Available online at www.sciencedirect.com
ScienceDirect
Gene 396 (2007) 75–83
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Analysis of tRNA abstract shapes of precursor/derivative amino acids in Archaea

Hernán A. Burbano [1], Eugenio Andrade *

Grupo de Biología Molecular Teórica y Evolutiva, Universidad Nacional de Colombia, Bogotá, D.C., Colombia

Received 14 August 2006; received in revised form 27 February 2007; accepted 27 February 2007
Available online 15 March 2007
Received by: M. Di Giulio

Abstract

Wong's theory of the genetic code's origin states that because of historical constraints, codon assignment depends on the relation between precursor and derivative amino acids, a result of the coevolutionary process between amino acids' biosynthetic pathways and tRNAs. Based on arguments supporting the assumption that natural selection favors more stable and thus functionally constrained structures, we tested whether precursor and derivative tRNAs are equally evolved by measuring their structural parameters, thermostability and molecular plasticity. We also estimated the extent to which precursor and derivative tRNAs differ within Archaea.

We used Archaea sequences of both precursor and derivative tRNAs in order to examine the plastic repertoires or sets of suboptimal structures at a defined free energy interval. We grouped secondary structures according to their helix nesting and adjacency using abstract shapes analysis. This clustering enabled us to infer a consensus sequence for all shapes that fit the clover leaf secondary structure [Giegerich, R., et al., Nucleic Acids Res 2004; 32 (16): 4843–51.]. This consensus sequence was then folded in order to retrieve a set of suboptimal structures. For each pair of precursor and derivative tRNAs, we compared these plastic repertoires based on the number of secondary structures, the thermostability of the minimum free energy structure and two structural parameters (base pair propensity (*P*) and mean length of helical stem structures (*S*)), which were measured for every representative secondary structure [Schultes, E.A., et al., J Mol Evol 1999; 49 (1): 76–83.]. We found that derivative tRNAs have fewer numbers of shapes, higher thermostability and more stable parameters than precursor tRNAs, a fact in full agreement with Wong's coevolution theory of the genetic code.

Keywords: Genetic code; RNA plasticity; Wong's coevolution theory; RNA evolution; Plastic map

1. Introduction

There are two divergent theories about the origin of the genetic code, the physicochemical and the historical (Di Giulio, 2005a). The physicochemical theory states that the code was shaped by stereochemical (Crick, 1968) and physicochemical (Sonneborn, 1965; Woese et al., 1966) interactions between codons, anticodons and/or amino acids, thereby leading to a minimization of physicochemical distances between the amino acids encoded by codons that differ in one base.

The historical theory is best represented by Wong's coevolution theory (CET) (Wong, 1975), which affirms that the genetic code was produced as a result of an expansion of a primordial one, by recruiting novel amino acids biosynthesized as metabolic derivatives of their precursors. Since the chemical modifications occurred while the precursor was still attached to tRNA (pretranslational mechanism or *pretrans*) (Wong, 1975, 1981, 1988), the assignment of codons to novel amino acids implies the selection of paralogs or duplicated tRNAs. These paralogs were selected because they exhibited a higher affinity for the derivative amino acids, which later became members of

Abbreviations: tRNA, transfer RNA; CET, coevolution theory; *P*-tRNA, precursor tRNA; *D*-tRNA, derivative tRNA; mfe, minimum free energy; LUCA, last universal common ancestor.

* Corresponding author. Departamento de Biología, Universidad Nacional de Colombia, Bogotá, Ciudad Universitaria Bogotá, Colombia. Tel.: +57 1 3165000x11313.
E-mail address: leandradep@unal.edu.co (E. Andrade).
[1] Present address: Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany.

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.02.024

the same biosynthetic family (Xue et al., 2003). Consequently, the assignment of codons to precursor tRNAs (*P*-tRNAs) historically constrained the assignment of newly produced amino acids to derivative tRNAs (*D*-tRNAs).

Sequence analysis (Di Giulio, 1995; Xue et al., 2003) has served to test the correlations between *P*-tRNA and *D*-tRNA. Nevertheless, very few studies have attempted to verify whether tRNA structural features support these correlations (Bermudez et al., 1999). Although Wuchty et al. (1999) and Wuchty (2003) studied the plasticity of these molecules, no one has studied how plasticity correlates with the CET of the genetic code. We raised the following questions: Are *D*-tRNAs more evolved molecules than *P*-tRNAs? If so, can we detect structural differences by measuring the structural parameters, plasticity and thermostability of sets of suboptimal structures?

We expect to detect structural differences between *P*-tRNA and *D*-tRNA similar to a footprint of the genetic code's origin. This footprint was created by two opposing processes: 1. The genetic code's evolution through assigning codons and recruiting duplicated tRNA genes caused paralogs to have insufficient time to accumulate differences. Therefore, these differences must be very small. 2. Since Archaea were selected in environments that best represent conditions close to the origin of life, it should have preserved this footprint. Accordingly, Xue et al. (2003) showed how Archaea reveals an ancestral condition, because both the mutation rate of tRNA genes and gene duplication events show a lower rate when they were compared with Bacteria and Eukarya. Therefore, we are concerned with the prominence of this genetic code footprint, and its presence in the structural differences between *P*-tRNA and *D*-tRNA within Archaea.

Based on Schultes et al.'s (1999) findings from comparing random molecules with evolved molecules, we expect, in general, that the differences between *P*-tRNA and *D*-tRNA might best be revealed through changes in structural parameters, however small these may be. Schultes et al. (1999) detected small differences between random molecules and evolved molecules. Nevertheless, we expect that with naturally evolved molecules, the differences in Archaea may be detectable and show footprints of the evolutionary process that gave rise to the extant genetic code.

RNA folding can be studied with algorithms that use thermodynamic parameters (Hofacker, 2003a) derived from experimental measures (Mathews et al., 1999). These algorithms work at the level of the secondary structure, which functions as the scaffolding for the tertiary structure (Fontana and Schuster, 1998). There are two approaches that map RNA genotype and phenotype. The first retrieves a simple map that yields only the optimal folding of an RNA sequence (Ancel and Fontana, 2000). However, the functional native structure is not always the optimal structure, though the energy of the native structure should be close to it (Giegerich et al., 2004). The second approach yields the plastic map or the repertoire of suboptimal structures. This approach is recommended, because it takes into account the minimum free energy structure and its suboptimal counterparts (Ancel and Fontana, 2000; Fontana, 2002).

Plastic repertoires are obtained using an algorithm that calculates all secondary structures within a defined energy range above the minimum free energy (Wuchty et al., 1999; Wuchty, 2003; Ancel and Fontana, 2000; Meyers et al., 2004). It is, however, difficult to deal with suboptimal structures, because their number rises exponentially with each additional increment of both length and energy range. Most secondary structures generated are nearly identical, and usually only structures that show fundamental differences are of interest (Giegerich et al., 2004). The abstract shape concept represents a class of similar structures sharing a common pattern of helix nesting and adjacency, and a useful tool for studying plastic repertoires is derived from it. These shapes are abstract in the sense that not all structural features are considered, so that many structures (in Vienna's dot-bracket representation) map to one abstract shape. The abstract shape concept serves as the basis for a useful tool when studying plastic repertoires. The advantage of this method is that one abstract shape represents a family of similar or suboptimal RNA structures (Fig. 1) consequently, the set of all available shapes provides a complete view of the folding space.

Both the folding space (the set of secondary structures) and the shape space (the set of abstract shapes) grow exponentially with energy range increments for fixed sequence length. However, the shape space grows more slowly than the folding space, which makes abstract shape analysis preferable (Giegerich et al., 2004).

The RNAshapes package (Steffen et al., 2006) combines different applications (Giegerich et al., 2004; Reeder and Giegerich, 2005; Voss et al., 2006; Reeder et al., 2006). Among them is one that identifies a consensus structure–and therefore a consensus sequence–in a way that eliminates the requirement of a multiple sequence alignment step (Reeder and Giegerich, 2005). It is especially useful for sequences with low conservation, where methodologies based on sequence conservation cannot be employed (Steffen et al., 2005). However, this consideration can be applied to all kinds of sequences, since the probability of obtaining gaps in a consensus sequence is high, and RNA folding algorithms cannot proceed if the input sequence contains gaps. It is assumed that structure is conserved more than sequence for equally functional homologous RNA molecules, because function depends more on structure than on sequence (Reeder et al., 2006). For a family of RNA sequences, the method, which is computationally fast, independently generates the near-optimal abstract shape space, and then predicts a consensus abstract shape common to all sequences (Reeder and Giegerich, 2005). In this paper we identify consensus amino acid specific tRNA sequences based on shape comparisons, and then generate their corresponding plastic repertoires. Fig. 2 shows an outline of our theoretical framework.

2. Methods

2.1. Data base

We used 503 tRNA sequences from 22 species of Archaea (Table 1) reported at GeneBank (Benson et al., 2005). We