How do synonymous mutations affect fitness?

Joanna L. Parmley and Laurence D. Hurst*

Summary

While it has often been assumed that, in humans, synonymous mutations would have no effect on fitness, let alone cause disease, this position has been questioned over the last decade. There is now considerable evidence that such mutations can, for example, disrupt splicing and interfere with miRNA binding. Two recent publications suggest involvement of additional mechanisms: modification of protein abundance most probably mediated by alteration in mRNA stability[1] and modification of protein structure and activity,[2] probably mediated by induction of translational pausing. These case histories put a further nail into the coffin of the assumption that synonymous mutations must be neutral. *BioEssays* 29:515–519, 2007. © 2007 Wiley Periodicals, Inc.

Introduction

It is seductive to think that, owing to the redundancy in the genetic code, a point mutation in a protein-coding exon that changes the DNA but not the protein sequence (a synonymous mutation), would have no discernible fitness consequences. Indeed, even a decade ago such an assumption looked relatively sound. Since then, however, there has been a plethora of evidence to indicate that synonymous mutations can, indeed, have important fitness consequences, with over 40 genetic diseases now associated with such ''silent'' mutations.[3] How do apparently innocuous base changes have such an effect?

Codon usage bias puts the neutral theory in retreat

Since the introduction of the neutral theory and the finding that synonymous substitutions happen much faster than non-synonymous ones,[4] the neutrality of synonymous mutations was initially widely assumed. For species with large population sizes (worms, flies, yeast, bacteria etc.), however, this position was gradually eroded through the 1980s by the finding that, especially in highly expressed genes, the choice of which synonymous codon is employed to specify a given amino acid was not random.[5,6] Rather the codon that matched the most-abundant iso-acceptor tRNA species was preferentially employed (see Refs. 7–9). Indeed, it was conjectured that the skew in tRNA pool and codon usage should co-evolve so as to ensure that the most highly expressed genes could be translated as fast and as accurately as possible.[10]

Deptartment of Biology and Biochemistry, University of Bath, Bath, BA2 7AY.
*Correspondence to: Laurence D. Hurst, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY.
E-mail: l.d.hurst@bath.ac.uk
DOI 10.1002/bies.20592
Published online in Wiley InterScience (www.interscience.wiley.com).

A mammal is not an invertebrate

In this translation rate modification model, selection on a synonymous mutation that specifies an un-preferred rather than a preferred codon is likely to be weak.[11] Given that, in the framework of the nearly neutral model of molecular evolution,[12] selection is less efficient in species with small effective population sizes, it was supposed that selection of this variety would be all but irrelevant in mammals.[13] This was given credence by a variety of studies that failed to find any evidence of the expected forms of codon bias in mice and humans.[14,15] More recently, however, with the vast datasets now available and improvements in methods to detect codon bias (for example, those that allow for the overwhelming influence of regional nucleotide differences around mammalian genomes, see Ref. 16), there have been some small indications of this mode of selection [17–20] and, more generally, of selection of some form on synonymous mutations.[21–24]

Selection for translational accuracy/rate appears, however, to be weak if at all present in humans,[25] and cannot obviously explain why large tracts in exons containing highly conserved synonymous positions exist.[26,27] What then might be the mechanism or mechanisms of selection on synonymous mutations in mammals?

One early clue came from the finding that alternatively spliced exons have unusually low rates of evolution at synonymous sites;[28] this has since been verified on numerous occasions.[29] Combining this with evidence that synonymous rates of evolution can be especially low in exonic domains associated with splice control,[30,31] has led to the understanding that most selection on synonymous mutations in mammals is associated with perturbation of splicing. Remarkably, in one well-studied example, exon 12 of CFTR, a quarter of synonymous variations result in exon skipping.[32] More generally, most of the 40 or so genetic diseases

associated with synonymous mutation appear owing to disruption of splicing.[3] Likewise, many of the large exonic tracts of low synonymous substitution rate are associated with alternative exons.[26,27]

An association with splicing need not simply reflect mutation in the few base pairs immediately adjacent to the intron–exon boundary. Rather, the role of exonic splice enhancer (ESE) domains has been highlighted in several incidences.[3] These are sequences necessary for the binding of SR proteins to the immature mRNA, which, in turn, are needed for specification of the location of the intron–exon boundary. Importantly, ESEs have low synonymous SNP densities[33,34] and synonymous sites in ESEs evolve significantly slower than the flanking non-ESE synonymous sites.[35] Selection favouring ESEs in the vicinity of the intron–exon boundary has striking effects both on genic synonymous[35] and non-synonymous[36] rates of evolution in mammals.

It would then be tempting to suppose that, in humans, with their very high density of introns, selection on synonymous mutations is different to that which occurs in yeast, fly and worm, and is all associated with control of splicing. It appears premature to suppose that, in mammals, splicing explains all of the selection on synonymous mutations. For one thing, miRNA binding within coding exons appear to impose selective constraint on synonymous mutations within the binding sites,[37] as might be expected. Importantly, two recent papers highlight further different modes of selection. In one instance, the stability of mRNA is affected, which, in turn, affects protein concentration and net enzymatic rate. In the other, the synonymous mutations appear to affect protein folding, possibly by causing translational pausing while rare tRNAs are recruited. This in turn affects the activity of the protein.

mRNA stability and the case of *COMT*

Nackley et al. focused on the single nucleotide polymorphisms (SNPs) that affect the activity of the cathechol-*O*-methyltransferase (*COMT*). This gene is responsible for the degradation of catecholamines and is associated with responsiveness to pain in humans. There are three common haplotypes that are associated with levels of pain sensitivity: low (LPS), average (APS) and high (HPS).[38] The three haplotypes are composed of varying combinations of four SNPs: one in the promoter (A/G), two synonymous changes (C/T and C/G) and one non-synonymous valine-to-methionine change (A/G). It had been widely accepted, in humans, that the cause of the variation in *COMT* activity is due only to the non-synonymous SNP. Evidence for this, however, is weak as the two haplotypes with the most-extreme phenotypes (LPS and HPS), aside from differing in the promoter, only differ within the coding sequence at one synonymous SNP. These differences between the haplotypes are paralleled by differences in enzyme activity (reduced in cells expressing the HPS haplotype, in comparison to the LPS haplotype). Importantly, it was shown that this was due to reduced protein abundance and not reduced mRNA abundance.[1] As mRNA abundance does not parallel enzyme activity levels,[1] it seems unlikely that the promoter in the SNP could explain the differences between haplotypes (see also Supplementary Table 3 of Ref. 38). It was thus proposed that the underlying cause of the pain phenotype was implemented at the mRNA/translation level, and that a change in mRNA secondary structure could lead to a perturbation in protein synthesis.

To test this, Nackley et al. computationally analysed the stability of the mRNA secondary structures across the three haplotypes. Results of such in silico analyses should always be taken with caution, as the methods of assessing mRNA structure tend not to allow for features such as the proteins that are left on the mature mRNA at splice junctions. Nonetheless, the team report that (1) the predicted least-stable structure was that of the LPS haplotype mRNA, forming the shortest stem-loop structure, (2) the most-stable structure was encoded by the HPS haplotype (Fig. 1) and (3) the APS haplotype formed a mRNA secondary structure of intermediate stability. This mechanism was given credence by site-directed mutagenesis analyses in which new nucleotide changes were introduced that disrupted the predicted stem-loop structure observed in HPS haplotype, creating the LPS-like mRNA structure. Then, a secondary compensatory nucleotide change was introduced that converted the LPS-like structure back to a HPS-like structure. These predicted structural changes resulted in the same protein expression and enzymatic activity levels associated with the newly acquired haplotype mRNA structure.

Assuming mRNA stability is at the heart of the different activity and expression levels, why might this be? Several hypotheses could be considered. Perhaps the ultra-stable mRNA is hard for the ribosome associated helicases[39] to unwind? Perhaps, the stable structures are more prone to attack by RNAases?[40] RNA levels and degradation rates did not parallel protein levels[1] suggesting that a mechanism more like the former than the latter probably applies. Whatever the mechanism, this result runs counter to that previously proposed by mRNA stability studies[41–43] which find that, generally speaking, more stable mRNAs are selectively favoured as they enable mRNA persistence, potentially increasing the rate of protein expression. However, ultra-stable structures have been suggested as a means to limit expression in special cases by limiting scanning the 5′ end of the mRNA.[44]

Perhaps what is more remarkable in this case history is that the impact of the synonymous changes on enzyme activity is vastly greater than the modest thermostability effect of the non-synonymous change. If this case history tells us anything, it is that we have probably been too fast in ascribing all phenotypic effects to non-synonymous changes simply because the only other SNPs in a haplotype are synonymous.

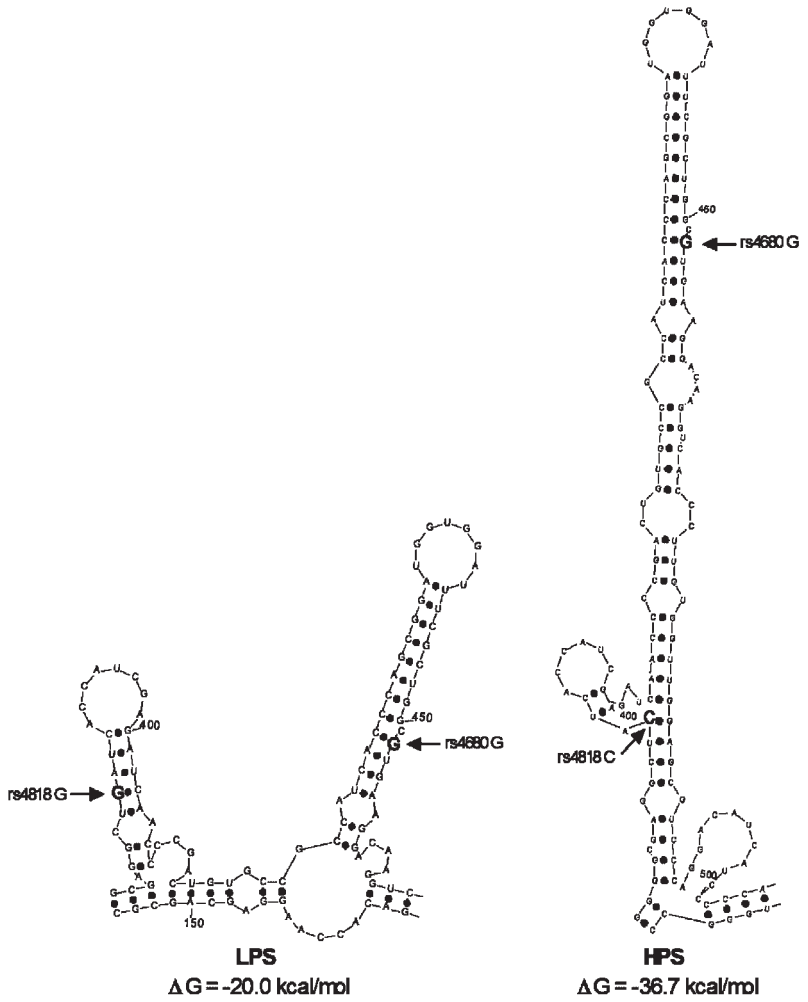

Figure 1. The predicted structures of the LPS and HPS haplotype mRNAs and their Gibbs free energy (ΔG). Note that these two extreme haplotypes differ only at the synonymous rs4818 SNP and not at the non-synonymous rs4680 SNP (courtesy of Andrea G. Nackley Neely).

Translational pausing, protein folding and the case of *MDR1*

The *Multidrug Resistance 1* (*MDR1*) gene encodes an ATP-driven efflux pump (P-gp) that has been associated with the multidrug resistance of cancer cells, though in many instances, the molecular mechanisms of such resistance are unknown. The variation within this gene is high, with over 50 reported SNPs. One synonymous SNP (C3435T) has been linked to a change in P-gp activity and is further associated, when present with a greater combination of SNPs, with reduced functionality. Kimchi-Sarfaty et al.[2] endeavoured to find out how the presence of a silent polymorphism can induce such a fitness effect.

The underlying mechanism for drug resistance is very complicated. Assays analysing the function of P-gp on single mutants, and haplotypes from combinations of these polymorphic variants, revealed no reduction in transporter function compared to the wild type. There was, however, an alteration in drug specificity in only those haplotypes containing the synonymous C3435T variant, even though this was not observed with this SNP alone. How, then, can this silent SNP cause altered drug specificity when in combination with other synonymous and non-synonymous variants? Neither mRNA nor protein levels were found to be diminished in these haplotypes and the protein sequence was as expected, ruling out the possibility that aberrant splice forms were involved.

Perhaps, then, a conformational change has occurred that allows P-gp to function, but inhibits the drug–protein interaction? Assays of trypsin digestion of the common (C1236T-G2677T-C3435T) P-gp haplotype required more than a three-fold increase in trypsin concentration, from the wild-type protein, to reach 50% degradation, indicating that the two

proteins have different tertiary structures. This conclusion was supported by the differential recognition of the haplotype protein compared to wild type using a conformation-sensitive monoclonal antibody. The mechanism by which these two isoforms were produced was attributed to the formation of a cluster of rare codons. One model supposes that rare codons are specified by rare tRNAs (which may[18–19] or may not[25] be the case) and that this ensures that the translational machinery must pause to enable tRNA recruitment. In line with both theory[45] and experiment,[46] such pausing in turn could enable the protein to find new structures. Closer inspection of the common haplotype associated with drug resistance revealed that all three SNPs involved a codon that was more rare than that of the wild type. The pausing mechanism was reinforced as the cause of drug resistance when an artificial haplotype was produced, employing a codon yet more rare than that originally found in this synonymous SNP, that reduced the sensitivity to the drugs yet further.

Conclusions

We have then, in mammals, at least four relatively well-resolved mechanisms by which synonymous mutations can have an effect on fitness: splice regulation, miRNA binding, mRNA folding and protein folding. If we add the possibility of weak effects of translational rate/accuracy and an otherwise mysterious effect of synonymous nucleotide content on mRNA levels,[47] mediated at either the transcriptional or RNA-processing level, that brings the current possible mechanisms to six. It is also likely that overlapping transcripts, which may well be much more common than once thought,[48] will impose some form of extra constraint[49] on mutations that are synonymous but only in one of the two genes.

The present-day predominance in the literature of the splice-associated mechanisms accounting for disease phenotypes[3] may reflect the relative ease of determining that an alternative splice form is found, as opposed to showing, for example, that a protein or mRNA structure is different (see Ref. 50). Suggestive of greater than previously recognized importance of the alternative mechanisms, we note that neither of the two new case histories is without precedent. A synonymous mutation was, for example, previously shown to be associated with disease mediated via its effects on mRNA stability.[51] More generally, several computational analyses have indicated a role for selection acting on synonymous mutations that affect mRNA stability,[41,42,52] although, as noted, these suggest that high stability is preferred. RNA half-life need be associated not only with stem-loop structures but also with residues that enable RNAases to digest mRNA, notably UA residues, which in turn are both avoided and are possibly under selection.[40] Likewise a role for usage of rare codons in enabling translational pausing, which alters protein folding, has been noted previously[53] and may indeed explain why stretches of rare codons correspond to turns, loops and links between protein domains.[54,55] It is notable that different protein structures may well be translated at different rates owing to skews in codon usage,[56] although these skews may instead relate to mRNA stability.[57] Preventing co-translational misfolding has been suggested to be especially important in mammals[58] and could explain why GAT is preferred over GAC at the N termini of alpha-helices in humans.[55,59]

Whether these new case histories are the tip of the iceberg or just rare curiosities remains to be seen. What is clear, however, is that in mammals not only are many synonymous mutations under selection, but the mechanisms by which selection acts on such changes are more diverse than commonly appreciated.

Acknowledgments

We thank Luda Diatchenko, Chava Kimchi-Sarfaty amd Michael Gottesman for comments on an earlier versions of the manuscript. We particularly thank Andrea Nackley Neely for providing us with Fig. 1.

References

1. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, et al. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314: 1930–1933.
2. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, et al. 2007. A ''silent'' polymorphism in the MDR1 gene changes substrate specificity. Science 315:525–528.
3. Chamary J-V, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet 7:98–108.
4. Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276.
5. Duret L. 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649.
6. Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in Drosophila. Proc Natl Acad Sci USA 94:7784–7790.
7. Ikemura T. 1992. Correlation between codon usage and tRNA content in microrganisms. In: Hatfield DL, Lee B, Pirtle PM, editors. Transfer RNA in protein synthesis. Boca Raton: CRC Press. p 87–111.
8. Duret L. 2000. tRNA gene number and codon usage in the C-elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet 16:287–289.
9. Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in Drosophila. J Mol Evol 45:514–523.
10. Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325:728–730.
11. McVean GAT, Vieira J. 1999. The evolution of codon preferences in Drosophila: A maximum- likelihood approach to parameter estimation and hypothesis testing. J Mol Evol 49:63–75.
12. Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol System 23:263–286.
13. Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA-sequence evolution—the sounds of silence. Philos Trans R Soc B-Biol Sci 349:241–247.
14. Eyre-Walker A. 1991. An analysis of codon usage in mammals: selection or mutation bias? J Mol Evol 33:442–449.
15. Smith NGC, Hurst LD. 1999. The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics 152:661–673.
16. Urrutia AO, Hurst LD. 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 159:1191–1199.

17. Urrutia AO, Hurst LD. 2003. The signature of selection mediated by expression on human genes. Genome Res 13:2260–2264.
18. Lavner Y, Kotlar D. 2005. Codon bias as a factor in regulating expression via translation rate in the human genome. Gene 345:127–138.
19. Comeron JM. 2004. Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. Genetics 167:1293–1304.
20. Comeron JM. 2006. Weak selection and recent mutational changes influence polymorphic synonymous mutations in humans. Proc Natl Acad Sci USA 103:6940–6945.
21. Chamary JV, Hurst LD. 2004. Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: Evidence for selectively driven codon usage. Mol Biol Evol 21:1014–1023.
22. Kondrashov FA, Ogurtsov AY, Kondrashov AS. 2006. Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites. J Theor Biol 240:616–626.
23. Lu J, Wu CL. 2005. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. Proc Natl Acad Sci USA 102:4063–4067.
24. Hellmann I, Zollner S, Enard W, Ebersberger I, Nickel B, et al. 2003. Selection on human genes as revealed by comparisons to chimpanzee cDNA. Genome Res 13:831–837.
25. dos Reis M, Savva R, Wernisch L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res 32:5036–5044.
26. Schattner P, Diekhans M. 2006. Regions of extreme synonymous codon selection in mammalian genes. Nucleic Acids Res 34:1700–1710.
27. Parmley JL, Hurst LD. 2007. How common are intragene windows with KA > KS owing to purifying selection on synonymous mutations? J Mol Evol: (in press).
28. Iida K, Akashi H. 2000. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. Gene 261:93–105.
29. Xing Y, Lee C. 2006. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. Nat Rev Genet 7: 499–509.
30. Hurst LD, Pal C. 2001. Evidence for purifying selection acting on silent sites in BR CA1. Trends Genet 17:62–65.
31. Orban TI, Olah E. 2001. Purifying selection on silent sites - a constraint from splicing regulation? Trends Genet 17:252–253.
32. Pagani F, Raponi M, Baralle FE. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci USA 102:6368–6372.
33. Fairbrother WG, Holste D, Burge CB, Sharp PA. 2004. Single nucleotide polymorphism-based validation of exonic splicing enhancers. PLoS Biol 2:E268.
34. Carlini DB, Genut JE. 2006. Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. J Mol Evol 62: 89–98.
35. Parmley JL, Chamary JV, Hurst LD. 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol 23:301–309.
36. Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD. 2007. Splicing and the evolution of proteins in mammals. PLoS Biol 5:343–353.
37. Hurst LD. 2006. Preliminary assessment of the impact of microRNA-mediated regulation on coding sequence evolution in mammals. J Mol Evol 63:174–182.
38. Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, et al. 2005. Genetic basis for individual variations in pain perception and the development of a chronic pain condition. Hum Mol Genet 14:135–143.
39. Takyar S, Hickerson RP, Noller HF. 2005. mRNA helicase activity of the ribosome. Cell 120:49–58.
40. Duan J, Antezana MA. 2003. Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. J Mol Evol 57:694–701.
41. Chamary JV, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol 6:R75.
42. Cohen B, Skiena S. 2003. Natural selection and algorithmic design of mRNA. J Comput Biol 10:419–432.
43. Seffens W, Digby D. 1999. mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res 27:1578–1584.
44. Kozak M. 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34.
45. Purvis IJ, Bettany AJ, Santiago TC, Coggins JR, Duncan K, et al. 1987. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. J Mol Biol 193:413–417.
46. Komar AA, Lesnik T, Reiss C. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett 462:387–391.
47. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol 4:933–942.
48. Sun M, Hurst LD, Carmichael GG, Chen J. 2006. Evidence for variation in abundance of antisense transcripts between multicellular animals but no relationship between antisense transcription and organismic complexity. Genome Res 16:922–933.
49. Lipman DJ. 1997. Making (anti)sense of non-coding sequence conservation. Nucleic Acids Res 25:3580–3583.
50. Shen LX, Basilion JP, Stanton VP Jr. 1999. Single-nucleotide polymorphisms can cause different structural folds of mRNA. Proc Natl Acad Sci USA 96:7871–7876.
51. Duan JB, Wainwright MS, Comeron JM, Saitou N, Sanders AR, et al. 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet 12:205–216.
52. Shabalina SA, Ogurtsov AY, Spiridonov NA. 2006. A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res 34:2428–2437.
53. Cortazzo P, Cervenansky C, Marin M, Reiss C, Ehrlich R, et al. 2002. Silent mutations affect in vivo protein folding in Escherichia coli. Biochem Biophys Res Commun 293:537–541.
54. Thanaraj TA, Argos P. 1996. Ribosome-mediated translational pause and protein domain organization. Protein Sci 5:1594–1612.
55. Oresic M, Shalloway D. 1998. Specific correlations between relative synonymous codon usage and protein secondary structure. J Mol Biol 281:31–48.
56. Thanaraj TA, Argos P. 1996. Protein secondary structural types are differentially coded on messenger RNA. Protein Sci 5:1973–1983.
57. Jia MW, Luo LF. 2006. The relation between mRNA folding and protein structure. Biochem Biophys Res Commun 343:177–182.
58. Netzer WJ, Hartl FU. 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature 388:343–349.
59. Oresic M, Dehn M, Korenblum D, Shalloway D. 2003. Tracing specific synonymous codon-secondary structure correlations through evolution. J Mol Evol 56:473–484.