BioSystems 141 (2016) 20–30

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Codon bias from minimization of codon–anticodon interaction

A. Sciarrino[a], P. Sorba[b,*]

[a] I.N.F.N., Sezione di Napoli, Complesso Universitario di Monte S. Angelo, Via Cinthia, I-80126 Napoli, Italy
[b] Laboratoire d'Annecy-le-Vieux de Physique Théorique LAPTH, CNRS, UMR 5108, associée à l'Université de Savoie, BP 110, F-74941 Annecy-le-Vieux Cedex, France

ARTICLE INFO

Article history:
Received 20 June 2015
Received in revised form
20 November 2015
Accepted 23 November 2015
Available online 14 December 2015

Keywords:
Genetic code
Crystal Basis Model
Codon
Anticodon
Codon bias
Evolution

ABSTRACT

Inequalities between codon usage probabilities for quartets of codons are derived using a minimum principle for codon–anticodon interaction and a probability sum rule in the framework of the Crystal Basis Model of the genetic code. Performing this study separately for the Early and for the Eukaryotic Genetic Code, we observe a consistency in the obtained results for the two codes as well as a good agreement with data from Kasuza Codon Usage Database. Then we analyze the coherent changes of sign occurring, in the evolution from the Early to the Eukaryotic code, in the two parameters regulating the codon–anticodon interaction potential.

1. Introduction

It is well known that the genetic code is degenerate in the sense that 64 codons encode for 20 amino-acids (a.a.), so almost all the a.a. are encoded by multiple codons (called synonymous codons). Degeneracy is found primarily in the third position of the codon, i.e. the nucleotide in the third position can change without changing the corresponding a.a. When the genetic code was discovered, a uniform codon usage inside the multiplets was assumed, but in the seventies-eighties it was realized that this assumption was not true. With the availability of more data, it was confirmed that some codons are used much more frequently than others to encode a particular amino-acid, i.e. there is a "codon usage bias". The non-uniform usage of synonymous codons is a widespread phenomenon and it is experimentally observed that the pattern of codon usage varies between species.

The codon usage bias is related to many biological processes, such as gene expression level, gene length and function, and tissue or organ specificity, as well as evolution. No clear indication comes out for the existence of one or more factors which universally engender the codon bias, on the contrary the role of some factors is controversial. The main reasons for the codon usage bias are believed to be, for a general discussion see (Karlin and Mrazek, 1996; Sharp et al., 2010) and for a review (Salim and Cavalcanti, 2008; Hershberg and Petrov, 2008; Plotkin and Kudla, 2011), the mRNA stability and its secondary structure, the mutational pressure, the translation efficiency, the natural selection, the DNA methylation, the genetic coding error minimization, the CG content, the abundance of specific anticodons in the tRNA and the environmental effects.

The aim of this paper is to discuss possible effects of the codon–anticodon interactions on the codon bias in the framework of the Crystal Basis Model, according to the approach introduced in (Sciarrino and Sorba, 2012), and to propose semi-quantitative predictions of the codon bias. Moreover we briefly analyze the codon usage bias variation along the evolution of the genetic code on the basis of the model developed in (Sciarrino and Sorba, 2013). Let us make precise that, in the following, we will be concerned about amino acids encoded by quartets. For the ones encoded by a sextet, that we consider as the sum of a quartet and a doublet, only the quartet will be considered. The method we developed is essentially based on the determination of the minimum values of an operator which can be seen as an interaction

* Corresponding author.
E-mail addresses: sciarrino@na.infn.it (A. Sciarrino), sorba@lapth.cnrs.fr (P. Sorba).

http://dx.doi.org/10.1016/j.biosystems.2015.11.006
0303-2647/

potential between a codon and its corresponding anticodon. A possible general pattern of the bias is searched by deriving inequalities for the codon usage probabilities.

The paper is organized as it follows. In Section 2 the main features of the minimum principle approach are summarized. In Section 3 and in Section 4 we apply the method of the minimization of the codon–anticodon interaction, respectively, to the pattern of codon–anticodon in the Early and in the Eukaryotic Genetic Code. Then, particularly for the latter, we compare and discuss in some detail in Section 5 our theoretically expected inequalities of the codon usage probabilities with the codon usage frequencies observed in suitable specimens of biological species. Finally in Section 6 we examine, through the variation of the interaction potential parameters, the scenario of the evolution from the Early to the Eukaryotic Genetic Code.

2. Tools and assumptions

Hereafter we recall briefly the main points of the minimum principle approach for codon–anticodon interaction elaborated in the framework of the so-called Crystal Basis Model. Such a method has been rather successfully used first to recognize the minimum set of anticodons allowing the translational transcription in the mitochondrial code for animals (Sciarrino and Sorba, 2012). Then it has appeared well adapted to analyze and reconcile in a unique path the different steps respectively denoted Ancient, Archetypal and Early genetic codes which are proposed and generally accepted for the evolution of the genetic code (Sciarrino and Sorba, 2013).

Let us first remind that in the framework of the Crystal Basis Model, nucleotides and codons appear as mathematical objects (Frappat et al., 1998). The four nucleotides are assigned to the fundamental representation of the quantum group $\mathcal{U}_q(SU(2)\oplus SU(2))$ in the limit of the parameter of deformation q going to 0. Then each codon, as a composite state of three nucleotides, belongs to an irreducible representation of the above introduced quantum group, obtained by performing the tensor product of three nucleotide representations. The position of each nucleotide in the fundamental representation (1/2, 1/2) of $SU(2)\oplus SU(2)$ is chosen in order to distinguish between the purine bases A and G and the pyrimidine C and U (first $SU(2)$ denoted $SU_H(2)$); to distinguish between the bases C and G related by three hydrogen bonds with respect to those related by two bonds, i.e. U and A (second $SU(2)$ denoted $SU_V(2)$). In this context, the eigenvalues of the $J_{H,3}$ and $J_{V,3}$ generators stand as follows:

$$\mathrm{C}\equiv\left(+\frac{1}{2},+\frac{1}{2}\right)\quad \mathrm{U}\equiv\left(-\frac{1}{2},+\frac{1}{2}\right)\quad \mathrm{G}\equiv\left(+\frac{1}{2},-\frac{1}{2}\right)\quad \mathrm{A}\equiv\left(-\frac{1}{2},-\frac{1}{2}\right) \tag{1}$$

We point out that the choice of a quantum group in the limit of the deformation parameter $q\to 0$ instead of an ordinary Lie algebra, is essential in order to obtain, in the tensor product, each codon as a unique ordered triplet of three nucleotides and not as a linear combination of triplets made by three nucleotides appearing in different orders. First applications of this model, among them the determination of codon usage probabilities correlations, can be found in (Frappat et al., 2001).

In order to modelize the codon–anticodon interaction, we construct, from the generators of the above introduced quantum group, an operator $\mathcal{T}$, acting on each codon–anticodon (c–a) pair and defined as follows:

$$\mathcal{T}=8c_H\vec{J}_H^c\cdot\vec{J}_H^a+8c_V\vec{J}_V^c\cdot\vec{J}_V^a \tag{2}$$

where:

- $c_H\,.\,c_V$ are parameters depending on the encoded amino-acid and of the biological species, the factor 8 being present for numerical convenience
- the three components of the vector operators J_H^c, J_V^c (resp. J_H^a, J_V^a) being the generators of $U_{q\to 0}(su(2)_H\oplus su(2)_V)$ in the representations of the codon and anticodon under consideration
- the scalar product $\vec{J}_\alpha^c\cdot\vec{J}_\alpha^a$ ($\alpha=H,V$) can be computed from:

$$\vec{J}_\alpha^c\cdot\vec{J}_\alpha^a=\frac{1}{2}\left\{\left(\vec{J}_\alpha^c\oplus\vec{J}_\alpha^a\right)^2-(\vec{J}_\alpha^c)^2-(\vec{J}_\alpha^a)^2\right\} \tag{3}$$

where $\vec{J}_\alpha^c\oplus\vec{J}_\alpha^a\equiv\vec{J}_\alpha^T$ stands for the irreducible representation containing the state of the codon–anticodon, the tensor product being performed (see (Frappat et al., 1998) for the rule) choosing the codon as first vector and the anticodon as second vector. Finally, $\vec{J}_\alpha^{\,2}$ should be read as the Casimir operator.

Let us make precise that to the codon XYN[1] (N denoting the nucleotides C,G,U and A) will be associated with the anticodon $N'Y''X''$ with nucleotides N', Y'', X'' associated to N,Y and X respectively. Let us make precise that we write both codons and anticodons in the 5' to 3' direction, and since an anticodon is antiparallel to a codon, its the first (respectively third) anticodon nucleotide will then be paired to the third (respectively first) codon nucleotide. In the following, we will be lead to consider the standard pairing (i.e. Watson-Crick pattern C-G, U-A) for the pairs X-X'' and Y-Y'' while N' may pair to more than one nucleotide (i.e., Crick wobble hypothesis (Crick, 1966)).

In the following, we denote: $<N'Y''X''|\mathcal{T}|XYN>$ the $\mathcal{T}$ eigenvalue with $\vec{J}_\alpha^c$ acting on XYN and $\vec{J}_\alpha^a$ acting on $N'Y''X''$. The main idea of the minimum principle we exploited in (Sciarrino and Sorba, 2012) and (Sciarrino and Sorba, 2013) (see also (Sciarrino and Sorba, 2014) for a review) is, given an amino acid encoded by codons XYN, N varying in the degeneracy quartet, to select among the possible associated anticodons $N'Y''X''$, N' varying, the one minimizing the expression $<N'Y''X''\mathcal{T}|XYN>$, weighted by the frequencies P_{XYN}, (see below) of the XYN codon usage for the amino-acid under study, as illustrated in the next section.

[1] As we consider only quartets $XY\in\{CC, UC, GC, AC, CU, GU, CG\,.\,GG\}$.