Review

Codon usage: Nature's roadmap to expression and folding of proteins

Evelina Angov

Division of Malaria Vaccine Development, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Biomedical and biotechnological research relies on processes leading to the successful expression and production of key biological products. High-quality proteins are required for many purposes, including protein structural and functional studies. Protein expression is the culmination of multistep processes involving regulation at the level of transcription, mRNA turnover, protein translation, and post-translational modifications leading to the formation of a stable product. Although significant strides have been achieved over the past decade, advances toward integrating genomic and proteomic information are essential, and until such time, many target genes and their products may not be fully realized. Thus, the focus of this review is to provide some experimental support and a brief overview of how codon usage bias has evolved relative to regulating gene expression levels.

1 Introduction

Biomedical and biotechnological research relies on processes leading to the successful expression and production of key biological products. High-quality proteins are required for many purposes, including protein structural and functional studies. Protein expression is the culmination of multistep processes involving regulation at the level of transcription, mRNA turnover, protein translation, and post-translational modifications leading to the formation of a stable product. Although significant strides have been achieved over the past decade, advances toward integrating genomic and proteomic information are essential, and until such time, many target genes and their synthetic potential may not be fully realized. Thus, the focus of this

review is to provide some experimental support and a brief overview of how codon usage bias has evolved relative to regulating gene expression levels.

Due to their apparent "silent" nature, synonymous codon substitutions have long been thought to be inconsequential. In recent years, this long-held dogma has been refuted by evidence that even a single synonymous codon substitution can have significant impact on gene expression levels, protein folding, and protein cellular function [1–4]. It is certainly conceivable that, by design, nature has provided the basic instructions to direct efficient protein synthesis and folding through the information encoded at the genetic code level. For most sequenced genomes, synonymous codons are not used at equal frequencies. Sixty-one codons specify the twenty amino acids found commonly in protein sequences; most of these are specified by more than one synonymous codon, with the exception of methionine and tryptophan. The redundancy in the genetic code may have evolved as a way to preserve structural information of proteins within the nucleotide content [5]. In unicellular organisms, high-frequency-usage codons correlate with abundant cognate isoacceptor tRNA molecules and have

Correspondence: Dr. Evelina Angov, Division of Malaria Vaccine Development, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD 20910, USA
E-mail: Evelina.angov@us.army.mil

Abbreviations: AU, adenine:thymidine; **CAI**, codon adaptation index; **GC**, guanidine:cytosine; **GFP**, green fluorescent protein; **ORF**, open reading frame; **RBS**, ribosomal binding site; **SD**, Shine and Dalgarno

evolved to optimize translational efficiency [6–10]. In bacteria, the co-evolution of tRNA isoacceptor abundance and codon bias is most evident for proteins from highly expressed genes involved in essential cellular functions, such as protein synthesis and cell energetics, and are likely to have co-developed as a result of a positive selective force to achieve faster translation rates and translational accuracy [7, 8, 11–13]. A similar relationship between codon usage bias and intracellular tRNA abundance levels on translation efficiency is also observed for some eukaryotes, *C. elegans* and *D. melanogaster* [14, 15]. Although somewhat controversial, evidence for the ubiquity of codon bias functionality in translational control was recently extended to the tissue-specific level in eukaryotes [16–19].

Codon bias has been extensively observed and varies widely within genomic DNA sequences of different organisms [6, 7, 20, 21]. Even with relatively limited genomic sequence data, early studies in prokaryotes and yeast established the fundamental existence of codon bias in encoded DNA [21–24]. From these earlier studies, a positive correlation between codon usage, gene expression level, and growth efficiency of prokaryotic cells [25] was established and used to generate codon adaptation indices (CAI) [26]. CAI define a relative "adaptiveness" to codons and are widely used to predict expression levels from genes and to approximate the success of heterologous gene expression. A caveat and limitation is that the predictive value of CAI is directly dependent on the genes used to establish the "reference set" and may have relatively limited value for predicting expression from genes not reflected by the codon bias found in the reference set. Recognizing these limitations, investigators have developed more "universal" CAI, which measure codon bias based on reference sets that are not necessarily derived from preferred, "highly" expressed genes, but from all known coding sequences for a specific organism [27]. Notwithstanding these limitations, CAI do not necessarily reflect all possible factors that influence gene expression levels per se, for example, the efficiency of ribosome binding and translation initiation [28]. Similar selective forces on codon co-adaptation to tRNA pools have also been seen for some eukaryotes, namely, yeast [21], Drosophila [29, 30], *Caenorhabditic elegans* [31], *Arabidopsis thaliana* [32], and *Xenopus laevis* [33]. Unlike prokaryotes and some eukaryotic organisms, the evolutionary pressure on codon bias in the human genome cannot be solely explained by forces on selection for translational efficiency [16, 34], but are also influenced by the extensive guanidine:cytosine (GC)

content found in isochore structures on chromosomes (GC-rich DNA <100 kb) [35] and by the effects of mRNA secondary-structure stability [36].

The recent volume of completed nucleotide sequence and protein structural data for many species has allowed for comprehensive analyses into the relationship between genomic GC content, codon usage bias, and gene expression levels. Knight et al. examined the codon usage pattern for a large set of species (311 prokaryotes, 28 archaea and 257 eukaryotes) using a relatively simple quantitative mutational model and predicted that the GC mutational bias on genomes rather than codon usage bias on GC content was the driving force for both codon and amino acid usage in prokaryotes and eukaryotes [37]. These conclusions were strengthened by the replication of the observations across all three cellular life domains, archaea, bacteria, and eukaryota, suggesting highly conserved mechanisms for mutational and selection equilibrium at the genome level. These findings were corroborated in a more recent study using a stochastic continuous Markov chain model for GC-biased synonymous point substitutions across hundreds of bacterial, plant, and human genes. The model established that GC mutational bias is indeed a dominant factor determining codon bias [38]. However, other factors could account for the codon bias of an organism not considered within the model, for instance, transcriptional efficiency in bacteria and the GC skew in mammals (GC isochores).

2 Role of mRNA structure on gene expression

It is widely held that mRNA secondary structure influences translational efficiency. In bacteria, formation of strong hairpin loops centered at the Shine-Dalgarno (SD) ribosome binding site (RBS) and the initiation codon (AUG) can significantly reduce expression levels [39]. With regards to heterologous protein expression, synonymous codon substitution at the 5'-end of mRNA can impact mRNA structure and stability and thus the relative kinetics of translation at both the level of translation initiation [40] and elongation [41]. As such, there are at least two influences at the nucleotide level that govern regulation of expression from the 5'-end of open reading frames (ORFs): the GC content and formation of mRNA secondary structure (discussed here), and codon usage bias and efficiency of translation initiation (discussed later in the context of slowing translation). Gu et al. reported that mRNA secondary-structure stability corre-

lated with both GC content and codon usage [42]. In their study of 340 genomes from bacteria, archaea, fungi, plants, insects, fishes, birds, and mammals, they revealed that, with the exception of birds and mammals, the 5′-end translation initiation sites had reduced mRNA secondary structure. The universality of their observations suggested that reduced mRNA stability at 5′-ends of ORFs may have evolved as a result of selective pressure toward efficient translation initiation. These observations were further expanded by Allert et al., who analyzed the nucleotide composition of 816 fully sequenced bacterial genomes [43]. They found that nucleotide composition at both the 5′- and 3′-ends of an ORF played an important role in gene expression levels. Their analysis revealed a bias toward higher adenine:thymidine (AU) content in the first and last 35 bases of an ORF relative to the central coding region of the ORF. Not surprisingly, the highest expression levels were observed when all three parameters (AU content, mRNA secondary-structure content, and CAI) were considered at the same time. By using a systematic approach, the impact of random, synonymous codon substitutions on mRNA secondary-structure stability was experimentally derived using 154 gene variants expressing green fluorescent protein (GFP) in *E. coli* [44]. These investigators concluded that synonymous codon substitutions that reduced mRNA structure stability, particularly in the first forty nucleotides of the transcript, were significantly correlated with GFP protein abundance. They attributed the majority of the effect on GFP expression to the local nucleotide content and not to codon usage bias or CAI. Supek and Smuc [45] recently refuted Kudla et al. [44] using nonlinear regression analysis on the same data and argued that the effects of CAI and codon bias were masked by the inherent strong mRNA structure found in GFP. They argued that codon usage may have a significant role in gene expression levels, particularly in cases in which 5′-mRNA is weakly associated with secondary structure [45].

3 Role of rare codons in gene translation

Mounting evidence exists that low-frequency-usage codons within a coding sequence can provide the genetic instruction that regulates the rate of protein synthesis to allow for some secondary and tertiary structure formation by the nascent polypeptide [2, 46]. Computational analysis of the available *E. coli* genome and protein structure databases identified that high-frequency-usage codons are mainly associated with structural elements, such as

alpha helices, whereas clusters of lower frequency usage codons are more likely to be associated with beta-strands, random coils, and structural domain boundaries [47]. These findings confirmed earlier speculations that the positioning and clustering of codons with different usage frequencies was non-random and played a role in gene expression [6, 8, 48].

Some experimental evidence for the role of synonymous codon substitutions, particularly in slow translating regions on mRNA, and their impact on protein structure or function are briefly summarized. Systematic single-codon substitutions of five low-frequency codons in a linker region of the *Echinococcus granulosus* fatty acid-binding protein 1 (EgFABP1) gene to synonymous higher frequency usage codons significantly impacted protein solubility [49]. These results suggest that codon substitution in a slower translation region alters the kinetics of translation and in vivo folding. In another example, synonymous substitution to eliminate rare codons in chloramphenicol acetyltransferase yielded significantly reduced specific enzyme activity [50], also suggesting a dramatic effect on folding. In the framework of heterologous gene expression, deleteriously placed translational pause sites have also been shown to lead to translational frame-shifting [51, 52] and to protein misfolding [53]. A single silent mutation in the human MDR1 gene caused the P-glycoprotein to have an altered folding pathway. Other studies also point to low-frequency-usage codons in "pausing" translation to allow local protein-structure formation [18, 54, 55].

Alternatively, targeted substitution to introduce low-frequency codons at the 5′-end of a coding sequence enhanced heterologous expression levels for streptokinase from the *src* gene of *Streptococcus equisimilis* in *E. coli* [56]. Arguably, similar to the recent findings by Tuller et al., inclusion of rare codons at key positions allowed for stabilization of the ribosomal initiation complex, or concomitantly, reduced mRNA secondary structure, thereby yielding higher levels of recombinant protein [41]. They reported that the positioning of low-frequency codons to the 5′-end of an ORF is a universally conserved phenomenon and is a putative mechanism for regulating gene expression. A pattern emerges that low-frequency codon bias, particularly at the 5′-end of the ORF, has evolved as a means to regulate the efficiency of translation initiation. In this model, the first 30–50 codons act as the "ramp" for slowing translation initiation to allow for efficient ribosome binding on mRNA and to avoid ribosome bottlenecking. Therefore, selection for "poorly" adapted codons at the 5′-end on mRNA is a mechanism to reduce deleterious ribosomal traffic jams
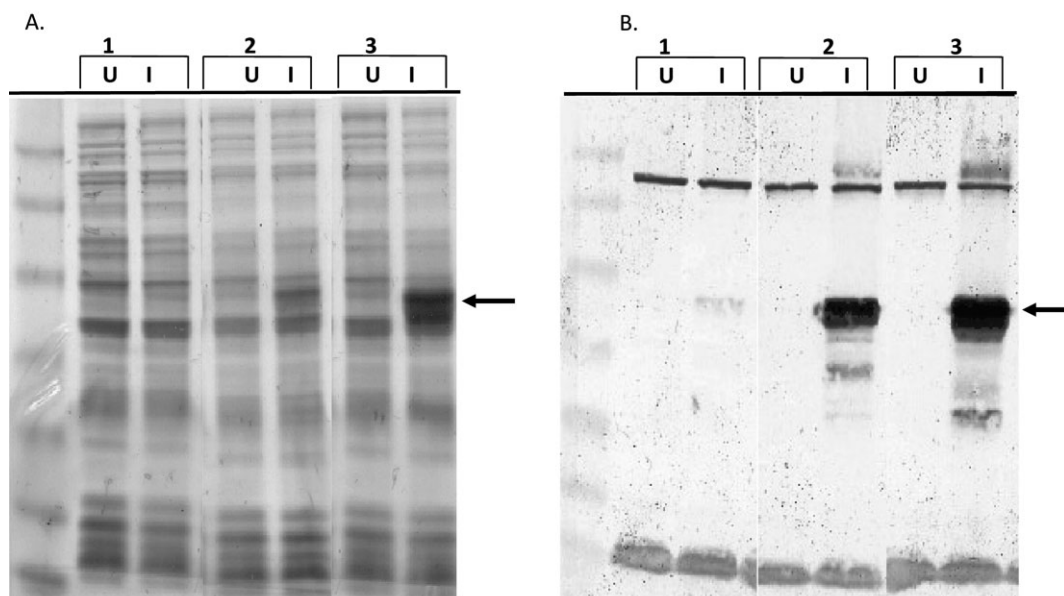
Figure 1. Codon harmonization improves expression level of *P. falciparum* malaria protein, MSP1$_{42}$ in *E. coli*. (**A**) Coomassie Blue-stained gel on total cell lysates, uninduced (U) and induced (I) cells (3 h with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG)), expressing *P. falciparum* MSP1$_{42}$ from various constructs. (**B**) Western blot on total cell lysates (same as above) probed with rabbit polyclonal anti-MSP1$_{42}$ antibodies for U and I cells (3 h with 0.1 mM IPTG), expressing *P. falciparum* MSP1$_{42}$ from various constructs. Lane 1: MSP1$_{42}$ with a single, synonymous codon substitution; lane 2: MSP1$_{42}$ codon-harmonized at the 5'-end, first thirty codons; and lane 3: MSP1$_{42}$ full gene sequence codon-harmonized. Arrow points to the expressed protein band.

on the messenger and control the rate of peptide elongation [57–59]. Interestingly, the second codon position immediately following the initiation codon, AUG, is more likely to be a high-frequency-usage codon, and thereby, would be translated more quickly. This codon organization would ensure efficient translation initiation, release, and re-cycling of the initiator tRNA for the next round of initiation. And finally, in the case of bacterial exported proteins, rare codon clusters at the 5'-end of the ORF may allow time for the nascent translocating peptide–ribosome complex to reach target cellular membranes [60]. In prokaryotes, these rare codon clusters have adapted to act in a similar fashion as the signal recognition particles (SRP) in eukaryotes.

From a practical perspective, we applied the codon harmonization approach iteratively to optimize expression levels in *E. coli* for the malaria *P. falciparum*, Merozoite Surface Protein 1, (MSP1$_{42}$). In the first instance, we targeted a putative translational pause site that was predicted to be in disharmony when expressed in *E. coli* [61]. Substitution to harmonize the codon frequency toward that of the native sequence codon resulted in an approximately ten-fold improvement in soluble-protein expression (Fig. 1). Partial purification by nickel-affinity chromatography allowed for a more quantitative comparison of expression levels relative to the native sequence, that is, to compare

native and single-pause-site mutant (Fig. 2A and B). Interestingly, in the second instance, when we codon-harmonized the first thirty codons at the 5'-end of the ORF, we saw another dramatic improvement in expression levels (Fig. 1, lane 2, and Fig. 2C). In the final instance, codon harmonization of the full gene sequence of MSP1$_{42}$ yielded the highest level of soluble protein compared with both the
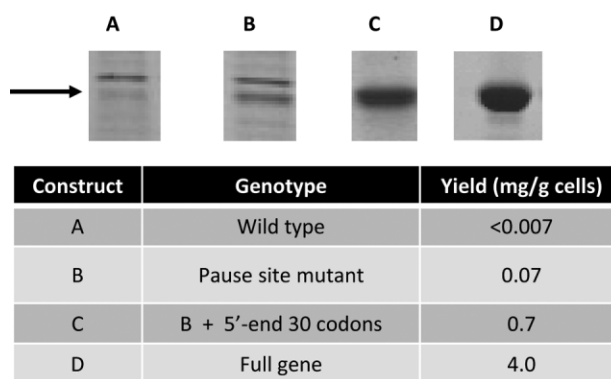

Construct	Genotype	Yield (mg/g cells)
A	Wild type	<0.007
B	Pause site mutant	0.07
C	B + 5'-end 30 codons	0.7
D	Full gene	4.0

Figure 2. Quantitation of expression levels of *P. falciparum* MSP1$_{42}$ following partial purification. Cells expressing various *P. falciparum* MSP1$_{42}$ constructs were purified by nickel-affinity chromatography. Construct A represents the yield from the native gene sequence of *P. falciparum* MSP1$_{42}$. Construct B represents the yield from the single, synonymous codon substitution (pause-site mutant). Construct C represents the yield from the 5'-end first thirty harmonized codons. Construct D represents the yield from the full gene codon-harmonized sequence. Arrow points to the partially purified band.

single-pause-site mutant and the 5'-end harmonized variants (Fig. 1, lane 3, and Fig. 2D) [62, 63]. These results demonstrate that incremental improvements in expression levels can be achieved, however, at least for the case of this malaria protein, maximal improvements were only achieved when the full gene sequence was recoded for heterologous expression, yielding approximately 1000-fold higher protein yields than the native gene sequence. An excellent review on the role of sequence codon bias and gene expression was recently published by Plotkin and Kudla [4].

4 Practical considerations for heterologous expression

Genetic redundancy at the codon level purportedly increases the organisms' resistance to mutations, however, even a single nucleotide change, leading to a synonymous codon substitution, can impact protein expression, protein structure, and/or function [3, 10, 49, 64–70]. From a practical standpoint, the common causes for failures in heterologous gene expression are primarily related to the disparities in codon bias, mRNA secondary structure and stability, gene product toxicity, and product solubility [71, 72]. Currently, it is accepted that nonoptimal codon content can limit expression of heterologous proteins due to limiting available cognate tRNAs in the expression host. Introduction of rare codons that are incongruent with the native gene sequence during heterologous expression can lead to reduced translation rates and overall expression levels [73, 74]. The disparities in codon usage can cause significant stress on host-cell metabolic and translational processes. Various strategies have been used to minimize the bias in codon usage for heterologous expression. In bacteria and simple eukaryotes, the observation that highly expressed genes have strong codon bias toward "preferred" codons led to the development of algorithms for "codon optimization" that substitute codons in a target sequence toward preferred high-frequency codons from the expression host. The premise for this approach is that by introducing the most abundant codons throughout the length of the sequence the resulting protein would be expressed at high levels, primarily because cognate isoacceptor tRNA molecules are not rate limiting [71, 72]. This approach has been successful for the heterologous production of some proteins [75], however, in some cases, the high levels of protein expressed have led to the formation of insoluble products sequestered in inclusion bodies [76]. An alternative approach has been to adjust the intracellular tRNA isoacceptor concentrations directly by co-expressing copies of rare tRNA molecules [71, 77–80]. This approach resolves some, but not necessarily all, codon bias issues. A third approach to recode a target gene sequence is to "match" the codon usage bias inherent in the native host more closely when expressed in the heterologous host and is referred to as "codon harmonization" [62]. In this approach, two features are primarily considered: first, that the expression host synonymous codon usage should more closely match that of the native gene host codon usage, and second, that putative nonstructural segments between local alpha-helical content are coded to translate more slowly. Predicting nonstructural segments on proteins without structures obtained by crystallography or NMR spectroscopy is highly empirical and is based on the earlier report by Thanaraj and Argos, which identified ten out of twenty amino acids with bulky hydrophobic side chains or side chains that can hydrogen bond to the peptide backbone as being more likely to be found in nonstructural segments for *E. coli* proteins [81]. Slowing ribosomal translation through these regions may allow co-translational folding by allowing nascent flanking structural elements to gain some structure prior to synthesis of the next element. The codon harmonization approach was successfully applied to express several *P. falciparum* malaria target antigens in *E. coli* [62, 82, 83]. A list of recent, although certainly not exhaustive, review articles focusing on heterologous expression of foreign proteins from various host systems is provided [72, 84–92].

Nascent polypeptide synthesis is complex and is influenced by many factors, such as the rate of tRNA binding, the kinetics of translation and protein folding, the environment within the ribosomal tunnel, and the interaction with chaperones [46, 48, 93, 94]. Efficient protein synthesis and peptide folding occurs co-translationally within the protective environment and is best described for the prokaryotic ribosomal tunnel [95, 96]. Direct interaction of the nascent polypeptide with the tunnel can initiate protein folding by briefly stalling or arresting translation, probably by charge-specific interactions between charged amino acids and the tunnel [97, 98]. In contrast to these direct interactions, the rate of protein translation is also influenced by the local mRNA structure and the presence of slowly translated codons. A crude model representing the kinetics of translation and nascent protein synthesis is shown in Fig. 3. The model depicts (Fig. 3A) a single ribosome binding at the translation initiation complex centered at the AUG codon; at the 5'-end of the mRNA, the double-lined region depicts the "ramp" or slowly translated re-
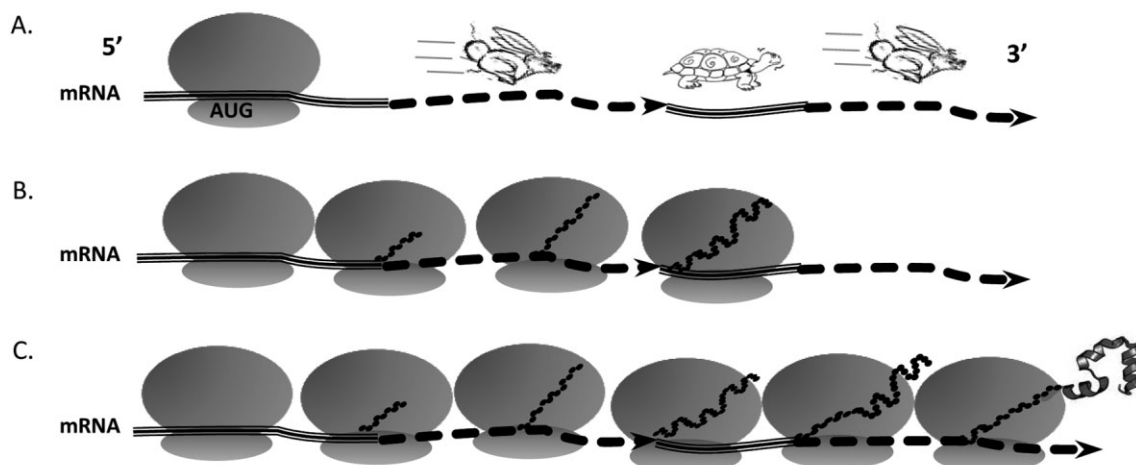
Figure 3. Schematic model of co-translational folding on mRNA by ribosomes. (**A**) Ribosomal complex centered on the translation initiation site, AUG (initiation codon). (**B**) Nascent polypeptide synthesis within the protective environment of the ribosomal tunnel. (**C**) Putative translational pause sites in conjunction with co-translational folding occur within the ribosomal tunnel. Differences in codon usage frequency are shown as thick dashed lines with arrowheads for areas representing high-frequency-usage codons, and therefore, translating rapidly (hare) and regions that are double lined represent segments of lower frequency usage codons (i.e., putative pause sites; tortoise) where translation proceeds more slowly to allow nascent polypeptide folding.

gion; high-frequency codons are translated quickly within the protective ribosomal tunnel (Fig. 3B; tunnel is not shown); and as the translocating ribosome reaches an mRNA segment encoded by low-frequency-usage codons, the rate of translation slows, and allows for the preceding nascent peptide

to gain some helical structure within the tunnel (Fig. 3C).

Over the past decade, several codon adaptation algorithms have been developed and are available through public website access (Table 1). Needless to say the success or failure of applying any ap-

Table 1. Codon usage analysis and optimization tools

Algorithm	Description	Citation
$O_{RF}O_{PT}$	Tunes regional nucleotide composition, codon choice, mRNA secondary structure	[43]
Gene Composer	Gene and protein engineering using PCR-based gene assembly and PIPE cloning.	[100]
Codon Harmonization	Adjusts codon usage by predicting translational pauses and matching codon usage on native gene hosts in heterologous hosts	[62, 63]
GASCO	Codon optimization based on host genome codon bias with the identification of desirable/undesirable motifs http://miracle.igib.res.in/gasco/	[101]
QPSO	Quantum-behaved particle swarm optimization	[102]
OPTIMIZER	Codons computed based on highly expressed prokaryotic genes, based on CAI http://genomes.urv.es/OPTIMIZER	[103]
Gene Designer (DNA 2.0 Inc.)	Synthetic biology workbench using advanced optimization algorithms and an intuitive drag-and-drop graphic interface	[104]
Synthetic Gene Designer	Enhanced functionality enabling users to work with nonstandard genetic codes, with user-defined patterns of codon usage, and an expanded range of methods for codon optimization	[105]
JCat	Codon adaptation with the avoidance of cleavage sites http://ww.prodoric.de/JCat	[106]
GeMS	Gene design functions, including restriction site prediction, codon optimization for expression, stem-loop determination, and oligonucleotide design	[107]
UpGene	SIV/HIV coding sequence adaptation for eukaryotic expression http://www.vectorcore.pitt.edu/upgene.htm	[108]

proach for optimal heterologous protein expression is likely to be sequence dependent and is not necessarily theoretically predictable. Achieving a level of understanding for how genomic codon usage bias has evolved to regulate gene expression in organisms will greatly facilitate the development of synthetic DNA design parameters for optimal heterologous protein production.

5 Summary

Significant advances have been made in the past decade toward revealing the role of codon bias and synonymous codon substitution and the impact on regulating native gene expression, mRNA secondary structure, and protein function and structure. Codon usage bias generally reflects a balance between mutational forces and natural selection, leading to optimal translational efficiency. Limiting tRNA isoacceptor pools during translation can have a significant, negative effect on the accuracy of translation by impacting the progression of ribosomes on mRNA, leading to ribosomal stalling or queuing, premature translational termination, translational frame-shifting, and amino acid misincorporation. Clearly, the extensive body of information included in genomic and proteomic databases has allowed for comprehensive surveys of genes and their proteins, and has redefined the roadmap that is used for efficient protein translation. Well-adapted codons, that is, preferred codons, could confer a metabolic advantage by selecting for translation efficiency and reducing the impact of misfolded proteins. Thus, with a view toward developing optimal strategies for synthetic gene design, increasing the relative AU codon content (i.e., lowering mRNA hairpin structure stability) in the termini of an ORF (particularly the 5′-end) can lead to dramatic improvements in expression levels [99]. The roadblocks to heterologous expression can be alleviated by considering sequence GC content, and thus, codon bias relative to the host expression system and the native gene sequence. Improvements can be augmented by avoiding strong mRNA secondary structures primarily at the 5′-end of coding sequences, allowing for more stable translation initiation complexes and ensuring impediment-free launches of ribosomes on mRNA required for efficient peptide translation and elongation.

6 References

[1] Chamary, J. V., Parmley, J. L., Hurst, L. D., Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* 2006, *7*, 98–108.

[2] Marin, M., Folding at the rhythm of the rare codon beat. *Biotechnol. J.* 2008, *3*, 1047–1057.

[3] Saunders, R., Deane, C. M., Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res.* 2010, *38*, 6719–6728.

[4] Plotkin, J. B., Kudla, G., Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* 2011, *12*, 32–42.

[5] Zull, J. E., Smith, S. K., Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem. Sci.* 1990, *15*, 257–261.

[6] Ikemura, T., 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.* 1981, *151*, 389–409.

[7] Ikemura, T., Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 1985, *2*, 13–34.

[8] Bulmer, M., Coevolution of codon usage and transfer RNA abundance. *Nature* 1987, *325*, 728–730.

[9] Dong, H., Nilsson, L., Kurland, C. G., Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 1996, *260*, 649–663.

[10] Lithwick, G., Margalit, H., Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res.* 2003, *13*, 2665–2673.

[11] Ehrenberg, M., Kurland, C. G., Costs of accuracy determined by a maximal growth rate constraint. *Q. Rev. Biophys.* 1984, *17*, 45–82.

[12] Rocha, E. P., Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* 2004, *14*, 2279–2286.

[13] Sharp, P. M., Emery, L. R., Zeng, K., Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc., B* 2010, *365*, 1203–1212.

[14] Duret, L., Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 2002, *12*, 640–649.

[15] 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*, 287–289.

[16] Plotkin, J. B., Robins, H., Levine, A. J., Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 12 588–12 591.

[17] Semon, M., Lobry, J. R., Duret, L., No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Mol. Biol. Evol.* 2006, *23*, 523–529.

[18] Parmley, J. L., Huynen, M. A., Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. *PLoS Genet.* 2009, *5*, e1000548.

[19] Waldman, Y. Y., Tuller, T., Shlomi, T., Sharan, R., Ruppin, E., Translation efficiency in humans: Tissue specificity, global optimization and differences between developmental stages. *Nucleic Acids Res.* 2010, *38*, 2964–2974.

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

[21] Bennetzen, J. L., Hall, B. D., Codon selection in yeast. *J. Biol. Chem.* 1982, *257*, 3026–3031.

[22] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., Mercier, R., Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 1981, *9*, r43–74.

[23] Grosjean, H., Fiers, W., Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 1982, *18*, 199–209.

[24] Gouy, M., Gautier, C., Codon usage in bacteria: Correlation with gene expressivity. *Nucleic Acids Res.* 1982, *10*, 7055–7074.

[25] Kurland, C. G., Ehrenberg, M., Growth-optimizing accuracy of gene expression. *Annu. Rev. Biophys. Biophys. Chem.* 1987, *16*, 291–317.

[26] Sharp, P. M., Li, W. H., The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987, *15*, 1281–1295.

[27] Carbone, A., Zinovyev, A., Kepes, F., Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 2003, *19*, 2005–2015.

[28] Berg, O. G., Kurland, C. G., Growth rate-optimised tRNA abundance and codon usage. *J. Mol. Biol.* 1997, *270*, 544–550.

[29] Moriyama, E. N., Powell, J. R., Codon usage bias and tRNA abundance in Drosophila. *J. Mol. Evol.* 1997, *45*, 514–523.

[30] Powell, J. R., Moriyama, E. N., Evolution of codon usage bias in Drosophila. *Proc. Natl. Acad. Sci. USA* 1997, *94*, 7784–7790.

[31] Stenico, M., Lloyd, A. T., Sharp, P. M., Codon usage in Caenorhabditis elegans: Delineation of translational selection and mutational biases. *Nucleic Acids Res.* 1994, *22*, 2437–2446.

[32] Chiapello, H., Lisacek, F., Caboche, M., Henaut, A., Codon usage and gene function are related in sequences of Arabidopsis thaliana. *Gene* 1998, *209*, GC1–GC38.

[33] Musto, H., Cruveiller, S., D'Onofrio, G., Romero, H., Bernardi, G., Translational selection on codon usage in Xenopus laevis. *Mol. Biol. Evol.* 2001, *18*, 1703–1707.

[34] Lavner, Y., Kotlar, D., Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2005, *345*, 127–138.

[35] Bernardi, G., Isochores and the evolutionary genomics of vertebrates. *Gene* 2000, *241*, 3–17.

[36] Chamary, J. V., Hurst, L. D., Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *GenomeBiology* 2005, *6*, R75.

[37] Knight, R. D., Freeland, S. J., Landweber, L. F., A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *GenomeBiology* 2001, *2*, research0010–research0010.13.

[38] Palidwor, G. A., Perkins, T. J., Xia, X., A general model of codon bias due to GC mutational bias. *PLoS One* 2010, *5*, e13431.

[39] Kubo, M., Imanaka, T., mRNA secondary structure in an open reading frame reduces translation efficiency in Bacillus subtilis. *J. Bacteriol.* 1989, *171*, 4080–4082.

[40] de Smit, M. H., van Duin, J., Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proc. Natl. Acad. Sci. USA* 1990, *87*, 7668–7672.

[41] Tuller, T., Carmi, A., Vestsigian, K., Navon, S., et al., An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 2010, *141*, 344–354.

[42] Gu, W., Zhou, T., Wilke, C. O., A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.* 2010, *6*, e1000664.

[43] Allert, M., Cox, J. C., Hellinga, H. W., Multifactorial determinants of protein expression in prokaryotic open reading frames. *J. Mol. Biol.* 2010, *402*, 905–918.

[44] Kudla, G., Murray, A. W., Tollervey, D., Plotkin, J. B., Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 2009, *324*, 255–258.

[45] Supek, F., Smuc, T., On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli*. *Genetics* 2010, *185*, 1129–1134.

[46] Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., et al., The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J. Mol. Biol.* 1987, *193*, 413–417.

[47] Thanaraj, T. A., Argos, P., Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci.* 1996, *5*, 1973–1983.

Evelina Angov received her MSc. and PhD. from the University of Maryland, College Park in Biochemistry. She completed a postdoctoral fellowship in the Genetics and Biochemistry Branch of NIDDK, NIH. She currently serves as the Chief of the Department of Molecular Parasitology at the Walter Reed Army Institute of Research, Division of Malaria Vaccine Development. She has worked on projects related to expression and purification of recombinant *P. falciparum* malaria proteins for their evaluation in preclinical models and in clinical trials. In conjunction with the development of these target antigens, she has worked to establish standardized immunological assays to characterize humoral and cellular immune responses. Recently, her laboratory has evaluated several novel platforms for expression and delivery of malaria vaccine candidates as ways to improve immunological responses.

[48] Wolin, S. L., Walter, P., Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* 1988, *7*, 3559–3569.

[49] Cortazzo, P., Cervenansky, C., Marin, M., Reiss, C., et al., Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 2002, *293*, 537–541.

[50] Komar, A. A., Lesnik, T., Reiss, C., Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* 1999, *462*, 387–391.

[51] Farabaugh, P. J., Programmed translational frameshifting. *Microbiol. Rev.* 1996, *60*, 103–134.

[52] Ivanov, I. P., Atkins, J. F., Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: Close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res.* 2007, *35*, 1842–1858.

[53] Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., et al., A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007, *315*, 525–528.

[54] Wen, J. D., Lancaster, L., Hodges, C., Zeri, A. C., et al., Following translation by single ribosomes one codon at a time. *Nature* 2008, *452*, 598–603.

[55] Komar, A. A., A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* 2009, *34*, 16–24.

[56] Thangadurai, C., Suthakaran, P., Barfal, P., Anandaraj, B., et al., Rare codon priority and its position specificity at the 5' of the gene modulates heterologous protein expression in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 2008, *376*, 647–652.

[57] Dittmar, K. A., Sorensen, M. A., Elf, J., Ehrenberg, M., Pan, T., Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.* 2005, *6*, 151–157.

[58] Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., Weissman, J. S., Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009, *324*, 218–223.

[59] Fredrick, K., Ibba, M., How the sequence of a gene can tune its translation. *Cell* 2010, *141*, 227–229.

[60] Clarke, T. F., Clark, P. L., Increased incidence of rare codon clusters at 5' and 3' gene termini: Implications for function. *BMC Genomics* 2010, *11*, 118.

[61] Darko, C. A., Angov, E., Collins, W. E., Bergmann-Leitner, E. S., et al., The clinical-grade 42-kilodalton fragment of merozoite surface protein 1 of Plasmodium falciparum strain FVO expressed in *Escherichia coli* protects Aotus nancymai against challenge with homologous erythrocytic-stage parasites. *Infect. Immun.* 2005, *73*, 287–297.

[62] Angov, E., Hillier, C. J., Kincaid, R. L., Lyon, J. A., Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One* 2008, *3*, e2189.

[63] Angov, E., Legler, P. M., Mease, R. M., Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol. Biol.* 2011, *705*, 1–13.

[64] Biro, J. C., Protein folding information in nucleic acids which is not present in the genetic code. *Ann. N. Y. Acad. Sci.* 2006, *1091*, 399–411.

[65] Hamano, T., Matsuo, K., Hibi, Y., Victoriano, A. F., et al., A single-nucleotide synonymous mutation in the gag gene controlling human immunodeficiency virus type 1 virion production. *J. Virol.* 2007, *81*, 1528–1533.

[66] Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., Silent polymorphisms speak: How they affect pharmacogenomics and the treatment of cancer. *Cancer Res.* 2007, *67*, 9609–9612.

[67] Komar, A. A., Silent SNPs: Impact on gene function and phenotype. *Pharmacogenomics* 2007, *8*, 1075–1080.

[68] Komar, A. A., Genetics. SNPs, silent but not invisible. *Science* 2007, *315*, 466–467.

[69] Zhang, G., Hubalewska, M., Ignatova, Z., Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 2009, *16*, 274–280.

[70] Zhou, M., Li, X., Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol. Biol. Rep.* 2009, *36*, 2039–2046.

[71] Makrides, S. C., Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 1996, *60*, 512–538.

[72] Gustafsson, C., Govindarajan, S., Minshull, J., Codon bias and heterologous protein expression. *Trends Biotechnol.* 2004, *22*, 346–353.

[73] Kane, J. F., Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 1995, *6*, 494–500.

[74] Goldman, E., Rosenberg, A. H., Zubay, G., Studier, F. W., Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. *J. Mol. Biol.* 1995, *245*, 467–473.

[75] Zhou, Z., Schnake, P., Xiao, L., Lal, A. A., Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein Expression Purif.* 2004, *34*, 87–94.

[76] Sorensen, H. P., Mortensen, K. K., Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 2005, *115*, 113–128.

[77] Brinkmann, U., Mattes, R. E., Buckel, P., High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene* 1989, *85*, 109–114.

[78] Cinquin, O., Christopherson, R. I., Menz, R. I., A hybrid plasmid for expression of toxic malarial proteins in *Escherichia coli*. *Mol. Biochem. Parasitol.* 2001, *117*, 245–247.

[79] Yu, Z. B., Jin, J. P., Removing the regulatory N-terminal domain of cardiac troponin I diminishes incompatibility during bacterial expression. *Arch. Biochem. Biophys.* 2007, *461*, 138–145.

[80] Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W., Zubay, G., Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J. Bacteriol.* 1993, *175*, 716–722.

[81] Thanaraj, T. A., Argos, P., Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* 1996, *5*, 1594–1612.

[82] Hillier, C. J., Ware, L. A., Barbosa, A., Angov, E., et al., Process development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for Plasmodium falciparum. *Infect. Immun.* 2005, *73*, 2109–2115.

[83] Chowdhury, D. R., Angov, E., Kariuki, T., Kumar, N., A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in *Escherichia coli*. *PLoS One* 2009, *4*, e6352.

[84] Francis, D. M., Page, R., Strategies to optimize protein expression in *E. coli*. *Current Protocols in Protein Science* 2010, Chapter 5, Unit 5.24, pp. 21–29.

[85] Murasugi, A., Secretory expression of human protein in the Yeast Pichia pastoris by controlled fermentor culture. *Recent Pat. Biotechnol.* 2010, *4*, 153–166.

[86] Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., Berger, I., New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* 2010, *172*, 45–54.

[87] Idiris, A., Tohda, H., Kumagai, H., Takegawa, K., Engineering of protein secretion in yeast: Strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* 2010, *86*, 403–417.

[88] Desai, P. N., Shrivastava, N., Padh, H., Production of heterologous proteins in plants: Strategies for optimal expression. *Biotechnol. Adv.* 2010, *28*, 427–435.

[89] Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., et al., Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression Purif.* 2008, *59*, 94–102.

[90] Sahdev, S., Khattar, S. K., Saini, K. S., Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* 2008, *307*, 249–264.

[91] Peti, W., Page, R., Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expression Purif.* 2007, *51*, 1–10.

[92] Terpe, K., Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 2006, *72*, 211–222.

[93] Berisio, R., Schluenzen, F., Harms, J., Bashan, A., et al., Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat. Struct. Biol.* 2003, *10*, 366–370.

[94] Varenne, S., Buc, J., Lloubes, R., Lazdunski, C., Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* 1984, *180*, 549–576.

[95] Lu, J., Deutsch, C., Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* 2005, *12*, 1123–1129.

[96] Kramer, G., Boehringer, D., Ban, N., Bukau, B., The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 2009, *16*, 589–597.

[97] Lu, J., Kobertz, W. R., Deutsch, C., Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* 2007, *371*, 1378–1391.

[98] Lu, J., Deutsch, C., Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* 2008, *384*, 73–86.

[99] Voges, D., Watzele, M., Nemetz, C., Wizemann, S., Buchberger, B., Analyzing and enhancing mRNA translational efficiency in an *Escherichia coli* in vitro expression system. *Biochem. Biophys. Res. Commun.* 2004, *318*, 601–614.

[100] Lorimer, D., Raymond, A., Walchli, J., Mixon, M., et al., Gene composer: Database software for protein construct design, codon engineering, and gene synthesis. *BMC Biotechnol.* 2009, *9*, 36.

[101] Sandhu, K. S., Pandey, S., Maiti, S., Pillai, B., GASCO: Genetic algorithm simulation for codon optimization. *In Silico Biol.* 2008, *8*, 187–192.

[102] Cai, Y., Sun, J., Wang, J., Ding, Y., et al., Optimizing the codon usage of synthetic gene with QPSO algorithm. *J. Theor. Biol.* 2008, *254*, 123–127.

[103] Puigbo, P., Guzman, E., Romeu, A., Carcia-Vallve, S., OPTIMIZER: A web server for optimizing the codon usage of DNA squences. *Nucleic Acids Res..* 2007, *35*, W126–W131.

[104] Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., Govindarajan, S., Gene Designer: A synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 2006, *7*, 285.

[105] Wu, G., Bashir-Bello, N., Freeland, S. J., The synthetic gene designer: A flexible web platform to explore sequence manipulation for heterologous expression. *Protein Expression Purif.* 2006, *47*, 441–445.

[106] Grote, A., Hiller, K., Scheer, M., Munch, R., et al., JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 2005, *33*, W526–531.

[107] Jayaraj, S., Reid, R., Santi, D. V., GeMS: An advanced software package for designing synthetic genes. *Nucleic Acids Res.* 2005, *33*, 3011–3016.

[108] Gao, W., Rzewski, A., Sun, H., Robbins, P. D., Gambotto, A., UpGene: Application of a web-based DNA codon optimization algorithm. *Biotechnol. Prog.* 2004, *20*, 443–448.