Acc. Chem. Res. **2004,** *37,* 784−797

Understanding Nucleic Acids Using Synthetic Chemistry

STEVEN A. BENNER*
Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200

Received January 13, 2004

ABSTRACT

This Account describes work done in these laboratories that has used synthetic, physical organic, and biological chemistry to understand the roles played by the nucleobases, sugars, and phosphates of DNA in the molecular recognition processes central to genetics. The number of nucleobases has been increased from 4 to 12, generating an artificially expanded genetic information system. This system is used today in the clinic to monitor the levels of HIV and hepatitis C viruses in patients, helping to manage patient care. Work with uncharged phosphate replacements suggests that a repeating charge is a universal feature of genetic molecules operating in water and will be found in extraterrestrial life (if it is ever encountered). The use of ribose may reflect prebiotic processes in the presence of borate-containing minerals, which stabilize ribose formed from simple organic precursors. A new field, synthetic biology, is emerging on the basis of these experiments, where chemistry mimics biological processes as complicated as Darwinian evolution.

It has been 50 years since Watson and Crick proposed the double-helical structure for duplex DNA.[1] The proposal was supported at first by little experimental data. Nevertheless, the double helix explained elegantly how DNA could be replicated and how it could support genetics and evolution through mutation. It was therefore obviously correct and soon accepted by most.

From a chemical perspective, this is remarkable, because the duplex structure appears to violate many of the rules that chemists might themselves use to design molecular recognition systems. For example, the double helix appears to disregard Coulomb's law. Cations bind anions, and anion bind cations. In DNA, however, a polyanion binds another polyanion. One might think (and indeed, many have thought)[2] that the duplex would be more stable if one strand were uncharged, or polycationic.[3] Likewise, chemists often exploit rigidity when designing receptors to fit ligands. DNA strands are floppy. One would think (and many have thought)[4] that conformationally rigid DNA analogues should be better at molecular recognition.

Another curious feature of the DNA duplex is its use of hydrogen bonding to achieve specificity. This would not be the chemist's preference in water, where hydrogen bonding to solvent competes with interstrand hydrogen bonding. Indeed, several groups have sought to dispense with hydrogen bonding between nucleobases entirely.[5,6]

Despite these evident violations of "rules" of molecular recognition, recognition in DNA and RNA is remarkably predictable. A high school student can design two DNA molecules that specifically recognize each other, simply by remembering that A pairs with T and G pairs with C. In no other chemical system is such simple rule-based molecular recognition known.

Elegance became familiarity in the 1970s, as these two rules came to support a burgeoning biotechnology industry. As a result, when analytical chemical and synthetic chemical tools allowed the sequencing and synthesis of DNA by 1980, only a few synthetic organic chemists were using their skills to explore these remarkable features of DNA.[7] Our opportunity to do so came in 1984, when the ETH invited my group to move from Harvard to Zurich. The grant support that we had at Harvard from the NIH and NSF was generous, but directed toward extensions of my postdoctoral research. There seemed to be little chance to fund a program intended to redesign nucleic acids. We are indebted to the ETH for the opportunity to undertake what was then regarded as risky research.

Today, we know that the nucleobases at the center of the double helix can be changed quite broadly. An artificially expanded genetic information system (AEGIS), where 12 nucleotides form six specific nucleobase pairs, has been implemented.[8] AEGIS now is used in the clinic to monitor the load of virus in over 400 000 patients, annually, infected by human immunodeficiency virus (HIV) and hepatitis C virus.[9] Other clinical and research applications are in development.

We also know that the repeating charge carried by the backbone phosphates is key to the ability of DNA, as a molecular system, to support Darwinian evolution. Indeed, a repeating charge may be a universal feature of genetic molecules that work in water, including those in (hypothetical) life forms from other planets.

Last, the backbone ribose is more than a simple scaffolding to hold the nucleobases.[10] Many sugars other than ribose have been explored.[11] These include analogues that may have served as primitive genetic molecules that were precursors of RNA and DNA for life on Earth, as ribose has been viewed as being too unstable to have emerged prebiotically.[12,13] Very few ribose replacements have been found to improve on duplex stability.[11] Further, we now know that borate minerals stabilize ribose under conditions that might have been present on early Earth.[14] This connects geology via chemistry to biology in an intriguing way and suggests that ribose is central to genetics because of chemical processes that occurred before the origin of life.

Together, these results are supporting the emerging field of synthetic biology.[15] This field moves past biomimetic chemistry, seeking artificial chemical systems that reproduce advanced biological behaviors, including replication, selection, and evolution. In this direction, a six letter artificial genetic system based on AEGIS has recently been shown to support evolution in a very primitive polymerase chain reaction.[16]

Steven Benner was schooled at Yale and Harvard Universities and served on the faculty at Harvard and the ETH Zurich before becoming a Distinguished Professor at the University of Florida.

* E-mail: benner@chem.ufl.edu. Tel: 352 392 7773. Fax: 352 392 7918.

10.1021/ar040004z CCC: $27.50 © 2004 American Chemical Society
Published on Web 08/31/2004

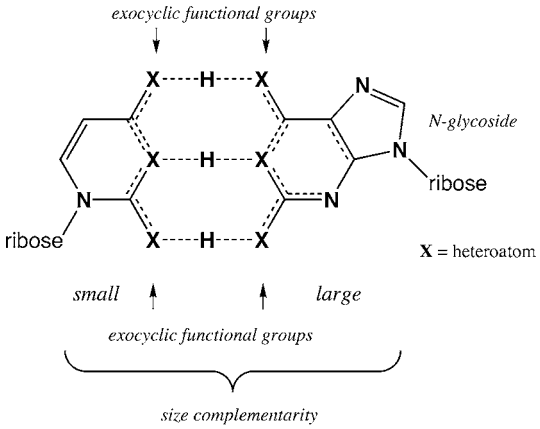

FIGURE 1. The generalized Watson−Crick nucleobase pair places three protons between a large heterocycle and a small heterocycle. Two internucleobase hydrogen bonds are formed between exocyclic functional groups; one is formed between heteroatoms in the heterocycle. Dotted lines indicate placement of double bonds to complete the valences and to make the heterocycles aromatic.

With the 50th anniversary of the Watson−Crick model for duplex DNA past, it seems timely to review some of synthetic work done in these laboratories to create artificial genetic systems. These may foreshadow the next 50 years of DNA evolution, not in Nature but in the laboratory and clinic.

The Nucleobases Can Be Changed

Watson−Crick nucleobase pairing follows two rules of complementarity. Size complementarity requires that large purines pair with small pyrimidines. Hydrogen-bonding complementarity requires that hydrogen bond donors from one nucleobase pair with hydrogen bond acceptors on the other. These rules can be generalized (Figure 1), where the large and the small components are joined by three hydrogen bonds. In this general structure, $2^3 = 8$ hydrogen-bonding patterns are conceivable. Six are readily accessible (Figure 2).

The pyDAD, pyDDA, and pyADD hydrogen-bonding patterns (the terminology is defined in the Figure 2 legend) presented the most interesting design and synthetic challenges. These patterns require a nucleobase joined to the sugar via a carbon−carbon bond. These C-glycosidic linkages are different from the N-glycosidic linkages found in natural nucleotides.

Different heterocycles can support individual hydrogen-bonding patterns, however (Figure 3). As the program developed, it became clear that some implementations of the same pattern would be more suitable in artificial genetic systems than others.

For example, when Lawrence MacPherson, Joseph Piccirilli, and Tilman Krauch placed the pyDAD pattern on a pyridine heterocycle, the result was a compound that was too basic and too oxidation-sensitive to be used in either chemical DNA synthesis or genetics.[17] To exploit the electronegativity of nitrogen, the pyDAD pattern was implemented on a pyrazine skeleton to obtain a less basic and less oxidizable structure. But this structure epimerized (Figure 3).[18] In Goldilocks fashion, the pyrimidine heterocycle proved to be "just right".[19]

Physical organic chemistry and organic synthesis were then used to obtain suitable implementations of the other nonstandard nucleobases. As shown by Ulrike von Krosigk, Johannes Voegel, and Daniel Hutter (Figure 4),[20,21] implementing the pyDDA and pyADD hydrogen-bonding patterns on a pyridine heterocycle gave easily oxidized compounds. The same patterns on a pyrazine suffered epimerization.[22,23] The patterns implemented on a pyrimidine heterocycle gave rise to tautomeric ambiguity. Finally, a suitable derivative was found where the pyDDA hydrogen-bonding pattern was implemented on a 5-nitropyridine heterocycle.[24] This compound was stable against both epimerization and oxidation.

An extra DNA letter implementing the pyAAD hydrogen-bonding pattern on the N-glycoside isocytidine was developed by Christopher Switzer (Figure 5).[25] Isocytidine displayed a propensity to deaminate in base and depyrimidinylate in acid. Thomas Battersby, Simona Jurczyk, and Janos Kodra,[26] as well as scientists at Chiron,[9] managed these problems through strategic use of 5-position substituents and optimized protecting groups.[27,28] Today, oligonucleotides containing multiple isocytosines are synthesized routinely for use in the clinic.

Nonstandard purines (puAAD, puDAA, puADA, and puDDA) were synthesized to complement the pyDAD, pyDDA, pyADD, and pyAAD nucleobases (Figure 6).[29,30] Isoguanine, implementing the puDDA pattern, was known from the work of Shugar and co-workers.[31] Xanthosine, implementing the puADA pattern complementary to pyDAD, was known from Nature. Implementations of the puAAD and puDAA hydrogen-bonding patterns were prepared by standard chemistry.

Here too, undesirable properties were encountered and managed by a combination of synthetic and physical organic chemistry (Figure 6). Xanthosine is acidic ($pK_a \approx 5.7$); the heterocycle is anionic at physiological pH.[32] Systematic analysis of AEGIS systems showed that anionic nucleobases destabilize the duplex.[8] Therefore, alternative implementations of the pyADA pattern were prepared by Joseph Piccirilli, using the methyloxoformycin B skeleton,[33] and Michael Lutz, Photon Rao, and Heike Held, who explored the 7-deazaxanthosine and the 5-aza-7-deazaxanthosine systems (Figure 6).[34,35] These created useable AEGIS components implementing the puADA bonding pattern.

Isoguanine proved to have a problem with tautomerism, however. In water, isoguanosine exists to about 10% in an enol form that presents a puDAD hydrogen-bonding pattern that is complementary to thymine (Figure 6).[29] Prior to 1989, some had thought that this would make isoguanosine useless in genetic systems.[36] Indeed, in early work, we found that polymerases preferentially incorporate the standard nucleobase U/T opposite the minor enolic tautomer, rather than isoC opposite the major keto tautomer.[37]

Fixing this problem began with a speculation that the enolic tautomer of isoG is favored because the ketonic

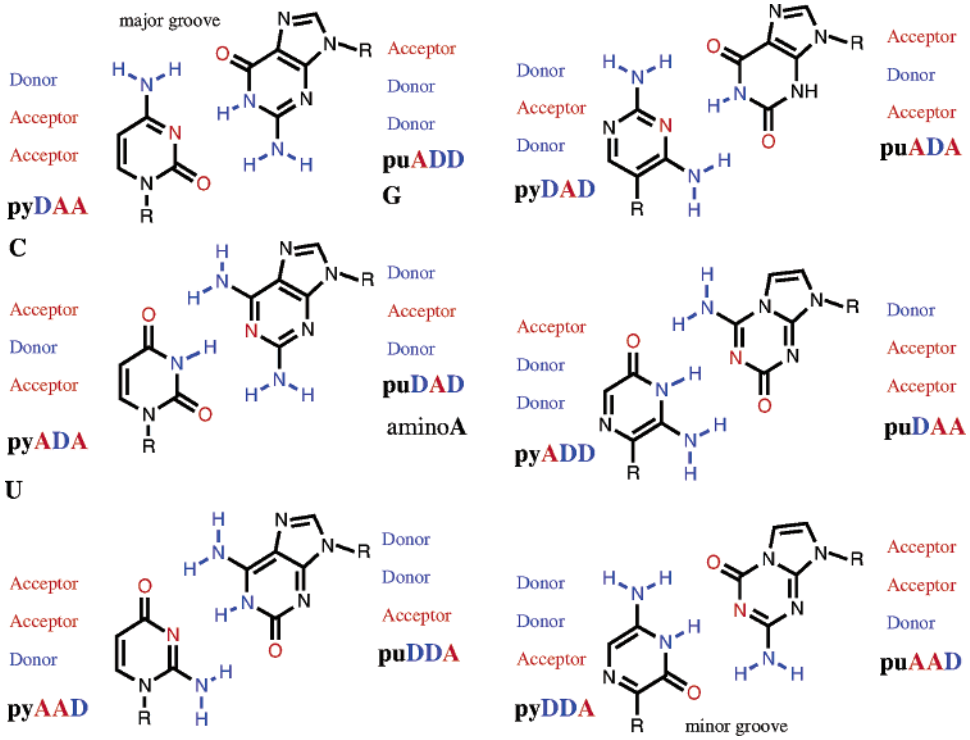

FIGURE 2. Six hydrogen-bonding patterns joining 12 different nucleobases with Watson−Crick specificity. Shown are the first implementations of the different hydrogen-bonding patterns. Nomenclature is as follows: pu = a 5−6 fused ring system; py = a six-membered ring. The hydrogen-bonding pattern of acceptor (A) and donor (D) groups from the major to the minor groove is indicated. Thus, the standard nucleobase cytosine is pyDAA. R indicates the point of attachment of the backbone.

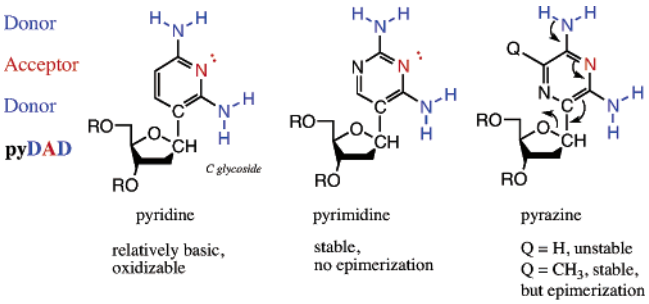

FIGURE 3. Different C-glycosidic heterocycles can implement the pyDAD hydrogen-bonding pattern. The pyridine implementation is basic and oxidizable. The pyrazine implementation suffers epimerization. The pyrimidine implementation is entirely satisfactory.

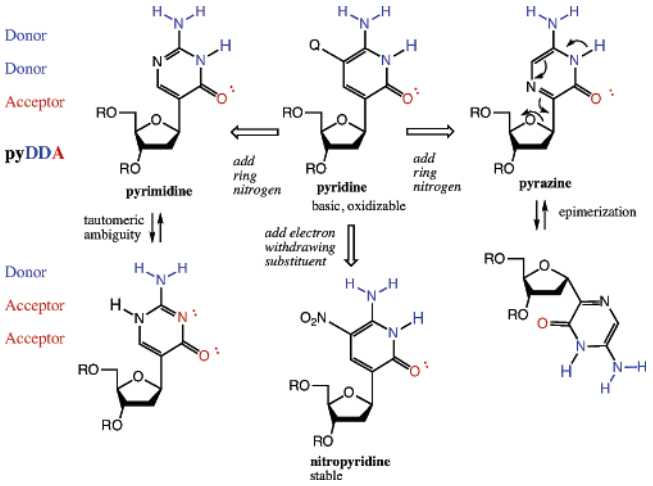

FIGURE 4. Different C-glycosides heterocycles implementing the pyDDA hydrogen-bonding scheme. Electron-withdrawing groups manage epimerization and oxidizability of these.

tautomer of isoG is "cross-conjugated"; this means that a double bond cannot be written between the two carbon atoms shared by the five and six membered rings. This is possible with the keto tautomer of G. Enolization of isoG restores formal aromaticity to the five-membered ring, because the enol can have a double bond between those two atoms (Figure 6). Reasoning that pyrrole is less aromatic than imidazole and that the drive to aromatization of pyrrole would be smaller than that to an imidazole, Theodore Martinot synthesized a 7-deaza-iso-G nucleoside implementing the puDDA pattern (Figure 6). The compound favored the keto form ca. 1000:1.[38]

Rules for Designing Artificial Genetic Alphabets. These syntheses delivered 12 different hydrogen-bonding patterns, 8 new within the context of the DNA duplex, in a form that made them likely to participate in molecular recognition processes, both enzymatic and nonenzymatic, through the formation of six mutually exclusive pairing schemes. To understand better the rules that might guide the expansion of genetic alphabets, Ronald Geyer and Thomas Battersby examined 150 nucleobase pairing combinations at two pHs (Figure 7).[8] Many possible pairing geometries were considered, including the classical Watson−Crick geometry, wobble geometries, and purine syn conformations (Figure 8). A summary of the data is shown in Figure 9. Rules for designing pairs within the Watson−Crick paradigm are collected in Figure 10. These studies suggest that hydrogen bond and size complementarity are approximately equally important to duplex stability. This

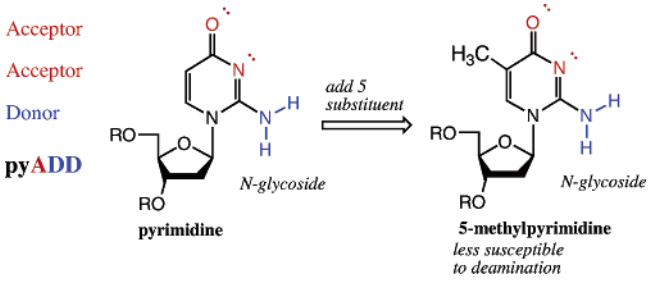

FIGURE 5. The pyAAD hydrogen-bonding pattern can be implemented on an N-glycoside with a pyrimidine heterocycle. Substituents manage deamination.

result was somewhat surprising, because it appeared to contradict the view that interbase hydrogen-bonding is dispensable.[39]

Rules emerged from this work about the subtle features of the structures as well. For example, an uncompensated amino group greatly diminishes the ability of a nucleobase pair to stabilize a duplex; an uncompensated carbonyl group does not.[8] The C-glycoside attachment is not particularly detrimental to duplex stability. Context dependence in the expanded alphabet remained subtle, influencing melting temperatures by perhaps 1−2 °C.[40]

Why Not More Letters in Natural Terran DNA? This work showed that the standard nucleobases in contemporary DNA/RNA are not unique chemical solutions to the problem of storing genetic information in DNA-like molecules. It is conceivable that a future NASA mission will encounter alien life that has an AEGIS-like genetic system. Popular culture has come to recognize this possibility. For example, an X-File (*The X Files*, 20th Century Fox, The Erlenmeyer Flask Episode 1X23) exists describing alien life having a genetic system built from six nucleotides. Art has followed science.

We might therefore ask why terran DNA does *not* have more letters. Szathmary has suggested, from a discussion of replication fidelity, that four is functionally optimal.[41] According to Darwinian theory, such functional explanations imply that, in the past, organisms on Earth experimented with larger alphabets but were less fit when they did so. Alternatively, a historical explanation might account for the absence of AEGIS components in modern genetics by hypothesizing that our four standard nucleobases were chosen by accident and never changed.

Vestigiality might also be invoked to explain the structure of modern DNA. Vestigiality includes a statement connecting structure to function, but implies that the function no longer exists, while the structure has survived. For example, standard nucleotides might have been more accessible to prebiotic chemistry than nonstandard nucleotides. This view is made somewhat persuasive by the degree of engineering needed to make nonstandard AEGIS components practical. It would be more persuasive, of course, if compelling prebiotic routes were known for *standard* nucleotides.[42]

Last, both C- and N-glycosides may not be easily compatible within a single genetic system. Polymerases, to be faithful, might need to specialize in one or the other. Likewise, a single genetic system might find difficult different nucleobases that present different functionality to the minor groove. These issues are discussed below.

Applications of Artificially Expanded Genetic Information Systems. AEGIS has had many practical applications, as expected for a chemical system that gives rule-based molecular recognition. For example, in the early 1990s, scientists at Chiron sought to detect viral DNA/RNA in biological samples using a sandwich assay exploiting a dendrimeric branched DNA (bDNA, Figure 11). Here, DNA/RNA from a virus forms a duplex with a capture strand, which pairs with a DNA molecule on a support. The analyte then captures a second strand that, in turn, captures a branched DNA molecule. The branched dendrimeric nanostructure carries multiple copies of yet another DNA sequence complementary to a DNA sequence carrying a signaling system. The branched DNA therefore directs the assembly of a signaling nanostructure having multiple signaling units to a support, if the analyte is present.

In its first implementation, the bDNA assay assembled the nanostructure using A, T, G, and C. So implemented, the assay was noisy; signal emerged from the support even when the analyte was absent. This was presumably due to the presence of nontarget nucleic acid molecules within the sample, also containing A, U/T, G, and C. With some frequency, these cross-hybridize with the amplifier and preamplifier strands, immobilizing the signaling dendrimer in the absence of analyte. Similar problems arise in DNA microarrays.

Incorporating AEGIS components into the DNA used to assemble the nanostructure helped solve the problem. The two sets of nucleobases, one natural for binding the analyte and the other artificial to build the nanostructure, allowed the two processes to run independently. This suppressed noise and raised sensitivity. Under ideal circumstances, the bDNA assay incorporating AEGIS detects as few as 30 molecules.[9]

Branched DNA assays exploiting AEGIS are now used to monitor the level of viral RNA in approximately 400 000 patients annually living with HIV and hepatitis C.[43] For HIV-infected patients undergoing antiretroviral therapy, for example, the level of HIV prognoses the future course of the disease. Thus, AEGIS chemistry has helped implement personalized medicine for these patients.

Artificial Genetic Alphabets Work with Terran Molecular Biology. Ligases, kinases, and other workhorses of molecular biology accept AEGIS components. So does the ribosome. In a collaboration with James Bain and Richard Chamberlin at the University of California at Irvine, we challenged the ribosome from *Escherichia coli* to translate a mRNA containing a 65th codon built using AEGIS components and synthesize a protein with a 21st amino acid. When given a charged tRNA carrying the anticodon incorporating the appropriate AEGIS complement, a nonstandard peptide was synthesized.[44] This study also showed how release factors work in natural systems to terminate translation.

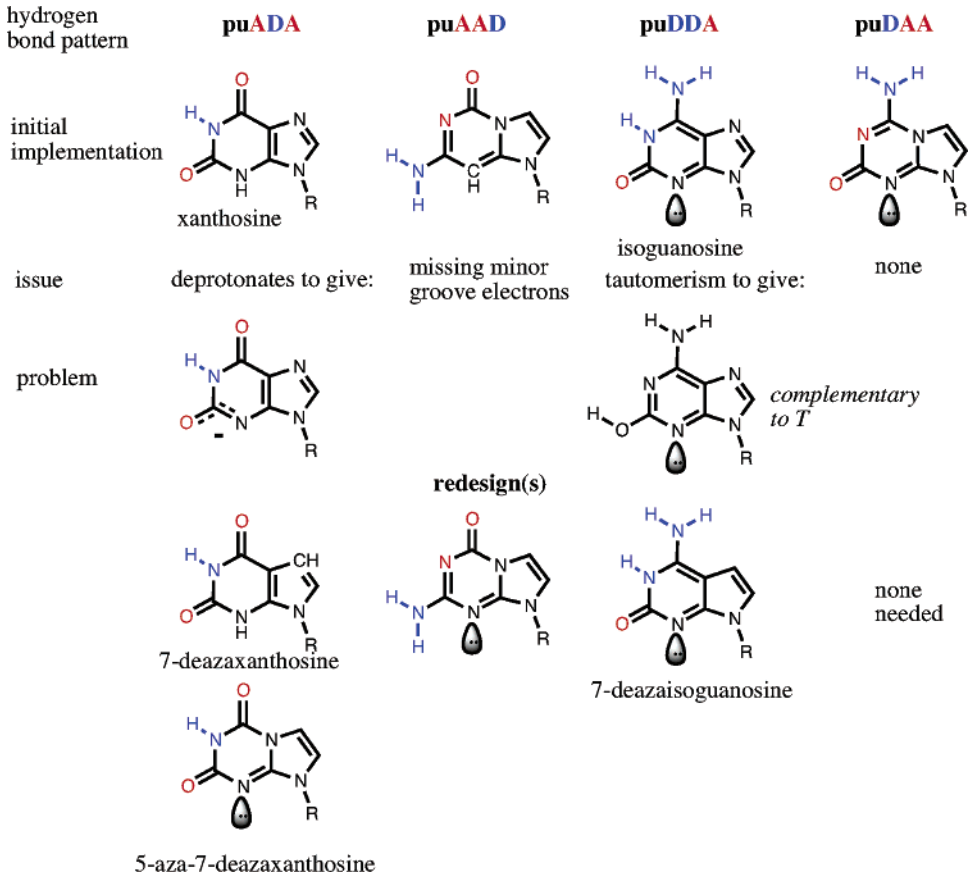

FIGURE 6. Engineering of purines to create analogues that complement AEGIS pyrimidines.

Polymerases proved to have some difficulty using AEGIS components, as Simon Moroney, Jennifer Horlacher, Christopher Switzer, and others in the group showed.[45] Our first experiments readily identified polymerases that accepted AEGIS components. It became apparent, however, that a polymerase that incorporates a single AEGIS component in a simple runoff polymerase experiment need not repeatedly incorporate the component with high fidelity. This level of performance would be necessary to implement a synthetic biology based upon AEGIS.

Contacts between polymerases and their nucleobase substrates have been explored by crystallography.[46] In particular, an unshared pair of electrons carried by N-3 of the purines and the exocyclic O-2 of the pyrimidines (Figure 12) appears to accept a hydrogen bond from many polymerases. Because standard nucleobases all have this unshared pair of electrons, several groups suggested that natural polymerases use these as a "site of common interaction", a place for a polymerase to contact all four standard nucleobases without discriminating between them. Several AEGIS components could not easily have this unshared pair of electrons (Figure 12).

These considerations suggested four strategies to obtain a polymerase−AEGIS combination to support synthetic biology. First, we altered the structure of the AEGIS components so that they presented, whenever possible, the unshared pair of electrons (Figure 12). Next, we screened natural polymerases to identify those most likely to accept AEGIS components.[47] Third, we used site-directed mutagenesis to engineer polymerases that could better accept AEGIS components.[48] Last, we initiated a program to evolve polymerases in vitro to accept AEGIS components.

Here, evolution helped. Polymerases come in families.[49] Within a family, amino acid sequence similarity makes homology indisputable. Between polymerase families, homology is inferred on the basis of the similarity in the polymerase fold. Although this inference is not strong, all polymerases may belong to a superfamily that shared a common ancestor several billion years ago.

Typically, homologous proteins have similar behaviors, reflecting behaviors that were displayed by their common ancestor.[50] If these behaviors have been important for function, they should have been conserved in the time since the descendent polymerases diverged. Based on this generalization, we expected polymerases to behave analogously, especially with respect to contacts that might be important for fidelity, including contacts in the minor groove.

Remarkably, Cynthia Hendrickson found that different polymerases react quite differently to the absence of minor groove electrons.[51] This was true between evolutionary families of polymerases. But even within a family, the minor groove electron pair was recognized differently. It is therefore difficult to make statements about what polymerases generally accept.

From a technological point of view, this is good. If one polymerase does not incorporate an AEGIS component well, another might. Further, laboratory mutation might

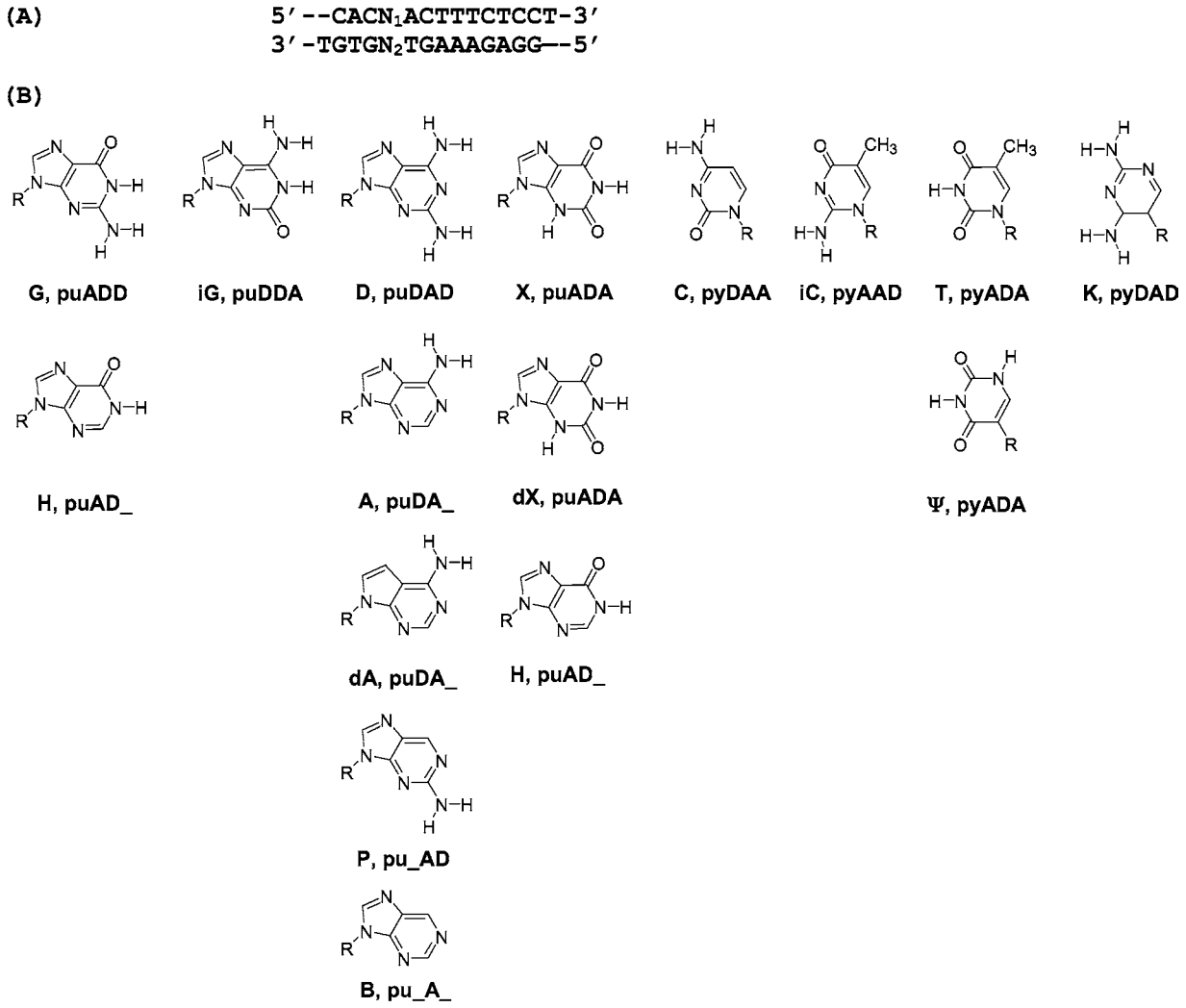

FIGURE 7. Nucleobase analogues used to develop rules for designing artificial genetic alphabets.

convert a polymerase that incorporates an AEGIS component poorly into one that incorporates the component well.

But the result is unexpected scientifically. The differences suggest either that the evolutionary families of polymerases do not share a common ancestor, that fidelity through minor groove contact was not a feature of that ancestor, or that this feature was not sufficiently critical to fitness to have been conserved. Each possible interpretation is significant to those interested in the planetary biology of polymerases.[52]

Artificial Genetic Information Systems and the Polymerase Chain Reaction. With this encouragement, we set out to obtain polymerases that could amplify six-letter genetic alphabets in a polymerase chain reaction (PCR). To obtain a suitable polymerase, we turned to the evolution of reverse transcriptase from HIV in the clinic. Here, the enzyme is targeted by inhibitors and has evolved in response. Reverse transcriptases emerging from *in clinico* evolution must retain high levels of fidelity and efficiency, because the virus itself needs these properties. They may, however, differ in their ability to accept AEGIS components. Thus, mutant reverse transcriptases from the clinic were the starting points for polymerase screening.

One variant from the clinic (Y188L) accepted the triphosphate of puADA opposite pyDAD rather well. In a step of rational protein engineering, Michael Sismour, Jeong Ho Park, Shuji Ikeda, and Stefan Lutz replaced another amino acid to kill a possible exonuclease activity. This generated the doubly changed reverse transcriptase (Y188L−E478Q).

This variant was able to amplify a DNA duplex containing a py(DAD)−pu(ADA) pair in a primitive PCR reaction. These AEGIS components, one lacking a minor groove unshared electron pair, were retained through multiple rounds of amplification. Though long incubation times were needed to ensure that full-length products were formed, this work gave the first example of a six-letter alphabet supporting a genetic process.[16]

Applications of Polymerase Incorporation. Polymerase-based incorporation of AEGIS components also has practical applications. One example comes from colleagues at EraGen Biosciences (www.eragen.com). Here, in a process that does not involve PCR amplification of the AEGIS pair itself, a primer having a fluorescently tagged isoC is targeted against a specific region of the SARS DNA sequence. A polymerase incorporates an isoG carrying a fluorescence quencher opposite isoC. In a real time PCR

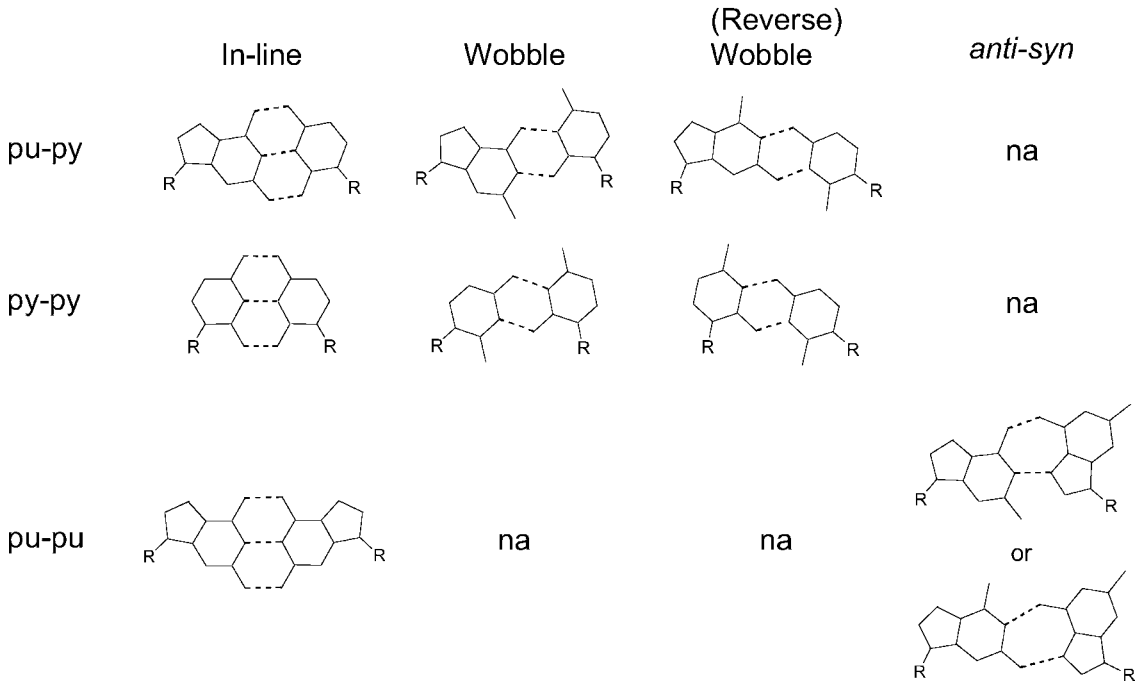

FIGURE 8. Alternatives to the in-line Watson–Crick geometry.

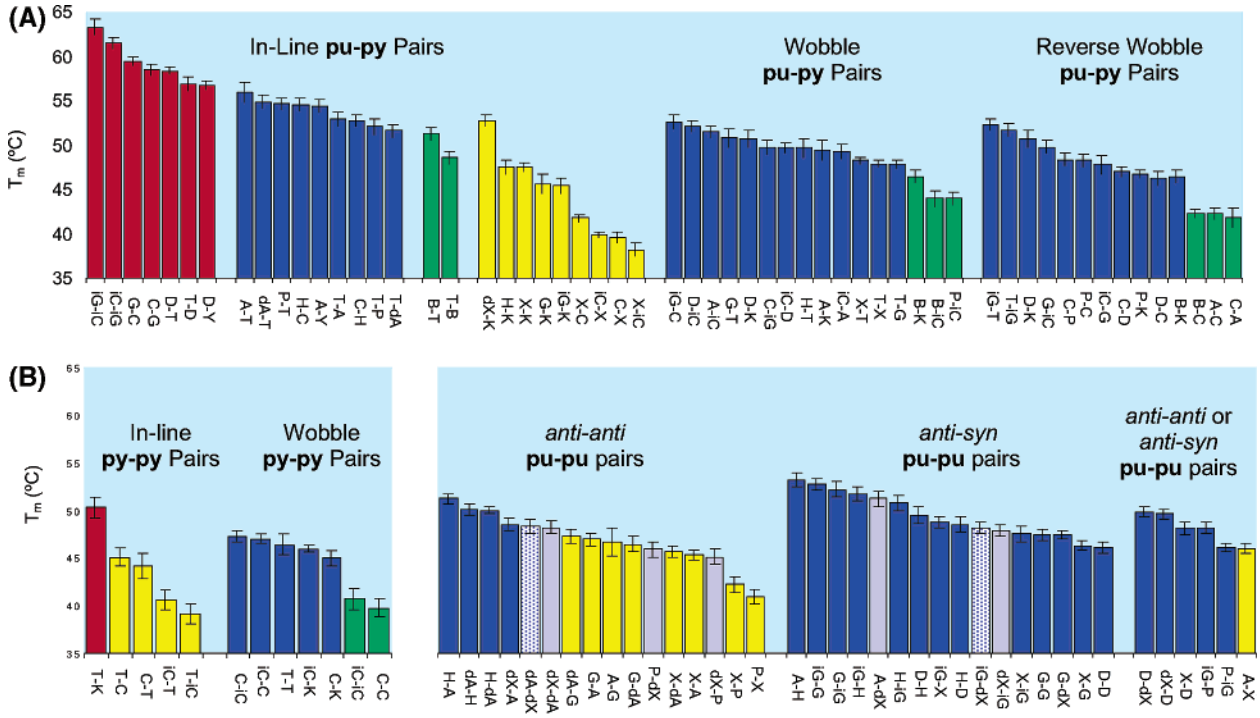

FIGURE 9. Hydrogen-bonding complementarity and size complementarity contribute similarly to duplex stability. Bars show melting temperatures of pairs in the reference duplex (Figure 8). In panel A, in-line purine–pyrimidine (pu–py) nucleobase pairs are grouped by the number of H-bonds: three H-bonds (red), two H-bonds (blue), and one H-bond (green). Pairs with destabilizing interactions are grouped together (yellow). Wobble and reverse wobble pu–py pairs are colored by number of H-bonds: two H-bonds (blue); one H-bond (green). In panel B, pyrimidine–pyrimidine pairs are grouped by conformation (in-line or wobble) and colored by number of H-bonds: three H-bonds (red), two H-bonds (blue); one H-bond (green). Pyrimidine–pyrimidine pairs with destabilizing interactions are grouped together (yellow). In panel C, purine–purine pairs are grouped according to conformation (anti–anti or anti–syn) and colored by number of H-bonds: two H-bonds (blue). Purine–purine pairs with destabilizing interactions are yellow; those with partially destabilizing interactions, such as nucleobases with anionic charge (dX), are shaded.

experiment, the presence of the SARS virus caused fluorescence from the primer to disappear.

The Future: A Synthetic Biology. The ultimate goal of artificial genetics is a synthetic biology, where high-level behaviors of living systems are mimicked by artificial chemical systems.[15] These include replication, selection, and evolution. In addition to their technological value, artificial Darwinian systems will help us understand how

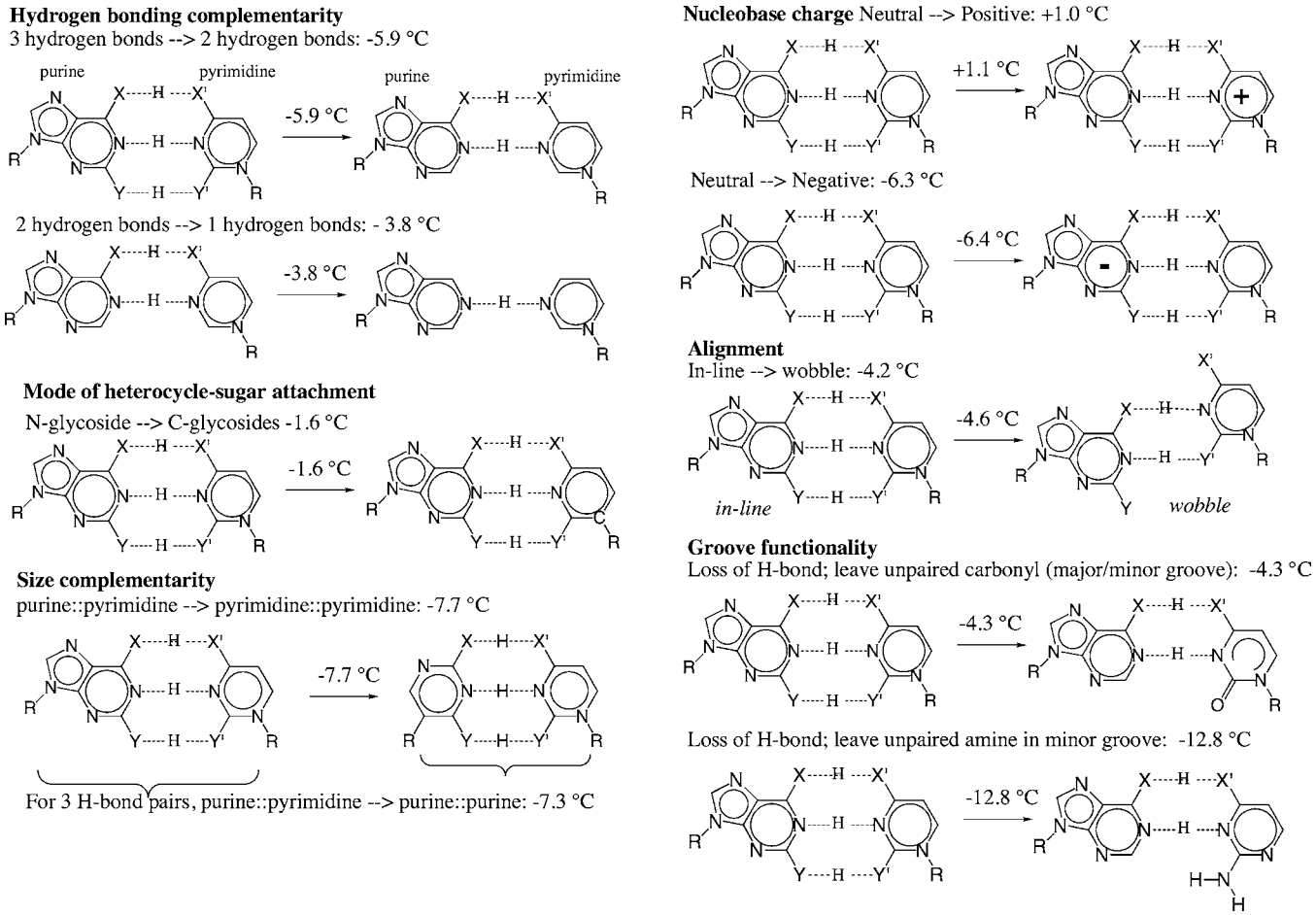

FIGURE 10. Abbreviated rules relating structural features of nonstandard nucleobases to their impact on the stability of the DNA duplex.

life emerged, model how natural biology works, and explain how systems-scale combinations of chemical reactions produce the behavior that we call "living". Building artificial genetic systems from the bottom up offers a way to learn more about life in general, one complementary to the deconstructive approach that dissects genetic systems from the top down.

The Charged Backbone May Be Universal in Water

We next turn to the phosphates, the linking groups that make DNA a polyanion. The repeated negative charges appear to be a nuisance in many ways. They appear to hinder the entrance of DNA into cells. The polyanion also appears to destabilize the DNA duplex through Coulombic repulsions. Many have asked: Can the charge be removed?

Christian Schneider and Zhen Huang began our effort to answer this question by replacing the anionic phosphate diester linker with the uncharged dimethylenesulfone linker (Figure 13).[53−55] Initial data suggested that sulfone-linked DNA analogues (SNAs) displayed molecular recognition of the Watson−Crick type. For example, the GSO_2C analogue of the self-complementary RNA dinucleotide GPO_2^-C forms a short Watson−Crick duplex in the crystal.[56]

These results encouraged us to prepare longer oligosulfones. Here, we encountered the first evidence that the loss of the phosphodiester charge damages rule-based molecular recognition. For example, the tetramer R-$T_{SO_2}T_{SO_2}$-T_{SO_2}C-SR′ (R and R′ were protecting groups) formed duplexes in tetrahydrofuran,[55b] even though the duplex could not be joined by Watson−Crick pairing. Longer SNAs deviated more significantly from Watson−Crick behavior. Particularly remarkable was the sulfone $A_{SO_2}U_{SO_2}G_{SO_2}G_{SO_2}U_{SO_2}C_{SO_2}A_{SO_2}U$, prepared by Clemens Richert and Andrew Roughton.[57] The molecule, by itself, displayed a melting transition at ca. 70 °C, although the sequence was not, in the Watson−Crick sense, self-complementary. This SNA folded.

SNAs differing by only one nucleobase often display different levels of solubility, aggregation, folding, and chemical reactivity. For example, Zhen Huang found evidence for nucleobase pairing between $d(U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}^-)$ and its DNA complement. Bernd Eschgfaeller and Juergen Schmidt found no evidence for such pairing for an octamer differing at only one position and having T instead of U.[58]

These results suggested three hypotheses for how charged phosphodiester linkages might be important to molecular recognition in DNA.

(a) Repeating Charges Force Interstrand Interactions away from the Backbone. Backbone charges direct in-

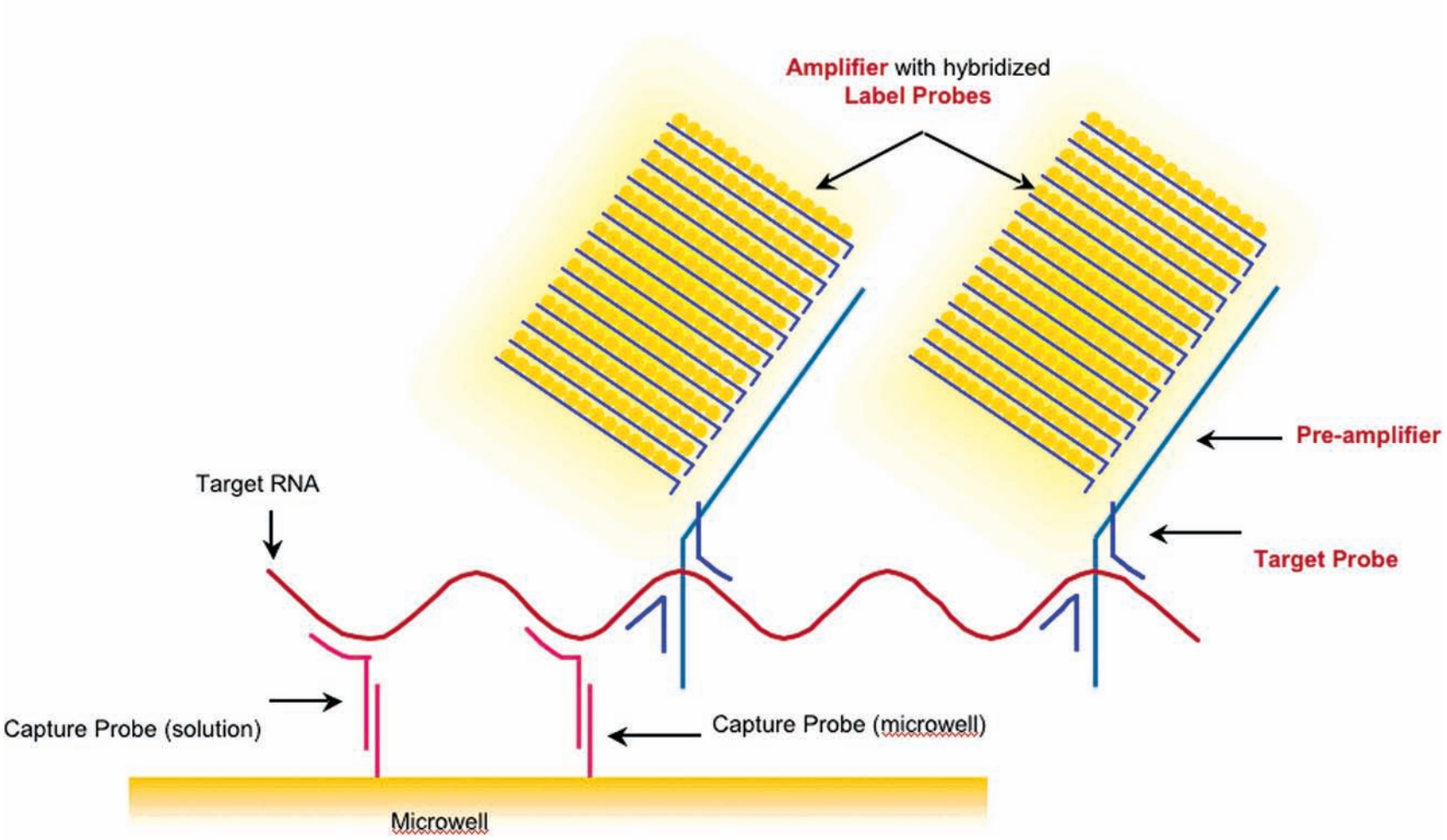

FIGURE 11. The branched DNA (bDNA) assay for HIV and hepatitis C exploits molecular recognition between AEGIS components to assemble the amplifier signaling nanostructure. The nonstandard nucleotides are present in the amplifier and preamplifier oligonucleotides. This allows the signaling nanostructure to be assembled without interference from nonanalyte DNA present in the sample, improving signal-to-noise ratios. Picture courtesy of Bayer Diagnostics.

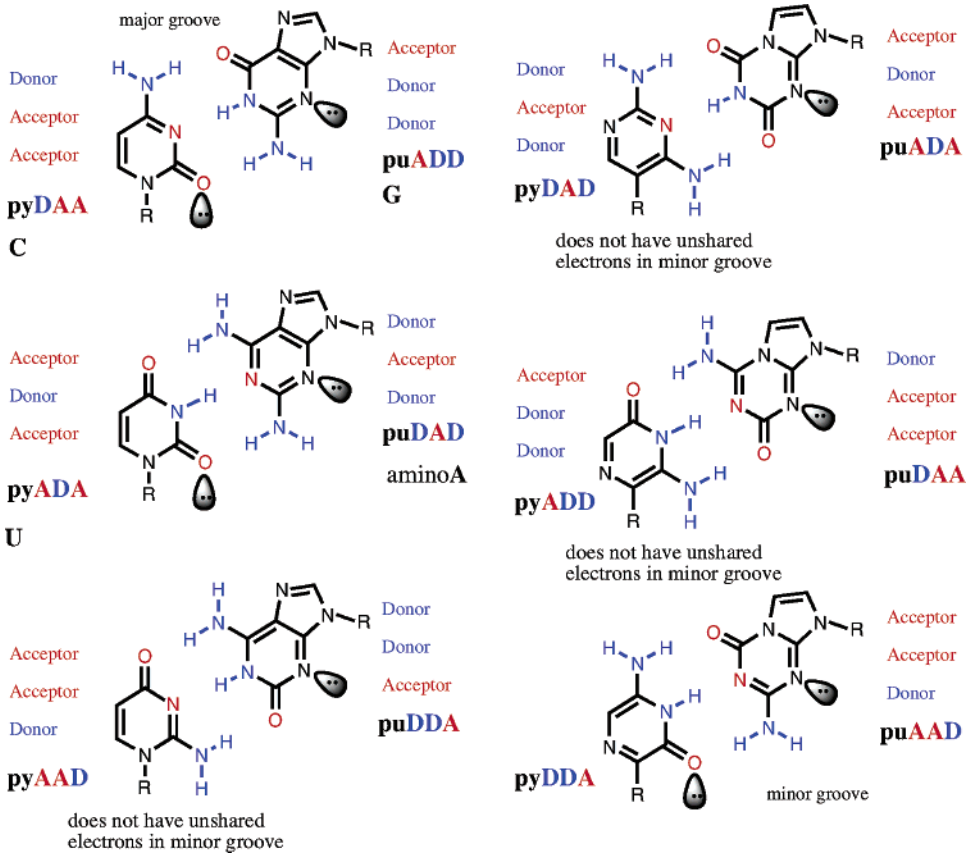

FIGURE 12. Standard nucleobases present a pair of electrons to the minor groove. The AEGIS compounds were designed, where possible, to present the pair at the same position. Three AEGIS pyrimidine analogues, implementing the pyAAD, pyDAD, and pyADD hydrogen-bonding schemes, cannot (if they have amino groups).

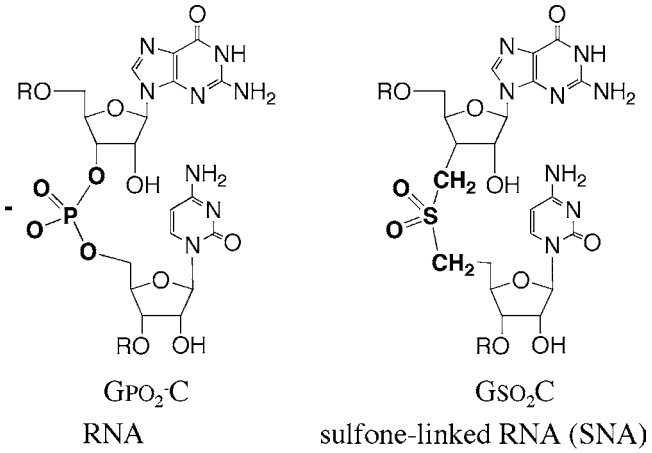

FIGURE 13. Sulfone-linked RNA (SNA) supports Watson–Crick pairing with short nonionic analogues but not with long analogues, which fold and aggregate. We speculate that a repeating charge is a universal feature of genetic molecules acting in water.

terstrand interactions to the Watson–Crick edge of the heterocycles (Figure 14). Without the polyanionic backbone, interstrand contacts can be anywhere. This was illustrated by the structure of the $U_{SO_2}C$ SNA.[59] In $CDCl_3$, this SNA forms a parallel duplex with a single U:U nucleobase pair and antiparallel orientation of the two ribose rings. Stabilizing this duplex is a hydrogen bond between the uridines, two hydrogen bonds from the ribose hydroxyls of one strand to O^2 of the cytosine bases of the other, and intrastrand hydrogen bonds from the 2′-hydroxyls of the 5′-terminal residues to hydroxyls of the 3′-terminal residues. Such contacts are possible because the neutral backbones do not repel each other.

(b) Repeating Charges Keep DNA Strands from Folding. Statistical theory of polymers holds that a randomly coiled hypothetical unidimensional polymer occupies a volume proportional to the length of the polymer raised to the 1.5 power.[60] This power is higher given the excluded volume of a real polymer and increases dramatically when the polymer is a polyelectrolyte. In short, a flexible polyanion is more likely to adopt an extended conformation suitable for templating than a flexible neutral polymer, which is more likely to fold. The polyanionic backbone thus allows DNA to be a template without a rigid scaffolding backbone.

(c) Repeating Charges Allow DNA To Support Darwinian Evolution. Replication is required, but not sufficient, for Darwinian evolution. The Darwinian system must also generate inexact replicates, descendants the chemical structures of which are different from those of their parents. These differences must then themselves be replicable.

While self-replicating systems are well-known in chemistry, those that generate inexact copies, the inexactness itself being replicable, are not. Indeed, small changes in molecular structure often lead to large changes in the physical properties of a system. This means that inexact

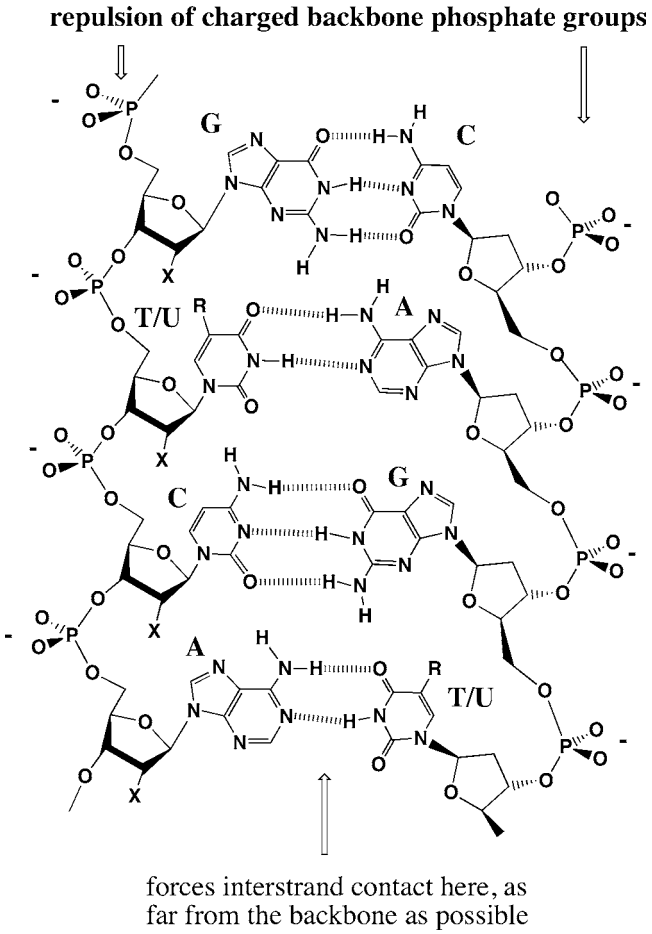

FIGURE 14. Backbone charges force interstrand contacts in a DNA duplex to the Watson−Crick edge of the heterocycles, hinder folding, and dominate physical behavior, allowing DNA/RNA to mutate and evolve.

replicates need not retain the properties of their ancestors, in particular, properties that are essential for replication.[53]

The polyanionic backbone dominates the physical properties of DNA. Replacing one nucleobase by another, therefore, has only a second-order impact on the physical behavior of the molecule. This allows nucleobases to be replaced during Darwinian evolution without losing properties essential for replication.

Molecules such as peptide-linked nucleic acid analogues (PNA) that lack the repeating charge illustrate how difficult it is to evolve uncharged DNA analogues.[61] PNAs differing only slightly in sequence can have very different physical properties; they may even precipitate. For this reason, we speculated that a repeating charge is needed for a genetic molecule to support Darwinian evolution in water.[48] Thus, if NASA missions do detect life in water on other planets, their genetics are likely to be based on polyanionic (or polycationic) backbones, even if their nucleobases differ from genetic molecules found on Earth.

Ribose as a Vestige of Prebiotic Synthesis

But what about the universality of the sugars? In the first generation model for duplex DNA, sugars were viewed as simple scaffolding.[1] To test the scope of this model, Christian Schneider replaced the ribose[62] by a flexible linker suggested by Joyce et al. as a molecule that was more likely to be made prebiotically.[63] The results were disappointing. The additional flexibility greatly diminished the stability of the duplex.

Since this work, others have prepared DNA analogues with novel sugars. Herdewijn introduced nucleic acid analogues based on hexose sugars.[64] Eschenmoser and his group have developed impressively the chemistry in this area, including analogues where ribose is replaced by threose.[65] In each case, the sugars proved to be more than simple scaffolding.

The search for a ribose replacement has been driven in part by prebiotic chemists. Compelling evidence suggests that modern life on Earth was preceded by a form of life where RNA was the only genetically encoded component of its biological catalysts.[66−68] Many have doubted, however, that life originated through the prebiotic formation of catalytic RNA. This doubt arises in part because of the instability of ribose[12] and the presumed difficulty of accumulating ribose on prebiotic Earth.

Ribose can be made abiologically. When incubated with technical calcium hydroxide, formaldehyde and glycolaldehyde, both known in the interstellar medium,[69] are converted into pentoses, including ribose, in a transformation known as the formose reaction (Figure 15).[70] This would be an interesting prebiotic route to ribose were it not for the fact that pentoses suffer further reaction under the conditions where they are formed to yield brown polymeric mixtures.

To generate ways to stabilize ribose, we turned to mineralogy. Borate is known to form a complex with organic molecules that carry 1,2-dihydroxyl groups. Borate is not expected to prevent the addition of the enediolate of glycolaldehyde (a nucleophile) to formaldehyde (an electrophile) to form glyceraldehyde (Figure 15). Glyceraldehyde should, however, complex borate. Because the complex is anionic, enolization of glyceraldehyde should be suppressed, preventing it from acting as a nucleophile. Glyceraldehyde nevertheless remains electrophilic, reacting with the enediolate of glycolaldehyde to give pentoses. Once formed, the cyclized forms of the pentoses should form largely unreactive borate complexes, because they lack C=O groups.

Experiments confirmed this reasoning. In the presence of $Ca(OH)_2$, a solution of glycolaldehyde and glyceraldehyde rapidly turns brown (Figure 16). When the same incubation is done in the presence of the borate minerals colemanite ($CaB_3O_4(OH)_3 \cdot H_2O$), ulexite ($NaCaB_5O_6(OH)_6 \cdot 5H_2O$), or kernite ($Na_2B_4O_6(OH)_2 \cdot 3H_2O$), however, the solution does not turn brown, even after months (Figure 16). The accumulation of ribose was confirmed chromatographically and by treating the mixture with $NaBH_4$ to generate ribitol, followed by enzymatic analysis with ribitol dehydrogenase.

Alkaline borate minerals are well-known on Earth. Serpentinization of olivine in mantle rocks generates alkaline solutions.[71] Borate is excluded from many silicate minerals, is enriched in igneous melts, and appears in

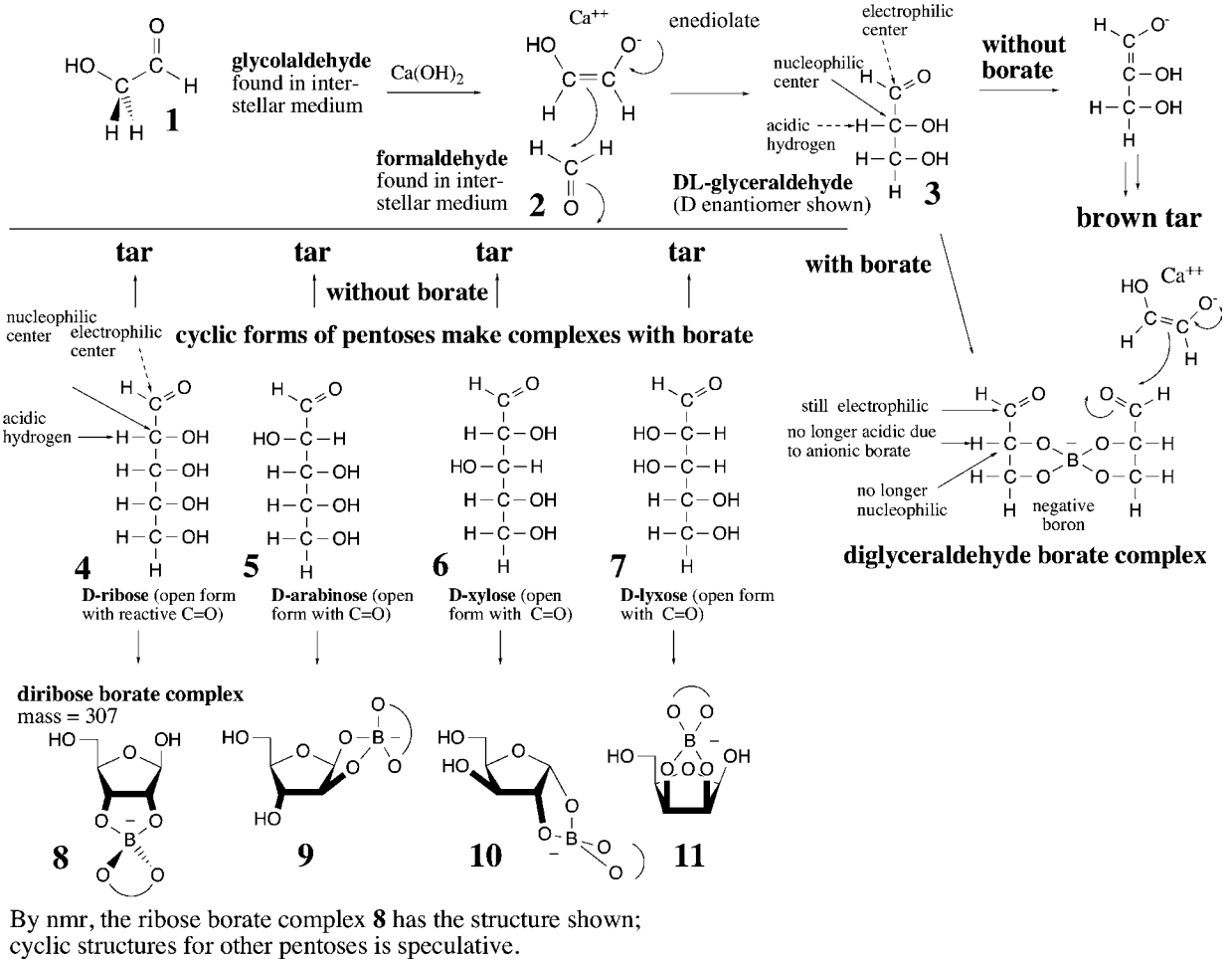

FIGURE 15. Ribose and its sister pentoses are formed from glycolaldehyde and formaldehyde in the formose reaction catalyzed by calcium hydroxide. Without borate minerals, pentoses continue to react to form brown tar. Borate captures the cis diol of the cycle of ribose, stabilizing it against degradation.

FIGURE 16. The stabilization of ribose by borate is shown by the lack of color, even after months. Shown are samples incubated under a variety of conditions with (white) and without (brown) borate.

igneous rocks on the surface as tourmalines. Weathering delivers borate to aqueous solution, the oceans, and to evaporite minerals.[72] Borate minerals so derived are well-known in Death Valley and as crusts on weathering rocks in Antarctic dry valleys.

The accumulation of ribose in the presence of borate minerals is therefore plausibly prebiotic. The structure of our genetic molecules might therefore reflect the intrinsic chemical interaction between organic molecules and geology.

Coincidence in Molecular Genetic Systems

The experiments discussed here are only a sample of those, in our laboratories and elsewhere, that have addressed the "What if?" and "Why not?" questions posed by the structure of DNA/RNA. In addition to work in other laboratories already mentioned, notable work has been done by Hirao,[73] Matsuda,[74] and Seela and their colleagues.[75]

Additional practical applications of artificial genetic systems are certain to emerge. AEGIS-containing sequences will soon be used to detect multiple nucleic acids, infectious agents, single nucleotide polymorphisms, and biowarfare agents.

If ribose turns out to be the intrinsically preferred product of interstellar molecules interacting in the presence of minerals on a prebiotic Earth, this will rank among one of the most fortunate coincidences in the cosmos. Not just any sugar will support Watson−Crick nucleobase pairing. Ribose does. Ribose is coincidentally stabilized by borate, and borate is available in the geosphere.

Further, the desert environments where borate-containing evaporite minerals are formed also have low water activities. These would help solve the problems associated with the synthesis of oligonucleotides that arise from their thermodynamic instability with respect to hydrolysis.

Synthetic and physical organic chemistry made these stories possible. Also essential was the diligence of many outstanding coworkers. These are mentioned throughout this article. I am indebted to them for their efforts, and grateful that they chose my laboratory in which to develop their careers.

References

(1) (a) Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **1953**, *171*, 737−738. (b) Watson, J. D.; Crick, F. H. C. General implications of the structure of deoxyribonucleic acid. *Nature* **1953**, *171*, 964−967.

(2) (a) Jayaraman, K.; McParland, K.; Miller, P.; Tso, P. O. P. Nonlonic Oligonucleoside Methylphosphonates. 4. Selective-Inhibition of *Escherichia coli* Protein-Synthesis and Growth by Non-Ionic Oligonucleotides Complementary to the 3′ End of 16S Ribosomal-RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 1537−1541. (b) Miller, P. S.; McParland, K. B.; Jayaraman, K.; Ts'o, P. O. P. Biochemical and biological effects of nonionic nucleic acid methylphosphonates. *Biochemistry* **1981**, *20*, 1874−1880.

(3) (a) Reddy, P. M.; Bruice, T. C. Solid-phase synthesis of positively charged deoxynucleic guanidine (DNG) oligonucleotide mixed sequences. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1281−1285. (b) Linkletter, B. A.; Szabo, I. E.; Bruice, T. C. Solid-phase synthesis of oligopurine deoxynucleic guanidine (DNG) and analysis of binding with DNA oligomers. *Nucleic Acids Res.* **2001**, *29*, 2370−2376.

(4) (a) Wengel, J.; Koshkin, A.; Singh, S. K.; Nielsen, P.; Meldgaard, M.; Rajwanshi, V. K.; Kumar, R.; Skouv, J.; Nielsen, C. B.; Jacobsen, J. P.; Jacobsen, N.; Olsen, C. E. LNA (Locked Nucleic Acid) *Nucleosides Nucleotides Nucleic Acids* **1999**, *18*, 1365−1370. (b) Renneberg, D.; Leumann, C. J. Watson−Crick base-pairing properties of Tricyclo-DNA. *J. Am. Chem. Soc.* **2002**, *124*, 5993−6002.

(5) Kool, E. T.; Morales, J. C.; Guckian, K. M. Mimicking the Structure and Function of DNA: Insights into DNA Stability and Replication. *Angew. Chem., Int. Ed.* **2000**, *39*, 990−1009.

(6) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable self-pairing hydrophobic base. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.

(7) (a) Pitha, J.; Pitha, M.; Ts'o, P. O. P. *Biochim. Biophys. Acta* **1970**, *204*, 39−48. (b) Eckstein, F.; Armstrong, V. W.; Sternbach, H. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2987−2990.

(8) Geyer, C. R.; Battersby, T. R.; Benner, S. A. Rules for designing artificial genetic systems. The nucleobases. *Structure* **2003**, *11*, 1485−1498.

(9) Collins, M. L.; Irvine, B.; Tyner, D.; Fine, E.; Zayati, C.; Chang, C. A.; Horn, T.; Ahle, D.; Detmer, J.; Shen, L. P.; Kolberg, J.; Bushnell, S.; Urdea, M. S.; Ho, D. D. A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/mL. *Nucleic Acids Res.* **1997**, *25*, 2979−2984.

(10) Schneider, K. C.; Benner, S. A. Oligonucleotides containing flexible nucleoside analogues. *J. Am. Chem. Soc.* **1990**, *112*, 453−455.

(11) Freier S. M.; Altmann K. The ups and downs of nucleic acid duplex stability: structure−stability studies on chemically modified DNA: RNA duplexes. *Nucleic Acids Res.* **1997**, *25*, 4429−4443.

(12) Larralde, R.; Robertson, M. P.; Miller, S. L. Rates of decomposition of ribose and other sugars. Implications for chemical evolution. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8158−8160.

(13) Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **1999**, *284*, 2118−2124.

(14) Ricardo, A.; Carrigan, M. A.; Olcot, A. N.; Benner, S. A. Borate minerals stabilize ribose. *Science* **2004**, *303*, 196−xxx.

(15) Benner, S. A. Synthetic biology. *Nature* **2003**, *421*, 118−xxx.

(16) Sismour, A. M.; Lutz, S.; Park, J.-H.; Lutz, M. J.; Boyer, P. L.; Hughes, S. H.; Benner, S. A. PCR Amplification of DNA containing nonstandard base pairs by mutants of reverse transcriptase from human immunodeficiency virus-1. *Nucleic Acids Res.*, in press.

(17) Piccirilli, J. A.; Krauch, T.; MacPherson, L. J.; Benner, S. A. A direct route to 3-(ribofuranosyl)-pyridine nucleosides. *Helv. Chim. Acta* **1991**, *74*, 397−406.

(18) Rao, P.; Benner, S. A. A fluorescent charge-neutral analogue of xanthosine: Synthesis of a 2′-deoxyribonucleoside bearing a 5-aza-7-deazaxanthine base. *J. Org. Chem.* **2001**, *66*, 5012−5015.

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

(20) Voegel, J. J.; Benner, S. A. Synthesis, molecular recognition and enzymology of oligonucleotides containing the non-standard base pair between 5-aza-7-deaza-iso-guanine and 6-amino-3-methylpyrazin-2-one, a donor-acceptor-acceptor purine analog and an acceptor-donor-donor pyrimidine analog. *Helv. Chim. Acta* **1996**, *79*, 1881−1898.

(21) von Krosigk, U.; Benner, S. A. pH-independent triple helix formation by an oligonucleotide containing a pyrazine donor-donor−acceptor base. *J. Am. Chem. Soc.* **1995**, *117*, 5361−5362.

(22) Voegel, J. J.; von Krosigk, U.; Benner, S. A. Synthesis and tautomeric equilibrium of 6-amino-5-benzyl-3-methylpyrazin-2-one. An acceptor-donor-donor nucleoside base analog. *J. Org. Chem.* **1993**, *58*, 7542−7547.

(23) Voegel, J. J.; Benner, S. A. Synthesis and characterization of nonstandard nucleosides and nucleotides bearing the acceptor−donor-donor pyrimidine analog 6-amino-3-methylpyrazin-2-one. *Helv. Chim. Acta* **1996**, *79*, 1863−1880.

(24) Hutter, D.; Benner, S. A. Expanding the genetic alphabet. Nonepimerizing nucleoside with the pyDDA hydrogen bonding pattern. *J. Org. Chem.* **2003**, *68*, 9839−9842.

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

(26) Kodra, J.; Benner, S. A. Synthesis of an N-alkyl derivative of 2′-deoxyisoguanosine. *Synlett* **1997**, 939−940.

(27) Jurczyk, S.; Kodra, J. T.; Rozzell, J. D., Jr.; Benner, S. A.; Battersby, T. R. Synthesis of oligonucleotides containing 2′-deoxy*iso*guanosine and 2′-deoxy-5-methyl*iso*-cytidine using phosphoramidite chemistry. *Helv. Chim. Acta* **1998**, *81*, 793−811.

(28) Jurczyk, S. C.; Battersby, T. R.; Kodra, J. T.; Park, J. H.; Benner, S. A. Synthesis of 2′-deoxyisoguanosine triphosphate and 2′-deoxy-5-methyl-isocytidine triphosphate. *Helv. Chim. Acta.* **1999**, *82*, 1005−1015.

(29) Voegel, J. J.; Altorfer, M. M.; Benner, S. A. 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* **1993**, *76*, 2061−2069.

(30) Voegel, J. J.; Benner, S. A. Non-standard hydrogen bonding in duplex oligonucleotides. The base pair between an acceptor-donor-donor pyrimidine analog and a donor-acceptor-acceptor purine analog. *J. Am. Chem. Soc.* **1994**, *116*, 6929−6930.

(31) Sepiol, J.; Kazimierczuk, Z.; Shugar, D. Tautomerism of Isoguanosine and Solvent-Induced Keto−Enol Equilibrium. *Z. Naturforsch.* **1976**, *31*, 361−370.

(32) Jurczyk, S. C.; Horlacher, J.; Devine, K. G.; Benner, S. A.; Battersby, T. R. Synthesis and characterization of oligonucleotides containing 2′-deoxyxanthosine using phosphoramidite chemistry. *Helv. Chim. Acta* **2000**, *83*, 1517−1524.

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

(34) Lutz, M. J.; Held, H. A.; Hottiger, M.; Huebscher, U.; Benner, S. A. Differential discrimination of DNA polymerases for variants of the nonstandard nucleobase pair between xanthosine and 2,4-diaminopyrimidine, two components of an expanded genetic alphabet. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.

(35) Rao, P.; Benner, S. A. A fluorescent charge-neutral analogue of xanthosine: Synthesis of a 2′-deoxyribonucleoside bearing a 5-aza-7-deazaxanthine base. *J. Org. Chem.* **2001**, *66*, 5012−5015.

(36) Jaworski, A.; Kwiatkowski, J. S.; Lesyng, B. Why isoguanine and isocytosine are not the components of the genetic code. *Int. J. Quantum Chem.* **1986**, *12*, 209−216.

(37) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. Enzymatic recognition of the base pair between *iso*-cytidine and *iso*-guanosine. *Biochemistry* **1993**, *32*, 10489−10496.

(38) Martinot, T. A.; Benner, S. A. Expanding the genetic alphabet: 7-Deaza-isoguanosine favors the 1N−H keto form by 10^3-to-1 over the enol. *J. Org. Chem.*, in press.

(39) Goodman, M. F. On the wagon. DNA polymerase joins "H-bonds anonymous". *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493−xxxxx.

(40) Packer, M. J.; Hunter, C. A. Sequence-Structure Relationships in DNA Oligomers: A Computational Approach. *J. Am. Chem. Soc.* **2001**, *123*, 7399−7406.

(41) Szathmary, E. What is the optimum size for the genetic alphabet? *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2614−2618.

(42) Shapiro, R. *Origins. A Skeptics Guide to the Creation of Life on Earth*; Penguin: London, 1986.

(43) (a) Elbeik, T.; Surtihadi, J.; Destree, M.; Gorlin, J.; Holodniy, M.; Jortani, S. A.; Kuramoto, K.; Ng, V.; Valdes, R.; Valsamakis, A.; Terrault, N. A. Multicenter evaluation of the performance char-

acteristics of the Bayer VERSANT HCV RNA 3.0 assay (bDNA). *J. Clin. Microbiol.* **2004**, *42*, 563−569. (b) Gleaves, C. A.; Welle, J.; Campbell, M.; Elbeik, T.; Ng, V.; Taylor, P. E.; Kuramoto, K.; Aceituno, S.; Lewalski, E.; Joppa, B.; Sawyer, L.; Schaper, C.; McNairn, D., Quinn, T. Multicenter evaluation of the Bayer VERSANT HIV-1 RNA 3.0 assay (bDNA), Analytical and clinical performance. *J. Clin. Virol.* **2002**, *25*, 205−216.

(44) Bain, J. D.; Chamberlin, A. R.; Switzer, C. Y.; Benner, S. A. Ribosome-mediated incorporation of nonstandard amino acids into a peptide through expansion of the genetic code. *Nature* **1992**, *356*, 537−539.

(45) Horlacher, J.; Hottiger, M.; Podust, V. N.; Huebscher, U.; Benner, S. A. Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with non-standard hydrogen bonding patterns. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333

(46) (a) Eom, S. H.; Wang, J.; Steitz, T. A. Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **1996**, *382*, 278−281. (b) Li, Y.; Korolev, S.; Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **1998**, *17*, 7514−7525. (c) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **1998**, *391*, 304−307. (d) Franklin, M. C.; Wang, J. M.; Steitz, T. A. Structure of the replicating complex of a Pol alpha family DNA polymerase. *Cell* **2001**, *105*, 657−667.

(47) Lutz, M. J.;. Horlacher, J.; Benner, S. A. Recognition of a non-standard base pair by thermostable DNA polymerases. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152.

(48) Lutz, M.; Horlacher, J.; Benner, S. A. Recognition of 2′-deoxy-isoguanosine triphosphate by HIV-1 reverse transcriptase and mammalian cellular DNA polymerases. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499−504.

(49) Braithwaite, D. K.; Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **1993**, *21*, 787−802.

(50) Benner, S. A.; Ellington, A. D. Interpreting the behavior of enzymes. Purpose or pedigree? *CRC Crit. Rev. Biochem.* **1988**, *23*, 369−426.

(51) Hendrickson, C.; Devine, K.; Benner, S. A. Probing the necessity of minor groove interactions with three DNA polymerase families using 3-deaza-2′-deoxyadenosine 5′-triphosphate. *Nucleic Acids Res.* **2004**, *32*, 2241−2250.

(52) Benner, S. A.; Caraco, M. D.; Thomson, J. M.; Gaucher, E. A. Planetary biology. Paleontological, geological, and molecular histories of life. *Science* **2002**, *293*, 864−868.

(53) Benner, S. A.; Hutter, D. Phosphates, DNA, and the search for nonterran life. A second generation model for genetic molecules. *Bioorg. Chem.* **2002**, *30*, 62−80.

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

(55) (a) Huang, Z.; Schneider, K. C.; Benner, S. A. Oligonucleotide analogs with dimethylene-sulfide, -sulfoxide and -sulfone groups replacing phosphodiester linkages. In *Protocols for oligonucleotides and analogs: synthesis and properties*; Agrawal, S., Ed.; Methods in Molecular Biology, Vol. 20; Humana Press Inc.: Totowa, NJ, 1993; pp 315−353. (b) Huang, Z.; Benner, S. A. Oligodeoxyribonucleotide analogs with bridging dimethylene-sulfide, sulfoxide, and sulfone groups. *J. Org. Chem.* **2002**, *67*, 3996−4013.

(56) Roughton, A. L.; Portmann, S.; Benner, S. A.; Egli, M. Crystal structure of a dimethylene sulfone linked ribodinucleotide analog. *J. Am. Chem. Soc.* **1995**, *117*, 7249−7250.

(57) Richert, C.; Roughton, A. L.; Benner, S. A. Nonionic analogs of RNA with dimethylene sulfone bridges. *J. Am. Chem. Soc.* **1996**, *118*, 4518−4531.

(58) (a) Schmidt, J. G.; Eschgfaeller, B.; Benner, S. A. A direct synthesis of nucleoside analogs homologated at the 3′ and 5′- positions. *Helv. Chim. Acta.* **2003**, *86*, 2937−2956. (b) Eschgfaeller, B.; Schmidt, J. G.; Benner, S. A. Synthesis and properties of oligodeoxynucleotide analogs with bis(methylene) sulfone-bridges. *Helv. Chim. Acta.* **2003**, *86*, 2957−2997.

(59) Steinbeck C.; Richert C. The role of ionic backbones in RNA structure: An unusually stable non-Watson-Crick duplex of a nonionic analog in an apolar medium. *J. Am. Chem. Soc.* **1998**, *120*, 11576−11580.

(60) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University, Ithaca, NY, 1953.

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

(62) Schneider, K. C.; Benner, S. A. Oligonucleotides containing flexible nucleoside analogs. *J. Am. Chem. Soc.* **1990**, *112*, 453−455.

(63) Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E.. The Case for an Ancestral Genetic System Involving Simple Analogues of the Nucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4398−4402.

(64) Augustyns, K.; Vanaerschot, A.; Herdewijn, P. 1992 Synthesis of 1-(2,4-dideoxy-beta-d-erythro-hexopyranosyl)thymine and its incorporation into oligonucleotides. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 945−948.

(65) Schoning, K. U.; Scholz, P.; Wu, X. L.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. The alpha-L-threofuranosyl-(3′-2′)-oligonucleotide system ('TNA'): Synthesis and pairing properties. *Helv. Chim. Acta* **2002**, *85*, 4111−4153.

(66) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* **2000**, *289*, 905−920.

(67) Bartel, D. P.; Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **1993**, *261*, 1411−1418.

(68) Benner, S. A.; Ellington, A. D.; Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7054−7058.

(69) Hollis, J. M.; Vogel, S. N.; Snyder, L. E.; Jewell, P. R.; Lovas, F. J. The spatial scale of glycolaldehyde in the galactic center. *Astrophys. J.* **2001**, *554* (Part 2), L81−L85.

(70) Breslow, R. On the mechanism of the formose reaction. *Tetrahedron Lett.* **1959**, *21*, 22−26.

(71) Moody, J. B. Serpentinization, Review. *Lithos* **1976**, *9*, 125−138.

(72) Kawakami, T. Tourmaline breakdown in the migmatite zone of the Ryoke metamorphic belt, SW Japan. *J. Metamorph. Geol.* **2001**, *19*, 61−75.

(73) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotech.* **2002**, *20*, 177−182.

(74) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. New base pairing motifs. The synthesis and thermal stability of oligodeoxynucleotides containing imidazopyridopyrimidine nucleosides with the ability to form four hydrogen bonds. *J. Am. Chem. Soc.* **2003**, *125*, 9970−9982.

(75) He, J.; Becher, G.; Budow, S.; Seela, F. Pyrazolo[3,4-d]pyrimidine nucleic acids: Ajustment of the dA-dT to the dG-dC base pair stability. *Nucleosides Nucleotides Nucl. Acids* **2003**, *22*, 573−576.

AR040004Z