A SYSTEM OF TWO POLYMERASES – A MODEL FOR THE ORIGIN OF LIFE

VICTOR KUNIN
Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

(Received 4 June, 1999)

Abstract. What was the first living molecule – RNA or protein? This question embodies the major disagreement in studies on the origin of life. The fact that in contemporary cells RNA polymerase is a protein and peptidyl transferase consists of RNA suggests the existence of a mutual catalytic dependence between these two kinds of biopolymers. I suggest that this dependence is a 'frozen accident', a remnant from the first living system. This system is proposed to be a combination of an RNA molecule capable of catalyzing amino acid polymerization and the resulting protein functioning as an RNA-dependent RNA polymerase. The specificity of the protein synthesis is thought to be achieved by the composition of the surrounding medium and the specificity of the RNA synthesis – by Watson – Crick base pairing. Despite its apparent simplicity, the system possesses a great potential to evolve into a primitive ribosome and further to life, as it is seen today. This model provides a possible explanation for the origin of the interaction between nucleic acids and protein. Based on the suggested system, I propose a new definition of life as a system of nucleic acid and protein polymerases with a constant supply of monomers, energy and protection.

Keywords: biogenesis, biological, coevolution, evolution, models, origin of life

1. Introduction

What was the earlier living molecule: RNA or protein? This question embodies a major controversy in theories on the origin of life. The essential role of proteins in the catalysis of the biochemical reactions, their structural role and compositional diversity gave rise to a 'protein world' theory (Dyson, 1982; Kauffman, 1986). In contrast, the RNA world theory (Rich *et al*., 1962; Raymond *et al*., 1993) was based on some important features of RNA molecules, including the universal distribution of these molecules and their participation in the most crucial functions for cell existence. The ability of RNA to fold and catalyze biological reactions (Cech *et al*., 1986) and its potential to serve as a template for self – replication were other important features of RNA that contributed to the establishment of the RNA world hypothesis. Both the RNA and protein world hypotheses assume that the interaction between these two kinds of biopolymers was a later development from a world consisted by RNA or protein respectively.

On the other hand, coevolution theories (Lahav, 1993) utilize the advantages of both approaches. They imply early coexistence and mutual dependence of both types of molecules. One of the most detailed hypotheses was proposed by Lahav

Origins of Life and Evolution of the Biosphere **30:** 459–466, 2000.
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

and Nir (1997). Their model describes the emergence of the sequence and template – directed synthesis of proteins. Nevertheless, this model was forced to start from a relatively complex system containing several types of RNA and catalytic peptides. A much simpler system, that could be a predecessor of the previous model, was suggested by Trifonov and Bettecken (1997). They proposed a system consisting solely of a mutually catalytic protein – RNA pair. They did not, however, suggest any details concerning the mechanism by which such a system could come into existence. In this study, such a mechanism will be considered, based on recent studies of RNA participation in protein synthesis.

2. Discussion

2.1. Achieving the components

The experiments of Miller and Urey (1959) make it hard to escape the conclusion that amino acids were present on the prebiotic Earth. While the origin of nucleotides remains unresolved, several possible mechanisms of nucleotide synthesis have been suggested and reviewed elsewhere (Joyce, 1989; Ferris, 1993; Orgel, 1998). After the appearance of the building blocks, non-biological polymerization of the monomers into polymers is supposed to occur. Oligonucleotides and oligopeptides were shown to form in number of different ways as a result of chemical catalysis (Lahav *et al*., 1978; Lawless *et al*., 1979; Odom *et al*., 1979; Fox *et al*., 1977; Saetia *et al*., 1993; Ferris *et al*., 1992; Hill *et al*., 1998). Several important questions concerning this process remain unanswered, notably, the differences between the conditions of monomer synthesis and polymerization, and the origin of chirality. Setting these challenges aside for a moment, chemical synthesis of random biopolymers from monomers appears to be viable and is assumed to have occurred by most theories concerning the origin of life.

A very primitive self-replicating system is thought to arise after the appearance of random biopolymers. Numerous efforts have been made to find self-replicating peptides (Lee *et al*., 1996), or self-replicating oligonucleotides (Sievers *et al*., 1994). In contrast, only a few studies demonstrating polymer synthesis from monomers have actually been attempted. Remarkable successes were achieved in isolation of RNA equivalents of oligonucleotide ligase (Ekland *et al*., 1995) and RNA polymerase (Ekland *et al*., 1996). Nevertheless, the reactions accomplished by described enzymes are still far off from polymerizing monomers to a functional enzyme.

The difficulty in these experiments comes from the following reason. In all extant organisms on the Earth, replication and synthesis of nucleic acids requires a protein polymerase, while synthesis of proteins requires the presence of RNA. Neither proteins, nor RNA self-replicate from monomers in modern cells. Thus, in today's world nucleic acids and protein are mutually catalytically dependent

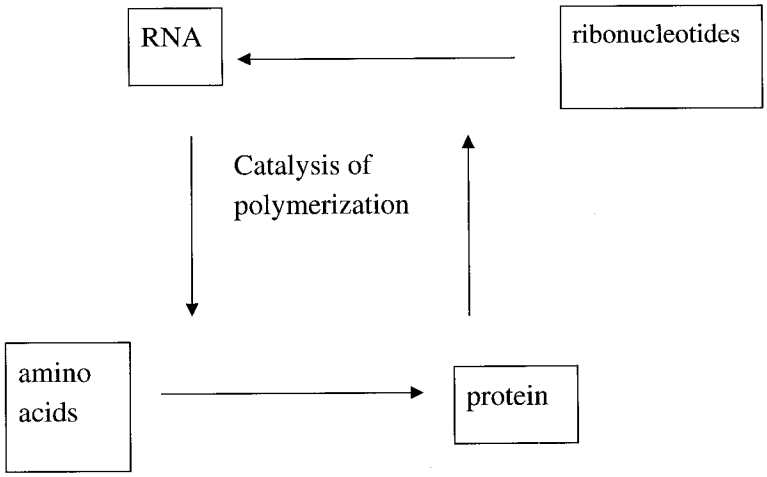

Figure 1. The scheme of mutual catalytic dependence: RNA catalyzes the polymerization of protein and vise versa.

(Figure 1). With this mutual dependence as the only data and invoking the principle of biological continuity (see Maizels *et al.*, 1994), I propose that this mutual dependence was a primary feature of the very first living systems. This hypothesis can explain the difficulties encountered in demonstrating self-replication of each component in laboratory studies, as well as the apparent disappearance of self-replication during subsequent evolution.

2.2. RNA SYNTHESIS

Consider now the primordial synthesis of RNA. The most natural and most parsimonious assumption would be that RNA synthesis was always catalyzed by protein. Lazcano and his collaborators (1988) hypothesized that RNA polymerase was one of the earliest proteins to appear and they attempted to identify vestiges of the 'original' enzyme. After studying several different RNA polymerases their conclusion, regarding the size of the hypothetical ancient enzyme, was: 'It is possible that the most ancient form of enzyme was that of relative simple, small catalytic polypeptide with the binding sites for the RNA template, the ribonucleotides, and the metallic cofactors'.

2.3. PROTEIN SYNTHESIS

On the other hand, the synthesis of proteins strongly depends on the presence of RNA. It was proposed that RNA has retained this catalytic function since the time when protein synthesis was solely catalyzed by RNA (Rich *et al.*, 1962; Green and Noller, 1997). It has been shown that rRNA alone, carefully cleaned from proteins, is capable of catalyzing the peptidyl transferase reaction (Noller *et al.*, 1992). However, in this experiment a small fraction of protein still remained, and

could possibly be a source of the catalyst for the reaction. In another experiment (Zhang *et al.*, 1997), a relatively short (196 nucleotides) synthetically synthesized ribozyme was able to accomplish the reaction. The synthesis of the ribozyme was accomplished without any ribosomal proteins, thereby excluding the possibility of protein catalysis.

Aminoacylated tRNA is another component necessary for the peptidyl transferase reaction. Modern tRNA molecules are too large to have arisen randomly in the primordial world due to combinatorial considerations. But is the entire tRNA molecule essential for the primitive reaction? Not at all. In fact, isolated minihelix domains, as short as seven – eight base pairs, are substrates for specific aminoacylation by many of the aminoacyl tRNA synthetases (Frugier *et al.*, 1996; Saks *et al.*, 1996; Francklyn *et al.*, 1990, 1992). In addition, minihelix domains could be the clue to the specific initiation of the primordial RNA replication (Maizels and Weiner, 1994).

Consider now the number of different RNA molecules essential for the peptidyl transferase catalysis. While in the modern world at least two different RNA molecules – tRNA and rRNA are essential, this was an unlikely event in the very first stage of life. At this stage, both molecules would likely to be present in extremely low copy numbers and thus, in almost null concentrations. Thus, in order to achieve the maximal simplicity in the proposed model, a single molecule is envisioned to be able to complete the whole reaction. It has been already suggested that rRNA and tRNA are homologous (Bloch *et al.*, 1985, 1989; Staves *et al.*, 1988, 1989; Ohnishi, 1992), yet the possibility that protein synthesis can be accomplished by a single RNA should be checked experimentally.

2.4. Composition – specific synthesis of peptides

From the discussion above, I conclude that peptide synthesis can be possibly achieved with a small RNA molecule functioning as both tRNA and rRNA. But at this stage, the polymerization would have been accomplished without any sequence specificity. I propose that under primordial conditions the specificity of peptide synthesis was achieved by the concentration of amino acids in the media. For comprehension of this concept, consider a medium containing one single type of amino acid. Polymerization of this amino acid would give rise to a homogenous population of oligopeptides, differing only in length. Since the polymerization would most probably stop when the chain reaches a sufficient length for folding, the length of the peptides synthesized under these conditions would be similar as well. Next, consider the addition of another amino acid to this system, in a lower concentration. The protein made in this new system would still consist of mostly one plentiful amino acid, with random rare insertions of the other. If the plentiful amino acid was a helix-former, e.g. alanine, and the rare amino acid was a loop-former, e.g. glycine, the peptide would probably exhibit a folded structure, perhaps similar to the helix-turn – helix motif. Alanine and glycine are the simplest amino

acids and hypothesized to be the most ancient (Trifonov and Bettecken, 1997, and references within). Adding other amino acids in small concentrations would produce a collection of proteins that are different in sequence, but possess a highly similar composition. These additional amino acids could possibly participate in the formation of an active site.

Considering the amino acid composition of the primordial soup, one could propose that it most probably contained one amino acid with a higher concentration than the others. Indeed, almost all the experiments that were performed to simulate the primordial synthesis of amino acids resulted in a mixture of one plentiful amino acid, with lower concentrations of the others (Miller, 1959; Fox *et al*., 1970; Harada *et al*., 1964; Bar-Nun *et al*., 1970). Thus, the model of medium – mediated specificity of protein synthesis can be applied to the prebiotic world with a high degree of confidence. The conclusion is that the first proteins were primary made of the plentiful amino acid and a few others, inserted at random rare positions. Thus, the synthesis of oligopeptides in this system would be partially specific, giving rise to various, but very similar, proteins. If among these proteins some possessed RNA–dependent RNA polymerase (RdRp) activity, the autocatalytic circle (Figure 1) would be complete.

Composition – specific synthesis would lead to the formation of a large protein library. Only a very small part of the synthesized proteins would have the desired RdRp activity. But consider the part of DNA replicase relatively to the total protein product produced in the modern cells. It is very small, and sometimes hardly detectable. This suggests that only a small amount of replicase is sufficient to maintain life and reproduction.

3. The Scheme of the Model

The following is the scheme of the primordial living system that I propose. This system consisted of two components. The first was an RNA molecule capable of catalyzing peptide bond formation. The second was a coexistent oligopeptide capable of catalyzing the replication of RNA (Figure 1).

The ability of the RNA molecule to catalyze the formation of peptide bonds allowed synthesis of oligopeptides from the amino acids, available in the media. Both the peptidyl transferase and the aminoacylation activities were not specific and were unable to distinguish between different amino acids. As a result, different oligopeptides could form. Nevertheless, the amino acid composition of the first proteins would be similar as a consequence of the amino acid composition of the medium. Since no specific aminoacyl tRNA synthetase existed, it is possible that the aminoacylation could occur with short peptides, as well as amino acids. The short peptides might also form peptide bonds with each other, due to RNA non-specific peptidyl transferase catalysis. The growing of the chain terminated due to

stereo hindrance, as the peptide chain grew longer and folded. Among the newly synthesized peptides, there likely would be some that possessed RdRp activity.

The RdRp activity of the oligopeptides would allow the proliferation of the RNA molecules. The polymerization reaction would use RNA as the template and ribonucleotides available in the environment. The reaction may have proceeded in two temperature-dependent stages. At elevated temperatures hydrogen bonds of folded RNA molecules melted, making them available for the RdRp that catalyzed the polymerization. When the temperature decreased, RNA molecules could refold and again be capable of catalyzing peptide bond formation. Day and night interchange could be the possible source of these temperature fluctuations.

The stage of RNA polymerization resembles a PCR reaction. The difference is that RNA polymerases do not require primers – an essential condition to accomplish the RNA polymerization reaction in primordial conditions where specific primers would be lacking. A PCR-like reaction would have the ability to provide exponential growth in the number of catalytic RNA molecules. These RNA molecules, in turn, would catalyze the appearance of peptides, including these with RdRp activity. This would help the system to retain a high enough reproduction rate in order to survive destruction and dilution, which would inevitably occur under primordial conditions.

The sequence of the RNA molecules could be influenced by selection. The heredity in the system was achieved by maintaining the sequence of RNA molecules. Imprecise replication and possibly other types of mutation could provide the material for the selection. Better catalytic RNA would bring about creation of a higher concentration of catalytic peptides in the closest microenvironment. The catalytic peptides, in turn, would replicate the closest available RNA, and most probably the one that had created them. As a result, a better RNA catalyst would have a better chance of being replicated. Thus, the system could be considered to be a primitive Darwinian entity.

The current model provides a challenge for experimentalists in various fields. Multiple questions could be asked. Is the composition-specific synthesis sufficient to create RNA polymerases? Would it also inevitably produce random enzymes that would dissipate the energy of the activated monomers? Is peptide synthesis possible with a single RNA molecule? Only experiments can answer these questions. The support for the model can come from RNA homology studies as well as combinatorial chemistry of peptides or RNA. *In silico* simplifying of RNA polymerase composition, with a subsequent synthesis of the catalytic peptides is an additional challenge. Other approaches to check the proposed model could be taken as well.

4. Evolution of the System

The described system could be initiated by the emergence of a single component. A single RNA catalyst would be able to synthesize a peptide RdRp, thereby establishing the whole system. The peptide RdRp, having no template, would possibly facilitate random synthesis of RNA molecules, with an occasional generation of functional ribozymes. This means that emergence of either component of the whole system could launch the autocatalytic cascade.

Once formed, such a system would possess great power to bring about further evolutionary development. One could speculate that evolutionary pressure would select those RNA molecules that create higher local concentrations of RdRp. This could be achieved by improving the catalytic activity of the RNA as well as synthesizing more peptides (or longer ones) with better yield of RdRp. Longer, and more specific sequences of peptides could possibly be achieved by aggregation of the RNA molecules on some RNA templates. This would drive the development of the system towards the evolution of template and sequence-directed synthesis of proteins (see Lahav and Nir, 1997), i.e. the primitive ribosome.

I believe that all living organisms today have retained this core system of two polymerases. All the known structures of nucleic acid polymerases are rather similar, suggesting divergence from a common ancestor (Hansen, 1997). Modern nucleic acid polymerases and peptidyl transferase have become much more efficient and controlled, so as not to run out of monomers and energy. They are still, however, the most crucial parts of any organism, and there is no life without them. Based on the suggested system, I propose the following definition of life. *Life is the system of nucleic acid and protein polymerases with a constant supply of monomers, energy and protection.* In the modern world, organisms achieve this supply by means of structure, metabolism, and behavior. Primordial organisms most probably did not have these properties, but obtained all the necessary components for replication from the environment. I believe that this is the only conceptual change from the appearance of life till these very days.

The system of two polymerases has survived billions of years and still exists in all living organisms on our beautiful Earth.

Acknowledgements

I would like to thank Dr. A. Shigaev and Prof. E. Trifonov for enlightening discussions during the preparation of the manuscript.

References

Bar-Nun, A., Bar-Nun, N., Bauer S. H. and Sagan, C.: 1970, *Science* **168**, 470.

Bloch, D. P., McArthur, B. and Mirrop, S.: 1985, *Biosystems* **17**, 209.
Bloch, D. P., McArthur, B., Guimaraes, R. C., Smith, J. and Staves, M. P.: 1989, *Braz. J. Med. Biol. Res.* **22**, 931.
Cech, T. R.: 1986, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4360.
Dyson, F. J.: 1982, *J. Mol. Evol.* **18**, 344.
Ekland, E. H., Szostak, J. W. and Bartel, D. P.: 1995, *Science* **269**, 364.
Ekland, E. H. and Bartel, D. P.: 1996, *Nature* **382**, 373.
Ferris, J. P. and Ertem, G.: 1992, *Science* **257**, 1387
Ferris, J. P.: 1993, *Origins Life Evol. Biosphere* **23**, 307.
Fox, S. W. and Windsor, C. R.: 1970, *Science* **170**, 984.
Fox, S. W. and Dose, K.: 1977, *Molecular Evolution and the Origin of Life*, Dekker, New York.
Francklyn, C. and Schimmel, P.: 1990, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8655.
Francklyn, C., Shi, J. P. and Schimmel, P.: 1992, *Science* **255**, 1121.
Frugier, M., Florentz, C. and Giege, R.: 1994, *EMBO J.* **13**, 2219.
Gesteland, R. F. and Atkins, J. F. (eds.): 1993, *The RNA World*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993.
Green, R. and Noller, H.: 1997, *Ann. Rev. Biochem.* **66**, 191.
Hansen, J. L., Long, A. M. and Schultz, S. C.: 1997, *Structure* **5**, 1109.
Harada, K. and Fox, S. W.: 1964, *Nature* **20**, 335.
Hill Jr., A. R., Bohler, C. and Orgel, L. E.: 1998, *Origins Life Evol. Biosphere* **28**, 235.
Joyce, G. F.: 1989, *Nature* **338**, 217.
Kauffman, S.: 1986, *J. Theor. Biol.* **119**, 1.
Lahav, N., White, D. and Chang, S.: 1978, *Science* **201**, 67.
Lahav, N.: 1993, *Origins Life Evol. Biosphere* **23**, 329.
Lahav, N. and Nir, S.: 1997, *Origins Life Evol. Biosphere* **27**, 377.
Lawless, J. G. and Levi, N.: 1979, *J. Mol. Evol.* **13**, 281.
Lazcano, A., Fastag, J., Gariglio, P., Ramirez, C. and Oro, J.: 1988, *J. Mol. Evol.* **27**, 365.
Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. and Ghadri, M. R.: 1996, *Nature* **382**, 525.
Maizels, N. and Weiner, A. M.: 1994, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6729.
Miller, S. L. and Urey, H. C.: 1959, *Science* **130**, 245.
Noller, H. F., Hoffarth, V. and Zimniak, L.: 1992, *Science* **256**, 1416.
Odom, D. G., Lahav, N. and Chang, S.: 1979, *J. Mol. Evol.* **12**, 259.
Ohnishi, K.: 1992, *Nucleic Acids Symp. Ser.* **27**, 145.
Orgel, L. E.: 1998, *Trends Biochem. Sci.* **23**, 491.
Rich, A.: 1962, in *Horizons in Biochemistry*, in Kasha, M. and Pullman (eds.), Academic Press, NY, pp. 103–126.
Saetia, S., Liedl, K. R., Eder, A. H. and Rode, B. M.: 1993, *Origins Life Evol. Biosphere* **23**, 167.
Saks, M. E. and Sampson, J. R.: 1996, *EMBO J.* **15**, 2843.
Sievers, D. and von Kiedrowski, O.: 1994, *Nature* **369**, 221.
Staves, M. P., Bloch, D. P. and Lacey, J. C.: 1988, *Z. Naturforsch.* **43**, 476.
Staves, M. P. and Lacey, J. C.: 1989, *Z. Naturforsch.* **44**, 296.
Trifonov, E. N. and Bettecken, T.: 1997, *Gene* **205**, 1.
Zhang, B. and Cech, T. R.: 1997, *Nature* **390**, 96.