ORIGIN OF GENETICALLY ENCODED PROTEIN SYNTHESIS: A MODEL BASED ON SELECTION FOR RNA PEPTIDATION

J. TZE-FEI WONG

Department of Biochemistry, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong

(Received September 17, 1990)

Abstract. The difficulty in explaining the origin of genetic coding centres on the need to identify selective advantages that could account for the synthesis of peptidyl-tRNA, the essential intermediate in genetically programmed translation. It is resolved by a recognition of the functional advantages derivable from the post-transcriptional addition of peptide cofactors to RNA apo-catalysts. This enables the formulation of a theory for the origin of the genetic encoding of protein synthesis by RNA.

Introduction

The genetic code assigns 64 triplet RNA codons to 20 amino acids and termination signals. It enables the translation of genetic information stored in nucleic acids into proteins. For years the relative primacy of proteins and nucleic acids in the origin of life has posed a dilemma. More recently the discovery of catalytic RNA has suggested that the embodying of both coding and catalytic functions in the same RNA macromolecules might enable these molecules to evolve prior to the development of the genetic code and proteins [1-3]. Since the amino acid side chains of proteins are much more proficient in catalysis than the structural components of RNA, the RNA coding of proteins not surprisingly has led to the superseding of RNA catalysts by proteins. However, the puzzling question is, under the circumstances of the RNA world, what could be the nature of the selective steps that gave rise to genetic coding, and in so doing established the cooperation between nucleates and polypeptides that is fundamental to life today?

Evolution acts only on the present, and not on the anticipated future. Accordingly, genetic coding could not emerge based on some distant promise held out by an RNA-encoded, protein-catalyst world that had yet to be constructed. Instead, it had to be accomplished through an unbroken chain of steps, each attended by an immediate selective advantage. Therefore it becomes necessary to enquire into the kind of immediately advantageous steps that, without aiming at genetic coding, nonetheless arrived ultimately at genetic coding. Previously, Orgel [4] proposed that attachment of an amino acid or a dipeptide to the 3'-hydroxyl group of RNA could help mark the site for initiation of transcription. This would provide an incentive to the synthesis of aminoacyl-tRNA and dipeptidyl-tRNA, but not RNA compounds containing longer peptides, or peptides of defined sequences. Indeed, the lability of the O-aminoacyl ester bond is such that the 3'-hydroxyls might not even be the initial sites of useful attachment of amino acids and peptides to RNA.

Therefore the object of this study is to examine the nature of catalytic and other functional advantages of RNA peptidation that could lead to the evolution of elongated peptides of defined sequences on RNA and bring about the modern process of genetic coding.

**Predisposing Factors**

It has been recognized that some of the features of the RNA world might furnish important predisposing factors toward the development of genetic coding. Some bacterial and plant RNA viruses have tRNA-like structures at their 3' ends which might function as an origin of replication and as a telomere. Accordingly tRNA-like structures could provide genomic tags in the RNA world to ensure completeness of RNA replication [5]. Moreover, many enzymes that utilise a nucleotide cofactor contain nucleotide-binding domains that could be derived from RNA-binding domains from precellular times [6]. Specific interactions between RNA and amino acids also have been observed [7, 8]. Bonding of chemical groups including amino acids to a polyanion such as RNA also could help to concentrate these groups on a cationic surface, and promote their participation in abiotic reactions [9].

The preexistence of tRNA-like structures and the specific interactions of RNA with amino acids and polypeptides all would favour a cooperation between the RNA and amino acid-polypeptide systems, and predispose the precellular formation of a genetic code. However, they fall short of defining a plausible chain of selective steps that would give rise to peptidyl-RNAs with defined sequences, and hence RNA encoded translation.

**Post-Translational Modifications**

One way to identify the factors within the RNA world that could confer selective advantages on the evolving RNA catalysts is to examine the factors that have been advantageous to present-day protein catalysts. In this regard, the most sustained source of improvement of protein function is evidently that furnished by an increased variety of amino acid side chains. Analysis of the structure of the genetic code has led to the coevolution theory that the early genetic code coevolved with amino acid biosynthesis. Primitive amino acids that were produced by prebiotic reactions gave rise to novel amino acids, and transferred part or all of their triplet codons to the latter. This accounts for the strong correlation between codon allocation and biosynthetic relationships among the amino acids, as well as the lack of prebiotic-type synthesis for many of the amino acids used in proteins to-day [10–14]. Even after the mainline genetic code was established, numerous amino acid side chains were added to proteins by post-translational modifications. There are more than 100 such additional side chains [15]. These developments attest to the powerful selective advantages constantly to be gained from having more amino acid side chains in the proteins.