

REVIEW

Open Access

# A code within the genetic code: codon usage regulates co-translational protein folding



Yi Liu

## Abstract

The genetic code is degenerate, and most amino acids are encoded by two to six synonymous codons. Codon usage bias, the preference for certain synonymous codons, is a universal feature of all genomes examined. Synonymous codon mutations were previously thought to be silent; however, a growing body of evidence now shows that codon usage regulates protein structure and gene expression through effects on co-translational protein folding, translation efficiency and accuracy, mRNA stability, and transcription. Codon usage regulates the speed of translation elongation, resulting in non-uniform ribosome decoding rates on mRNAs during translation that is adapted to co-translational protein folding process. Biochemical and genetic evidence demonstrate that codon usage plays an important role in regulating protein folding and function in both prokaryotic and eukaryotic organisms. Certain protein structural types are more sensitive than others to the effects of codon usage on protein folding, and predicted intrinsically disordered domains are more prone to misfolding caused by codon usage changes than other domain types. Bioinformatic analyses revealed that gene codon usage correlates with different protein structures in diverse organisms, indicating the existence of a codon usage code for co-translational protein folding. This review focuses on recent literature on the role and mechanism of codon usage in regulating translation kinetics and co-translational protein folding.

**Keywords:** Codon usage, Translation elongation, Co-translational protein folding, Intrinsically disordered protein

## Background

Of the 20 standard amino acids, 18 can be encoded by two to six synonymous codons. Preferential use of certain synonymous codons, a phenomenon called codon usage bias, has been found in all genomes evaluated [1–4]. Because synonymous codons encode the same amino acid, they were previously considered to be functionally redundant, and synonymous codon mutations that do not change protein sequences were regarded as silent mutations. However, a large body of evidence now demonstrates that codon usage plays multiple roles regulating gene expression and protein structure through translation-dependent and

translation-independent mechanisms [4–8] (Fig. 1a). Synonymous codons are recognized with different efficiencies by cognate tRNAs. In different eukaryotic and prokaryotic organisms, codon usage bias correlates with levels of cognate tRNAs or with tRNA gene copy numbers [1, 9–13]. Codons with strong bias are found to be strongly enriched in highly expressed protein encoding genes, and codon optimization increases endogenous and heterologous gene expression in diverse eukaryotes and prokaryotes [14–23]. Moreover, genome-wide correlations between codon usage bias and protein levels have been observed [17, 24].

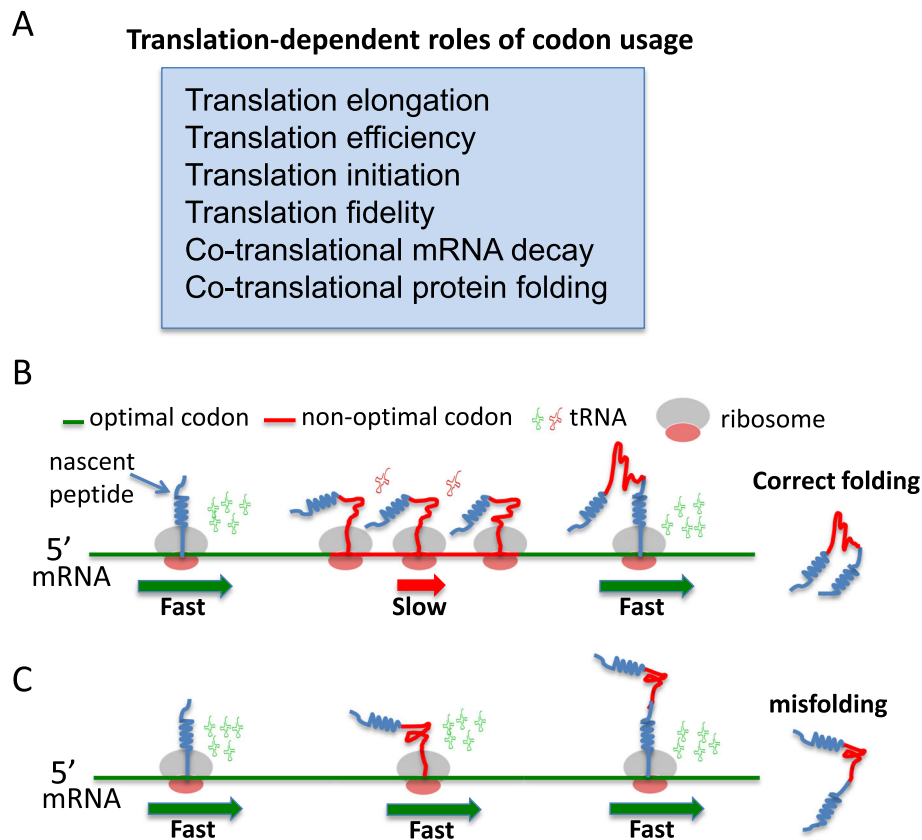
Whether codon usage bias can regulate translation kinetics had been the subject of intense debate [25–28], but it is now firmly established that codon usage plays an important role in controlling the speed of translation

Correspondence: [Yi.Liu@UTSouthwestern.edu](mailto:Yi.Liu@UTSouthwestern.edu)

Department of Physiology, ND13.214A, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9040, USA



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



**Fig. 1 a** The known translation-dependent roles of codon usage. **b** A model depicting the role of codon usage in regulating translation elongation speed and co-translational protein folding. **c** A model depicting that when codon usage is not optimally adapted to the co-translational protein folding process, protein misfolding increases

elongation during mRNA translation [29–32] and that an important role of codon usage bias is to promote translation efficiency [4, 8, 22, 23, 33–35]. Rare codons cause ribosome stalling on an mRNA during translation, which, in eukaryotes, can result in premature translation termination mediated by termination factors [29, 31, 36]. Importantly, codon usage bias also impacts mRNA stability in diverse organisms and can impact on translation fidelity [6, 37–44]. In budding yeast, codon optimality has been shown to regulate the interaction between the mRNA deadenylation complex CCR4-NOT and ribosome during translation due to its effect on translation kinetics, which then affecting co-translational decay of mRNAs [45].

In addition to its translation-dependent roles, codon usage bias has also many translation-independent roles that influence gene expression. In fungal and mammalian cells, codon usage and GC content (which is tightly correlated with codon usage) have been shown to affect mRNA levels by regulating transcription at the level of chromatin structure [17, 23, 46, 47]. Moreover, codon usage can also influence transcription by regulating premature transcription termination and splicing [48, 49]. Therefore, codon usage can influence gene expression

through multiple mechanisms at translational, transcriptional and post-transcriptional levels.

In addition to the roles of codon usage in regulating gene expression, a large body of biochemical and bioinformatic evidence has established that co-translational protein folding is influenced by codon usage [5, 13, 20, 29, 50, 51]. These studies led to the proposal that gene codon usage bias represents a new “code” that defines protein structures and protein expression levels. Codon usage regulates the speed of translation elongation, resulting in non-uniform ribosome decoding rates on mRNAs during translation that is adapted to co-translational protein folding. Mutations that do not alter the protein sequence can cause protein structure changes in proteins that are implicated in human disease [23, 52–54]. Thus, codon usage has a potentially underappreciated role in human disease pathogenesis. This review will focus on recent literature on the role and mechanism of codon usage in regulating translation kinetics and co-translational protein folding.

#### Codon usage regulates translation elongation speed

Because codon usage biases have been found to correlate with tRNA copy numbers in different organisms, codon

usage has been previously proposed to regulate mRNA translation elongate speed because rare codons are hypothesized to take longer to be recognized when their corresponding tRNAs are present at low concentrations [1, 4, 55, 56]. Most studies concerning this issue have relied on indirect measurements and protein overexpression, which led to conflicting conclusions [28, 57–59]. The development of the ribosome profiling technique provided a powerful tool to study ribosome translation dynamics with codon-level resolution in many organisms [26, 60]. Ribosome densities at individual codons should in principle inversely correlate with translation elongation speed; however, early studies based on ribosome profiling data in several organisms found no correlations between codon usage and ribosome density, leading to the conclusion that codon usage does not play a significant role in modulating translation kinetics [25–28].

Because ribosome profiling results could be influenced by experimental conditions, cloning and sequencing biases, methods of bioinformatic analysis, and experimental noise [61–63], we previously re-examined this issue using *Neurospora* and *Drosophila* cell-free translation systems in which translation elongation speed can be directly measured [29, 36]. We measured the time of first appearance of luciferase signal from mRNA templates encoding luciferase with different codon usage profiles (luciferase folding is co-translational in these systems), our results demonstrated clearly that preferred codons speed up rate of translation elongation, whereas rare codons slow translation elongation in both systems (Fig. 1b). This indicated that the effect of codon usage on elongation rate is a phenomenon conserved from fungi to animals. In addition, we found that codon usage controls ribosome traffic on mRNAs and that rare codons result in ribosome pausing and accumulation of nascent peptides of the expected sizes during translation in both amino-acid-context dependent and independent manners [29, 31, 36]. Furthermore, using luciferase-encoding mRNAs with high signal to noise, we convincingly demonstrated the impact of codon usage bias on ribosome occupancy in vitro and in vivo [29]. Using an improved ribosome profiling technique, clear genome-wide correlations between codon usage and ribosome occupancy were also observed [29–32], indicating that the lack of correlation between codon usage and ribosome occupancy in the early ribosome profiling studies was due to the lack of sensitivity and technical issues. Finally, the role of codon usage in translation elongation speed was also confirmed by imaging of single mRNAs during translation in vivo [64]. Together, these studies firmly established a role for codon usage in regulating translation elongation rate and demonstrated that codon usage biases can result in uneven translation elongation speed during mRNA translation (Fig. 1b).

### The role of translation kinetics in protein folding

After synthesis, ordered proteins become functional once they fold properly into stable tertiary structures. Protein misfolding results in impaired protein functions and aggregation or rapid degradation and contributes to many human diseases [65, 66]. Although protein structures are determined by amino acid sequences, it is now widely accepted that folding of most proteins in vivo occurs co-translationally (Fig. 1b). Folding is a modular process that begins on the ribosome as nascent peptide chains are synthesized. Proteins fold from the N-terminal end to the C-terminal end [67–71]. The folding of some protein structure elements, such as  $\alpha$  helices, can occur inside the ribosome exit tunnel. Once the newly synthesized nascent chains emerge from the ribosome, their interactions with the ribosome, with protein factors known as chaperones and with other folding catalysts mediate the folding process [67–72]. Translation kinetics was long proposed to influence co-translational protein folding [73, 74]. By using strains with mutant ribosomes that have slow translation speed, by altering tRNA levels, and by growth at low temperature, it was shown that the changes in elongation rate affect folding of proteins overexpressed in *Escherichia coli* [75, 76]. Consistent with this, expression of heterologous proteins often causes protein aggregation due to misfolding, a phenomenon that can be corrected by growth at a low temperature, which presumably slows translation and enhances correct folding.

### The role of codon usage in protein folding

Because codon usage bias often differs throughout a gene coding region, codon usage is expected to result in variation of elongation speed during mRNA translation. As a result, codon usage-influenced translation kinetics should affect the time available for different co-translational folding events. Analyses of in vitro translation and of protein overexpression in *E. coli* cells showed that replacing rare codons with common ones can result in modest decreases of protein activity or solubility [77, 78]. Moreover, synonymous substitutions of common codons with highly abundant tRNAs for codons predicted to slow translation impair the folding of the multidomain protein SufI in vitro and in cells [76]. In addition, using a metric to predict relative translation rates of codons based on tRNA availability, it was shown that synonymous codon harmonization designed to mimic ribosome movement in *E. coli* could significantly increase the specific activity of firefly luciferase overexpressed in *E. coli* [12]. Furthermore, codon usage influences the folding of an artificially designed fluorescent protein expressed in *E. coli* [79]. Importantly, the impact of codon usage on co-translational protein folding was confirmed by monitoring Förster resonance energy transfer in real time and by analysis of

fluorescence intensity changes when the mammalian gamma-B crytallin was expressed in *E. coli* [80]. The folding differences due to the presence of different synonymous codon variants was also visualized by NMR spectroscopy and shown to correlate with altered in vivo stability and in vitro protease sensitivity [80]. Recently, analysis of chloramphenicol acetyltransferase with various codon substitution in *E. coli* revealed that codon usage affects protein folding and cell growth [81].

The first suggestion of the role of codon usage in eukaryotic protein folding came from a study that showed that transient expression of variant with a synonymous single-nucleotide polymorphism in the *MDR1* gene, which encodes a protein involved in multidrug resistance, in human cells resulted in altered drug and inhibitor interactions [52]. The genetic role of the codon usage in eukaryotic protein folding was later firmly established by studying genes critical for circadian clock functions in *Neurospora* and *Drosophila* [13, 20, 51]. Unlike most genes in *Neurospora*, the core circadian clock gene *frequency* (*frq*) is enriched for rare codons across its entire open reading frame. Codon optimization of the parts of the *frq* gene with preference for rare codons abolished both overt and molecular circadian rhythms, impaired FRQ activity, and altered FRQ stability and phosphorylation profiles [13, 20]. Such changes are due to changes in FRQ structure, as indicated by altered trypsin sensitivity and resistance to freeze-thaw cycles in vitro. Importantly, the codon optimization of different parts of *frq* can result in opposite molecular phenotypes that can be predicted by local protein structures, indicating that the codon usage-mediated changes in FRQ structures are due to a local effect of codon usage [20]. Although *frq* and the *Drosophila* circadian clock gene *Period* (*Per*) are not sequence homologs, codon optimization of the parts of *Per* enriched for rare codons also results in abolishment of circadian locomotor rhythm and abnormal molecular rhythmicity due to severe impairment of PER activity in the circadian negative feedback loop and reduction of PER phosphorylation in a location- and codon usage-dependent manner [51]. The changes in PER structure due to codon usage manipulation were demonstrated by altered PER trypsin sensitivity and changes in thermal denaturation and aggregation temperatures. It is important to note that both of the *Neurospora* and *Drosophila* studies did not rely on protein overexpression, and the changes in protein activity and protein structure are not due to protein overexpression as overexpression of the wild-type genes lead to robust circadian rhythms [20, 51]. The roles of codon usage in co-translational protein folding in the *Neurospora* and *Drosophila* are further supported by in vitro translation assays in cell-free translation systems [29, 36]. In addition to the role of codon optimality in

determining the speed of translation elongation and translation efficiency of *luciferase* mRNAs, codon usage also affects luciferase folding and activity in vitro and in vivo [29, 36]. Although the codon-optimized *luciferase* mRNA was translated more rapidly and resulted in more full-length protein than wild-type mRNA, the luciferase protein produced was not as functional as the protein synthesized from the wild-type mRNA. These results indicate that when codon usage is not adapted to co-translational folding process, protein misfolding increases (Fig. 1c).

Interestingly, codon optimization of the cyanobacterium *Synechococcus* circadian clock genes *kaiBC* did not impair clock function [19]. On the contrary, the clock function of cyanobacterium *Synechococcus* became more robust upon codon optimization due to an increase in KaiC protein level, which results in decreased cell growth at low temperatures. The different outcomes of circadian clock gene codon optimization in cyanobacterium and eukaryotes suggest that differential structures have different sensitivities to codon usage-mediated co-translation folding (see below).

In human cells, the co-translational folding of the NBD1 domain of the conductance regulator CFTR, which is the transmembrane protein mutated in cystic fibrosis patients, occurs through sequential compaction of different subdomains [54]. The codon optimization of a region of NBD1 domain with non-optimal codon usage increases the protein synthesis rate but also results in aggregation of full-length CFTR, suggesting that translation elongation kinetics of the *CFTR* mRNA are modulated by codon usage in a manner that influences the co-translational folding of different protein regions [54]. In addition, the synonymous single nucleotide polymorphism (SNP) T2562G SNP in the *CFTR* gene introduces a codon recognized by a tRNA that is of low abundance specifically in human bronchial epithelia cells [53]. This SNP leads to changes in CFTR protein stability and channel activity in HeLa cells. Importantly, overexpression of the tRNA that recognizes this rare codon results in wild-type levels of expression, stability, and function of CFTR in HeLa cells, suggesting that alterations in translation elongation kinetics due to the SNP cause impaired CFTR co-translational folding [53]. Codon usage in *KRAS*, which encodes an oncogenic Ras GTPase family member, was shown to regulate the expression and protein structure of *KRAS* protein when expressed in human cells [22, 23]. More recently, codon usage optimization of the coagulation factor IX and a single SNP of the *ADAMTS13* has been shown to affect protein folding or activity [82, 83]. Together, these studies firmly established the universal role of codon usage in protein folding in both prokaryotes and eukaryotes.

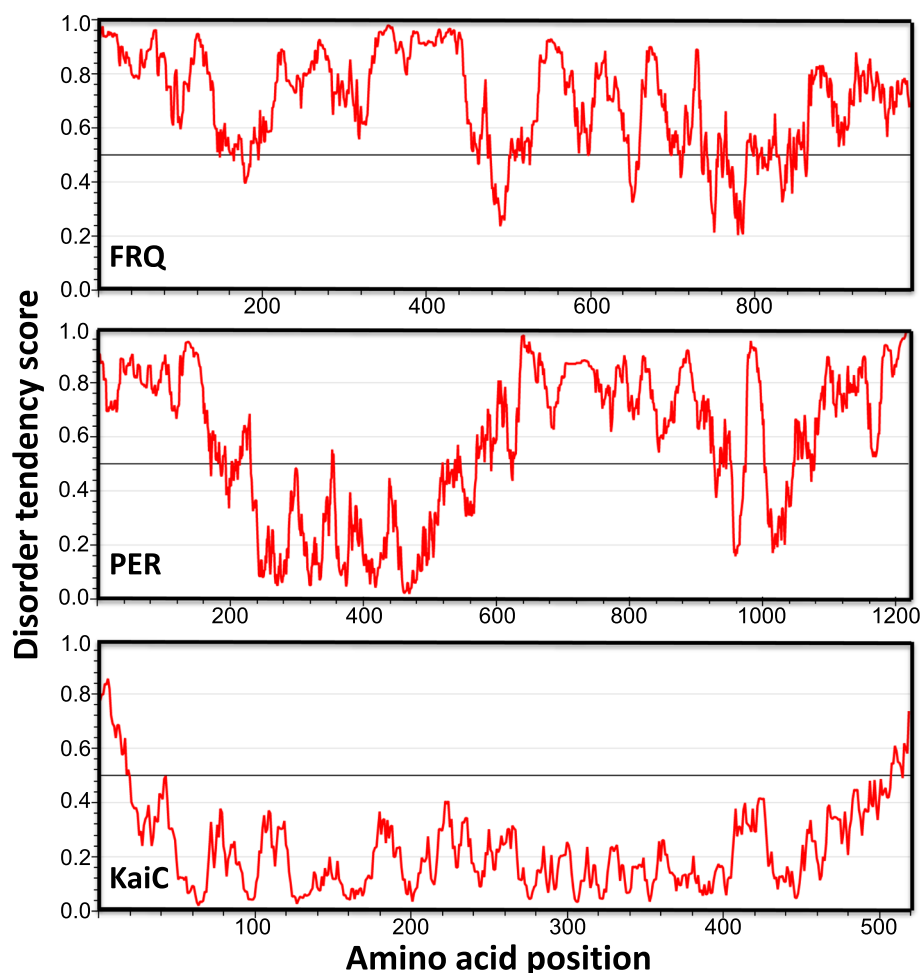
#### **Predicted intrinsically disordered domain structures are sensitive to codon usage**

Intrinsically disordered protein (IDP) domains are protein domains that are predicted to lack stable or ordered

three-dimensional structures. IDPs have pivotal roles in many biological processes [84–86]. Although IDPs may not form stable structures by themselves, they are often serve as sites critical for post-translational modifications that are important for the overall protein structures and as platforms for inter- or intra-molecular protein-protein interactions [87, 88]. Although both the *Neurospora* and *Drosophila* clock proteins FRQ and PER are large proteins (989 and 1224 amino acids, respectively), most regions of these proteins are predicted to be intrinsically disordered, and these proteins are extensively phosphorylated at many sites along the entire proteins [13, 20, 51, 87, 88]. In contrast, the cyanobacterium *Synechococcus* circadian clock protein KaiC is a highly structured protein for which a high-resolution crystal structure available [89]. The disorder tendency plots of the *Neurospora* FRQ, *Drosophila* PER and *Synechococcus* KaiC are shown in Fig. 2. The opposite circadian phenotypes result from codon optimization of *kaiC* versus *frq* and *Per* suggest that different proteins or protein

domains have different sensitivities to codon usage and intrinsically disordered domains are sensitive to change usage changes [13, 19, 20, 51]. This conclusion is further supported by codon optimization of different regions of the *frq* open reading frame: Codon optimization of the IDP regions result in impaired clock functions, whereas codon optimizations of the structured regions have little or no effects on rhythmicity [13].

In the *Neurospora* in vitro translation system, translation of the *luciferase* mRNA with optimal codon usage resulted in markedly reduced luciferase activity compared to the wild-type mRNA despite production of more luciferase protein [29]. By swapping different regions of the *luciferase* mRNA between the wild-type and optimized mRNAs, the region of open reading frame that is most sensitive to codon usage changes was identified [29]. The identified region is highly conserved among luciferase homologs and forms an unstructured loop in the luciferase crystal structure. Thus, both the in vitro and in vivo results suggested that compared to



**Fig. 2** Protein disorder tendency plots of the *Neurospora* FRQ, *Drosophila* PER, and *Synechococcus* KaiC proteins. Protein disorder tendency is predicted by the IUPred2A web interface (<https://iupred2a.elte.hu/>)



folding of well-structured proteins, co-translational folding of proteins with IDPs is more sensitive to codon usage changes. It is likely that the co-translational folding of IDP-containing domains requires more time and is more sensitive to changes in translation elongation speed than the folding of well-structured domains. Alternatively, fast translation elongation likely interferes with co-translational folding of the structural domains surrounding IDPs.

#### **Correlations between codon usage and protein structures: the codon usage code for co-translational protein folding**

The differential sensitivity of different protein domains to codon usage suggest that protein structures and codon usage has co-evolved to fine-tune the ribosome decoding rates on mRNAs to facilitate the co-translational folding process [19, 29, 74]. Supporting this conclusion, bioinformatics analyses of the relationship between codon usage and protein secondary and tertiary structures uncovered wide-spread correlations [13, 90–97]. Consistent with a role for non-optimal codons in folding of protein domains with IDPs, multiple studies using known structures and predicted secondary structures found that the predicted unstructured and predicted coil domains were found to be enriched for rare codons whereas optimal codons are preferentially found in  $\alpha$ -helical regions in proteins from *E. coli*, *Neurospora*, yeast, *C. elegans*, and *Drosophila* [13, 90, 94]. In addition, the translational optimal codons are enriched in buried residues in protein structures, and the association is highest in proteins encoded by highly expressed genes in several organisms [98]. It should be noted that regions with buried residues in highly expressed proteins mostly are found in well-structured protein domains. These bioinformatic studies suggest that well-structured regions are usually encoded by optimal codons whereas unstructured and structurally flexible regions are encoded by rare codons that can induce pauses in the co-translational folding process.

Codon usage also influences other co-translational folding-related processes. Rare codons are enriched in 5' ends of coding sequences of secreted proteins, and this enrichment was hypothesized to promote membrane targeting and secretion efficiency [99]. In *Saccharomyces cerevisiae*, codon usage in genes encoding some membrane proteins regulates the interaction of nascent polypeptides with the signal recognition particle (SRP), which assists protein translocation across membranes [100]. Co-translational recognition of nascent polypeptides by SRP is enhanced by an elongation pause that is mediated by non-optimal codon clusters 35 to 40 codons downstream of the SRP-binding site. In human cells, co-translational arginylation of  $\gamma$ -actin, which regulates stability of the protein by influencing ubiquitination, is influenced by rare codons that slow elongation rate in the 5' end of the gene [101].

In *Aspergillus nidulans*, a pair of rare codons in the gene encoding the urea transporter are important for the production and localization of the UreA protein [102].

Despite the observed correlations, the codon usage to protein structure relationship is a complex one, likely reflecting the complexity of protein folding for diverse protein structures. Although nonoptimal codons were found to be enriched at domain boundaries in some studies [76, 90, 97], a later study failed to find such a correlation and even found a small enrichment of optimal codons [95]. By analyzing homologous coding sequences across different eukaryotic and prokaryotic species, some rare codon clusters were found to be conserved, suggesting a conserved functional role for codon optimality in regulating the folding of homologous proteins [96]. Unexpectedly, the identified conserved rare codon clusters are preferentially located within conserved protein domains. Although the use of different methodologies and different protein structure predictions and databases underlie some apparent contradictions, differences in effects of codon optimization in different protein types and organisms also indicate the complex nature of the co-translational folding processes. The role of codon usage in different proteins and structure types should be dependent on co-translational folding kinetics of different proteins under different cellular environments.

#### **Conclusions**

A growing body of biochemical, genetic, and bioinformatic evidence indicates that the non-uniform decoding rate across mRNAs mediated by codon usage represents a “code” within genetic codons that promote optimal co-translational protein folding process. The adaptation of codon usage to influence the co-translational folding process is the result of evolutionary selection for protein function. Although there are now a few clear genetic examples showing the importance of codon usage in protein folding, more genetic evidence are needed to demonstrate the broad impact of this mechanism in protein folding. The lack of more genetic examples is largely due to the lack of sensitivity of the laboratory assays used. During evolution, even a minute change in protein function that cannot be detected by currently employed assays can have a major effect on organism survival due to the long time frame. Although codon usage influences co-translational protein folding as demonstrated in bacterial and eukaryotic systems, how it facilitates folding of various types of structures is still unclear. Moreover, although synonymous SNPs are associated with many human diseases [103, 104], more future studies will be needed to reveal whether silent SNPs are broad contributors to human disease development due to the role of codon usage in protein structure and gene expression.

**Abbreviations**

FRQ: FREQUENCY protein; IDPs: Intrinsically disordered domains; PER: Period protein; SNP: Single nucleotide polymorphism; SRP: Signal recognition particle

**Acknowledgements**

This work is supported by grants from the National Institutes of Health (R35GM118118), and the Welch Foundation (I-1560) to Yi Liu. I thank all members of Liu laboratory for assistance and discussions.

**Author's contributions**

Yi Liu wrote the paper and approve the final manuscript. All authors read and approved the final manuscript.

**Funding**

This work is supported by grants from the National Institutes of Health (R35GM118118), and the Welch Foundation (I-1560) to Yi Liu.

**Availability of data and materials**

This study does not contain any original data and materials.

**Ethics approval and consent to participate**

This study does not involve human participants, human data or human tissue.

**Consent for publication**

This study manuscript does not contain any individual person's data.

**Competing interests**

The author has no competing interests.

Received: 13 March 2020 Accepted: 10 August 2020

Published online: 09 September 2020

**References**

- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 1985;2(1):13–34.
- Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 1986;14(13):5125–43.
- Cameron JM. Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics.* 2004;167(3):1293–304.
- Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* 2011;12(1):32–42.
- Chaney JL, Clark PL. Roles for synonymous codon usage in protein biogenesis. *Annu Rev Biophys.* 2015;44:143–66.
- Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol.* 2018;19(1):20–30.
- Komar AA. Synonymous Codon Usage—a Guide for Co-Translational Protein Folding in the Cell. *Mol Biol (Mosk).* 2019;53(6):883–98.
- Quax TE, Claassens NJ, Soll D, van der Oost J. Codon Bias as a means to fine-tune gene expression. *Mol Cell.* 2015;59(2):149–61.
- Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 2000;16(7):287–9.
- Moriyama EN, Powell JR. Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol.* 1997;45(5):514–23.
- Sabi R, Tuller T. Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res.* 2014;21(5):511–26.
- Spencer PS, Siller E, Anderson JF, Barral JM. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol.* 2012;422(3):328–35.
- Zhou M, Wang T, Fu J, Xiao G, Liu Y. Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol Microbiol.* 2015; 97(5):974–87.
- Friberg M, von Rohr P, Gonnet G. Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast.* 2004;21(13):1083–93.
- Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci U S A.* 1999;96(8):4482–7.
- Carlini DB, Stephan W. In vivo introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. *Genetics.* 2003;163(1):239–43.
- Zhou Z, Dang Y, Zhou M, Li L, Yu CH, Fu J, Chen S, Liu Y. 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(41):E6117–25.
- Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends Biotechnol.* 2004;22(7):346–53.
- Xu Y, Ma PJ, Shah P, Rokas A, Liu Y, Johnson CH. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature.* 2013; 495(7439):116–20.
- Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature.* 2013;495(7439):111–5.
- Hense W, Anderson N, Hutter S, Stephan W, Parsch J, Carlini DB. Experimentally increased codon bias in the *Drosophila Adh* gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. *Genetics.* 2010;184(2):547–55.
- Lampson BL, Pershing NL, Prinz JA, Lacsina JR, Marzluff WF, Nicchitta CV, MacAlpine DM, Counter CM. Rare codons regulate KRAS oncogenesis. *Curr Biol.* 2013;23(1):70–5.
- Fu J, Dang Y, Counter C, Liu Y. Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J Biol Chem.* 2018; 293(46):17929–40.
- Jeacock L, Faria J, Horn D. Codon usage bias controls mRNA and protein abundance in trypanosomatids. *eLife.* 2018;7:e32496.
- Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature.* 2012;484(7395):538–41.
- Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147(4):789–802.
- Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 2012; 8(3):e1002603.
- Charneski CA, Hurst LD. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 2013;11(3):e1001508.
- Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, Liu Y. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol Cell.* 2015;59(5):744–54.
- Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* 2016;14(7):1787–99.
- Yang Q, Yu CH, Zhao F, Dang Y, Wu C, Xie P, Sachs MS, Liu Y. eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic Acids Res.* 2019;47(17):9243–58.
- Hussmann JA, Patchett S, Johnson A, Sawyer S, Press WH. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.* 2015;11(12):e1005732.
- Gingold H, Pilpel Y. Determinants of translation efficiency and accuracy. *Mol Syst Biol.* 2011;7:481.
- Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell.* 2016;166(3):679–90.
- Lyu X, Yang Q, Li L, Dang Y, Zhou Z, Chen S, Liu Y. Adaptation of codon usage to tRNA I34 modification controls translation kinetics and proteome landscape. *PLoS Genet.* 2020;16(6):e1008836.
- Zhao F, Yu CH, Liu Y. Codon usage regulates protein structure and function by affecting translation elongation speed in *Drosophila* cells. *Nucleic Acids Res.* 2017;45(14):8484–92.
- Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, et al. Codon optimality is a major determinant of mRNA stability. *Cell.* 2015;160(6):1111–24.
- Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff JD, Valecha M, Everett JK, Acton TB, et al. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature.* 2016;529(7586):358–63.
- Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, Giraldez AJ. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* 2016;35:2087–103.

40. Wu Q, Medina SG, Kushawah G, DeVore ML, Castellano LA, Hand JM, Wright M, Bazzini AA. Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife*. 2019;8:e45396.
41. Mishima Y, Tomari Y. Codon usage and 3' UTR length determine maternal mRNA stability in Zebrafish. *Mol Cell*. 2016;61(6):874–85.
42. Kramer EB, Farabaugh PJ. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *Rna*. 2007;13(1):87–96.
43. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008;134(2):341–52.
44. Mordret E, Dahan O, Asraf O, Rak R, Yehonadav A, Barnabas GD, Cox J, Geiger T, Lindner AB, Pilpel Y. Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation Fidelity. *Mol Cell*. 2019;75(3):427–41 e425.
45. Buschauer R, Matsuo Y, Sugiyama T, Chen YH, Alhusaini N, Sweet T, Ikeuchi K, Cheng J, Matsuki Y, Nobuta R, et al. The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science*. 2020;368(6488).
46. Kudla G, Lipinski L, Caffin F, Helwak A, Zylitz M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol*. 2006;4(6):e180.
47. Newman ZR, Young JM, Ingolia NT, Barton GM. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci U S A*. 2016;113(10):E1362–71.
48. Zhou Z, Dang Y, Zhou M, Yuan H, Liu Y. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *eLife*. 2018;7:e33569.
49. Chamary JV, Hurst LD. Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends Genet*. 2005;21(5):256–9.
50. Komar AA. A pause for thought along the co-translational folding pathway. *Trends Biochem Sci*. 2009;34(1):16–24.
51. Fu J, Murphy KA, Zhou M, Li YH, Lam VH, Tabuloc CA, Chiu JC, Liu Y. Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev*. 2016;30(15):1761–75.
52. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*. 2007;315(5811):525–8.
53. Kirchner S, Cai Z, Rauscher R, Kasteic N, Anding M, Czech A, Kleizen B, Ostedgaard LS, Braakman I, Sheppard DN, et al. Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLoS Biol*. 2017;15(5):e2000779.
54. Kim SJ, Yoon JS, Shishido H, Yang Z, Rooney LA, Barral JM, Skach WR. Translational tuning optimizes nascent protein folding in cells. *Science*. 2015;348(6233):444–8.
55. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol*. 1996;260(5):649–63.
56. Percudani R, Pavesi A, Ottonello S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol*. 1997;268(2):322–30.
57. Sorensen MA, Kurland CG, Pedersen S. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol*. 1989;207(2):365–77.
58. Bonekamp F, Dalboge H, Christensen T, Jensen KF. Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in *Escherichia coli*. *J Bacteriol*. 1989;171(11):5812–6.
59. Chevance FF, Le Guyon S, Hughes KT. The effects of codon context on in vivo translation speed. *PLoS Genet*. 2014;10(6):e1004392.
60. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009;324(5924):218–23.
61. Artieri CG, Fraser HB. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res*. 2014;24(12):2011–21.
62. Nakahigashi K, Takai Y, Shiwa Y, Wada M, Honma M, Yoshikawa H, Tomita M, Kanai A, Mori H. Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. *BMC Genomics*. 2014;15:1115.
63. Gerashchenko MV, Gladyshev VN. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res*. 2014;42(17):e134.
64. Yan X, Hoek TA, Vale RD, Tanenbaum ME. Dynamics of translation of single mRNA molecules in vivo. *Cell*. 2016;165(4):976–89.
65. Chiti F, Dobson CM. Protein Misfolding, amyloid formation, and human disease: a summary of Progress over the last decade. *Annu Rev Biochem*. 2017;86:27–68.
66. Hartl FU. Protein Misfolding diseases. *Annu Rev Biochem*. 2017;86:21–6.
67. Thommen M, Holtkamp W, Rodnina MV. Co-translational protein folding: progress and methods. *Curr Opin Struct Biol*. 2017;42:83–9.
68. Komar AA. Unraveling co-translational protein folding: concepts and methods. *Methods*. 2018;137:71–81.
69. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574–81.
70. Holtkamp W, Kocik G, Jager M, Mittelstaet J, Komar AA, Rodnina MV. Cotranslational protein folding on the ribosome monitored in real time. *Science*. 2015;350(6264):1104–7.
71. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. *Curr Opin Struct Biol*. 2010;20(1):33–45.
72. Shiber A, Doring K, Friedrich U, Klann K, Merker D, Zedan M, Tippmann F, Kramer G, Bukau B. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*. 2018;561(7722):268–72.
73. Purvis IJ, Bettany AJ, Santiago TC, Coggins JR, Duncan K, Eason R, Brown AJ. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol*. 1987;193(2):413–7.
74. O'Brien EP, Vendruscolo M, Dobson CM. Prediction of variable translation rate effects on cotranslational protein folding. *Nat Commun*. 2012;3:868.
75. Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J Mol Biol*. 2010;396(5):1310–8.
76. Zhang G, Hubalewska M, Ignatova Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*. 2009;16(3):274–80.
77. Komar AA, Lesnik T, Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett*. 1999;462(3):387–91.
78. Cortazzo P, Cervenansky C, Marin M, Reiss C, Ehrlich R, Deana A. Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem Biophys Res Commun*. 2002;293(1):537–41.
79. Sander IM, Chaney JL, Clark PL. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J Am Chem Soc*. 2014;136(3):858–61.
80. Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, Rodnina MV, Komar AA. Synonymous codons direct Cotranslational folding toward different protein conformations. *Mol Cell*. 2016;61(3):341–51.
81. Walsh IM, Bowman MA, Soto Santarriaga IF, Rodriguez A, Clark PL. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc Natl Acad Sci U S A*. 2020;117(7):3528–34.
82. Alexaki A, Hettiarachchi GK, Athey JC, Katneni UK, Simhadri V, Hamasaki-Katagiri N, Nanavaty P, Lin B, Takeda K, Freedberg D, et al. Effects of codon optimization on coagulation factor IX translation and structure: implications for protein and gene therapies. *Sci Rep*. 2019;9(1):15449.
83. Hunt R, Hettiarachchi G, Katneni U, Hernandez N, Holcomb D, Kames J, Alnifaidy R, Lin B, Hamasaki-Katagiri N, Wesley A, et al. A Single Synonymous Variant (c.354G>A [p.P118P]) in ADAMTS13 Confers Enhanced Specific Activity. *Int J Mol Sci*. 2019;20(22):5734.
84. Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol*. 2008;18(6):756–64.
85. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005;6(3):197–208.
86. Tompa P. Unstructural biology coming of age. *Curr Opin Struct Biol*. 2011; 21(3):419–25.
87. Tang CT, Li S, Long C, Cha J, Huang G, Li L, Chen S, Liu Y. Setting the pace of the *Neurospora* circadian clock by multiple independent FRQ phosphorylation events. *Proc Natl Acad Sci U S A*. 2009;106(26):10722–7.
88. Baker CL, Kettenbach AN, Loros JJ, Gerber SA, Dunlap JC. Quantitative proteomics reveals a dynamic interactome and phase-specific phosphorylation in the *Neurospora* circadian clock. *Mol Cell*. 2009;34(3):354–63.
89. Pattanayek R, Wang J, Mori T, Xu Y, Johnson CH, Eglı M. Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol Cell*. 2004;15(3):375–88.
90. Thanaraj TA, Argos P. Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci*. 1996;5(10):1973–83.
91. Oresic M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. *J Mol Biol*. 1998;281(1):31–48.
92. Gu W, Zhou T, Ma J, Sun X, Lu Z. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems*. 2004;73(2):89–97.
93. Xie T, Ding D. The relationship between synonymous codon usage and protein structure. *FEBS Lett*. 1998;434(1–2):93–6.
94. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*. 2013;20(2):237–43.



95. Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res.* 2010;38(19):6719–28.
96. Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, Li J, Emrich S, Clark PL. Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol.* 2017;13(5):e1005531.
97. Zhang G, Ignatova Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS One.* 2009;4(4):e5036.
98. Zhou T, Weems M, Wilke CO. Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol Biol Evol.* 2009;26(7):1571–80.
99. Clarke TF, Clark PL. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. *BMC Genomics.* 2010;11:118.
100. Pechmann S, Chartron JW, Frydman J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol.* 2014;21(12):1100–5.
101. Zhang F, Saha S, Shabalina SA, Kashina A. Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science.* 2010;329(5998):1534–7.
102. Sanguinetti M, Iriarte A, Amillis S, Marin M, Musto H, Ramon A. A pair of non-optimal codons are necessary for the correct biosynthesis of the *Aspergillus nidulans* urea transporter, UreA. *R Soc Open Sci.* 2019;6(11):190773.
103. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet.* 2011;12(10):683–91.
104. Chen R, Davydov EV, Sirota M, Butte AJ. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One.* 2010;5(10):e13574.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

