FEBS Letters 584 (2010) 334–341

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Development of the genetic code: Insights from a fungal codon reassignment

Gabriela R. Moura, João A. Paredes, Manuel A.S. Santos *

Department of Biology and CESAM, University of Aveiro, 3810-193 Aveiro, Portugal

ARTICLE INFO

Article history:
Received 29 October 2009
Revised 17 November 2009
Accepted 18 November 2009
Available online 24 November 2009

Edited by Michael Ibba

Keywords:
Genetic code
Protein synthesis
tRNA
Mistranslation
Codon reassignment
Codon usage
Evolution
Candida

ABSTRACT

The high conservation of the genetic code and its fundamental role in genome decoding suggest that its evolution is highly restricted or even frozen. However, various prokaryotic and eukaryotic genetic code alterations, several alternative tRNA-dependent amino acid biosynthesis pathways, regulation of tRNA decoding by diverse nucleoside modifications and recent in vivo incorporation of non-natural amino acids into prokaryotic and eukaryotic proteins, show that the code evolves and is surprisingly flexible. The cellular mechanisms and the proteome buffering capacity that support such evolutionary processes remain unclear. Here we explore the hypothesis that codon misreading and reassignment played fundamental roles in the development of the genetic code and we show how a fungal codon reassignment is enlightening its evolution.

1. Introduction

Life is based on the extraordinary capacity of cells to translate the nucleic acids information of their genomes into the amino acids information of their proteomes. The genetic code determines how gene words (codons) are translated into protein words (amino acids), highlighting the fundamental role of 20 aminoacyl-tRNA synthetases (aaRSs) in genome decoding [1]. Each aaRS binds and activates a specific amino acid and transfers it to a cognate tRNA, producing aminoacylated tRNAs (aa-tRNAs) [2,3]. The latter read mRNA codons translating the nucleic acids alphabet into the amino acids alphabet through specific ribosome dependent decoding rules [4]. The genetic code is therefore established by specific attachment of amino acids onto tRNA adaptor molecules by aaRSs and by direct reading of mRNA codons by aa-tRNA anticodons in the ribosome. This suggests that reconstruction of the evolutionary pathways that established the genetic code requires deep structural, biochemical, functional and evolutionary knowledge of aaRSs, tRNAs, mRNAs and of the ribosome. To date, many crystal structures of these molecules have been obtained, and detailed biochemical and biophysical characterization of the tRNA aminoacylation and decoding reactions [2,5–7], as well as large scale phylogenetic analysis of the various components of the genetic code have been carried out [8]. Despite these extraordinary advances, the evolution of the genetic code remains an open biological question.

The Frozen Accident Theory proposed by Crick in 1968 postulates that the code is immutable because any alteration to it would be lethal or highly detrimental to life [9]. However, a number of genetic code alterations discovered over the last 40 years indicate that the code has intrinsic flexibility and can evolve (reviewed in [10,11]). We discuss below how these genetic code alterations are enlightening the evolution of the genetic code and we raise the hypothesis that codon reassignment processes played an important role in the code development. The origin of the genetic code, i.e., the origin of tRNAs, aaRSs, the ribosome and the mechanisms of incorporation of the first 10 prebiotic amino acids into the code, which mediated the transition of life from the RNA to the protein worlds, are beyond the scope of this review and will not be addressed. We mention briefly the main theories that have been proposed to explain the origin of the genetic code in order to provide an integrated view of the code evolution.

2. Origin and early evolution of the genetic code

There are three main theories to explain the origin and structure of the genetic code, namely: (i) the Stereochemical Theory, (ii) the Adaptive Theory and (iii) the Coevolution Theory (reviewed in [12]). The Stereochemical Theory posits that codon and amino acid assignments were determined by physicochemical affinities

* Corresponding author. Fax: +351 234426408.
E-mail address: msantos@ua.pt (M.A.S. Santos).

0014-5793/$36.00
doi:10.1016/j.febslet.2009.11.066

between amino acids and nucleic acids [13,14]. This theory is supported by experimental data arising from selection-amplification of small RNAs (SELEX) which show that at least 8 of the 20 natural amino acids select RNA sequences enriched in cognate codon or anticodon binding motifs [15,16]. Indeed, RNA aptamers selected in the presence of Trp contained Trp CCA anticodons while small RNAs selected in the presence of Ile were enriched in Ile UAU anticodons [17–19], however the statistical significance and the strength of the associations between RNA aptamers and amino acids has been questioned and the Stereochemical Theory requires further validation [20].

The Adaptive Theory postulates that the evolution of the genetic code is mainly driven by the selective forces that minimize the effects of protein synthesis errors, being them from mutational origin or from mRNA misreading [21,22]. The observation that amino acids with similar chemical properties are assigned to similar codons plus statistical and computational evidence for a strong bias towards error minimization pressure in the code provide important support for this theory [12,23,24].

The Coevolution Theory postulates that the structure of the genetic code reflects directly the evolution of amino acid biosynthetic pathways [25]. This theory assumes that the number of amino acids that existed in the prebiotic earth was small (10 or so) and that the other amino acids of the genetic code were derived from the prebiotic ones through biosynthetic processes. The theory is supported by the identification of precursor-product pairs of amino acids and by the discovery of tRNA-dependent biosynthesis of Gln, Asn, Cys and Sec in various prokaryotes and eukaryotes (see below) [26].

The evolutionary scenarios described above, in particular the one proposed by the Coevolution Theory, suggest the existence of three critical moments (steps) in the development of the genetic code (Fig. 1A). An initial step (Phase-1) characterized by the incorporation of the prebiotic amino acids Gly, Ala, Ser, Asp, Glu, Val, Leu, Ile, Pro and Thr. An intermediate step (Phase-2) involving the incorporation of 7 additional amino acids derived from the prebiotic ones through biosynthetic means, namely Phe, Tyr, Arg, His, Trp, Lys and Met. And, a final step (Phase-3) where the five amino acids whose synthesis is tRNA-dependent or is mediated through non-canonical biosynthetic pathways, namely Asn, Gln, Cys, selenocysteine (Sec) and pyrrolysine (Pyl), were incorporated into the genetic code [12,25,26].

We discuss below the mechanistic and structural implications of this stratified evolution of the genetic code under the assumption of the following postulates for the Phase-1 of the code development:

1. The triplet nucleotide nature of codons and the translational machinery were largely established during the incorporation of the first 10 prebiotic amino acids into the genetic code.
2. The basic structure of the tRNA molecule and the codon–anticodon decoding principles were defined.
3. An essential proteome was synthesized with the 10 prebiotic amino acids.

The simultaneous existence of only 10 prebiotic amino acids and 64 codons suggests that some codons were initially unassigned (did not code for any amino acid) or that the 10 prebiotic amino acids were assigned to more than one codon family box (Fig. 1B), as is the case for Leu, Ser and Arg, in extant organisms (Fig. 1D). Indeed, Leu is still encoded by the CUN (N = any nucleotide) codon family plus the UUA/G codons of the UUN codon family box (Fig. 1C and D). The other two codons of the UUN codon family box (UUU/C codons) encode Phe, which was incorporated late into the genetic code [26]. Therefore, during Phase-1 of the code development Leu must have been assigned to both the CUN and the UUN

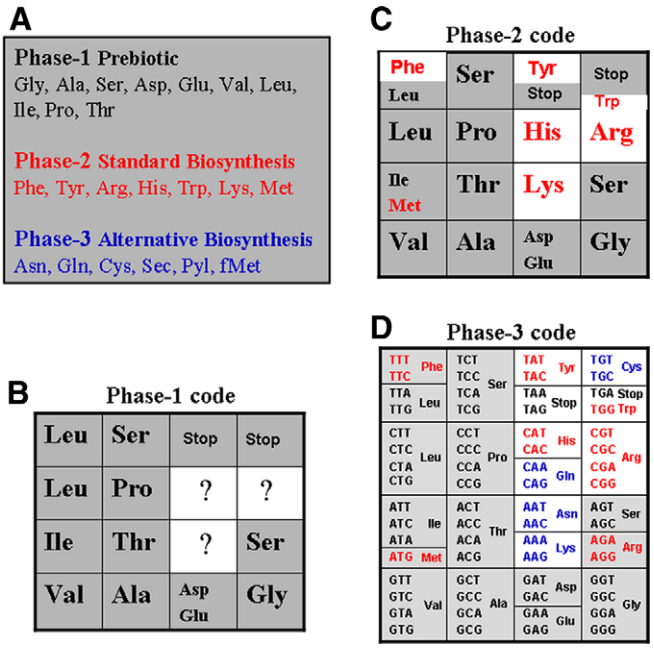

Fig. 1. Scheme outlining the putative evolution of the genetic code. The tables highlight the gradual incorporation of amino acids into the genetic code, according to the Coevolution Theory. (A) The development of the genetic code into three phases follows the evolution of amino acids biosynthetic processes and highlights the requirement of codon reassignments to accommodate new amino acids into the code, beyond the 10 prebiotic ones. (B and C) The tables show the incorporation of the Phase-2 and 3 amino acids into the primordial genetic code. The distribution of codons in the Phase-1 and 2 tables follow that indicated in the Phase-3 table. (D) Genetic code of most extant organisms. The boxes in white colour highlight codon boxes where incorporation of the indicated amino acids involved capture and reassignment of codons. The UUC/U codons were reassigned from Leu to Phe, the AUG codon was reassigned from Ile to Met, the UAU/C codons were reassigned from stop to Tyr and the UGG codon was reassigned from stop to Trp. The full set of codon reassignments was completed with the incorporation of Phase-3 amino acids into the code.

codon family boxes (Fig. 1B). Phe addition to the code required a new (mutant) $tRNA^{Phe}$ to capture the UUU/C codons from Leu. Complete reassignment of these codons to Phe required the loss of the ancestral $tRNA^{Leu}$ that decoded them (Fig. 1C). The same principle of codon capture followed by reassignment can be applied to the incorporation of the other Phase-2 amino acids (Fig. 1C and D). An alternative explanation would be that UUU/UUC, as well as the other codons of split codon families, were initially unassigned and that their late assignment to new amino acids escaped reassignment from one amino acid to another. However, tRNAs with U at the wobble position are able to decode the four codons of codon family boxes and it is likely that these rather than more sophisticated tRNAs bearing nucleoside modifications or expanded sets of tRNA isoacceptors were originally used to decode the 61 sense codons of the genetic code. Furthermore, the pairs of codons of split codon family boxes end with a purine or a pyrimidine and consequently cannot be unassigned simultaneously by genome G + C pressure alone. Therefore, it is unlikely that codon unassignment played a relevant role in the early amino acid assignments.

The Phase-3 amino acids (Asn, Gln, Cys, Sec, Pyl, fMet) are particularly interesting because their alternative biosynthesis suggests that they were incorporated rather late into the genetic code [27,28]. In various bacterial and archaeal species, Asn is still synthesized on a $tRNA^{Asn}$ which is charged with Asp by a non-discriminating AspRS, generating a mischarged Asp-$tRNA^{Asn}$ [29]. A similar mechanism is used in archaea, in most bacteria and in chloroplasts for the synthesis of Gln. In this case, a $tRNA^{Gln}$ is charged