translation becomes proportionately easier, which in turn increases the ratio of two-biopolymer living systems to single-biopolymer living systems in the galaxy, at the same time as increasing the likelihood that those two biopolymers are indeed proteins and nucleic acids. Indeed, it would argue that the genetic code is truly universal (in the cosmologic sense of the term).

[4] The phrase "adaptive function" is redundant in the Darwinian view, which holds that function can arise only through selection placed atop random variation. We use the redundancy here because so many biochemists confuse "function" with "behavior," "structure," or other observables that may or may not be adaptive.

[5] We do not regard this problem as hopeless. An active planet has abundant opportunity to fractionate, crystallize, and separate, as crystalline mineral deposits illustrate.

chemists to pronounce a compound as "not prebiological" simply because it is not (yet) known to be generated abiologically.[6]

Constraint 2: Attempt to limit the range of possibilities by identifying molecular systems that undergo template-directed reproduction

Reproduction is, of course, the first property that a molecular system must display to partake in Darwinian evolution. Many molecular systems make copies of themselves, of course, including some in quite trivial ways. For example, a crystal of sodium chlorate ($NaClO_3$), if fragmented, will nucleate (or "seed") the formation of more crystals of $NaClO_3$. The surface of the crystal is a template for the self-assembly of freely wandering ions into organized matter. In this particular case, the system spontaneously generates homochirality, a property widely believed to be unique to living systems (Macdermott et al. 1996). Rectangular prisms of $NaClO_3$, in either the right-hand or left-hand form of the chiral space group P2(1)3, form from the solutions of the achiral ions. If the solution is stirred, only right-handed or only left-handed crystals are formed (McBride and Carter 1991). If the solution is *not* stirred, crystals of both enantiomorphic forms precipitate in equal amounts. This highly reproducible result is a consequence of a "self-reproduction." Stirring breaks apart the first crystal formed with random chirality, whose fragments then seed the formation of many "daughter" crystals with the same chirality.

Templated replication is widespread in chemistry, and many examples are now known in bio-organic chemistry as well. Rebek (Wintner et al. 1994; cf. the challenge by Menger et al. [1995] over the validity of Rebek's interpretation and Rebek's rebuttal [Wintner et al. 1995]), Ghadiri (Lee et al. 1997), and von Kiedrowski et al. (1991) have shown how small molecules, peptides, and oligonucleotides undergo template-directed synthesis. It seems not to be excessively difficult to design such systems, and many are being examined in many laboratories.

Just as the success of prebiotic chemistry diminishes its value to constrain the search for structures that are universal molecular features of life, so has the widespread success of experiments to obtain templated reproduction. If self-templating systems made with nucleic acids work, with peptides, with salts, and with small organic molecules of many shapes and sizes, no single system has the "universality" attribute that demands its selection as a system for further study, let alone as a system that might be sought on Mars or Europa.

[6] Our favorite example of this is the discovery by Albert Eschenmoser of a relatively efficient abiological synthesis of ribose-2,4-diphosphate in the context of a literature that had all but excluded ribose as a "prebiotic" substance. Whether the Eschenmoser synthesis was in fact used prebiotically remains uncertain. However, the point remains: just because we do not know of a "prebiotic" synthesis of a compound does not mean that one does not exist.

Constraint 3: COSMIC-LOPER biopolymers

Mutation is the second property that a molecular system must display to partake in Darwinian evolution. In some usage, "mutation" denotes any structural change. Thus, a molecular system of $NaClO_3$ might be said to "mutate" if the sodium is replaced by potassium to yield $KClO_3$.

In practice, mutation cannot be this drastic if it is to be effectively coupled with natural selection. This is both a practical matter and a theoretical one. At the very least, mutation in a molecular system that can self-replicate must not change the physical properties of the molecules involved in a way that obviates mechanisms used for self-replication. The replacement of Na by K in the example above does.

To support a self-sustaining chemical system capable of undergoing Darwinian evolution, a biopolymer must be capable of searching mutation-space independent of concern that it will lose properties essential for replication. We will designate biopolymers that have this property as COSMIC-LOPER biopolymers (the acronym stands for "Capable Of Sustaining Mutation Independent of Concern over Loss Of Properties Essential for Replication").

The need for a single-biopolymer form of life to be based on a COSMIC-LOPER biopolymer is nearly axiomatic. If a substantial fraction of the mutations encompassed within an encoding system cause a biopolymer to precipitate, unfold, form a new folded structure and therefore no longer be recognized by a catalyst for replication, or acquire some other physical trait that makes it unable to guide template-directed replication, the biopolymer is not suited for Darwinian chemistry. Thus, the COSMIC-LOPER property is expected to be a universal in single-biopolymer living systems. We might conjecture that two-biopolymer forms of life must contain one COSMIC-LOPER polymer.

In contemporary terrestrial life, it is clear that proteins are not COSMIC ("Capable of Sustaining Mutation Independent of Concern") polymers — even in cases where they can direct template-based replication (as the splendid work of Ghadiri has found). The physical properties of proteins can change dramatically upon point mutation within the mutation space allowed by the twenty standard amino acids. The textbook case is sickle-cell hemoglobin, where a single amino acid substitution creates a variant that undergoes partial precipitation.

It is well known to those who work with peptides that precipitation is a common property; a peptide of random sequence chosen from the twenty standard amino acid is most likely to be insoluble, a property also known to those who boil eggs (after heating to escape the metastable state where the proteins are soluble). Thus, it is not remarkable that sickle-cell hemoglobin precipitates; what is remarkable is that native hemoglobin is soluble at the nearly gram-per-milliliter concentrations found in red blood cells. One of the most characteristic features of the divergent evolution of the hemoglobin gene family is the extent to which substitution has been constrained to avoid insolubility.

Experimental work illustrates the non-COSMIC properties of peptides. Small (14 amino acid) polypeptides that formed nearly exclusively a four-helix bundle have been designed (Johnsson et al. 1990, 1993) and their structure determined by NMR in solution, one of the first times that this was done for a designed peptide. The identical peptide was then prepared, but with an N-terminal acetyl group. The small change was sufficient to cause the peptide to aggregate. Still other point mutations caused the helix not to form. If solubility and/or helix formation are essential to the replicatability of a peptide template, a large range of plausible mutation would destroy it, at least in this system.

Synthetic organic chemists have long been aware of the anti-COSMIC behavior of many small molecules. The most common situation is encountered in natural product synthesis. Frequently, synthetic efforts on the natural product are preceded by exploratory work with a "model system," a simpler molecule that is more accessible than the natural product itself, but which is presumed to be representative of the kinds of problems that will be encountered when the natural product is synthesized. As often as not, reactivities of the model and the real natural product are sufficiently different that chemistry developed on the first is defeated by the second. A single methyl group can make the difference, with the practical implications of real chemistry unanticipated by models.

The molecular biologist relies on the fact that natural oligonucleotides with different sequences nevertheless behave similarly. Every (or almost every) oligonucleotide will precipitate in ethanol. Every (or almost every) 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 molecule.

Even standard nucleic acids are not entirely capable of searching mutation space independent of concern that they will lose properties essential to replication. Best known to the molecular biologist are RNA molecules that have G-rich sequences. These adopt tertiary structures around a G-quartet, and these structures often disrupt the templating ability of an RNA sequence (Wang and Patel 1994). This means that if an RNA molecule searches G-rich regions of sequence space, it runs the risk that it will lose certain properties essential for replication. In comparison with virtually every other class of organic molecule, however, nucleic acids are the most COSMIC-LOPER polymer known.

The notion of COSMIC-LOPER behavior is virtually absent from theoretical models of self-replicating systems. These nearly always simply assume that some molecular system (generally with undefined structure) exists that can be a template, a catalyst, and undergo COSMIC mutation to improve fitness, all at the same time. In the real world with real molecules displaying real reactivity, it is not so simple. Indeed, many theoretical discussions of self-replication can be criticized because they are disconnected from the chemical realities of real organic molecules.

In "What if?" and "Why not?" studies, some of the structural features of nucleic acids that give them their COSMIC-LOPER properties have been defined. One of the most important of these studies involved the generation of oligonucleotide analogs with nonionic backbones. One of the best studied of these took a small step from the natural backbone, replacing the phosphate diester linkers in DNA and RNA by nonionic dimethylenesulfone linking units. The sulfone group is an "isosteric" and "isoelectronic" replacement for a phosphate. The sulfone functional group is dipolar but nonionic, intrinsically soluble in water, stable to alkaline degradation, and not stereogenic (Richert et al. 1996).

Nevertheless, oligomers built from these units display some remarkable properties. First, they fold: for example, the octamer $AsO_2UsO_2GsO_2GsO_2UsO_2CsO_2$ AsO_2U folds in solution to give a folded form in water having a high melting temperature (ca. 87°C). More remarkably, 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, such behavior is not a general property of oligosulfones. For example, the sequence $d(TsO_2TsO_2TsO_2CsO_2TsO_2TsO_2TsO_2T)$ displays in water no particularly stable folded conformation, judging by a multidimensional NMR spectrum (Greenbaum, Eschgfäller, Benner, unpublished). The sequence GsO_2C in the crystal gives a antiparallel duplex approximately isomorphous with the analogous RNA. The AsO_2T gives an open structure with no Watson–Crick pairing in the crystal. The UsO_2C dinucleotide gives a complex featuring backbone-to-backbone and backbone-to-nucleobase hydrogen bonds (Steinbeck and Richert, unpublished).

Even within a relatively small search of mutation space, these nonionic oligonucleotide analogs retained no conformational or physical property that would offer a common basis for replication. Indeed, the concern that the next variant with a new sequence would display some unusual physical or catalytic behavior remains the central concern of the graduate students working with these molecules. In this respect, oligosulfone analogs of DNA and RNA have much the same behavior as peptides and small organic molecules, not the nucleic acids upon which they are modeled.

These results show that the need for COSMIC-LOPER behavior is a strong constraint on what types of biopolymers need to be considered as the basis for single-biopolymer life. They also suggest that a polyelectrolyte (polyanion or polycation) structure is important for the regular, largely sequence-independent behaviors that we see in nucleic acids. In particular, the anionic phosphate groups confer several properties important for COSMIC-LOPER behavior independent:

(a) Phosphate groups force the interaction surface between strands as far distant from the backbone as possible. This is the Watson–Crick "edge" of the nucleobases. As a consequence, 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, Watson–Crick–Hoogsteen interstrand interactions, all become important without

interstrand phosphate–phosphate repulsion. Non-ionic oligonucleotide analogs should (and do) have rich intermolecular conformational properties (just like peptides).

(b) Phosphates keep the DNA molecule from folding on itself, allowing it to act as a template. 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) have rich *intra*molecular conformational properties. They should (and do) fold just like peptides. The repeating charge in the backbone is responsible for the rule-based pairing of complementary DNA strands.

(c) Electronic distribution in a molecule is described as an infinite series (monopole + dipole + quadrapole + ...). The first nonvanishing term dominates the behavior of the molecule, when viewed at a distance. The repeating monopole (charge) in DNA renders dipolar interactions (hydrogen bonding) secondary. The physical behavior of a DNA molecule is the same regardless of its sequence, to a first approximation. An encoding molecule needs physical properties that are largely independent of its sequence. One does not want to mutate a gene to get a better protein, only to discover that the mutant DNA gene precipitates.

From these experiments has come a working hypothesis that may guide our search for universal chemical structures in single-biopolymer systems, and in encoding biopolymers in all forms of life:

Hypothesis 1: As a universal chemical characteristic, living systems must contain at least one biopolymer having a repeating charge, either polyanionic or polycationic,[7] as these are structures that are mutable without creating dysfunctional physical properties (such as precipitation). This biopolymer will perform both catalytic and repository roles (in a one-biopolymer system) and the information repository role (in a two-biopolymer system).

This hypothesis constrains our studies from the universe of all molecules that are known to undergo template-directed replication (including salts, small organic molecules, peptides, and nucleic acids) to a subset of these, those with repeating charges. This directs us to nucleic acids at a first level, to nucleic acid analogs at the next, to cationic analogs at the next (Dempcy et al. 1994), and then to other polycationic and polyanionic molecules. While an extraordinarily large number of structures remain, the scope of exploration is limited.

This hypothesis is particularly useful because if true, it suggests a way to detect chemical remnants of life in extraterrestrial samples: One searches for biopolymers or their fragments with regularly spaced positive or negative charges. This turns out to be a rather simple structural feature to search for experimentally. A chip having an

[7] In water.

absorbent with a regularly spaced negative or positive charge will bind the polycation or polyanion (respectively) more tightly than competing monocations or anions.

A molecular system also exists that can challenge the hypothesis. Peptide-linked Nucleic-acid Analogs (PNAs) are a nonionic oligonucleotide analog that is known to support rule-based molecular recognition of oligonucleotides following Watson–Crick rules. Structural studies with PNAs suggest a novel type of interstrand interaction that makes these molecules behave so differently from other nonionic oligonucleotide analogs. Miller (1997) recently argued on prebiotic grounds that PNAs might be a predecessor to RNA as the first living biopolymer. PNA is therefore a likely candidate for a biopolymer that violates Hypothesis 1.

Those who work experimentally with PNAs recognize signs of anti-COSMIC behavior. Their propensity to self-aggregate is recorded in the literature (Egholm et al. 1992), as is self-structure that appears to interfere with PNA–DNA duplex formation (Dueholm et al. 1994). G-rich sequences of PNAs are again problematic, although this phenomenon is not well explored. Many subtle modifications on the backbone destroy Watson–Crick like pairing. Not surprisingly, modification of the ends on the PNA polymer that might improve the COSMIC properties of PNAs involve the addition of charges (see, for example, Gangamani et al. 1997).

Constraint 4: Demands placed on a biopolymer under direct selection pressure for fitness

Perhaps the most useful constraint in the search for universal and unique structural features of Darwinian biopolymers arises from their need to contribute fitness. In a single-biopolymer system, fitness is more than replicatability and mutability. Fitness also requires a chemical feature that confers self-sustaining properties on the system, in particular, to enable the internal synthesis of components from external energy and nutrients. Most simply, fitness can arise if the biopolymer can have catalytic power, and this will be the primary focus of the work proposed here. A biopolymer that can catalyze reactions essential to its own self-reproduction at a higher rate or with greater efficiency is presumably fitter than a competing biopolymer.

Following the work of Cech and Altman, which demonstrated the catalytic power of RNA in contemporary terrestrial organisms (work foreshadowed by reports by Usher and others of RNA catalysis in a variety of model systems [Usher and McHale 1976]), RNA became the system of choice for seeking a single-biopolymer system capable of Darwinian evolution. Coupling affinity separations with PCR amplification led to the development of *in vitro* selection in the pioneering work of Joyce, Szostak, Cech, Gold, and Ellington. This technology will come, we believe, to be recognized as one of distinctive scientific advances of the late twentieth century.

At one level, these experiments have been fantastically successful. Aptamers, catalysts, ligands, and receptors have all been prepared. At another, these experiments have raised serious doubts that catalytic RNA formed the first living system. Some carefully quantitative *in vitro* selection experiments have shown that RNA displays

notably poor fitness as a catalytic biopolymer relative (for example) to proteins (Bartel and Szostak 1993).[8]

One measure of the intrinsic fitness of a biopolymer for a particular role is given by the fraction of the sequence space that performs this role to a defined level of specification. For example, to have a population of antibodies with a 50% chance of containing one that has an affinity for a specified ligand with a disassociation constant of approximately 1 mM, approximately 1 million antibody molecules must be prepared. This gives one a measure of the intrinsic ability of the antibody scaffolding to be a receptor. If the population size needed to be larger to have the same result, the scaffolding would be poor.

In vitro selection experiments are capable of measuring fitness in this way with some accuracy. For example, elegant studies by Bartel and Szostak (1993) have shown that to find a single RNA molecule that increases the rate of a templated ligation (a simple reaction) by a modest four orders of magnitude, one must examine ca. 2×10^{13} random RNA sequences. This number must be normalized over the size of the sequence space possible (only a small fraction of the total sequence space is searched in the Bartel–Szostak experiment, of course).

This measure of intrinsic fitness can be contrasted with the intrinsic capacity of a peptide for catalysis. For example, short peptides designed to catalyze the decarboxylation of oxaloacetate produced a rate enhancement of between three and four orders of magnitude for the rate-determining step. In this case, the sequence space contained approximately 10^{18} molecules. A combination of design (Johnsson et al. 1990, 1993), combinatorial experiments (Perezpaya et al. 1996), and refinement (Baltzer, unpublished) were performed, with the estimate that only 10^7 random sequences must be searched to obtain this degree of rate enhancement (Johnsson et al. 1990, 1993). This suggests that peptides as a biopolymer is intrinsically 1 to 10 *million*-fold fitter as a catalyst than RNA.

This comparison is "unfair" in a variety of ways. The reactions being compared are different, each better suited to its own catalyst. The nucleic acid is, in fact, not alone the catalyst; it requires divalent metal ion as a cofactor. Nevertheless, the point is clear: nucleic acids are orders of magnitude worse catalysts than proteins.

[8] This conclusion has been clouded by literature that paradoxically refers to RNA as a "perfect" catalyst (Herschlag and Cech 1990). The argument was based on the fact that once bound, substrate nearly always turns over to product. This was proposed as a criterion for perfection in a catalyst based on a paper of Albery and Knowles. In fact, the paper of Albery and Knowles (which itself is flawed; see Ellington and Benner 1987) does not assign an enzyme the status of "perfect" if every substrate molecule, once bound, turns over to product. If substrate binding is very tight, then unbinding is very slow. Through Watson–Crick interactions, the binding of RNA molecules to other RNA molecules can be fantastically tight. Thus, very slow RNA catalysts are easily viewed as "perfect," catalysis is very, very slow. However, the release of bound substrate is still slower. So every molecule bound eventually goes to product.

The superiority of proteins as catalysts compared with RNA reflects (at the very least) the fact that proteins have more building blocks and more functional groups available to serve catalytic roles than RNA. Further, it is clear that the increased catalytic fitness of proteins as a biopolymer more than compensates for the increased size of protein sequence space (for a given molecular weight).

This has consequences for how we view the origin of terrestrial Darwinian chemistry (Benner 1987; Benner et al. 1987). Prebiotic conditions could conceivably generate a library of 10^7 random peptide sequences. It is, however, more difficult to imagine a collection of 10^{13} random RNA sequences emerging under abiotic conditions.

More important to the search for extraterrestrial life is what these and analogous experiments imply about the robustness of single-biopolymer systems when challenged to improve their fitness under Darwinian selection pressure. Joyce and his coworkers have done a fascinating series of experiments to explore the mutation space of the *Tetrahymena* ribozyme challenged in various ways (Beaudry and Joyce 1992; Tsang and Joyce 1994; Lehman and Joyce 1993). Remarkable about these studies is the breadth of the sequence space explored (5% mutagenesis of the catalytic core) and the improvement in "fitness" (substrate binding, for example, by 25-fold; catalytic rate by 50-fold; metal ion dependency by 170-fold), when compared to analogous experiments in directed protein evolution *in vitro* (Johnsson et al. 1993; Perezpaya et al. 1996), by phage display, or *in vivo* (a particularly relevant case being the adaptation of HIV reverse transcriptase when challenged with azidothymidine). By any measure, the proteins are better — much better.

A decade ago, Benner et al. (1987) discussed the intrinsic limitations of standard nucleic acids as a matrix for obtaining functional behavior under conditions of Darwinian selection. These observations generated the second hypothesis:

Hypothesis 2: If RNA is to serve as a scaffolding for a single-biopolymer chemical system capable of undergoing robust Darwinian evolution, it must have more functionality than it presently has.

This functionality can be delivered by either extra letters in the genetic alphabet or by cofactors designed to bind to oligonucleotides. Interestingly, Miller has recently shown that 5-position functionalized pyrimidines are accessible under "prebiotic" conditions (Miller 1997). Further, functionalized ribonucleosides found in tRNA are very ancient and carry much of the functionality needed for catalysis (reviewed in Benner, Ellington et al. 1989).

The need for the polymer in a single-biopolymer Darwinian system to be both capable of suffering mutation independent of concern over the loss of properties essential for replication (COSMIC-LOPER) and to confer fitness through catalytic activity generates competing (and occasionally contradictory) demands on the structure of the biopolymer. 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. The inverse relation between fidelity and the number of building blocks is seen in both theory (Szathmary 1992) and experiment (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 does not fold easily, so that it can serve as a template. This is one of the clearest contradictions between the demands for a catalyst and the demands for an information repository.
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 have physical properties largely unchanged even after substantial change in its sequence, so that the polymer remains acceptable to the enzymes required for replication (the COSMIC-LOPER property).

The contradiction between the structures/properties required for catalysis and those required for information storage creates problems. At the very least, any biopolymer forming a single-biopolymer Darwinian system must be a compromise between these goals. Under this model, the advantage of a two-biopolymer system is that it allows the two biopolymers to specialize for genetic and catalytic roles, respectively. Thus, if an RNA world existed and if it had more than the four standard building blocks, one would have expected the number of building blocks to have been reduced after the breakthrough to translation to allow specialization of nucleic acids for encoding function. Further, once invented, proteins would have been expected to acquire functionality rapidly to match that not already available in RNA cofactors. The "palimpsest" of modern metabolism can be read exactly as such, if one is inclined to do so (Benner, Glasfeld, et al. 1989).

Yet no law requires that any real compound be able to make this compromise in a satisfactory way. 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 could achieve a suitable compromise. This would make the single-biopolymer model for the origin of life 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, not robust, and easily extinguished.

Conversely, it is conceivable that many biopolymeric systems could make an acceptable compromise between the demands of catalysis and the demands of genetics. If so, life would be abundant in the universe, and in many forms. There would be a less pressing need to invent two-biopolymer systems once a one-biopolymer system was established. Small cells would flourish with many different types of biopolymers, with only a few making the breakthrough to two-biopolymer systems.

Organic chemistry seems, in reality, to offer neither extreme. Hypothesis 1 holds that the information repository molecule must be a polyelectrolyte (polyanionic or polycationic). It is clear that some (albeit poor, and not very diverse catalysis) can emerge from RNA and DNA as a representative polyelectrolyte. It is also clear that the number of building blocks in DNA and RNA can be increased without destroying COSMIC-LOPER behavior (Switzer et al. 1989; Piccirilli et al. 1990). Thus, the problem becomes experimentally interesting. The more we can do in the laboratory to increase the structural diversity of polyanionic biopolymers (such as RNA and DNA) without destroying their COSMIC-LOPER properties, the more likely life is to emerge under the single-biopolymer model, the more likely it is to have been robust after it emerged, and the more likely we are to encounter it in planetary exploration.

It could be that many biopolymer systems retain COSMIC-LOPER behavior with a wide range of building blocks. If so, we may return to prebiotic analysis. With "What if?" and "Why not?" experiments showing the capability of a wide range of structures to be self-templating, and prebiotic chemistry experiments showing a wide range of structures accessible abiologically, the next phase of exobiological research must learn whether or not these chemical contradictions are resolvable by any real biopolymer (the "existence" question) and whether the fitness of this biopolymer is in any way sufficient to provide "self-sustaining" properties required by life.

With these thoughts in mind, we discuss briefly possible chemical embodiments of a possible biopolymer that combines the replicatability of nucleic acids with the functional group diversity found in proteins, without obviously destroying the COSMIC-LOPER properties necessary for the biopolymer to be a substrate for Darwinian evolution. One embodiment is based on the recognition that the Watson–Crick base pair could easily accommodate 12 bases forming 6 distinct base pairs, rather than the two found in natural RNA, by exploiting more of the hydrogen bonding patterns conceivable with three hydrogen bond acceptor/donor groups (Figure 19.1) (Switzer et al. 1989; Piccirilli et al. 1990). Additional letters in the genetic alphabet could carry a richer diversity of functionality; indeed (Figure 19. 2), one could readily conceive of a new type of biopolymer, one carrying functionalization like proteins but that can be copied like nucleic acids (Kodra and Benner 1997).

Some details of modern biochemistry support the view that life on earth might have at one point early in its history exploited an expanded alphabet of more than four nucleobases. Furthermore, the RNA world might have expanded its functional diversity by incorporating modified standard nucleobases. Contemporary tRNA and rRNA are known to contain much of the functionality lacking in contemporary encoded RNA, including amino, carboxylate, and aliphatic hydrophobic groups (Figure 19.3) (Limbach et al. 1994). Some of these might even be placed by parsimony in the protogenome, the reconstructable genome at the three fold trifurcation point joining the archaebacterial, eubacterial, and eukaryotic kingdoms (Benner, Ellington et al. 1989). Could these be vestiges of a functionalized RNA world?

To obtain information relating to this question, experiments must be done. First, we must learn whether the structures drawn on paper (Figures 19.1 and 19.2) indeed can

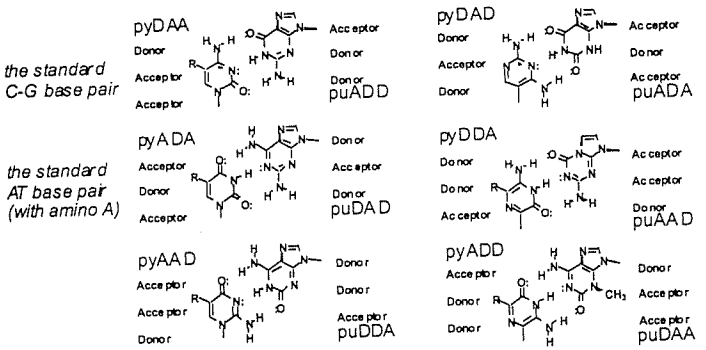

Figure 19.1 Twelve bases that are possible in a DNA- or RNA-based "alphabet" within the constraints of the Watson–Crick base pair geometry. Pyrimidine base analogs are designated by "py," purine by "pu." The upper case letters following the designation indicate the pattern of hydrogen-bonding acceptor (A) and donor (D) 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.

serve as effective components of a genetic information system. Then, we must ask whether *in vitro* selection based on an expanded genetic alphabet indeed improves the catalytic versatility of RNA. Only if both answers are affirmative would the RNA world have the motive to use an expanded genetic alphabet.

Then, we must ask whether the RNA world had the opportunity to use an expanded genetic alphabet. If the components of an expanded genetic alphabet are accessible

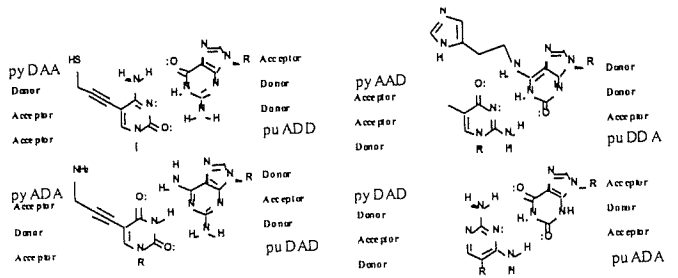

Figure 19.2 Nonstandard and standard nucleobases with functionality. Note that the pyDAD nucleobase can be protonated below pH 7 ($pK_a = 7.4$)

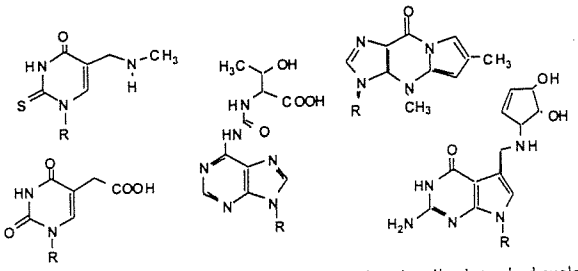

Figure 19.3 Transfer RNA contains a rich collection of functionalized standard nucleobases, created by post-transcriptional modification, that deliver functional groups (amino groups, carboxylic acid groups, aliphatic hydrophobic groups) not found within unmodified RNA. Could these be vestiges of functionalized RNA originating in the RNA world?

under abiological (prebiotic) conditions reconstructed in the laboratory, then it becomes more plausible that the RNA world exploited them.

Organic synthesis generated the nonstandard nucleobases by synthesis (Switzer et al. 1989; Piccirilli et al. 1990; Vögel, Altorfer, et al. 1993; Vögel and Benner 1994; Kodra and Benner 1997; Tarasow et al. 1997). Their structures were optimized for stability and pairing properties (Piccirilli et al. 1990; Vögel, von Krosigk, et al. 1993). New protecting group chemistry was developed to permit automated synthesis of oligonucleotides containing them (Huang and Benner 1993; von Krosigk and Benner 1995). Polymerases were found that would catalyze their incorporation into oligonucleotides by the polymerase chain reaction (Horlacher et al. 1995; Lutz et al. 1996).

Once these nonstandard nucleotides were in hand, it was possible to show that the RNA world had both the motive and the opportunity to exploit nonstandard and functionalized nucleobases. Three key advances have supported this conclusion:

(a) The first case where functionalized standard RNA has been exploited to generate a catalyst was reported by Tarasow et al. (1997). These researchers used Selex experiment with oligonucleotide libraries that incorporated a functionalized standard pyADA nucleobase (Figure 19.4). The result was a ribozyme that catalyzes a Diels-Alder reaction. As noted above, earlier Selex work with libraries that did not contain functionalized nucleotides had failed to generate a catalyst (Morris et al. 1994). For this reason, it is difficult to state quantitatively how much better this particular functionalized RNA is than standard RNA as a matrix for generating catalysis for this particular reaction. However, it appears as if the improvement is by several orders of magnitude.

(b) The first experiments where functionalized standard DNA was used to generate receptors were recently completed (Battersby et al. 1999). A functionalized 2′-deoxyuridine (trivially designated "J", Figure 19.4) was incorporated into an *in vitro* selection experiment that followed a recipe developed by Huizenga

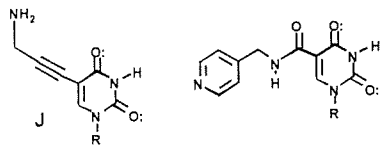

Figure 19.4 Functionalized standard bases that have been used in *in vitro* selections.

and Szostak (1995), who selected for adenosine receptors using a standard DNA library.

The results were intriguing. Most of the J-containing sequences were not analogous to those obtained from a standard DNA library, and presumably adopted a fold different from the fold adopted by the standard DNA receptors. A few of the sequences emerging from the J-containing library appeared to be analogous to the sequences obtained from a standard DNA library, however, and may fold analogously as well. Further, when one of the J-containing sequences was resynthesized with T replacing J, it still bound adenosine derivatives, only more weakly. Thus, the functionalization contributed by J influences modestly, rather than dramatically, the affinity-structure landscape, at least for receptors that bind adenosine derivatives selected under non-demanding conditions. More recent experiments (Ang, Westerman-Clark, and Benner, unpublished) show a larger impact of the functionality under more demanding conditions, especially in catalysis.

Careful analysis by capillary electrophoresis of the affinity of the J-receptors for adenosine showed that both the natural and the functionalized motifs bound not one, but two ATP molecules cooperatively. Likewise, the affinity of the receptors for ATP had converged; in both cases, the receptors are 50% at the 3 mM concentrations of ATP presented during the selection. The convergence of phenotype suggests that the outcome of this selection experiment was determined by features of the environment during the selection, in particular, a highly loaded affinity resin used in the selection step, and that an optimal molecular phenotype has been achieved by both selections for the selection conditions. This interplay between environment demanding a function from a biopolymer and the ability of the biopolymer to deliver that function is strictly analogous to that observed during natural selection.

The third advance 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) found that the intrinsic nucleophilicity of the 5-position of pyrimidines such as uracil can be exploited to generate functionalized uracil derivatives that carry positive charges at the 5-position under abiological conditions (Robertson and Miller 1995). Analogous chemistry can be used to generate other functionalized derivatives. The products resemble the amino group functionalized uracils found in some tRNA molecules (Figure 19.3). This suggests that the RNA world may have had the opportunity to, indeed, may have needed to, use some functionalized nucleosides when life first emerged on Earth.

A "SECOND GENERATION" MODEL FOR NUCLEIC ACIDS

It is intriguing to note that the "first generation" theory of nucleic acid structure proposed nearly fifty years ago by Watson and Crick permitted the design of alternative nucleobases, but failed to support the design of altered backbones. The Watson–Crick model is only barely reductionist, at least compared to quantum mechanics, molecular mechanics, or other models that are used to interpret the result of chemical modification. Nevertheless, it supports vast amounts of research effort in molecular biology, as well as the modest efforts to redesign genetic biopolymers discussed here.

There is clearly the opportunity for the development of a "second generation" model for nucleic acids, one that is supported by the synthesis of alternative nucleic acid structures. The model would certainly incorporate a more substantial role for the backbone, and especially the negative charge in the backbone. It would also incorporate the growing understanding of the flexibility of base pairing in strand–strand recognition process.

Above all, the second generation model for nucleic acid structure must incorporate our growing understanding of the ability of nucleic acids to have selectable function (phenotype) as well as performing genetic tasks. In particular, we must reconsider more seriously whether a single polymer can robustly carry information and catalyze reactions, in light of the contradicting chemical demands for performing each role. Chemical experiments attempting to resolve these contradictions in single-biopolymer systems will, we believe, provide the most important results in the coming decade to understand life's origins on Earth and how best to find life elsewhere.

ACKNOWLEDGMENTS

Work described in this paper was supported in part by the Office of Naval Research (N00025-96-1-0362), the National Aeronautics and Space Administration, and the National Institutes of Health (GM 54048).

REFERENCES

Alberti, S. 1997. The origin of the genetic code and protein synthesis. *J. Mol. Evol.* **45**:352–382.

Bain, J.D., A.R. Chamberlin, C.Y. Switzer, and S.A. Benner. 1992. Ribosome-mediated incorporation of non-standard amino acids into a peptide through expansion of the genetic code. *Nature* **356**:537–539.

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

Beaudry, A.A., and G.F. Joyce. 1992. Directed evolution of an RNA enzyme. *Science* **257**:635–641.

Benner, S.A. 1987. Redesigning life. Organic chemistry and the evolution of proteins. *Chimia* **41**:142–148.

Benner, S.A. 1994. Expanding the genetic lexicon. Incorporating non-standard amino acids into proteins by ribosome-based synthesis. *Trends Biotech.* **12**:158–163.

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

Benner, S.A., M.A. Cohen, G.H. Gonnet, D.B. Berkowitz, and K. Johnsson. 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, N.Y.

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

Benner, S.A., A. Glasfeld, and J.A. Piccirilli. 1989. Stereospecificity in enzymology. Its place in evolution. *Top. Stereochem.* **19**:127–207.

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

Cech, T.R., A.J. Zaug, and P.J. Grabowski. 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.

Dempcy, R.O., O. Almarsson, and T.C. Bruice. 1994. Design and synthesis of deoxynucleic guanidine. A polycationic analog of DNA. *Proc. Natl. Acad. Sci. USA* **91**:7864–7868.

DiGiulio M. 1997. On the origin of the genetic code. *J. Theor. Biol.* **187**:573–581.

Dueholm, K.L., K.H. Peterson, D.K. Jensen, P.E. Nielsen, M. Egholm, and O. Buchardt. 1994. Peptide nucleic acid (PNA) with a chiral backbone based on alanine. *Biomed. Chem. Lett.* **4**:1077–1081.

Egholm, M., O. Buchardt, P.E. Nielsen, and R.H. Berg. 1992. Peptide nucleic acids. Oligonucleotide analogs with an achiral peptide backbone. *J. Am. Chem. Soc.* **114**:1895–1897.

Ellington, A.D., and S.A. Benner. 1987. Free energy differences between enzyme bound states. *J. Theor. Biol.* **127**:491– 506.

Flory, P. 1953. Principles of Polymer Chemistry. Ithaca, N.Y.: Cornell Univ. Press.

Freier, S.M., and K.H. Altmann. 1997. The ups and downs of nucleic acid duplex stability, structure-stability studies on chemically modified DNA, RNA duplexes. *Nucl. Acids Res.* **25**:4429–4443.

Gangamani, B.P., V.A. Kumar, and K.N. Ganesh. 1997. Spermine conjugated peptide nucleic acids (spPNA), UV and fluorescence studies of PNA–DNA hybrids with improved stability. *Nucl. Acids Res.* **240**:778–782.

Herschlag, D., and T.R. Cech. 1990. Catalysis of RNA cleavage by the Tetrahymena thermophila enzyme. *Biochemistry* **29**:10159–10171.

Horlacher, J., M. Hottiger, V.N. Podust, U. Hübscher, and S.A. Benner. 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 S.A. Benner. 1993. Selective protection and deprotection procedures for thiol and hydroxyl groups. *Syn. Lett.*:83–84.

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

Johnsson, K., R.K. Allemann, and S.A. Benner. 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. Cambridge: Royal Society of Chemistry,.

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

Joyce, G.F. 1994. Forward. In: Origins of Life, The Central Concepts, ed. D.W. Deamer and G.R. Fleischaker. Boston: Jones and Bartlett.

Kalamaraki, M., G.M. McGeehan, and S.A. Benner. 1991. Site-directed mutagenesis of bovine pancreatic ribonuclease A. Serine 16, threonine 17, alanine 19 and alanine 20. In: Structure, Mechanism, and Function of Ribonucleases, Proc. 2nd Intl. Conf. on Ribonuclease, ed. C.M. Cuchillo, R. de Llorens, M.V. Nogués, and X. Parés, pp. 145–150. Barcelona: IBF Publications.

Kamaike, K., M. Takahashi, K. Utsugi, K. Tomizuka, and Y. Ishido. 1995. An efficient method for the synthesis of [4–^{15}N]cytidine and [6–^{15}N]adenosine derivatives from uridine and inosine. *Tetrahedron Lett.* **36**:91–94.

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

Khare, B.N., W.R. Thompson, L. Cheng, C. Chyba, C. Sagan, E.T. Arakawa, C. Meisse, and P.S. Tuminello. 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 S.A. Benner. 1997. Synthesis of an N-alkyl derivative of 2'-deoxyisoguanosine. *Syn. Lett.* :939–940.

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

Lehman, N., and G.F. Joyce. 1993. Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature* **361**:182–185.

Li, A.G., and J.M. Greenberg. 1997. A unified model of interstellar dust. *Astron. Astr.* **323**:566–584.

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

Luisi, P.L. 1998. About various definitions of life. In: Origins Life, in press.

Lutz, M.J., S.A. Benner, S. Hein, G. Breipohl, and E. Uhlmann. 1997. Recognition of uncharged polyamide-linked nucleic acid analogs by DNA polymerases and reverse transcriptases. *J. Am. Chem. Soc.* **119**:3177–3178.

Lutz, M.J., H.A. Held, M. Hottiger, U. Hübscher, and S.A. Benner. 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.

Macdermott, A.J., L.D. Barron, A. Brack, T. Buhse, A.F. Drake, R. Emery, G. Gottarelli, J.M. Greenberg, R. Haberle, et al. 1996. Homochirality as the signature of life. The Seth cigar. *Plan. Space Sci.* **44**:1441–1446.

McBride, J.M., and R.L. Carter. 1991. Spontaneous resolution by stirred crystallization. *Angew. Chem. Intl. Edn.* **30**:293–295.

Menger, F.M., A.V. Eliseev, N.A. Khanjin, and M.J. Sherrod. 1995. Evidence for an alternative mechanism to a previously proposed self-replicating system. *J. Org. Chem.* **60**:2870–2878.

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

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

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

Orgel, L.E. 1992. Molecular replication. *Nature* **358**:203–209.

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

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

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. New York: Academic.

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

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

Schmidt, J.G., L. Christensen, P.E. Nielsen, and L.E. Orgel. 1997. Information transfer from DNA to peptide nucleic acids by template-directed syntheses. *Nucl. Acids Res.* **25**:4792–4796.

Switzer, C.Y., S.E. Moroney, and S.A. Benner. 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. USA* **89**:2614–2618.

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

Thompson, L.A., and J.A. Ellman. 1996. Synthesis and applications of small molecule libraries. *Chem. Rev.* **96**:555–600.

Tsang, J., and G.F. Joyce. 1994. Evolutionary optimization of the catalytic properties of a DNA-cleaving enzyme. *Biochemistry* **33**:5966–5973.

Usher, D.A., and A.H. McHale. 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 R.M. Kellogg. 1978. Biotin. Its place in evolution. *J. Mol. Evol.* **11**:171–178.

Vögel, J.J., M.M. Altorfer, and S.A. Benner. 1993. 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., U. von Krosigk, and S.A. Benner. 1993. 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.

Vögel, J.J., and S.A. Benner. 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.

von Kiedrowski, G., B. Wlotzka, J. Helbing, M. Matzen, and S. Jordan. 1991. Parabolic growth of a self-replicating hexadeoxynucleotide bearing a 3',5'-phosphoramidate linkage. *Angew. Chem.* **30**:423–426.

von Krosigk, U., and S.A. Benner. 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 D.J. Patel. 1994. Solution structure of the Tetrahymena telomeric repeat DT2G44 G-tetraplex. *Structure* **2**:1141–1156.

Watson, J.D., N.H. Hopkins, J.W. Roberts, J.A. Steitz, and A.M. Weiner. 1987. In: Molecular Biology of the Gene, p. 1115. Menlo Park: Benjamin Cummings.

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

Wintner, E.A., M.M. Conn, and J. Rebek, Jr. 1994. Autocatalysis and the generation of self-replicating systems. *Accts. Chem. Res.* **27**:198–203.

Wintner, E.A., B. Tsao, and J. Rebek, Jr. 1995. Evidence against an alternative mechanism for a self-replicating system. *J. Org. Chem.* **60**:7997–8001.

Woese, C.R. 1967. The Genetic Code. The Molecular Basis for Genetic Expression. New York: Harper and Row.

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

Zent, A.P., and C.P. Mckay. 1994. The chemical-reactivity of the Martian soil and implications for future missions. *Icarus* **108**:146–157.