molecular
systems
biology

NEWS AND VIEWS

Chaperonin activity modulates codon adaptation

Sangjo Han[1] and Philip M Kim[1,2,3,4,*]

[1] Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada, [2] Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada, [3] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada and [4] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada
* Corresponding author. Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Room 606, Toronto, Ontario, Canada M5S 3E1. Tel: + 1 416 946 3419; Fax: + 1 416 978 8287; E-mail: pi@kimlab.org

Since the discovery of the genetic code by Holley, Khorana, and Nirenberg, researchers have been fascinated by the apparent redundancy of synonymous codons that code for the same amino acid. Subsequent work led to the discovery that synonymous codons are not used randomly in coding sequences. Rather, classes of genes that are characterized by a specific codon usage bias can be identified. Thus, 'optimal codons' were defined as codons that are enriched in highly expressed genes. Based on this notion, bioinformatics researchers have developed the Codon Adaptation Index, which has been adopted as an approximate measure for overall gene expression level. The classical theory of codon bias posits that optimal codons correlate with the abundances of cognate tRNAs, and thereby increase translational efficiency and accuracy (Ikemura, 1981) by reducing the rate of ribosomal stalling and decreasing the chances of incorporating the wrong amino acid. Consequently, optimal codons are thought to prevent mistranslation-induced misfolding (Drummond and Wilke, 2008).

In a recent study published in *Molecular Systems Biology*, Warnecke and Hurst (2010) report evidence for a novel mechanism underlying optimal codon usage. They hypothesize a novel link between codon bias and the cellular folding machinery; namely, the GroEL chaperonin complex. Specifically, proteins that only sporadically rely on chaperones (in *trans*) to prevent misfolding are hypothesized to be more dependent on the use of optimal codons (in *cis*) to mitigate errors. Conversely, genes encoding obligate chaperone substrates are anticipated to be less biased towards optimal codon composition. Although a connection between codon adaptation and protein misfolding had been previously drawn (Drummond and Wilke, 2008), this study presents the first evidence for a direct involvement of molecular chaperones. To test the validity of their hypothesis, the authors focus on the *Escherichia coli* GroEL chaperonin system and assemble supporting evidence from a variety of sources, including structural biology and comparative genomics and control for a number of confounding factors.

The *E. coli* chaperonin GroEL is the best-characterized chaperone system of any organism, and has two kinds of substrates: sporadic and obligate clients (Kerner *et al*, 2005). The authors show that codon bias, which is calculated as the frequency of optimal codons, is significantly different for these two classes of clients: obligate GroEL substrates exhibit lesser codon bias than sporadic ones, supporting the hypothesis that GroEL activity can compensate for mistranslation-induced misfolding resulting from non-optimal codon usage (Figure 1). Importantly, this effect is statistically separable from the relationship between codon bias and expression level.

The authors extend their analysis at the structural level by investigating codon usage at sites of restricted solvent accessibility. Such 'buried sites' are thought to be structurally sensitive regions. Supporting their relevance for misfolding and extending previous findings (Zhou *et al*, 2009), 'buried sites' turn out to be enriched for optimal codons in sporadic GroEL clients as compared with obligate ones, independently of mRNA expression level or protein length. This represents an optimal adaptation to evolutionary pressures, which can be understood in simple energy economics.

For additional support, the authors make elegant use of comparative genomics and analyze the loss of codon adaptation in a related species, *Shigella dysenteriae*. Since Shigella species underwent a substantial reduction in effective population size, the impact of purifying selection diminished after their split from *E. coli*, whereas genetic drift became the major factor in their evolutionary history. Although their chaperonin machinery is identical, the selective pressure on optimal codons should be markedly reduced. Confirming the initial hypothesis, the authors find that the loss of 'codon optimality' is most pronounced in the obligatory GroEL clients.

Finally, to generalize this finding of *cis–trans* complementarity in eukaryotes, the authors apply their analysis to a recent dataset of yeast chaperonin substrates: 303 interactors of the eukaryotic chaperonin CCT/TRiC identified in a physical and genetic interaction screen (Dekker *et al*, 2008). Although distinction between sporadic and obligate substrates is not possible in this case owing to the limitations of the experimental dataset, the authors can nevertheless show that CCT/TRiC substrates exhibit higher codon bias than non-
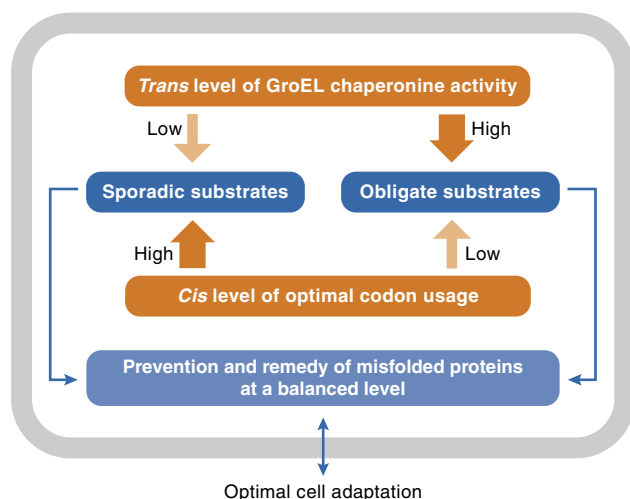
Figure 1 *Cis–trans* complementarity in a protein folding pathway. Optimal codons as a *cis* factor on a genome prevent misfolding by decreasing mistranslation at a translation level, and molecular chaperones such as *E. coli* GroEL as a *trans* factor in a cytoplasm also help their substrates to fold correctly at a post-translational level. Here, the above substrates of GroEL chaperonin are categorized according to their dependency on the chaperonin into sporadic and obligate ones.

substrates. Following the same strategy as in *E. coli*, they again find stronger enrichment at structurally sensitive sites for CCT interactors, albeit only for a subset of amino acids.

Owing to the current data quality on chaperonin dynamics, similar investigations in higher eukaryotes are difficult at this point. However, if misfolding-related selection on codon usage would prove to be a widespread phenomenon, further genome-scale analysis may provide new insights into the interplay between codon usage and protein folding and the respective constraints imposed by the structure of multi-domain proteins. In this regard, it will be fascinating to investigate whether the impact of codon usage extends beyond error reduction and might also be involved in the active regulation of protein folding in mammals.

The authors present a compelling story that introduces a novel idea into the field of codon bias. Nevertheless, codon bias remains a poorly understood topic and is likely a product of a manifold variety of different mutational and selective forces. For example, cell fitness in response to environmental and physiological changes could be fine-tuned by codon bias affecting global rather than local gene expression through increasing translational efficiency by increasing the elongation rate, as recently pointed out (Kudla *et al*, 2009). In addition, recent work has shown that synthetic design of codon usage can be used practically to produce attenuated vaccines rapidly and efficiently (Coleman *et al*, 2008). We believe that closer investigations of codon bias across more genomes are needed to shed more light on this emerging kaleidoscope of effects. It is likely that the picture will be more complex in mammalian genomes. As more systematic experimental and computational approaches emerge, we will begin to reach a more integrated view of the molecular mechanisms and evolutionary processes that shape codon usage at the genome scale.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S (2008) Virus attenuation by genome-scale changes in codon pair bias. *Science* **320:** 1784–1787

Dekker C, Stirling PC, McCormack EA, Filmore H, Paul A, Brost RL, Costanzo M, Boone C, Leroux MR, Willison KR (2008) The interaction network of the chaperonin CCT. *EMBO J* **27:** 1827–1839

Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134:** 341–352

Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol* **151:** 389–409

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* **122:** 209–220

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324:** 255–258

Warnecke T, Hurst LD (2010) GroEL dependency affects codon usage—support for a critical role of misfolding in gene evolution. *Mol Syst Biol* **6:** 340

Zhou T, Weems M, Wilke CO (2009) Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol* **26:** 1571–1580