1-2009

DNA Before Proteins? Recent Discoveries in Nucleic Acid Catalysis Strengthen the Case

Aaron Steven Burton
Portland State University

Niles Lehman
Portland State University

Citation Details

Burton, A.S.; Lehman, N. (2009). DNA before proteins? Recent discoveries in nucleic acid catalysis strengthen the case. Astrobiology 9:125-130.

ASTROBIOLOGY
Volume 9, Number 1, 2009

DOI: 10.1089/ast.2008.0240

Hypothesis Paper

DNA Before Proteins? Recent Discoveries in Nucleic Acid Catalysis Strengthen the Case

Aaron S. Burton and Niles Lehman

Abstract

An RNA-DNA World could arise from an all-RNA system with the development of as few as three ribozymes—a DNA-dependent RNA polymerase, an RNA-dependent DNA polymerase, and a catalyst for the production of DNA nucleotides. A significant objection to DNA preceding proteins is that RNA has not been shown to catalyze the production of DNA. However, RNA- and DNAzymes have been recently discovered that catalyze chemical reactions capable of forming deoxyribose, such as mixed aldol condensation of 5′-glyceryl- and 3′-glycoaldehyde-terminated DNA strands. Thus, the only remaining obstacles to RNA-catalyzed *in vitro* DNA synthesis are alterations of substrate and template specificities of known ribozymes. The RNA-DNA World lessens genomic size constraints through a relaxed error threshold, affording the evolutionary time needed to develop protein synthesis. Separation of information from catalyst enables genotype and phenotype to be readily discriminated by absence or presence, respectively, of the 2′-OH. Novel ribozymes that arise through mutation can be preserved in DNA by reverse transcription, which makes them much more likely to be retained than in an RNA-genome milieu. The extra degree of separation between protein and mRNA, in terms of identifying and then retaining a useful enzyme, may have in fact necessitated storing information in DNA prior to the advent of translation. Key Words: RNA World—Origin of DNA—Origin of protein—Ribozyme—Ribonucleotide reduction. Astrobiology 9, 125–130.

Introduction

The RNA World is generally accepted as a period during the origins of life on Earth in which RNA served as both the primary catalytic and informational macromolecule. Evidence supporting this hypothesis has been described elsewhere (Rich, 1962; Kuhn, 1972; Orgel, 1986; Joyce, 1991, 1998, 2002; Szostak and Ellington, 1993; Gesteland *et al*., 2006; Boussau *et al*., 2008), but it is worth mentioning a few reasons why RNA is believed to have preceded both DNA and proteins. In extant organisms, DNA synthesis is entirely dependent on RNA. For example, the monomer units for DNA synthesis, 2′-deoxyribonucleotides, are formed by the modification of ribonucleotides, and the primers used to initiate DNA polymerization are oligoribonucleotides. Regarding proteins, the strongest argument that they are preceded by RNA is the observation that the catalytic portion of the ribosome, which makes proteins, is composed entirely of RNA. Also, given that RNA is a poorer and less versatile catalyst than proteins, had proteins arisen first, what selective pressure could have existed to cause RNA catalysts to evolve?

Transitioning from RNA to DNA as the hereditary molecule greatly improved genomic stability, which increased the likelihood that a given organism (or molecule) would be around long enough to reproduce. Turning over the role of primary catalyst to proteins offered significant advantages as well—a wider array of chemical reactions could be catalyzed at a much faster rate, again contributing to a heightened probability that an organism survives to reproduce. Either transition affords obvious benefits to a ribo-organism, though in fundamentally different ways, and would come about through very different evolutionary pathways. Assuming RNA was the first of the three macromolecules, an unresolved dilemma is which came next, DNA or protein? Figure 1 outlines these two alternatives. Arguments have

Department of Chemistry, Portland State University, Portland, Oregon.

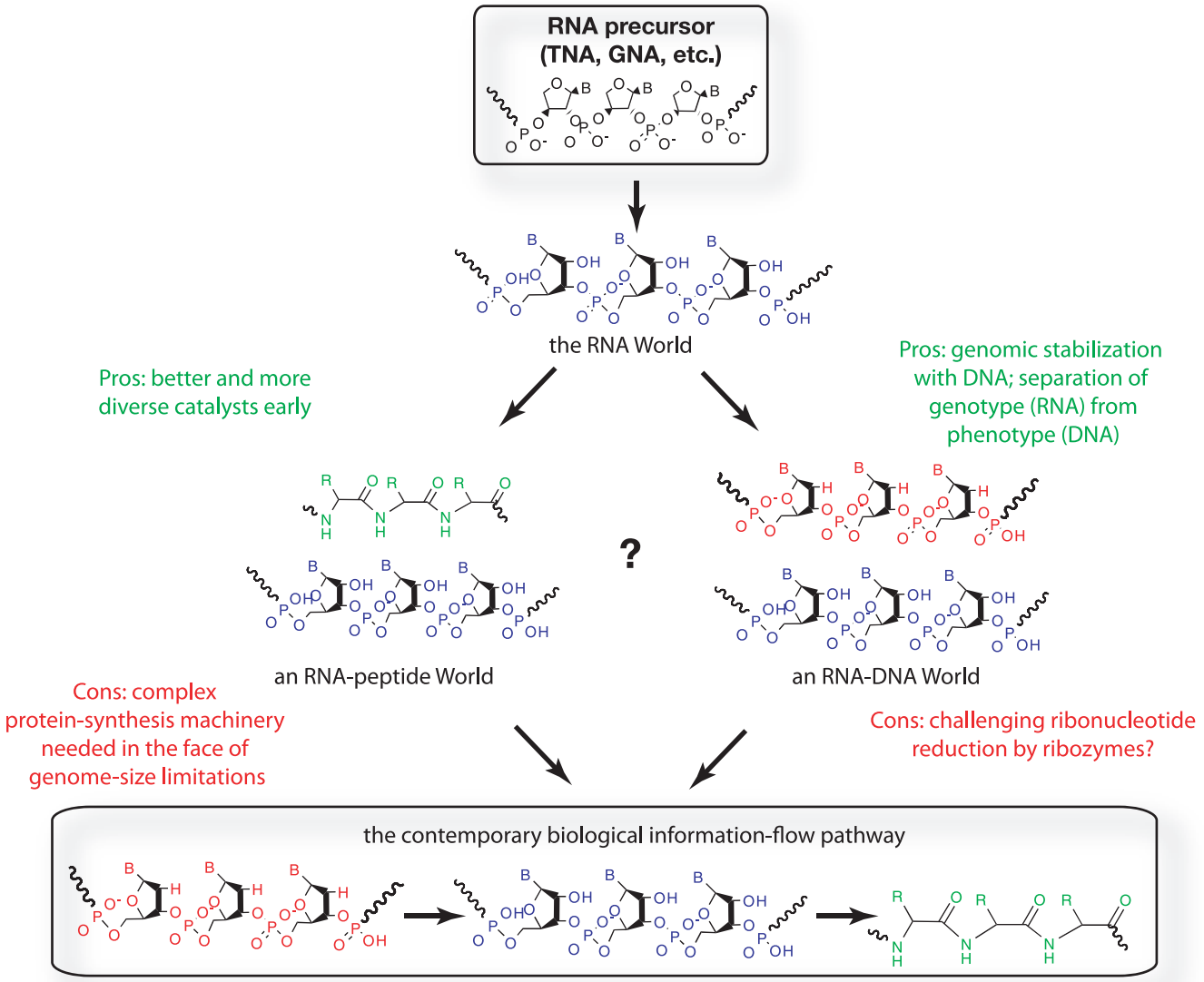

FIG. 1. Alternative pathways leading from the first information-rich macromolecule, such as glycerol nucleic acid (GNA) or threose nucleic acid (TNA), to present-day information flow. The order of appearance of DNA and proteins is currently unresolved.

been made for both sides. Simplicity favors DNA next (Benner *et al.*, 1989; Dworkin *et al.*, 2003), as fewer genes would be required to evolve DNA than protein synthesis. On the other hand, a compelling argument for proteins preceding DNA was made based on the difficulty of ribonucleotide reduction and homology of protein ribonucleotide reductases (Freeland *et al.*, 1999). However, recent discoveries in nucleic acid catalysis (summarized in Table 1) have led us to the conclusion that DNA likely came about before proteins, an opinion that we will defend below.

RNA-Catalyzed DNA Production

A key argument against DNA preceding protein revolves around ribonucleotide reduction (Freeland *et al.*, 1999; Poole *et al.*, 1999). The chemistry performed by ribonucleotide reductase proteins invariably involves a free-radical mechanism (Reichard, 1993; Sun *et al.*, 1993), which ribozymes have not yet been shown to catalyze. Furthermore, these radicals are manipulated by sulfhydryl groups, a functionality to which naturally occurring RNAs would have no access unless sulfur-containing cofactors were employed. However, ribozymes identified in nature and through *in vitro* selection have demonstrated a diverse catalytic repertoire, including redox chemistry (Tsukiji *et al.*, 2003), so it seems premature to declare RNA incapable of ribonucleotide reduction. In fact, it may prove that devising an effective *in vitro* selection scheme for a ribonucleotide reductase ribozyme is more difficult than the mechanism such a ribozyme ends up catalyzing. Moreover, it should not be assumed that a ribozyme would be required to utilize the same mechanism as protein ribonucleotide reductases; for example, alkane hydroxylation (essentially the reverse of ribonucleotide reduction), as catalyzed by methane monooxygenase protein enzymes, proceeds via a concerted two-step mechanism without the generation of a free radical (Yoshizawa, 1998). This raises the

Table 1. Catalytic Nucleic Acid Discoveries in the Last Decade Relevant to RNA-Catalyzed DNA Synthesis

Reaction	*Potential function in DNA synthesis*	*Citation*
RNA-catalyzed alcohol oxidation	Manipulation of aldol condensation substrates	Tsukiji *et al*. (2003)
RNA-catalyzed aldol condensation	Formation of deoxyribose	Fusz *et al*. (2005)
DNA-catalyzed aldol condensation of 2′-4′ linked DNA	Formation of deoxyribose	Oberhuber and Joyce (2005)
RNA-catalyzed attachment of bases to activated ribose	Nucleotide synthesis	Unrau and Bartel (1998)
RNA-catalyzed attachment of bases to activated ribose	Nucleotide synthesis	Lau *et al*. (2004)
RNA-catalyzed RNA polymerization	Evolutionary precursor for DNA polymerase	Johnston *et al*. (2001)
RNA-catalyzed RNA polymerization	Evolutionary precursor for DNA polymerase	Zaher and Unrau (2007)

possibility that a ribozyme could perform ribonucleotide reduction through a relatively simple radical-free mechanism.

More importantly, the enzymatic route to DNA nucleotides need not have begun with ribonucleotide reduction. The protein enzyme D-2-deoxyribose-5-phosphate aldolase catalyzes the formation of this product from acetaldehyde and D-glyceraldehyde-3-phosphate by the aldol reaction (Heine *et al*., 2001). Because RNA can catalyze aldol condensation chemistry (Fusz *et al*., 2005) and phosphate transfer (Lorsch and Szostak, 1994), and acetaldehyde and D-glyceraldehyde can be formed abiotically, aldol reactions present an alternative route for the enzymatic synthesis of deoxyribose prior to the evolution of a ribonucleotide reductase. Potentially, the synthesis of deoxyribonucleotides could share many of the catalytic steps (and enzymes, perhaps) needed for the production of ribonucleotides (Fig. 2). It has already been demonstrated that, within the context of double-stranded DNA, 2′-5′-phosphoester-linked ribose can be made as a consequence of the aldol condensation of phosphoglyceryl- and phosphoglycoaldehyde-terminated oligonucleotides (Oberhuber and Joyce, 2005). It is likely that RNA can catalyze this reaction as well. Artificial ribozymes have been discovered that are capable of attaching purine and pyrimidine bases to activated ribose (Unrau and Bartel, 1998; Lau *et al*., 2004), so all that remains is to confirm the ability of RNA to catalyze DNA-forming aldol reactions, find a ribozyme for activating ribose and deoxyribose for nucleobase addition (which presumably had to exist to make ribonucleotides), and verify that nucleotide synthase ribozymes can act on deoxyribose. While not trivial, these feats seem well within the catalytic repertoire of RNA, given the diversity of ribozymes discovered to date. Thus, at least two discrete chemical paths exist for the RNA-catalyzed formation of DNA: ribonucleotide reduction and synthesis from small molecules.

It has been noted that deoxyribose is unlikely to have been sufficiently abiotically available to have arisen prior to its enzymatic synthesis. It should be remembered that this same argument is made for ribose. Recent research has presented myriad abiotic routes from formaldehyde and its oligomers to nucleosides (Anastasi *et al*., 2007). Analogous routes could potentially exist for increased production of deoxyribonucleosides relative to other molecules, which simply have not yet been found. Indeed, having one less chiral center should reduce stereochemical complexity. The question is, regardless of whether its source is abiotic or RNA-catalyzed, how much DNA does an organism need? Cells generally have only a single copy of their genome in DNA. When expression occurs, multiple copies of RNA are made, and multiple proteins can be made per mRNA template. In a putative RNA-DNA organism that consists of one DNA copy of each gene and several RNA gene products per DNA gene, the deoxyribonucleotide demand would have been much smaller than that of their ribonucleotide counterparts, which must have been sufficiently available for the RNA World to have been a reality. Adding in the greater stability of DNA relative to RNA, the demand for deoxyribonucleotides was reduced even further. Thus, the required efficiency of DNA synthesis might have been significantly less than that needed for RNA production.

It has already been demonstrated that ribozymes are capable of acting on DNA. Group II introns have been demonstrated to be capable of DNA "polymerization," generating longer DNA oligos via successive transesterification reactions, as well as DNA ligation (Hetzer *et al*., 1997). In addition, a minimal number of mutations allow RNA enzymes to perform catalysis on either RNA or DNA (Beaudry and Joyce, 1992). It is conceivable then that existing RNA-acting ribozymes could have evolved to catalyze reactions on DNA substrates. Transitioning to an RNA-DNA World could have occurred with the development of as few as three ribozymes in the all-RNA World: (1) a DNA-producing catalyst; (2) an RNA-dependent DNA polymerase to convert the RNA genomic molecules into a more permanent storage form; and (3) a DNA-dependent RNA polymerase ribozyme to enable gene expression. The latter two, at least, would have had significant evolutionary head starts in RNA-dependent RNA polymerase ribozymes, and all three would have head starts if aldolase catalysts were responsible for ribose production. Thus, the advent of DNA was within evolutionary reach of the RNA World.

DNA Makes Protein Possible

The advent of protein synthesis by translation was one of the most challenging evolutionary transitions in the history of life. In its simplest modern form, the ribosome itself has three or more rRNAs and >60 proteins, ~40 tRNAs are charged with amino acids by 18–20 amino-acyl tRNA synthetases, and 7–8 translation factors are required; thus a total of ~130 discrete genes are needed for translation (Wolf and Koonin, 2007). A minimal translation system would re-

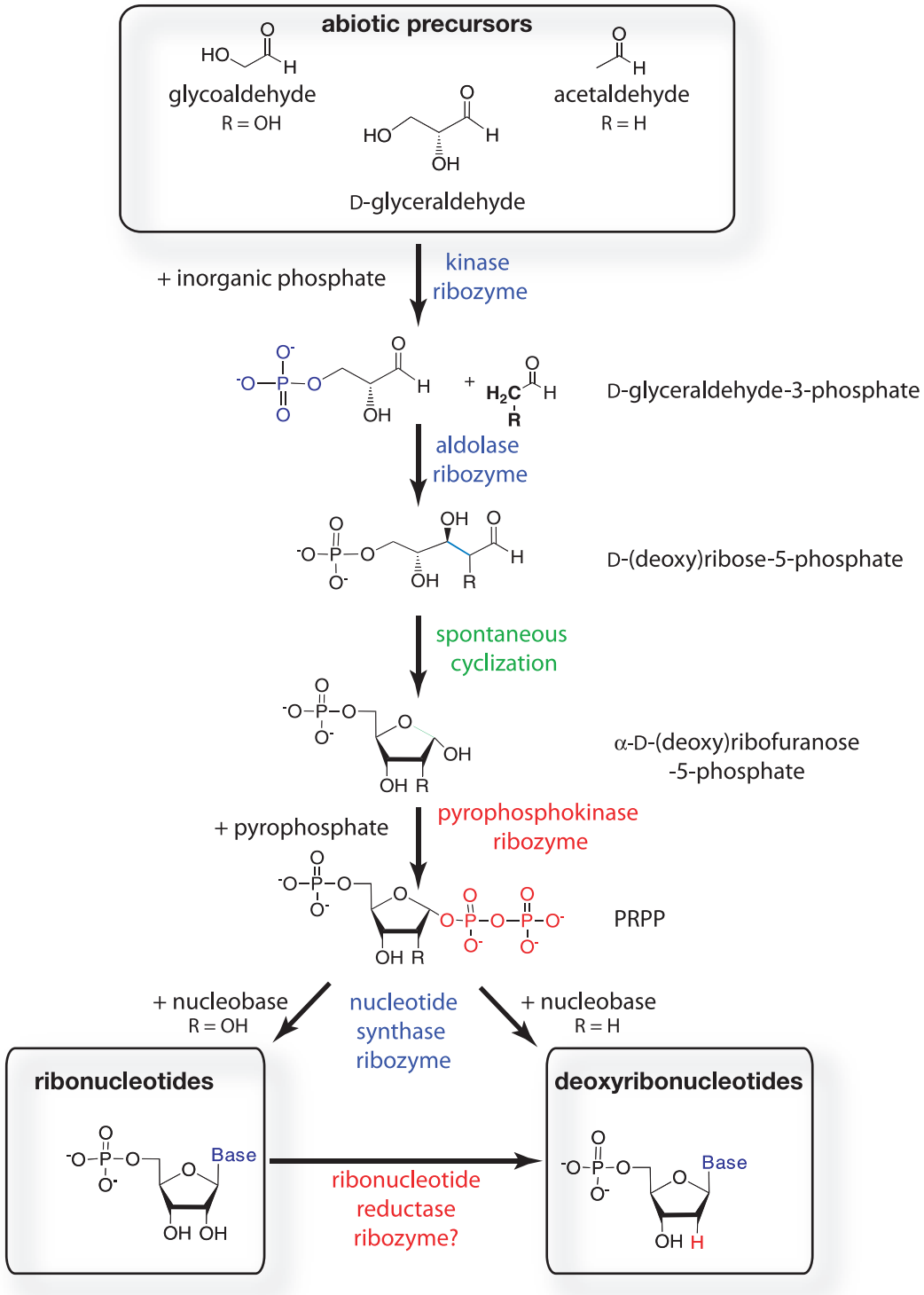

FIG. 2. Plausible metabolic routes to DNA and RNA from abiotic precursors. Ribo- and deoxyribonucleotides can be synthesized by the same chemistries and, possibly, the same enzymes simply by changing the initial abiotic molecules. Catalysts are color-coded: blue, chemistries already demonstrated in ribozymes; green, a chemistry that does not necessarily require a catalyst; red, chemistries yet to be demonstrated in ribozymes.

quire at least one ribosome ribozyme to catalyze peptide bond formation, and one tRNA (or equivalent structure for amino acid recognition) per amino acid, assuming all tRNAs were self-charging. Even with a primitive version of the genetic code that used 5–10 amino acids, this minimal set would have required at least 6–11 genes, without even considering the protein-coding genes themselves. It has been hypothesized that the minimum number of nucleotides a ribo-organism would need in its genome was at least 10,000–15,000 prior to the advent of translation (Jeffares *et*

al., 1998). Estimations of maximum genome size are based on polymerase fidelity. Because ribozymes can tolerate some mutations as truly neutral, they may not strictly fit the Eigen (1971) model where every mutation from the master sequence is considered deleterious (Kun *et al*., 2005; Takeuchi *et al*., 2005). The possibility remains, however, that the ribo-organism genome had reached the maximum allowed by the fidelity of RNA polymerase ribozymes (Johnston *et al*., 2001; Zaher and Unrau, 2007). Space constraints would further complicate the evolution of proteins, as translation-related genes would have to replace only non-essential genes. Any increase in maximum genome size would certainly make this evolutionary step easier. One solution to genome size constraints could be achieved through the development of error correction in RNA polymerase ribozymes, but ascertaining whether RNA can perform this function will be a formidable task. Alternatively, information could be stored in a more stable genomic polymer, which would also increase the genomic length maximum. For example, Poole *et al*. (2000) proposed a model by which RNA genomes could be stabilized through methylation of certain 2′-hydroxyls. However, any such methylation would have had to occur shortly after the molecule was synthesized, as the hydroxyls most in need of protection, by definition, would be the ones that promoted the fastest self-cleavage. Polymers of DNA are not faced with this same pressure and confer additional advantages to a primitive organism.

Early RNA genomes were likely double stranded, as this duplication of information allows for longer informational polymers; it is almost certain, then, that DNA-containing genomes were double stranded as well. The most advantageous way of incorporating DNA would have been in duplexes composed of one strand of RNA and one of DNA. In addition to the improved stability of DNA relative to RNA, the lack of 2′-OHs in the template strand clearly identifies it as such, which allows the information in that strand to be selectively used for gene expression. Prior to DNA, this discrimination could have been based on primitive sequence recognition or perhaps never occurred. This affords a side benefit of resource conservation, as unnecessary genome copies would not be made during the course of normal gene expression. The reduced number of hydroxyls also limits the complexity of structures that DNA can adopt, making it easier to keep the DNA in a double-stranded state and, thus, easier from which to extract information. Another improvement in fidelity granted by DNA stems from the spontaneous deamination of cytidine to uridine. If this occurs in the context of dsRNA, it is impossible to know whether the strand containing the guanosine originally paired to the C or the U resulting from the deamination is correct. Methylation of uridines in ribozymes would add steric hindrance, which would compromise catalytic activity but not affect the structure of typical double-stranded helices. Thus, the advent of DNA allows increased genome size by exploiting multiple facets of subtle chemical differences from RNA. The extra genomic space and stability granted by this transition would allow the more difficult evolution of translation to occur.

Conclusion

The RNA-DNA World was, at the very least, a possible stage on the road to life, a possibility largely overlooked because of the difficulty of ribonucleotide reduction. The 2′-OH of RNA is clearly a two-edged sword—while facilitating the adoption of catalytically active tertiary structures, it also renders RNA much more susceptible to hydrolysis. Chemistries newly observed in nucleic acid catalysts have outlined a plausible path by which DNA could be made by RNA, with no requirement for amino acid cofactors. Employing DNA makes gene expression more efficient, as it is now clear which molecules are genotypes and which are phenotypes. In addition, the increased stability of DNA polymers relative to RNA improves the lifespan of a genomic molecule and allows for larger overall genome size. The ability to store more genetic information allows for the inclusion of more genes, an important advance for an organism operating at or near its Eigen error threshold; indeed, this ability may have been essential for the development of translation and protein-coding genes.

The RNA-DNA World model presented here will be bolstered by the demonstration of a few catalytic feats: showing that the aldolase ribozyme can assemble deoxyribose, confirming that nucleotide synthase ribozymes can attach bases to deoxyribose, and demonstrating that in ribozyme catalysis RNA can serve as a template for DNA synthesis or vice versa. Because all these demonstrations begin with known ribozymes, however, they do not appear to be intractable. A plausible scenario for the ancient relationships among RNA, DNA, and proteins may be within reach.

Acknowledgments

This work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program, grant NNX07AR36H to A.S.B.

References

Anastasi, C., Buchet, F.F., Crowe, M.A., Parkes, A.L., Powner, M.W., Smith, J.M., and Sutherland J.D. (2007) RNA: Prebiotic, or biotic invention? *Chem. Biodivers*. 4:721–739.

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

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

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., and Gouy, M. (2008) Parallel adaptations to high temperatures in the Archaean eon. *Nature* 456:942–946.

Dworkin, J.P., Lazcano, A., and Miller, S.L. (2003) The roads to and from the RNA world. *J. Theor. Biol*. 222:127–134.

Eigen, M. (1971) Self-organization of matter and evolution of biological macromolecules. *Naturwissenschaften* 58:465–523.

Freeland, S.J., Knight, R.D., and Landweber, L.F. (1999) Do proteins predate DNA? *Science* 286:690–692.

Fusz, S., Eisenführ, A., Srivatsan, S.G., Heckel, A., and Famulok, M. (2005) A ribozyme for the aldol reaction. *Chem. Biol*. 12:941–950.

Gesteland, R.F., Cech, T.R., and Atkins, J.F., editors. (2006) *The RNA World*, 3rd ed., Cold Spring Harbor Laboratory Press, New York.

Heine, A., DeSantis, G., Luz, J.G., Mitchell, M., Wong, C.H., and Wilson, I.A. (2001) Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science* 294: 369–374.

Hetzer, M., Schweyen, R.J., and Mueller, M.W. (1997) DNA polymerization catalysed by a group II intron *in vitro*. *Nucleic Acids Res*. 25:1825–1829.

Jeffares, D.C., Poole, A.M., and Penny, D. (1998) Relics from the RNA World. *J. Mol. Evol*. 46:18–36.

Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* 292:1319–1325.

Joyce, G.F. (1991) The rise and fall of the RNA World. *New Biol*. 3:399–407.

Joyce, G.F. (1998) Nucleic acid enzymes: playing with a fuller deck. *Proc. Natl. Acad. Sci. U.S.A.* 95:5845–5847.

Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* 418:214–221.

Kuhn, H. (1972) Self-organization of molecular systems and evolution of the genetic apparatus. *Angew. Chem. Int. Ed. Engl*. 11: 798–820.

Kun, A., Santos, M., and Szathmáry, E. (2005) Real ribozymes suggest a relaxed error threshold. *Nat. Genet*. 37:1008–1011.

Lau, M.W.L., Cadieux, K.E.C., and Unrau, P.J. (2004) Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc*. 126:15686–15693.

Lorsch, J.R. and Szostak, J.W. (1994) *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* 371:31–36.

Oberhuber, M. and Joyce, G.F. (2005) A DNA-templated aldol reaction as a model for the formation of pentose sugars in the RNA World. *Angew. Chem. Int. Ed. Engl*. 44:7580–7583.

Orgel, L.E. (1986) RNA catalysis and the origin of life. *J. Theor. Biol*. 123:127–149.

Poole, A., Jeffares, D., and Penny, D. (1999) Early evolution: prokaryotes, the new kids on the block. *Bioessays* 21:880–889.

Poole, A., Penny, D., and Sjöberg, B.M. (2000) Methyl-RNA: an evolutionary bridge between RNA and DNA? *Chem. Biol*. 7:R207–R216.

Reichard, P. (1993) From RNA to DNA, why so many ribonucleotide reductases. *Science* 260:1773–1777.

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

Sun, X., Harder, J., Krook, M., Jörnvall, H., Sjöberg, B.M., and Reichard, P. (1993) A possible glycine radical in anaerobic ribonucleotide reductase from *Escherichia coli*: nucleotide sequence of the cloned *nrdD* gene. *Proc. Natl. Acad. Sci. U.S.A*. 90:577–581.

Szostak, J.W. and Ellington, A.D. (1993) *In vitro* selection of functional RNA sequences. In *The RNA World: the Nature of Modern RNA Suggests a Prebiotic RNA World*, edited by R.F. Gesteland and J.F. Atkins, Cold Spring Harbor Laboratory Press, New York, pp 511–533.

Takeuchi, N., Poorthuis, P.H., and Hogeweg, P. (2005) Phenotypic error threshold: additivity and epistasis in RNA evolution. *BMC Evol. Biol*. 5:9.

Tsukiji, S., Pattnaik, S.B., and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nature Struct. Biol*. 10:713–717.

Unrau, P.J. and Bartel, D.P. (1998) RNA-catalyzed nucleotide synthesis. *Nature* 395:260–263.

Wolf, Y.I. and Koonin, E.V. (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol. Direct* 2:14.

Yoshizawa, K. (1998) Two-step concerted mechanism for alkane hydroxylation on the ferryl active site of methane monooxygenase. *J. Biol. Inorg. Chem*. 3:318–324.

Zaher, H.S. and Unrau, P.J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* 13:1017–1026.

Address reprint requests to:
Aaron S. Burton
Department of Chemistry
Portland State University
PO Box 751
Portland, OR 97207

E-mail: burtona2001@msn.com

This article has been cited by:

1. Niles Lehman. 2010. RNA in evolution. *Wiley Interdisciplinary Reviews - RNA* **1**:2, 202-213. [CrossRef]