Refining the Genetic Alphabet: A Late-Period Selection Pressure?

Andro C. Rios and Yitzhak Tor

Abstract

The transition from genomic ribonucleic acid (RNA) to deoxyribonucleic acid (DNA) in primitive cells may have created a selection pressure that refined the genetic alphabet, resulting from the global weakening of the *N*-glycosyl bonds. Hydrolytic rupture of these bonds, termed deglycosylation, leaves an abasic site that is the single greatest threat to the stability and integrity of genomic DNA. The rates of deglycosylation are highly dependent on the identity of the nucleobases. Modifications made to the bases, such as deamination, oxidation, and alkylation, can further increase deglycosylation reaction rates, suggesting that the native bases provide optimum *N*-glycosyl bond stability. To protect their genomes, cells have evolved highly specific enzymes called glycosylases, associated with DNA repair, that detect and remove these damaged bases. In RNA, however, the occurrence of many of these modified bases is deliberate. The dichotomous behavior that cells exhibit toward base modifications may have originated in the RNA world. Modified bases would have been advantageous for the functional and structural repertoire of catalytic RNAs. Yet in an early DNA world, the utility of these heterocycles was greatly diminished, and their presence posed a distinct liability to the stability of cells' genomes. A natural selection for bases exhibiting the greatest resistance to deglycosylation would have ensured the viability of early DNA life, along with the recruitment of DNA repair. Key Words: DNA—Nucleic acids—RNA world—Asteroid—Chemical evolution—Ribozymes. *Astrobiology* 12, 884–891.

Introduction: A Selection Pressure in the Early DNA World

DEOXYRIBONUCLEIC ACID (DNA) CONSISTS OF four letters that comprise the genetic alphabet: adenine (A), thymine (T), guanine (G), and cytosine (C). What were the underlying mechanisms of natural selection that favored these specific nucleobases during the course of biochemical evolution? This question has occupied investigators for decades given its fundamental relation to understanding the origin and evolution of nucleic acids (Rich, 1962; Westheimer, 1987; Eschenmoser, 1999; Szathmary, 2003). Here, we hypothesize that a refinement of the genetic alphabet could have taken place after the transition from genomic ribonucleic acid (RNA) to DNA in primitive cells. We describe a selection pressure that emerged in the early DNA world and is based on the difference in *N*-glycosyl bond stabilities between RNA and DNA. Stronger *N*-glycosyl bonds may have contributed to the utility of a wide variety of bases employed in RNA for diverse structural and functional capabilities. But with the invention of DNA, a global weakening to these bonds emerged and their hydrolysis became the governing threat to

genomic stability. Because the stability of *N*-glycosyl bonds is directly related to the identity of the linked heterocycles, a selection pressure for nucleobases most resistant to deglycosylation would have been imposed, resulting in the refinement of the genetic alphabet.

Prebiotic Chemistry and Alternative Bases

Studies in prebiotic chemistry routinely suggest that the native bases, while synthetically accessible and common, were likely to have been present in a mixture with numerous other nucleobases (Orgel, 2004; Borquez *et al.*, 2005; Benner *et al.*, 2010). Reactions such as deamination, aromatic substitution, oxidation, and alkylation of exocyclic amines are examples of modifications that readily occur with purines and pyrimidines under prebiotic conditions (Robertson and Miller, 1995a; Robertson and Miller, 1995b; Shapiro, 1995; Levy and Miller, 1998; Shapiro, 1999; Siegel and Tor, 2005; Powner *et al.*, 2009; Barks *et al.*, 2010). Analyses of carbonaceous meteorites also suggest that both the extraterrestrial and early terrestrial environment may have been diversely populated with related heterocycles (Botta and Bada, 2002;

Martins *et al.*, 2008; Callahan *et al.*, 2011). Furthermore, contributions illustrating that alternative bases and base pairs can both replace and expand the genetic alphabet continue to underscore the question as to why nature selected the native letters (Piccirilli *et al.*, 1990; Benner, 2004; Benner and Sismour, 2005; Chiba and Inouye, 2010). The discovery of bacteriophages that employ modified bases completely replacing one of the native letters (Fig. 1) is testament to the utility of modified letters in a functional DNA alphabet. Pyrimidine derivatives appear to be the most common modifications (Warren, 1980), but there is at least one case where 2,6-diaminopurine (Dap) was found to completely replace adenine (Kirnos *et al.*, 1977).

Selection pressures favoring the native bases, such as increased photochemical stability (Abo-Riziq *et al.*, 2005; Serrano-Andres and Merchan, 2009), decreased susceptibility to tautomerization (Roberts *et al.*, 1997; Robinson *et al.*, 1998), and greater comparative stability against decomposition (Levy and Miller, 1998), in comparison to other accessible heterocycles, have been discussed. Consequently, many investigations have operated under the assumption that the selection of the native bases could have been made during

the prebiotic epoch (Powner *et al.*, 2009; Powner *et al.*, 2010; Sutherland, 2010), pre-RNA (Joyce, 2002; Bean *et al.*, 2009; Engelhart and Hud, 2010), and/or the RNA world (Joyce, 1989; Bean *et al.*, 2007). However, with the emergence of DNA, another opportunity for base selection or refinement seems plausible.

The Emergence of Labile *N*-Glycosyl Bonds and DNA Repair

The transition from genomic RNA to DNA is widely accepted to be a result of a selection pressure for early forms of life to overcome the kinetic instability of the 3',5'-phosphodiester bond in ribonucleotides (Fig. 2) (Lazcano *et al.*, 1988; Li and Breaker, 1999). The removal of the 2'-OH group in the deoxyribonucleotides has weakened, however, the *N*-glycosyl bonds (Lindahl, 1993). DNA is vulnerable to a specific type of hydrolytic damage called deglycosylation, which involves the loss of a nucleobase via rupture of the *N*-glycosidic bond. Unlike RNA, where the 3',5'-phosphodiester bonds are subjected to transesterification reactions by the 2'-OH group, the loss of genetic information and backbone stability in DNA is

FIG. 1. Examples of bacteriophages that employ a modified base (HmU=5-hydroxymethyluracil, HmC=5-hydroxymethylcytosine, mC=5-methylcytosine, Dap=2, 6-diaminopurine, Hyp=Hypoxanthine) completely replacing one of the native bases in their genomes (Warren, 1980). All of the modified bases shown here are also considered prebiotically relevant heterocycles, and the purines have recently been identified in meteorites (Callahan *et al.*, 2011). It is interesting that while hypoxanthine (commonly known as inosine in nucleic acids) is routinely employed as a guanine letter in RNA, it has not been identified in bacteriophage functional DNA genomes.

FIG. 2. Half-life values for spontaneous damage to RNA and DNA. Both polymers are highly susceptible to spontaneous hydrolysis and subsequent chemistry, as indicated by the color (red: most vulnerable, blue: vulnerable, and black: minimally vulnerable) and arrow width of experimentally determined kinetics for single-stranded polymers at neutral pH and extrapolated to 25–37°C. While only one 5'-phosphodiester bond is indicated by an arrow, note that any other RNA backbone linkage is vulnerable to cleavage. Phosphodiester bonds in DNA appear to be essentially stable within the lifetime of any organism. However, it is the formation of abasic sites resulting from deglycosylation that exposes the Achilles Heel of DNA. Abasic sites in DNA, being hemi-acetals, are in equilibrium with their open chain aldehydes (about 1%) and are prone to β elimination reactions and strand cleavage. Experimental conditions: (a) pH 7.4, 37°C (Lindahl, 1993; Friedberg *et al.*, 2006; Gates, 2009); (b) pH 7, 23°C (Li and Breaker, 1999); (c) pH 7, 25°C (Schroeder *et al.*, 2006; Wolfenden, 2011); (d) comparison of abasic stability determined under alkaline pH, 37°C (Küpfer and Leumann, 2006); and (e) comparison of adenosine *N*-glycosyl stability determined at pH 7, extrapolated to 25°C (Stockbridge *et al.*, 2010).

dependent on the specific identity of the bases (Gates, 2009). Among these, the purines (A/G) deglycosylate under physiological conditions more frequently than the pyrimidines (C/T) (Fig. 2), and heat, divalent metal ions, or a low pH can accelerate these reactions even further (Lindahl, 1993). While RNA can also suffer from depurination, the occurrence of these reactions takes place at significantly reduced rates and low pH (Kochetkov, 1972) (Fig. 2). Importantly, even slight modifications made to the DNA bases, such as deamination, methylation, or oxidation, typically result with *increased* deglycosylation rates, suggesting that the native bases may provide optimum stability of their *N*-glycosyl bonds (Schroeder and Wolfenden, 2007; Gates, 2009).

The product of deglycosylation is called an apurinic or apyrimidinic site (AP), also known as an abasic site (Fig. 2) (Lindahl, 1993). The formation of an AP site is the single greatest threat to the integrity and stability of DNA (Lindahl, 1993; Gates, 2009). AP sites are both powerful mutagenic lesions (Loeb and Preston, 1986) and cytotoxic species given their reactive nature, which could lead to strand cleavage (Shapiro, 1981; Lhomme *et al.*, 1999; Boiteux and Guillet, 2004). For RNA, however, abasic sites are less reactive than

the DNA counterparts since RNA AP sites maintain enhanced stability against degradation (Fig. 2) (Küpfer and Leumann, 2006). It is important to note that while DNA is extremely resistant to direct phosphodiester bond cleavage (Williams *et al.*, 1999; Schroeder *et al.*, 2006), it can readily suffer from the same problem that plagues RNA upon deglycosylation (Eigner *et al.*, 1961; Sugiyama *et al.*, 1994).

The generation of abasic sites would have posed a formidable barrier to the persistence of DNA life had it not been for the recruitment of repair proteins (Jensen, 1976; Friedberg *et al.*, 2006). Modern cells devote substantial resources to the surveillance and maintenance of their genome, which includes AP sites and damaged bases (Lindahl and Wood, 1999). The base excision repair (BER) pathway is a preeminent defense mechanism used by cells that uses a variety of enzymes to detect and remove lesions and others to repair AP sites (Baute and Depicker, 2008). The need for cells to recruit enzymes that repair AP sites could have been the primary pressure that helps explain the origin of this highly sophisticated pathway. This activity has recently been discovered to also exist within the enzymatic capabilities of a DNA polymerase, hinting at a possible early pressure for

maintaining genomic stability combined with proper replication (Banos *et al.*, 2010). Yet, particularly advantageous to the evolution of BER was the recruitment of a class of enzymes known as glycosylases that specifically undertake the task of finding and removing damaged bases as a form of preventive measure to ensure the “health” of a cell’s genome (O’Brien, 2006). Primitive glycosylases may have evolved to differentiate bases simply by the relative ease of glycosidic bond excision. This feature, being able to differentiate between normal and damaged bases due to differences in *N*-glycosyl stability, has been observed even in modern glycosylases (O’Brien and Ellenberger, 2004; Bennett *et al.*, 2006; O’Brien, 2006). It is proposed, however, that many extant glycosylases use other methods for the detection and removal of lesions (Friedman and Stivers, 2010). But in an early DNA world in which less sophisticated forms of BER were used, glycosylases that exploit the difference in glycosidic bond stability would seem to be the simplest.

Greater *N*-Glycosyl Stability May Have Aided in the Utility of Diverse Bases in the RNA World

Intriguingly, many of the damaged DNA bases so diligently removed by glycosylases are the exact modifications created by proteins in transfer RNA (tRNA), ribosomal RNA

(rRNA), and messenger RNA (mRNA) (Table 1). Cells seem to exploit the enhanced stability of *N*-glycosyl bonds in RNA. Modified bases such as alkylated purines and pyrimidines (e.g., 7-methylguanine, 3-methylcytosine) that are unstable lesions in DNA are found to reside in RNAs (Limbach *et al.*, 1994; Gates, 2009). Hypoxanthine, having a weaker glycosidic bond compared to A or G and being a particularly potent mutagenic lesion in DNA (Schroeder and Wolfenden, 2007), is a ubiquitous modification in RNA, which is employed as a reliable guanine surrogate in RNA editing (Fig. 1, Table 1) (Nishikura, 2010). Even uracil, the base excluded from the DNA alphabet, is known to exhibit higher deglycosylation rates under neutral pH in comparison to thymine (Shapiro and Kang, 1969). Not all modifications found in RNA are necessarily excluded from DNA. Two examples of nucleobases that are used in DNA are 5-methylcytosine and 5-hydroxymethylcytosine (Table 1, Fig. 1), but they present interesting peculiarities. These bases, along with the parent cytosine, are notorious for their rapid rates of spontaneous deamination in comparison to the other letters (Fig. 2) (Levy and Miller, 1998). Yet cytosine is the one base that nature has selected to exploit, and its modifications can make up a substantial presence in DNA (Poole *et al.*, 2001; Nabel *et al.*, 2011). From the viewpoint of genetic fidelity, deamination reactions of C, mC, and HmC are the most problematic, but

TABLE 1. EXAMPLES OF THE DICHOTOMOUS OCCURRENCE OF BASES IN RNA AND DNA

Heterocycle	Identified RNA occurrence	Identified DNA occurrence
Hypoxanthine (Hyp)	All three domains in tRNA; used in eukarya mRNA; product of RNA editing; found in eukarya rRNA	Potent mutagenic lesion that arises from adenosine deamination; is a glycosylase substrate
Uracil (U)	Native to all RNA, is also generated from C in mRNA as part of the eukarya RNA editing process	Potent mutagenic lesion that arises from cytosine deamination and misincorporation by DNA polymerase; is a glycosylase substrate
5-methyluracil (T)	All three domains in tRNA, also found in rRNA of domain bacteria	Native to DNA (thymine) but also can arise from deamination of 5-methylcytosine and is a glycosylase substrate
5-methylcytosine (mC)	tRNA in domain archaea, eukarya rRNA all three domains	Used as an epigenetic marker in domain eukarya and completely replaces C in some bacteriophage DNA (Fig. 1)
5,6-Dihydrouracil (D)	All three domains in tRNA, and rRNA in the domain bacteria	An oxidative-deaminated lesion of cytosine; is a glycosylase substrate
Isoguanine (isoG)	Only known as a naturally occurring ribonucleoside	A lesion resulting from oxidative damage to adenine; is a glycosylase substrate
5-Hydroxyuracil (HoU)	tRNA in domain bacteria, eukarya	An oxidative-deaminated lesion of cytosine; is a glycosylase substrate
5-hydroxymethylcytosine (HmC)	Generated in rRNA of domain eukarya	Associated with epigenetic pathways, and completely replaces C in some bacteriophage DNA (Fig. 1)
7-methylguanine (7mG)	All three domains in tRNA, rRNA; used in mRNA as the 5' cap	Lesion resulting from alkylation; is a glycosylase substrate
1-methyladenine (1mA)	All three domains in tRNA, and in rRNA of the domain bacteria	Lesion resulting from alkylation; is a glycosylase substrate
N ⁶ -methyladenine (N ⁶ mA)	All three domains in rRNA, and in tRNA of archaea and eukarya	Lesion resulting from alkylation; is a glycosylase substrate
3-methylcytosine (3mC)	tRNA of domain bacteria and eukarya	Lesion resulting from alkylation; is a glycosylase substrate

References: Fuhrman *et al.*, 1981; Limbach *et al.*, 1994; Rozenski *et al.*, 1999; Ushijima *et al.*, 2005; O'Brien, 2006.

FIG. 3. Hypothesis diagram illustrating a refinement of the genetic alphabet.

with regards to DNA stability, cytosine retains one of the strongest glycosidic bonds (Fig. 2).

It was further shown in a detailed study in which deoxynucleosides were used to measure spontaneous deglycosylation rates that cytosine contributes the strongest *N*-glycosyl bond above all of the native bases (Schroeder and Wolfenden, 2007). It could be that one explanation for nature's particular selection of modifying the C5 position of cytosine in DNA is because the global impact to *N*-glycosyl stability is comparatively minimal. Many other modifications highlighted in Table 1 (D, HoU, 1mA, N⁶mA, isoG), while employed in RNA, are necessarily removed from DNA given their inability to maintain or exhibit genetic fidelity or function. Lacking, however, are thorough investigations of relative deglycosylation rates of these and other modified bases compared to the native letters that could provide a quantitative perspective as to how various modifications also affect the *N*-glycosyl bonds.

What is the evolutionary origin of the diversity seen in RNA bases? Although all these heterocycles result from post-transcriptional modifications, their structures resemble side chains of amino acids (Lazcano, 1994; Robertson and Miller, 1995a; Levy and Miller, 1999). The greater stability of RNA *N*-glycosyl bonds thus may have been an advantageous feature in the RNA world (Fig. 3). The utility of modified and exotic bases could have expanded the repertoire of catalytically competent RNA oligomers (Robertson and Miller, 1995a; Cermakian and Cedergren, 1998; Benner *et al.*, 1999; Levy and Miller, 1999; Forterre and Grosjean, 2009; Nguyen and Burrows, 2011) without the consequences of rapid deglycosylation. However, with the emergence of DNA and the greater utility of proteins, the functionalized bases would have become obsolete and a detriment to the survival of life (Fig. 3). While the selection of the DNA bases may have been aided by those exhibiting the greatest glycosidic bond stability, the eventual refinement of RNA bases would have largely mirrored the selection process based on energetic costs to the cell. With the increasing takeover by proteins, only the most essential bases for structural and functional roles in RNA would have continued to persist. Modified bases identified in contemporary RNA may actually have been part of the original larger family of diverse bases used in the RNA world (Cermakian and Cedergren, 1998).

Conclusion

The selection of the native bases did not occur in any one hypothetical period. It is more likely that a continuous process of refinement directed their selection throughout prebiotic and early biotic epochs. As suggested here, the fundamental change to the differences in hydrolytic susceptibility between RNA and DNA may have contributed to this refinement process. While life in the RNA world may have been challenged by the nature of the sugar moiety and its impact on backbone stability, in an early DNA world the governing pressure came from the identity of the attached nucleobase.

In this sense, the arrival of DNA should not be considered just a later modification of RNA; rather, it is a unique biopolymer in its own right that challenged life to adapt to its specific chemical vulnerabilities, to further refine the genetic

alphabet, and to evolve repair pathways that allowed for the ubiquity of DNA as we know it.

Acknowledgments

We are grateful to the NIH for support (via grant number GM 069773), Dr. Ulrich Muller, and the reviewers for their helpful comments to our manuscript.

Author Disclosure Statement

No competing financial interests exist.

Abbreviations

AP, apurinic or apyrimidinic; BER, base excision repair; Dap, 2,6-diaminopurine; DNA, deoxyribonucleic acid; RNA, ribonucleic acid.

References

- Abo-Riziq, A., Grace, L., Nir, E., Kabelac, M., Hobza, P., and de Vries, M.S. (2005) Photochemical selectivity in guanine-cytosine base-pair structures. *Proc Natl Acad Sci USA* 102:20–23.
- Banos, B., Villar, L., Salas, M., and de Vega, M. (2010) Intrinsic apurinic/apyrimidinic (AP) endonuclease activity enables *Bacillus subtilis* DNA polymerase X to recognize, incise, and further repair abasic sites. *Proc Natl Acad Sci USA* 107:19219–19224.
- Barks, H.L., Buckley, R., Grieves, G.A., Di Mauro, E., Hud, N.V., and Orlando, T.M. (2010) Guanine, adenine, and hypoxanthine production in UV-irradiated formamide solutions: Relaxation of the requirements for prebiotic purine nucleobase formation. *ChemBioChem* 11:1240–1243.
- Baute, J., and Depicker, A. (2008) Base excision repair and its role in maintaining genome stability. *Crit Rev Biochem Mol Biol* 43:239–276.
- Bean, H.D., Sheng, Y., Collins, J.P., Anet, F.A.L., Leszczynski, J., and Hud, N.V. (2007) Formation of a β -pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. *J Am Chem Soc* 129:9556–9557.
- Bean, H.D., Lynn, D.G., and Hud, N.V. (2009) Self-assembly and the origin of the first RNA-like polymers. In *Chemical Evolution II: From the Origins of Life to Modern Society*, Vol. 1025, edited by L. Zaikowski, J.M. Friedrich, and S.R. Seidel, American Chemical Society, Washington DC, pp 109–132.
- Benner, S.A. (2004) Understanding nucleic acids using synthetic chemistry. *Acc Chem Res* 37:784–797.
- Benner, S.A., and Sismour, A.M. (2005) Synthetic biology. *Nat Rev Genet* 6:533–543.
- Benner, S.A., Burgstaller, P., Battersby, T.R., and Jurczyk, S. (1999) Did the RNA world exploit an expanded genetic alphabet? In *The RNA World*, Vol. 37, edited by R.F. Gesteland, T.R. Cech, and J.F. Atkins, Cold Spring Harbor Monograph Series, Cold Spring Harbor, New York, pp 163–181.
- Benner, S.A., Kim, H.J., Kim, M.J., and Ricardo, A. (2010) Planetary organic chemistry and the origins of biomolecules. *Cold Spring Harb Perspect Biol* 2:a003467.
- Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L., and Drohat, A.C. (2006) Specificity of human thymine DNA glycosylase depends on *N*-glycosidic bond stability. *J Am Chem Soc* 128:12510–12519.
- Boiteux, S., and Guillet, M. (2004) Abasic sites in DNA: Repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* 3:1–12.

- Borquez, E., Cleaves, H.J., Lazcano, A., and Miller, S.L. (2005) An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers. *Origins Life Evol Biospheres* 35:79–90.
- Botta, O., and Bada, J.L. (2002) Extraterrestrial organic compounds in meteorites. *Surv Geophys* 23:411–467.
- Budke, B., and Kuzminov, A. (2006) Hypoxanthine incorporation is nonmutagenic in *Escherichia coli*. *J Bacteriol* 188:6553–6560.
- Callahan, M.P., Smith, K.E., Cleaves, H.J., Ruzicka, J., Stern, J.C., Glavin, D.P., House, C.H., and Dworkin, J.P. (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc Natl Acad Sci USA* 108:13995–13998.
- Cermakian, N., and Cedergren, R. (1998) Modified nucleosides always were: an evolutionary model. In *Modification and Editing of RNA*, edited by H. Grosjean and R. Benne, ASM Press, Washington DC, pp 535–541.
- Chiba, J., and Inouye, M. (2010) Exotic DNAs made of nonnatural bases and natural phosphodiester bonds. *Chem Biodiversity* 7:259–282.
- Eigner, J., Boedtker, H., and Michaels, G. (1961) Thermal degradation of nucleic acids. *Biochim Biophys Acta* 51:165–168.
- Engelhart, A.E., and Hud, N.V. (2010) Primitive genetic polymers. *Cold Spring Harb Perspect Biol* 2:a002196.
- Eschenmoser, A. (1999) Chemical etiology of nucleic acid structure. *Science* 284:2118–2124.
- Forterre, P., and Grosjean, H. (2009) The interplay between RNA and DNA modifications: back to the RNA world. In *DNA and RNA modification enzymes: Structure, Mechanism, Function and Evolution*, edited by H. Grosjean, Landes Bioscience, Austin, TX, pp 259–274.
- Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T. (2006) *DNA Repair and Mutagenesis*, 2nd edition, ASM Press, Washington, DC, pp 3–69.
- Friedman, J.I., and Stivers, J.T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* 49:4957–4967.
- Fuhrman, F., Fuhrman, G., Nachman, R., and Mosher, H. (1981) Isoguanosine: isolation from an animal. *Science* 212:557–558.
- Gates, K.S. (2009) An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol* 22:1747–1760.
- Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30:409–425.
- Joyce, G.F. (1989) RNA evolution and the origins of life. *Nature* 338:217–224.
- Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* 418:214–221.
- Kirnos, M.D., Khudyakov, I.Y., Alexandrushkina, N.I., and Vanyushin, B.F. (1977) 2-aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* 270:369–370.
- Kochetkov, N.K. (1972) *Organic Chemistry of Nucleic Acids, Part B*, edited by N.K. Kochetkov and E.I. Budovskii, Plenum Press, New York, pp 425–448.
- Küpfer, P.A., and Leumann, C.J. (2006) The chemical stability of abasic RNA compared to abasic DNA. *Nucleic Acids Res* 35: 58–68.
- Lazcano, A. (1994) The transition from nonliving to living. In *Nobel Symposium: Early Life on Earth*, edited by S. Bengtson, Columbia University Press, New York, pp 60–69.
- Lazcano, A., Guerrero, R., Margulis, L., and Oró, J. (1988) The evolutionary transition from RNA to DNA in early cells. *J Mol Evol* 27:283–290.
- Levy, M., and Miller, S.L. (1998) The stability of the RNA bases: Implications for the origin of life. *Proc Natl Acad Sci USA* 95:7933–7938.
- Levy, M., and Miller, S.L. (1999) The prebiotic synthesis of modified purines and their potential role in the RNA world. *J Mol Evol* 48:631–637.
- Lhomme, J., Constant, J.F., and Demeunynck, M. (1999) Abasic DNA structure, reactivity, and recognition. *Biopolymers* 52:65–83.
- Li, Y., and Breaker, R.R. (1999) Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J Am Chem Soc* 121:5364–5372.
- Limbach, P.A., Crain, P.F., and McCloskey, J.A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res* 22:2183–2196.
- Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362:709–715.
- Lindahl, T., and Wood, R.D. (1999) Quality control by DNA repair. *Science* 286:1897–1905.
- Loeb, L.A., and Preston, B.D. (1986) Mutagenesis by apurinic/apyrimidinic sites. *Annu Rev Genet* 20:201–230.
- Martins, Z., Botta, O., Fogel, M.L., Sephton, M.A., Glavin, D.P., Watson, J.S., Dworkin, J.P., Schwartz, A.W., and Ehrenfreund, P. (2008) Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet Sci Lett* 270:130–136.
- Nabel, C.S., Manning, S.A., and Kohli, R.M. (2011) The curious chemical biology of cytosine: Deamination, methylation, and oxidation as modulators of genomic potential. *ACS Chem Biol* 7:20–30.
- Nguyen, K.V., and Burrows, C.J. (2011) A prebiotic role for 8-oxoguanosine as a flavin mimic in pyrimidine dimer photorepair. *J Am Chem Soc* 133:14586–14589.
- Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 79:321–349.
- O'Brien, P.J. (2006) Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem Rev* 106:720–752.
- O'Brien, P.J., and Ellenberger, T. (2004) The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J Biol Chem* 279:26876–26884.
- Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 39:99–123.
- Piccirilli, J.A., Benner, S.A., Krauch, T., and Moroney, S.E. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343:33–37.
- Poole, A., Penny, D., and Sjöberg, B.M. (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2:147–151.
- Powner, M.W., Gerland, B., and Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242.
- Powner, M.W., Sutherland, J.D., and Szostak, J.W. (2010) Chemoselective multicomponent one-pot assembly of purine precursors in water. *J Am Chem Soc* 132:16677–16688.
- Rich, A. (1962) On the problems of evolution and biochemical information transfer. In *Horizons in Biochemistry*, edited by M. Kasha and B. Pullmans, Academic Press, New York, pp 103–126.
- Roberts, C., Bandaru, R., and Switzer, C. (1997) Theoretical and experimental study of isoguanine and isocytosine: base pairing in an expanded genetic system. *J Am Chem Soc* 119:4640–4649.
- Robertson, M.P., and Miller, S.L. (1995a) Prebiotic synthesis of 5-substituted uracils: A bridge between the RNA world and the DNA-protein world. *Science* 268:702–705.
- Robertson, M.P., and Miller, S.L. (1995b) An efficient prebiotic synthesis of cytosine and uracil. *Nature* 375:772–774.
- Robinson, H., Gao, Y.-G., Bauer, C., Roberts, C., Switzer, C., and Wang, A.H.J. (1998) 2'-deoxyisoguanosine adopts more than

- one tautomer to form base pairs with thymidine observed by high-resolution crystal structure analysis. *Biochemistry* 37:10897–10905.
- Rozenski, J., Crain, P.F., and McCloskey, J.A. (1999) The RNA modification database: 1999 update. *Nucleic Acids Res* 27:196–197.
- Schroeder, G.K., and Wolfenden, R. (2007) Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry* 46:13638–13647.
- Schroeder, G.K., Lad, C., Wyman, P., Williams, N.H., and Wolfenden, R. (2006) The time required for water attack at the phosphorus atom of simple phosphodiester and of DNA. *Proc Natl Acad Sci USA* 103:4052–4055.
- Serrano-Andres, L., and Merchan, M. (2009) Are the five natural DNA/RNA base monomers a good choice from natural selection? A photochemical perspective. *J Photochem Photobiol C-Photochem Rev* 10:21–32.
- Shapiro, R. (1981) Damage to DNA caused by hydrolysis. In *Chromosome Damage and Repair*, edited by E. Seeberg and K. Kleppe, Plenum Press, New York, pp 3–18.
- Shapiro, R. (1995) The prebiotic role of adenine: a critical analysis. *Origins Life Evol Biospheres* 25:83–98.
- Shapiro, R. (1999) Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life. *Proc Natl Acad Sci USA* 96:4396–4401.
- Shapiro, R., and Kang, S. (1969) Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine. *Biochemistry* 8:1806–1810.
- Siegel, J.S., and Tor, Y. (2005) Genetic alphabetic order: What came before A? *Org Biomol Chem* 3:1591–1592.
- Stockbridge, R.B., Schroeder, G.K., and Wolfenden, R. (2010) The rate of spontaneous cleavage of the glycosidic bond of adenosine. *Bioorg Chem* 38:224–228.
- Sugiyama, H., Fujiwara, T., Ura, A., Tashiro, T., Yamamoto, K., Kawanishi, S., and Saito, I. (1994) Chemistry of thermal degradation of abasic sites in DNA. Mechanistic investigation on thermal DNA strand cleavage of alkylated DNA. *Chem Res Toxicol* 7:673–683.
- Sutherland, J.D. (2010) Ribonucleotides. *Cold Spring Harb Perspect Biol* 2: a005439
- Szathmari, E. (2003) Why are there four letters in the genetic alphabet? *Nat Rev Genet* 4:995–1001.
- Ushijima, Y., Tominaga, Y., Miura, T., Tsuchimoto, D., Sakumi, K., and Nakabeppu, Y. (2005) A functional analysis of the DNA glycosylase activity of mouse MUTYH protein excising 2-hydroxyadenine opposite guanine in DNA. *Nucleic Acids Res* 33:672–682.
- Warren, R.A.J. (1980) Modified bases in bacteriophage DNAs. *Annu Rev Microbiol* 34:137–158.
- Westheimer, F. (1987) Why nature chose phosphates. *Science* 235:1173–1178.
- Williams, N.H., Takasaki, B., Wall, M., and Chin, J. (1999) Structure and nuclease activity of simple dinuclear metal complexes: Quantitative dissection of the role of metal ions. *Acc Chem Res* 32:485–493.
- Wolfenden, R. (2011) Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu Rev Biochem* 80:645–667.

Address correspondence to:

Andro C. Rios
Department of Chemistry and Biochemistry
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093

E-mail: androrios@gmail.com

Submitted 23 May 2012

Accepted 4 August 2012