Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 243 (2006) 393–406
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

A molecular model for the origin of protein translation in an RNA world

William R. Taylor

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Received 7 February 2006; received in revised form 9 June 2006; accepted 14 June 2006
Available online 30 June 2006

Abstract

The RNA world hypothesis requires a ribozyme that was an RNA-directed RNA polymerase (ribopolymerase). A model for this, based on the core of the large subunit of the ribosome, is developed further. The geometry of a potential active site for this ribopolymerase suggests that it contained a cavity (now occupied by the aminoacyl-tRNA) and that an amino acid binding in this might have ''poisoned'' the ribopolymerase by cross-reacting with the nucleoside triphosphate before polymerization could occur. Based on a similarity to the active site components of the class-I tRNA synthetase enzymes it is proposed that the amino acid could become attached to the nascent RNA transcript producing a variety of amino-acylated tRNA-like products. Using base-pairing interactions, it is suggested that some of these molecules might cross-link two ribopolymerases giving rise to a precursor of the modern ribosome with two subunits linked by tRNA. A hybrid dimer, half polymerase and half proto-ribosome, could account for mRNA translocation before the advent of protein elongation factors. Some implications for the genetic code are discussed.

Keywords: RNA world; Ribosome; Translation; Genetic code

1. Introduction

The origin of life has been a topic for speculation probably for as long as people have had the ability to wonder. With the ideas of Darwin, it became possible to imagine that the evolution of life might be traced backwards to its simplest beginnings. Darwin himself took this as far as unspecified events in a ''warm little pond'' but it was only with the development of molecular biology that we have been able to speculate on the nature of the molecular events below the surface of his pond.

The central problem in the molecular biology of the origin of life is that the current system has two components: a nucleic acid archive and a largely protein-based mechanism. Like mind and body, the two are so deeply interdependent that it is difficult to see how one could have emerged without the other. Of course, both could have co-evolved from a simpler joint origin and while this cannot be discounted, the possibility that both roles were once embodied in a single molecule offers an alternative, and potentially simpler, solution.

It was suggested in some of the earliest speculations that RNA might be able to fulfil both rôles (Crick, 1968): clearly it can encode genetic information but it was only relatively recently that it has been demonstrated to have the catalytic ability that might let it take on, at least part of, the role of proteins (Cech and Bass, 1986). This allowed the firmer statement that life might once have been based entirely on RNA. The ''RNA world'' hypothesis (Gilbert, 1986) is now widely accepted despite having some considerable remaining difficulties. These predominately occur at either end of the epoch: at the earlier end, the prebiotic synthesis of the required ribonucleosides is problematic (Joyce, 1989), while at the later end it is difficult to imagine how a transition to a protein-based world could have been effected. Central to this latter problem is the origin of translation and the establishment of a genetic code. (For further reviews, see Refs. Smith and Szathmáry, 1995; Gesteland et al., 1999).

E-mail address: wtaylor@nimr.mrc.ac.uk.

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.06.015

1.1. Replication in the RNA world

The replication of RNA in an RNA world, unaided by proteins or DNA, is very likely to have involved the transmission of genetic messages through Watson–Crick (WC) base pairing. Although alternative schemes can be postulated, there is little reason to avoid this mechanism, given that, at some point in time, there must have been a primitive system that was a forerunner of the current universal mechanism used by all known RNA and DNA encodings.

1.1.1. Synthesis of a parallel complement

Transmission of a genetic message by WC pairing with a single-stranded molecule involves the creation of a complementary daughter strand that is the reverse complement of the template strand. However, in a world without any proteins, there would seem to be little to stop this copy hybridizing to form a double-stranded complex (duplex) with its template. Even if this is prevented during transcription by the shielding effect of the replicase, the complementary strand must be sufficiently accessible to be recopied by another replicase which means that it will also be equally accessible to hybridize with any other complementary daughter strand. Since, by definition, both copies must occur with equal frequency in the population of molecules, it would seem to be impossible that such encounters could have been avoided.

It can be argued that the double-stranded molecules will form a "genome" pool from which copies are sometimes teased apart and replicated into functional molecules—just as that which occurs with modern DNA. However, in the RNA world, the replicase itself must also be an RNA molecule: a ribozyme that is an RNA-dependent RNA polymerase (or ribopolymerase for short). This molecule must also be represented in the molecular population along with its reverse complement and will similarly be sequestered into the inactive pool of double-stranded molecules. While such a system may be viable, it does not appear as a very promising starting point for life.

From arguments based on the relative stability of a folded single strand and its antiparallel duplex, it has been suggested that an RNA replicase might have synthesized a complementary strand that ran in the same direction (Taylor, 2005). The same chemistry can be used in this process as with modern protein polymerases with the only difference being that the ribopolymerase would start at the beginning of a template and move towards the end (3′), unlike the protein polymerases that start at the 3′ end. It is known that with DNA, a parallel complementary strand can still hybridize with its complement but would form a parallel duplex (Rippe and Jovin, 1992; Parvathy et al., 2002). Since we know also from DNA that the parallel duplex is less stable than the antiparallel duplex (Germann et al., 1988) this would shift the balance of stability closer to that of the folded single strand. With a smaller difference in stability, it becomes more likely that the duplex will, by chance, become sufficiently accessible for copying under conditions in which the ribopolymerase is functional (Taylor, 2005).

1.1.2. A model for the ribopolymerase

In the course of its function, the ribopolymerase must be able to bind two polynucleic acids (template and transcript) and a nucleoside triphosphate. All of these must be brought together in a site that is sufficiently shielded from the solvent to prevent the wasteful hydrolysis of the triphosphate. This argues for quite a large ribozyme; yet if it were too large, then replication errors would make it difficult for the system to maintain a viable population. A potential model for a ribopolymerase that can accommodate the structural requirements with a minimal sequence length was found in the large subunit (LSU) of the ribosome (Taylor, 2004). The core of the LSU is the site of the peptidyl transferase reaction, linking the amino acids delivered by the binding of the tRNA molecules. This core consists of a continuous sequence of 180 nucleotides with a non-contiguous segment that can provide a model for the template, while the tRNA can be taken as a model for the transcript (Fig. 1).

It was further suggested that this model, when combined with the reversed direction of transcription, would function most efficiently as a dimer in which one copy of the polymerase could pass its transcript directly to the other as a template. This minimizes the exposure of the unfolded single-strand transcript/template and may also help to reduce template/transcript loss from the complex. Taken as a unit, the dimer takes a sequence and generates an identical copy (the complement of a complement) (Fig. 2, see also Fig. 3).

The model of a dimeric ribopolymerase can be further extended as the emerging copies (of both sense) can be directly fed into yet more ribopolymerase molecules. The resulting network would take the form of a quasi-crystal, or sponge-like matrix, possibly growing to macroscopic size.

1.1.3. The transition to proteins

At some point, proteins came and usurped the function of the ribopolymerase. There has been much speculation on how this might have occurred. Some view the peptides as useful helpers (cofactors) that gradually assumed an ever increasing role until now all we have left are a few ribozymes and nucleotide cofactors (Szathmáry, 1999). (See Ref. Smith and Szathmáry, 1995 for a review). An alternative is that the original peptides were not quite so benevolent but coexisted alongside the dominant RNA molecules—much as the mammals existed in the shadow of the dinosaurs—and, like our ancestors, seized their chance when the time was right (Dyson, 1985).

If the ribopolymerase made use of an intermediate parallel complementary strand, then any protein that could act as a primitive polymerase and synthesize an antiparallel strand would have caused some disruption to