Synonymous Codon Usage Based on Codon-Anticodon Interaction Energy and Gene Expression Level

Mamoru Kato [1] **Akira Suyama** [2]

mkato@genzo.c.u-tokyo.ac.jp suyama@dna.c.u-tokyo.ac.jp

[1] Department of Physics, Graduate School of Science, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan

[2] Department of Life Science, Graduate School of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-Ku, Tokyo 153, Japan

1 Introduction

In unicellular organisms, synonymous codon bias is correlated with the level of gene expression [1]. The codon bias becomes stronger in genes with higher expression levels. It becomes weaker but still remains in genes with lower expression levels. The bias means that synonymous codons are not used at random. In highly expressed genes, the synonymous codon usage is mainly determined by an abundance of iso-accepting tRNAs [2]. In weakly expressed genes, on the other hand, constraints imposed by base composition mainly affect the codon usage.

We have hypothesized that independently of the iso-accepting tRNA abundance and base composition, codon-anticodon interaction energy affects the synonymous codon usage. To verify this hypothesis, we investigated 1664 function-known genes in *E. coli* genome and 2575 function-known genes in *S. cerevisiae* genome.

2 Methods

2.1 Classification of synonymous codons

We classified synonymous codons according to their codon-anticodon interaction energy. A synonymous codon with both C/G and A/U bases is defined as a "moderate" codon, and with either C/G or A/U bases as a "extreme" codon. Moderate and extreme codons should promote the codon-anticodon interaction with moderate and extreme strength, respectively. We confirmed the codon-anticodon stability by calculating the free energy changes with the nearest neighbor model [3].

2.2 The Zo score

To statistically evaluate synonymous codon usage affected by only codon-anticodon interaction energy, Zo score is defined as

$$Z_o = \{D_o - E(D_r)\}/\sigma(D_r)$$

D_o is the difference between the number of moderate codons and that of extreme codons in an observed sequence. $E(D_r)$ and $\sigma(D_r)$ are the expected number and the standard deviation of the difference in random sequences, respectively. For each gene, 3000 random sequences were generated under the following constraints: (a) they have the same amino acid sequence and the same frequency of occurrence of optimal codons as the observed sequence; (b) they have the same base composition as that of the genome sequence. The constraint (a) leads to the exclusion of synonymous codon usage affected by an amino acid sequence and iso-accepting tRNA content. The constraint (b) eliminate the GC content contribution to synonymous codon usage.

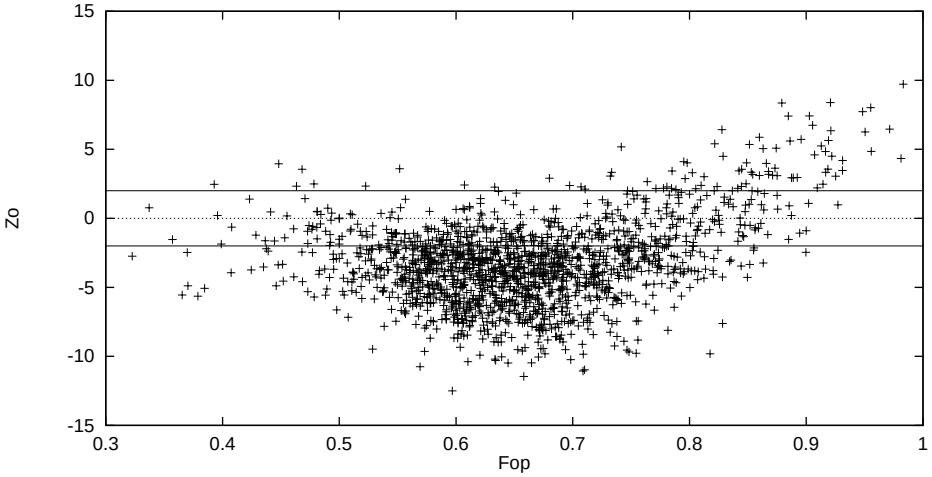

Figure 1: Correlation between F_{op} and Z_o for 1664 genes in *E. coli*.

3 Results and Discussion

The statistical analysis proved that our hypothesis was correct. In Fig. 1, the Z_o score for each of 1664 genes in *E. coli* is plotted against the frequency, F_{op}, of occurrence of optimal codons. F_{op} is positively correlated with the gene expression level in *E. coli* [2]. Therefore, the positive correlation between Z_o and F_{op} results in the positive correlation between Z_o and the level of gene expression. It is thus concluded that in highly expressed genes, moderate codons are significantly preferred to extreme codons; in weakly expressed genes, extreme codons are significantly preferred to moderate codons. The same tendency was confirmed for 2575 genes in *S. cerevisiae*.

The codon preference is based on molecular mechanism of translational process. In highly expressed genes, the preference of moderate codons would lead to binding or being released of tRNAs with moderate strength. As a result, translational process is flowing smooth. In weakly expressed genes, the preference of extreme codons would lead to binding or being released of tRNAs with extremely strong or weak strength. When with extremely strong strength, tRNAs are easily binding but not easily released. When with extremely weak strength, tRNAs are easily released but not easily binding. As a result, translational process is flowing unsteady. The molecular mechanism is universal among organisms ranging from prokaryotes to eukaryotes.

References

[1] Sharp, P.M. and Li, W.-H., The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.*, 15:1281–1295, 1987.

[2] Ikemura, T. and Ozeki, H., Codon usage and transfer RNA Contents: organism-specific codon-choice patterns in reference to the isoacceptor contents, *Cold Spring Harbor Symp. Quant. Biol.*, 47:1087–1097, 1983.

[3] Xia, T., SantaLucia, J.Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., and Turner, D.H., Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs, *Biochemistry*, 37:14719–14735, 1998.