J. theor. Biol. (2001) **209,** 181–187
doi:10.1006/jtbi.2000.2253, available online at http://www.idealibrary.com on IDEAL®

The RNA/Protein Symmetry Hypothesis: Experimental Support for Reverse Translation of Primitive Proteins

MASAYUKI NASHIMOTO*

Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602, *U.S.A. and National Food Research Institute, Genetic Engineering Laboratory, Tsukuba, Ibaraki* 305-8642, *Japan*

(*Received on* 13 *September* 2000, *Accepted in revised form on* 15 *December* 2000)

Although the "RNA-world" theory, or the RNA-first theory is renowned for a promising theory of biogenesis, it is also possible that both RNAs and proteins have coevolved forming a stable metabolic complex from the very beginning. I investigated this possibility assuming that the genetic information flowed symmetrically in the era of the origin of life, i.e. the primitive translation machinery worked in both directions (from RNA to protein and from protein to RNA). According to this RNA/protein symmetry theory, the genetic information would have come from existing cellular proteins via reverse translation. This process would have been completed in a short period of time without searching an enormous RNA sequence space. Furthermore, reverse translation would have ensured biological continuity; proteins that were essential for cellular metabolism would have been utilized in the same way as before the protein sequence information would have been transferred into the RNA sequences. I also propose a possible mechanism for the process of reverse translation. The reverse translation would proceed in the 3′ to 5′ direction using a set of at least 20 reverse transfer RNAs (rtRNAs) that can recognize their specific amino acid residue and carry their corresponding codon. A source of genetic information would be a primary sequence of a protein molecule. Several basic steps of reverse translation were demonstrated using $rtRNA^{Arg}$.

1. Introduction

Life is thought to have appeared on Earth within 0.8 billion years after the formation of Earth around 4.6 billion years ago (Schidlowski, 1988). However, little is known about how life was generated on Earth. Biochemical building blocks such as amino acids, pyrimidines and purines might have been synthesized abiotically in the primitive atmosphere through electric discharges (Lazcano & Miller, 1996). Alternatively, prebiotic CO_2 fixation might have taken place on the surfaces of pyrite crystals, where energy might have been supplied by the pyrite formation reaction (Wächtershäuser, 1997). Both suppositions seem to be useful as the basis for further investigation of biogenesis, but they are still premature and controversial. Because there are so many difficulties in explaining the terrestrial origin of life, Crick & Orgel (1973) and Hoyle & Wickramasinghe (1981) have proposed the Panspermia theory in which life was formed elsewhere in the universe and came to Earth.

One of the major difficulties is a contemporary version of the chicken or egg problem: which came first, nucleic acids or proteins? In extant living organisms, nucleic acids carry heritable information that can be copied and expressed

*Author to whom correspondence should be addressed. National Food Research Institute, Genetic Engineering Laboratory, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8642, Japan. E-mail: nashimo@nfri.affrc.go.jp

0022–5193/01/060181 + 07 $35.00/0

only with the help of catalytic proteins, while proteins function in metabolism as biocatalysts that can be generated only through the information carried by nucleic acids.

In the early 1980s, it was discovered that RNA not only carries genetic information but also can function as a catalyst (Kruger *et al*., 1982; Guerrier-Takada *et al*., 1983). Following the first two examples, various other RNA enzymes (ribozymes) have been discovered in living organisms or created by artificial *in vitro* evolution (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990), and the versatility of RNA in terms of catalytic abilities has been confirmed. These observations have led many molecular biologists to believe that the first self-replicating system on Earth might have been made mainly of RNA molecules, which might have functioned both as information carriers and as biocatalysts (Gilbert, 1986). This era named "RNA world" is thought to have been followed by "RNP world" in which proteins produced by the translation machinery might have cooperated with RNAs by taking over RNAs' catalytic roles. According to this RNA-world scenario, DNA that plays a central role in storing the genetic information might have come on the scene last of all ("DNA world") (Freeland *et al*., 1999). Although the RNA-world hypothesis is attractive, it is interpreted in different ways by various scientists and is still controversial.

Protein-first scenarios are also likely and would be supported by the observation that protein-like substances ("proteinoids"), which have a variety of catalytic activities, can be easily formed by thermal polymerization of amino acids (Fox, 1988). The catalytic reactions include phosphatolysis, esterolysis, decarboxylation, deamination, oxidation and photochemical decarboxylation. A theoretical consideration by Dyson (1985) has suggested that even such a system that contains metabolizers but no replicators can jump from chaos to organized complexity. According to this model, $\sim$10 different kinds of building blocks for the metabolizers are needed for the transition to life, suggesting that nucleic acids were not appropriate for the primitive metabolism. In protein-first theories, nucleic acids might have been generated by autocatalytic sets of proteins as a part of accidental byproducts and subsequently might have become genetic material (Dyson, 1985).

The third possibility is that both nucleic acids and proteins have coevolved from the very beginning. In this paper, I investigate this possibility assuming that the genetic information flow was symmetric in the era of the origin of life.

2. Theory

2.1. SYMMETRY IN THE GENETIC INFORMATION FLOW

The contemporary central dogma on the genetic information flow forbids the transmission of information from proteins to nucleic acids (Crick, 1970). Generally, DNA makes RNA which then makes protein. In some biological systems, RNA also makes DNA. However, protein never makes RNA in any known organisms [Fig. 1(a)]. The symmetry with respect to the genetic information flow is broken in the extant organisms. What if, however, the genetic information flow was symmetric in the era of the origin of life [Fig. 1(b)]? What if the information flowed from proteins to nucleic acids? I examine how plausible this hypothesis is and what it implies for the early evolution of life. Furthermore, I present a model of the reverse flow of the genetic information.

2.2. THE RNA/PROTEIN SYMMETRY THEORY

I assume that nucleic acids and proteins cooperated from the beginning of life on Earth. Presumably, the nucleic acid players may have been mainly RNAs because the experiments for prebiotic synthesis of biochemical materials have suggested that ribonucleotides are much more easily generated than deoxynucleotides (Lazcano & Miller, 1996). The primordial cell might have been created through a transition from disordered to well-organized systems composed of relatively small RNAs and proteins at "the edge of chaos" (Kauffman, 1993). At the early stage of evolution where metabolism without genetic information processing might have been the central phenomenon for the cells, both RNAs and proteins might have functioned indistinguishably to maintain the system's complexity. The structural similarity between tRNA and elongation factor G in a contemporary organism might be a relic of this stage (Nissen *et al*., 1995). In addition, both

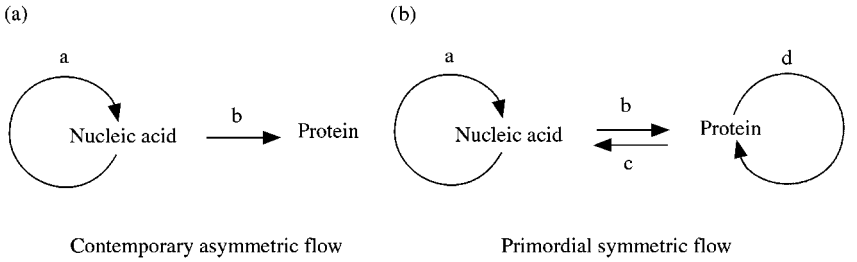

FIG. 1. Information flow in contemporary organisms (a) and in hypothetical primordial organisms (b). The arrows a, b, c, and d indicate nucleic acid replication, translation, reverse translation, and protein replication, respectively.

RNA and protein might have self-replicated before cell division (Doudna *et al.*, 1991; Lee *et al.*, 1996) (Fig. 1(b)).

The advent of genetic code and translation machinery might have been the most striking evolutionary breakthrough, although admittedly little is known about how they were invented. I assume that the primitive translation machinery worked in both directions of information flow using the same genetic code: from RNA to protein and from protein to RNA. Presumably, the proto-ribosomes for forward and reverse processes, which must have differed in a fundamental way so that each proceeds toward completion, might have been made of both RNAs and proteins from the beginning. The primary sequence information of useful primitive proteins might have been stored in RNA sequences through reverse translation, and the proteins might have been produced by translation of the RNAs carrying their information when they were necessary for cellular metabolism. Through the reverse translation and translation processes, the original RNA/protein symmetry might have been broken, resulting in role separation of RNA and protein: RNAs functioned mainly as information carriers, while proteins primarily played roles in metabolism as enzymes. Subsequently, the RNAs' role as an information storehouse might have been taken over by DNA molecules.

2.3. ADVANTAGES OF THE RNA/PROTEIN SYMMETRY HYPOTHESIS

What would be the advantages of the RNA/protein symmetry theory? One of the prominent advantages is that the theory can naturally incorporate the origin of genetic information in itself. Nothing is known about how genetic information was generated on Earth. By genetic information, I mean nucleotide sequences that can create functional proteins when translated using the contemporary genetic code. It seems that scientists tacitly think that the genetic information was generated by the Darwinian selection of random RNA sequences after the advent of translation apparatus; RNA molecules that code advantageous proteins for cells to survive might have been fixed as a part of genome. If this process was how primordial cells obtained their genetic information, it would have taken tremendous time.

According to the RNA/protein symmetry theory, the genetic information would have come from existing cellular proteins, which might have been generated by thermal polymerization, via reverse translation. This process would have been completed in a short period of time without searching an enormous RNA sequence space. More importantly, reverse translation would have ensured biological continuity; proteins that were essential for cellular metabolism would have been utilized in the same way as before the protein sequence information would have been transferred into the RNA sequences through reverse translation. We would not have had to wait until a monkey types Hamlet by chance.

Why did cells lose such an advantageous system, if it had ever evolved in the history of life? During the early evolution of life, the number of RNA genes might have continued to increase through reverse translation until an elaborate cell system had been built up, but later to build new RNA genes created by reverse translation into such an established system might have become difficult and a very rare event, resulting in the loss

of reverse translation machinery. In addition, organisms with the reverse translation system might have been overly susceptible to invasion by devastating parasites containing proteins that were very good at being reverse-translated.

3. Model

3.1. A MODEL FOR REVERSE TRANSLATION

I propose one possible mechanism for the process of reverse translation below (Fig. 2). In this model, reverse translation would proceed in the 3′ to 5′ direction using a set of at least 20 rtRNAs that can recognize their specific amino acid residue and carry their corresponding codon. A source of genetic information would be a primary sequence of a protein molecule.

First of all, the 3′ terminus of a precursor mRNA (pre-mRNA) that contains an anchor sequence (e.g. CCA) at its 3′ region and stop codon(s) (e.g. UGA and UAG) at its 5′ terminus would dock with the C terminus of a protein molecule [Fig. 2(a) and (b)]. Second, an rtRNA would specifically bind to the protein/pre-mRNA complex by recognizing the C-terminal amino acid residue corresponding to its codon and the anchor sequence, and would subsequently transfer its three-nucleotide codon to the 5′ end of the pre-mRNA [Fig. 2(c) and (d)]. Then, the rtRNA would leave the protein/pre-mRNA complex, resulting in the formation of a new complex of the pre-mRNA that obtained one additional codon and the protein that lost the C-terminal amino acid [Fig. 2(e)]. The new C-terminal amino acid would be recognized likewise by a corresponding rtRNA, and each step would proceed in the same way until the N-terminal amino acid is encoded into its codon sequence [Fig. 2(f–h)].

The following three points that are critical for the model should be noted. The machinery needs to read protein sequences by terminal degradation so that the terminal amino acid will be much

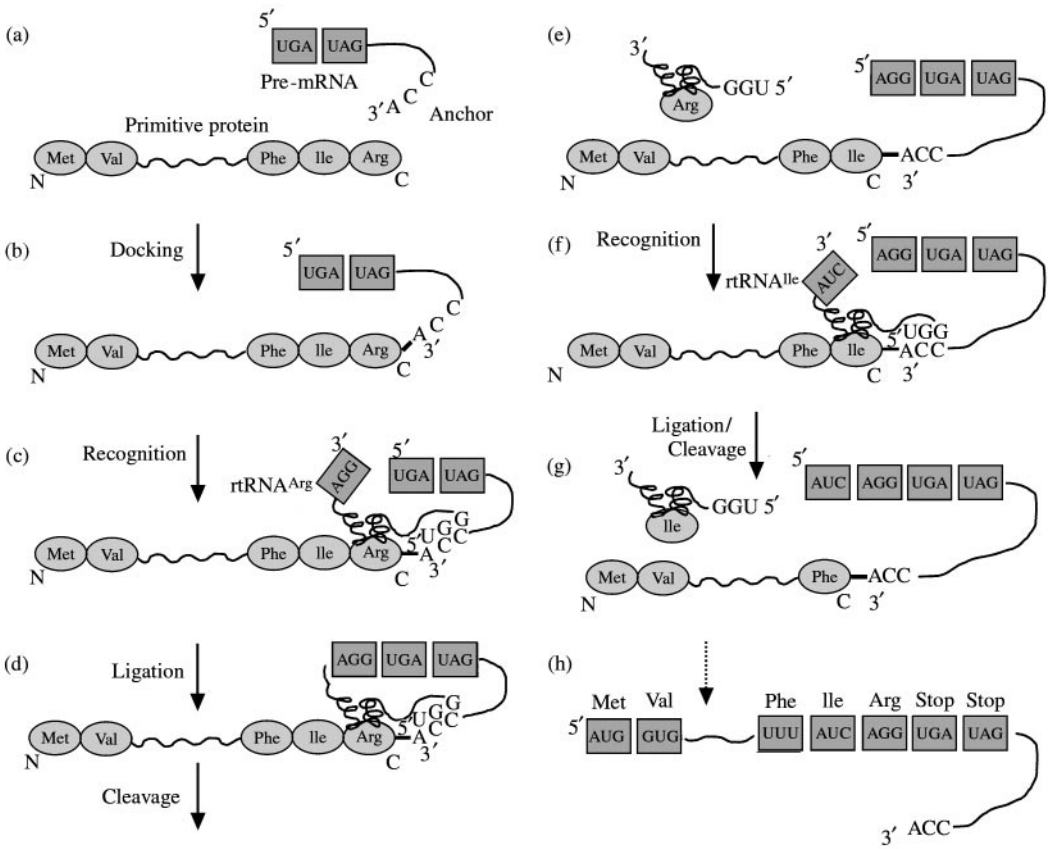

FIG. 2. A model for reverse translation. Each amino acid and each codon are denoted by oval and rectangle, respectively. See the text for details.

more similar to a free amino acid than the side-chain in a random-sequence context. The reverse translation must be processive, otherwise partially degraded proteins would be released and be used as templates of reverse translation for partially elongated pre-mRNAs, resulting in mRNA shuffling. Each pre-mRNA and each rtRNA must have a common anchor sequence (e.g. CCA) at the 3′ terminus and a common complementary anchor sequence (e.g. UGG) at

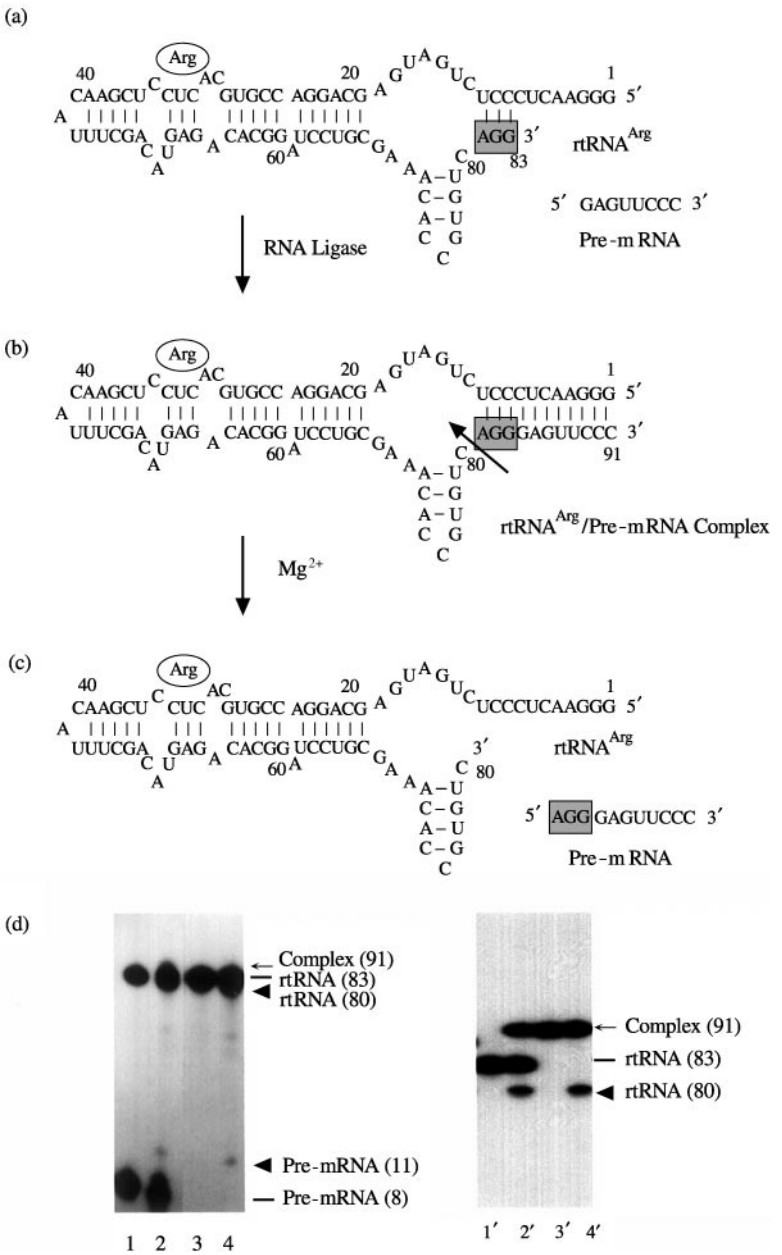

FIG. 3. Several basic steps of reverse translation. (a)–(c) Rectangles and an arrow indicate an arginine codon and the cleavage site of hammerhead ribozyme, respectively. An arginine is located tentatively on $rtRNA^{Arg}$. (d) An 83-nt $rtRNA^{Arg}$ (bar) and an 8-nt pre-mRNA (bar), both of which were ^{32}P-labeled, were ligated to produce a 91-nt rtRNA/pre-mRNA complex (arrow), and subsequently self-cleaved to generate an 11-nt pre-mRNA (arrowhead) and an 80-nt $rtRNA^{Arg}$ (arrowhead). A 91-nt $rtRNA^{Arg}$ and its self-cleavage products were used as size standards (lanes 3 and 4). Lanes 1 and 3 show the RNAs before the reactions, and lanes 2 and 4 denote the RNAs after the reactions. The numbers in parentheses indicate nt sizes. The right-hand side panel shows an autoradiograph with a higher resolution.

the 5′ terminus, respectively, so that the common rtRNAs can be utilized for any pre-mRNA.

3.2. DEMONSTRATION OF SEVERAL BASIC STEPS OF REVERSE TRANSLATION

I demonstrated that several basic steps of the reverse translation system are available in a laboratory using RNA engineering. An 83-nt $rtRNA^{Arg}$ was synthesized *in vitro* with T7 RNA polymerase [Fig. 3(a)]. The $rtRNA^{Arg}$ is composed of an arginine-binding domain (Connell *et al*., 1993) and a hammerhead ribozyme domain (Symons, 1992) that contains an arginine codon, AGG, at the 3′ terminus. An 8-nt pre-mRNA that is complementary to the 5′ sequence of the $rtRNA^{Arg}$ was synthesized with a DNA/RNA synthesizer. These two RNA molecules were first covalently joined via a phosphoester bond between the 3′ end of the $rtRNA^{Arg}$ and the 5′ end of the pre-mRNA with T4 RNA ligase [Fig. 3(b)]. Subsequently, this combined intermediate molecule was subjected to self-cleavage reaction [Fig. 3(c)]. This reaction resulted in the generation of two new RNA products [Fig. 3(d)]. One is an 80-nt $rtRNA^{Arg}$ lacking its codon sequence, and the other is an 11-nt pre-mRNA containing the AGG codon (Fig. 3). The sizes of the new products were confirmed with self-cleavage products of a 91-nt intermediate RNA molecule as reference standards (Fig. 3d). These results indicate that this system can transmit the codon AGG to another RNA molecule using the $rtRNA^{Arg}$ that can potentially recognize arginine.

4. Perspectives

4.1. TOWARD THE COMPLETE REVERSE TRANSLATION SYSTEM

Although this work is only a starting point to establish the complete reverse translation system, I expect that this system will be accomplished in the next decade by using the state-of-the-art RNA technology. rtRNAs corresponding to the other 19 amino acids should also be available relatively easily. A binding domain for each amino acid could be created by using *in vitro* RNA selection, or SELEX method (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990). Indeed, the arginine-binding domain used here and a tryptophan-binding domain have been identified using this method (Connell *et al*., 1993; Famulok & Szostak, 1992). However, there are three major hurdles to overcome: (i) to join the C-terminal residue of a protein to the 3′-end nucleotide of a pre-mRNA, (ii) to cleave out the C-terminal amino acid and rejoin a newly generated C-terminal residue to the 3′-end nucleotide, and (iii) to regenerate a codon sequence of each rtRNA. Although these problems seem to be tough, the *in vitro* RNA selection would be a promising method to solve them; ribozymes selected from a group I intron sequence pool have been shown to cleave an amide bond (Dai *et al*., 1995). It is also known that in Nature, poliovirus RNA is covalently combined to the VPg protein (Tobin *et al*., 1989).

4.2. APPLICATION

We have no idea whether reverse translation really existed in primordial cells, and even if it did, the ancient reverse translation system may have been far different from the system proposed here. Nevertheless, a successful reverse translation system should provide a widely applicable novel technology for molecular biologists: e.g. in proteomics, we would be able to reverse-translate thousands of different proteins from whole-cell extracts to mRNAs simultaneously, reverse-transcribe them to cDNAs, and subsequently analyse the cDNA mixtures using DNA arrays to obtain information on protein-expression patterns. Furthermore, this system could be introduced into living cells to make them evolve by incorporating new genetic information through reverse translation of exogenous proteins. Hopefully, future studies might reveal relics of reverse translation in contemporary organisms.

I would like to thank the referees who reviewed this paper for their helpful comments.

REFERENCES

CONNELL, G. J., ILLANGESEKARE, M. & YARUS, M. (1993). Three small ribooligonucleotides with specific arginine sites. *Biochemistry* **32,** 5497–5502.

CRICK, F. H. C. (1970). Central dogma of molecular biology. *Nature* **227,** 561–562.

CRICK, F. H. C. & ORGEL, L. E. (1973). Directed panspermia. *Icarus* **19,** 341–346.
DAI, X., MESMAEKER, A. D. & JOYCE, G. F. (1995). Cleavage of an amide bond by a ribozyme. *Science* **267,** 237–240.
DOUDNA, J. A., COUTURE, S. & SZOSTAK, J. W. (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251,** 1605–1608.
DYSON, F. (1985). *Origins of Life*. Cambridge: Cambridge University Press.
ELLINGTON, A. D. & SZOSTAK, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822.
FAMULOK, M. & SZOSTAK, J. W. (1992). Stereospecific recognition of tryptophan agarose by *in vitro* selected RNA. *J. Am. Chem. Soc.* **114,** 3990–3991.
FOX, S. (1988). *The Emergence of Life*: *Darwinian Evolution from the Inside*. New York: Basic Books.
FREELAND, S. J., KNIGHT, R. D. & LANDWEBER, L. F. (1999). Do proteins predate DNA? *Science* **286,** 690–692.
GILBERT, W. (1986). The RNA world. *Nature* **319,** 618.
GUERRIER-TAKADA, C., GARDINER, T., MARSH, N., PACE, N. & ALTMAN, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857.
HOYLE, F. & WICKRAMASINGHE, N. C. (1981). *Space Travelers*. Cardiff: University College Cardiff Press.
KAUFFMAN, S. A. (1993). *The Origins of Order*: *Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.
KRUGER, K., GRABOWSKI, P. J., ZAUG, A. J., SANDS, J., GOTTSCHLING, D. E. & CECH, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31,** 147–157.
LAZCANO, A. & MILLER, S. L. (1996). The origin and early evolution of life: prebiotic chemistry, the pre-RNA-world, and time. *Cell* **85,** 793–798.
LEE, D. H., GRANJA, J. R., MARTINEZ, J. A., SEVERIN, K. & GHADRI, M. R. (1996). A self-replicating peptide. *Nature* **382,** 525–528.
NISSEN, P., KJELDGAARD, M., THIRUP, S., POLEKHINA, G., RESHETNIKOVA, L., CLARK, B. F. C. & NYBORG, J. (1995). Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270,** 1464–1472.
ROBERTSON, D. L. & JOYCE, G. F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344,** 467–468.
SCHIDLOWSKI, M. (1988). A 3800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature* **333,** 313–364.
SYMONS, R. H. (1992). Small catalytic RNAs. *Annu. Rev. Biochem.* **61,** 641–671.
TOBIN, G. J., YOUNG, D. C. & FLANEGAN, J. B. (1989). Self-catalyzed linkage of poliovirus terminal protein VPg to poliovirus RNA. *Cell* **59,** 511–519.
TUERK, C. & GOLD, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510.
WÄCHTERSHÄUSER, G. (1997). The origin of life and its methodological challenge. *J. theor. Biol.* **187,** 483–494.

APPENDIX

RNA Synthesis

The 83- and 91-nt rtRNA^{Arg}s were synthesized with T7 RNA polymerase (Takara) from double-stranded DNA templates containing a T7 promoter in the presence of [α-^{32}P]UTP (Amersham). The transcription reactions were performed at 37°C for 1 hr in a buffer containing 40 mM Tris-HCl (pH 8.0), 20 mM $MgCl_2$, and 5 mM DTT. The 8-nt pre-mRNA was chemically synthesized with a DNA/RNA synthesizer, and subsequently 5′-end-labeled at 37°C for 30 min with T4 polynucleotide kinase (Takara) in the presence of [γ-^{32}P]ATP (Amersham) in a buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, and 5 mM DTT. These RNAs were gel-purified before assays.

RNA Ligation and Cleavage

The RNA ligation reaction of the 83-nt rtRNA^{Arg} and the 8-nt pre-mRNA was carried out with T4 RNA ligase (Takara) at 4°C for 1 hr in a buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, and 1 mM ATP). Subsequently, the RNA self-cleavage reaction was performed at 50°C for 1 hr in the same buffer. The reaction products were resolved in a 10% polyacrylamide–8 M urea gel, and the gel was autoradiographed.