Available online at www.sciencedirect.com

SCIENCE DIRECT®

Biochemical and Biophysical Research Communications 324 (2004) 1–2

BBRC

www.elsevier.com/locate/ybbrc

Breakthroughs and Views

How reliable re-adjustment is: correspondence regarding A. Fuglsang, “The ‘effective number of codons’ revisited”

Sayed-Amir Marashi*, Hamed Shateri Najafabadi

Department of Biotechnology, Faculty of Science, University of Tehran, Enghelab Ave., Tehran, Iran

Received 22 August 2004
Available online 17 September 2004

Abstract

A. Fuglsang [Biochem. Biophys. Res. Commun. 317 (2004) 957–964] suggested that effective number of codons for individual amino acids (*Nc*-values) should be re-adjusted to the number of synonymous codons of those amino acids, in order to prevent the overestimation of the effective number of codons. Here, it is shown that re-adjustment at the level of individual amino acids results in loss of considerable amounts of information. Furthermore, we have shown that theoretical *Nc*-values are functions of GC3s (and GC1s); as a result, when an amino acid *Nc*-value exceeds the related theoretical *Nc*-value, the implication of re-adjustment depends on the GC composition of the gene.

Keywords: Effective number of codons; *Nc*; Re-adjustment; Codon usage bias

In a paper pertinent to March 2004 by Fuglsang [1], an altered form of ‘effective number of codons’ [3] as a measure for codon bias is established, declared as

$$\hat{N}c^{*} = \hat{N}c_{\mathrm{Ala}} + \hat{N}c_{\mathrm{Arg}} + \cdots + \hat{N}c_{\mathrm{Val}}, \qquad (1)$$

where each of the individual values represents the effective number of codons for the related amino acid, calculated according to Wright [3], and where each individual *Nc*-value is re-adjusted if it exceeds the number of synonymous codons of the related amino acid. $\hat{N}c^{*}$ was proposed as an alternative to $\hat{N}c$ since it does not overestimate the effective number of codons, as $\hat{N}c$ does.

We would like to draw attention to the consequence of application of $\hat{N}c^{*}$ in the calculation of the effective number of codons for *Escherichia coli* K12 genes. As shown in Table 1, when individual *Nc*-values are calculated for 4390 *E. coli* genes (GenBank Accession No. NC000913), in so many cases re-adjustment should be applied before computing $\hat{N}c^{*}$, showing that this is a more common phenomenon than an exception, like what Fuglsang [1] exemplified with *lrp* gene. Furthermore, if we assume that the data presented in Table 1 can be generalized to other organisms with intermediate GC (which is not an implausible assumption) then, using average relative frequencies of occurrence of amino acids among genomes of different organisms [2], it can be simply calculated that in about 25% of cases, when an individual *Nc*-value is calculated, re-adjustment is needed. This is more likely to lose a lot of information rather than correcting the previous method of calculation.

In addition, this method shows more need to be revised when the effect of re-adjustment of individual *Nc*-values of amino acids on the position of a gene with respect to the plot of $\hat{N}c$ vs. GC3s under H_0 of no selection [3] and A = U, G = C is considered. *Nc*-plots for individual amino acids can be derived using Eqs. (1) and (2) in [1] for large *n*’s, as

[*] Corresponding author. Fax: +98 21 6491622.
E-mail address: marashie@khayam.ut.ac.ir (S.-A. Marashi).

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.08.213

Table 1
The percents of *E. coli* genes in which individual *Nc*-values need re-adjustment

Amino acid	Percent of *E. coli* genes which need re-adjustment
Ala	26.59
Arg	7.04
Asn	47.77
Asp	40.34
Cys	58.74
Gln	39.89
Glu	29.36
Gly	19.77
His	51.55
Ile	9.40
Leu	5.23
Lys	18.75
Met	—
Phe	45.83
Pro	21.56
Ser	30.96
Thr	21.66
Trp	—
Tyr	51.76
Val	27.31

$$\hat{N}c\ (\mathrm{aa}) = \frac{\left(\sum n_i\right)^2}{\sum n_i^2}, \tag{2}$$

where n_i's are the actual usage of synonyms of amino acid (aa). Equations listed in Table 2 have resulted from Eq. (2) under H_0 of no selection and A = U, G = C. It should be mentioned that Arg and Leu each have six codons which can be divided into two groups, four codons beginning with C and two codons beginning with A/U. Therefore, as Table 2 shows, their individual *Nc*-values depend on both base compositions at first and third codon positions ($r =$ GC1s and $s =$ GC3s, respectively). Note that the combination of equations listed in Table 2 with Eq. (3) in [1], assuming a linear relationship between r and s, as seen in *E. coli* (data not presented), results in a bell-shaped theoretical $\hat{N}c$ vs. GC3s plot to which the approximation used in [3] shows an acceptable proximity.

Table 2
Theoretical *Nc*-values for individual amino acids as functions of r and s, which represent GC1s and GC3s, respectively, under H_0 of no selection and A = U, G = C

Amino acid	Theoretical *Nc*-value
SF type 2	$1/(s^2+(1-s)^2)$
Ile	$(2-s)^2/(2(1-s)^2+s^2)$
SF type 4	$2/(s^2+(1-s)^2)$
Ser	$3/(s^2+(1-s)^2)$
Arg, Leu	$(1+r)^2/([s^2+(1-s)^2][2r^2+(1-r)^2])$

SF type *i* is the abbreviation for Synonymous Family type *i*, which stands for the group of amino acids having a degeneracy of *i* [3].

Consider two different *E. coli* genes, *cdsA* and *yegG*, with GC3s of about 0.5 and 0.9, respectively. Individual *Nc*-value of amino acid lysine exceeds the number of synonymous codons in both genes. Therefore, both genes locate above the *Nc*-plot of lysine under H_0 of no selection. However, after re-adjustment, *cdsA* locates on the *Nc*-plot, while *yegG* still locates above the *Nc*-plot. This example simply shows that the re-adjustment has different implications in different GC3s and results in false subsuming when comparison of genes with different GC3s is considered. This methodological problem exists in re-adjustment of $\hat{N}c$ as well as individual amino acid *Nc*-values.

References

[1] A. Fuglsang, The 'effective number of codons' revisited, Biochem. Biophys. Res. Commun. 317 (2004) 957–964.

[2] D. Gilis, S. Massar, N.J. Cerf, M. Rooman, Optimality of the genetic code with respect to protein stability and amino acid frequencies, Genome Biol. 2 (2001) 49.1–49.12.

[3] F. Wright, The 'effective number of codons' used in a gene, Gene 87 (1990) 23–29.