

Synonymous Codon Usage Controls Various Molecular Aspects

Eu-Hyun Im, Sun Shim Choi*

Division of Biomedical Convergence, College of Biomedical Science, and Institute of Bioscience & Biotechnology,
Kangwon National University, Chuncheon 24341, Korea

Synonymous sites are generally considered to be functionally neutral. However, there are recent contradictory findings suggesting that synonymous alleles might have functional roles in various molecular aspects. For instance, a recent study demonstrated that synonymous single nucleotide polymorphisms have a similar effect size as nonsynonymous single nucleotide polymorphisms in human disease association studies. Researchers have recognized synonymous codon usage bias (SCUB) in the genomes of almost all species and have investigated whether SCUB is due to random nucleotide compositional bias or to natural selection of any functional exposure generated by synonymous mutations. One of the most prominent observations on the non-neutrality of synonymous codons is the correlation between SCUB and levels of gene expression, such that highly expressed genes tend to have a higher preference toward so-called optimal codons than lowly expressed genes. In relation, it is known that amounts of cognate tRNAs that bind to optimal codons are significantly higher than the amounts of cognate tRNAs that bind to non-optimal codons in genomes. In the present paper, we review various functions that synonymous codons might have other than regulating expression levels.

Keywords: optimal codon, single nucleotide polymorphism, synonymous codon usage bias

Introduction

According to the molecular evolutionary theory, mutations occurring in coding regions involved in amino acid changes, called nonsynonymous mutations, are basically harmful and deleterious to organisms and are subject to strong purifying selection [1]. Changes in amino acids in protein sequences by any type of mutation—particularly, changes from original or ancestral amino acids into new amino acids with very different physicochemical properties—can cause severe problems in protein structure and function [2]. In contrast, mutations in coding regions but not related to amino acid changes are called synonymous mutations and are generally considered to be functionally neutral [3-5]. In fact, most recent research on identifying disease-causing variants in genetic diseases, including Mendelian diseases and common complex diseases, such as diabetes and cancers, has been focusing on searching for nonsynonymous variants [6-12]. In most of these studies,

synonymous variants have been ignored and even filtered out for further functional validation processes. Another difficulty in studying synonymous variants is that there are no good or established tests for functional synonymous codons.

One clue suggesting the non-neutrality of synonymous codons could be drawn from the study on synonymous codon usage bias (SCUB) [13-15]. SCUB means that the uses of synonymous codons involved in encoding the same amino acids are not equivalent to each other, depending on the kinds of proteins or the species carrying those proteins. In other words, different proteins in different species have different synonymous codon preferences when determining amino acids [15, 16]. No such biased codon preference among different synonymous codons is expected under the assumption of the neutrality of synonymous codons. Researchers have been recognizing this intriguing phenomenon for a long time and trying to figure out what factors lead to SCUB [17].

Two different explanations have been provided so far

Received August 18, 2017; Accepted September 25, 2017

*Corresponding author: Tel: +82-33-250-6487, Fax: +82-33-259-5644, E-mail: schoi@kangwon.ac.kr

Copyright © 2017 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

about what causes the SCUB phenomenon. One is based on uneven nucleotide compositions throughout genomes [18]. Simply, the research groups supporting the idea of biased nucleotide compositions think that genes located in genomic regions with abundant G or C content tend to prefer codons with G- or C-endings [19]. Consistently, it has been observed that there is a correspondence of G/C content between introns and exons in mammals, which means that genes harboring introns with higher G/C content are likely to carry exons with higher G/C content. The second scenario posits that SCUB is basically a consequence of natural selection of the roles of synonymous codons in the regulation of the expression of genes [20-22]. The research groups supporting the second scenario insist that genes with higher levels of expression are expected to have higher evolutionary constraints on the requirement of codons necessary for improving translational efficiency and accuracy than genes with lower levels of expression [23, 24]. In fact, numerous studies have shown that the codon usage of highly expressed genes is biased toward optimal codons than lowly expressed genes and that this biased usage is linked to the enhancement of translational speed and accuracy [25, 26]. In relation, the amounts of cognate tRNAs carrying anticodons are higher against optimal codons than against non-optimal codons.

An evolutionary explanation of SCUB, based on natural selection, was investigated under a hypothesis, called Hill-Robertson (HR) interference. The HR hypothesis posits that the efficiency of natural selection of one site will weaken when the site is linked to adjacent sites and does not segregate independently, wherein recombination can play a role in relieving the interference. There is some agreement among researchers that SCUB is positively correlated with recombination rates, although there are opposite observations on the effect of HR [27-29].

In the present work, we thus decided to summarize other facets of the functions that synonymous codons might have, other than the functions that are related to translational efficiency and accuracy in the regulation of gene expression (Fig. 1).

Results

Function of nonoptimal codons

SCUB is considered to be a general phenomenon that can be observed in the genomes of almost all species [13-15]. However, patterns of SCUB are not equivalent to each other among different genomes in different species. Some genomes have a higher preference toward optimal codons, while other genomes do not show this preference [13]. The amounts of optimal codons vary, depending on the kinds of genes [13].

SCUB patterns in genomes are more similar in closely related species than in distantly related species [30]. Under the natural selection scenario, genes with low levels of bias in codon usage are generally thought to be the result of a lack of natural selection toward optimal codons.

However, some researchers have tried to provide an opposing explanation regarding why some genes avoid optimal codons within genes. They think that genes tend to harbor codons that are rarely used in translation, because rare codons are beneficial in checking the step right before ribosomes start translation or the protein folding step before secreting nascent protein product in the endoplasmic reticulum [31, 32]. Additionally, recently, Zhou *et al.* (2013) [33] provided another interesting observation in *Neurospora*, suggesting that the usage of non-optimal codons in the *Frq* gene is essential in regulating circadian rhythms.

Splicing regulation

Intron-exon boundary regions, also known as limited by the GU-AG rule, are important in carrying out splicing events [34]. Additionally, potential splicing enhancers (ESEs) or splicing silencers residing near intron-exon boundaries—i.e., DNA sequence motifs within an exon—are known to enhance or suppress splicing. Thus, some synonymous codons can participate as constituents of these motifs. Therefore, a synonymous mutation of a gene can affect the splicing of that gene, without any influence on amino acid changes in that gene. In fact, Takahashi (2009) [35] has shown in *Drosophila* that translationally optimal codons tend to be avoided within the ESE motifs, and compared codon usage biases of exons with those of ESE regions in the Down syndrome cell adhesion molecule (*Dscam*) gene using codon bias indices, called CBI [35, 36]. *Dscam* is known as one of the genes with the largest number of alternatively spliced exons [35]. Furthermore, another study showed that almost none of the synonymous codons residing in ESEs was an optimal

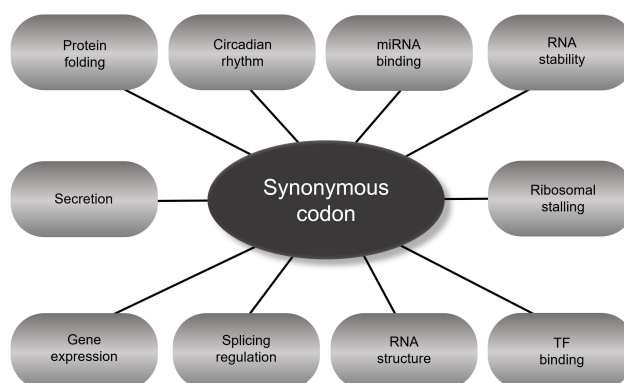


Fig. 1. Schematic representation of functional aspects in which synonymous codons might be involved.

codon in the regulation of translational efficiency [37]. The same study showed that this conflict on the roles of synonymous codons between translational regulation and splicing control was larger in highly expressed genes [37].

Regulation in transcription

It is quite unexpected that synonymous codons are somehow linked to the regulation of transcription. Transcription is a process of copying genic sequences into mRNAs, which requires various specific and delicate controls, usually conducted by the combinatorial actions of cis-acting DNA elements and transacting regulatory factors, called transcription factors (TFs). Promoters are well-established controlling cis-acting elements and are generally located upstream of transcription start sites, which are thought to be separate regions from the coding regions of genes.

Recently, Stergachis *et al.* (2013) [38] reported a surprising observation that approximately 15% of coding regions play roles both as coding sequences for determining amino acids and, astonishingly, as TF binding sites, by DNase footprint analysis in more than 80 different human cell types. Additionally, according to the paper, synonymous codon sites with dual functions are evolutionarily conserved, the binding sites of TFs recognizing stop codons are selectively depleted, and single-nucleotide variants residing within synonymous codons with dual functions can alter TF binding [38]. Therefore, it is plausible that mutations in synonymous sites can lead to obstruction of normal transcription, despite no amino acid alterations being driven.

Regulation of RNA secondary structure

The secondary structure of mRNA is important for controlling translational speed and timing, on which synonymous codon mutations might have crucial effects. For instance, removing rare synonymous codons from an expression construct decreases the enzyme activity of chloramphenicol acetyltransferase [15], which is due to alteration of the mRNA secondary structure by the removal of rare codons. The secondary structure of mRNA is also known to be associated with ribosome pausing, which is important for the formation of correct protein folding for insertion into the lipid bilayer in yeast [39].

The same observation has also been made in *Escherichia coli*, such that the translational rate is influenced by the protein folding efficiency and that the folding efficiency is associated with an abundance of codons with a low concentration of cognate tRNAs. Consistently, Zhou *et al.* [40] found that optimal codons for high efficiency of expression are located in structurally sensitive sites in proteins. Zhang *et al.* [41] also demonstrated that synonymous codon mutations significantly perturb folding efficiency. Saunders and

Deane [42] observed that synonymous codon usage is related to the secondary structure of local mRNA. All of these studies consistently suggest that synonymous codons might be involved in the formation of mRNA secondary structures that, ultimately, control translational speed and timing.

Conclusion

Researchers have long been searching for important disease-associated variants, mainly by investigating functional perturbations caused by nonsynonymous mutations. On the other hand, synonymous mutations have not been considered to be variants that are responsible for showing disease phenotypes. Often, synonymous mutations are filtered out and are not even considered, because candidate variants need to be functionally validated for the possibility of causing diseases or phenotypes. In that sense, it is an interesting revelation that there are various functional activities conducted by synonymous codons in molecular processes. Synonymous mutations do not change the amino acid sequence in proteins but can interrupt the formation of correct mRNA secondary structures, reduce translational accuracy and speed, and even alter the start of transcription. We think that advancing our understanding of the functions of synonymous codons will contribute to the identification of all disease-associated genes and mutations.

ORCID: Eu-Hyun Im: <http://orcid.org/0000-0002-3323-3957>; Sun Shim Choi: <http://orcid.org/0000-0001-6289-4220>

Authors' contribution

Conceptualization: SSC
 Formal analysis: EHI
 Funding acquisition: SSC
 Writing – original draft: SSC
 Writing – review & editing: EHI, SSC

Acknowledgments

This research was supported by a 2016 research grant from Kangwon National University (No. 120131854) to S.S.C.

References

1. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. *Nat Genet* 2008;40:340-345.

2. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812-3814.
3. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418-426.
4. King JL, Jukes TH. Non-Darwinian evolution. *Science* 1969;164:788-798.
5. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;24:1586-1591.
6. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;30:3894-3900.
7. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, *et al.* High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 2007;39:347-351.
8. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007;39:207-211.
9. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, *et al.* Genetic variation in *PNPLA3* confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2008;40:1461-1465.
10. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* 2009;30:1237-1244.
11. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30-35.
12. Li MX, Kwan JS, Bao SY, Yang W, Ho SL, Song YQ, *et al.* Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet* 2013;9:e1003143.
13. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011;12:32-42.
14. Doherty A, McInerney JO. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. *Mol Biol Evol* 2013;30:2263-2267.
15. Komar AA. The Yin and Yang of codon usage. *Hum Mol Genet* 2016;25:R77-R85.
16. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985;2:13-34.
17. Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 2006;7:98-108.
18. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci U S A* 2004;101:3480-3485.
19. Nabiyouni M, Prakash A, Fedorov A. Vertebrate codon bias indicates a highly GC-rich ancestral genome. *Gene* 2013;519:113-119.
20. Lavner Y, Kotlar D. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2005;345:127-138.
21. Quax TE, Claassens NJ, Soll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell* 2015;59:149-161.
22. Zhou Z, Dang Y, Zhou M, Li L, Yu CH, Fu J, *et al.* Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc Natl Acad Sci U S A* 2016;113:E6117-E6125.
23. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 2010;141:344-354.
24. Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 2001;53:290-298.
25. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* 2015;160:1111-1124.
26. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 2008;134:341-352.
27. Comeron JM, Kreitman M, Aguadé M. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* 1999;151:239-249.
28. Marais G, Mouchiroud D, Duret L. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc Natl Acad Sci U S A* 2001;98:5688-5692.
29. Zhou T, Lu ZH, Sun X. The correlation between recombination rate and codon bias in yeast mainly results from mutational bias associated with recombination rather than Hill-Robertson Interference. *Conf Proc IEEE Eng Med Biol Soc* 2005;5:4787-4790.
30. Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. A role for tRNA modifications in genome structure and codon usage. *Cell* 2012;149:202-213.
31. Zalucki YM, Beacham IR, Jennings MP. Biased codon usage in signal peptides: a role in protein export. *Trends Microbiol* 2009;17:146-150.
32. Clarke TF 4th, Clark PL. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. *BMC Genomics* 2010;11:118.
33. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, *et al.* Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 2013;495:111-115.
34. Michel F, Dujon B. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J* 1983;2:33-38.
35. Takahashi A. Effect of exonic splicing regulation on synonymous codon usage in alternatively spliced exons of *Dscam*. *BMC Evol Biol* 2009;9:214.
36. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986;14:5125-5143.
37. Warnecke T, Hurst LD. Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. *Mol Biol Evol* 2007;24:2755-2762.
38. Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Rey-

- nolds A, *et al.* Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* 2013;342:1367-1372.
39. Képès F. The “+70 pause”: hypothesis of a translational control of membrane protein assembly. *J Mol Biol* 1996;262:77-86.
40. Zhou T, Weems M, Wilke CO. Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol* 2009;26:1571-1580.
41. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol* 2009;16:274-280.
42. Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* 2010;38:6719-6728.