



## Review

# Manipulating the genetic code for membrane protein production: What have we learnt so far?

Morten H.H. Nørholm <sup>a,\*</sup>, Sara Light <sup>b</sup>, Minttu T.I. Virkki <sup>a</sup>, Arne Elofsson <sup>a,c</sup>,  
Gunnar von Heijne <sup>a,c</sup>, Daniel O. Daley <sup>a,\*\*</sup>

<sup>a</sup> Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91, Sweden

<sup>b</sup> Science for Life Laboratory, Bioinformatics Infrastructure for Life Sciences, Department of Biochemistry and Biophysics, Stockholm University, SE-17 121, Sweden

<sup>c</sup> Stockholm Bioinformatics Center and Science for Life Laboratory, Stockholm University, SE-106 91, Sweden

## ARTICLE INFO

## Article history:

Received 20 June 2011

Received in revised form 4 August 2011

Accepted 15 August 2011

Available online 22 August 2011

## Keywords:

Membrane protein overexpression

Membrane protein folding

Codon use

Membrane protein engineering

## ABSTRACT

With synthetic gene services, molecular cloning is as easy as ordering a pizza. However choosing the right RNA code for efficient protein production is less straightforward, more akin to deciding on the pizza toppings. The possibility to choose synonymous codons in the gene sequence has ignited a discussion that dates back 50 years: Does synonymous codon use matter? Recent studies indicate that replacement of particular codons for synonymous codons can improve expression in homologous or heterologous hosts, however it is not always successful. Furthermore it is increasingly apparent that membrane protein biogenesis can be codon-sensitive. Single synonymous codon substitutions can influence mRNA stability, mRNA structure, translational initiation, translational elongation and even protein folding. Synonymous codon substitutions therefore need to be carefully evaluated when membrane proteins are engineered for higher production levels and further studies are needed to fully understand how to select the codons that are optimal for higher production. This article is part of a Special Issue entitled: Protein Folding in Membranes.

© 2011 Elsevier B.V. All rights reserved.

## Contents

1. Introduction . . . . .	1091
2. Synonymous codon usage can affect membrane protein expression and biogenesis. . . . .	1092
3. Codon use in membrane protein mRNAs . . . . .	1093
4. Implications for gene design . . . . .	1094
5. Conclusions . . . . .	1095
Acknowledgements . . . . .	1095
References . . . . .	1095

## 1. Introduction

The nature of the genetic code was deciphered 50 years ago [1]. As RNA is made of four different nucleotides, there are 64 possible combinations of codons for the 20 different amino acids. Different

synonymous codons can therefore encode for the same amino acid. For example, serine, arginine, and leucine are each encoded by six different synonymous codons (Fig. 1).

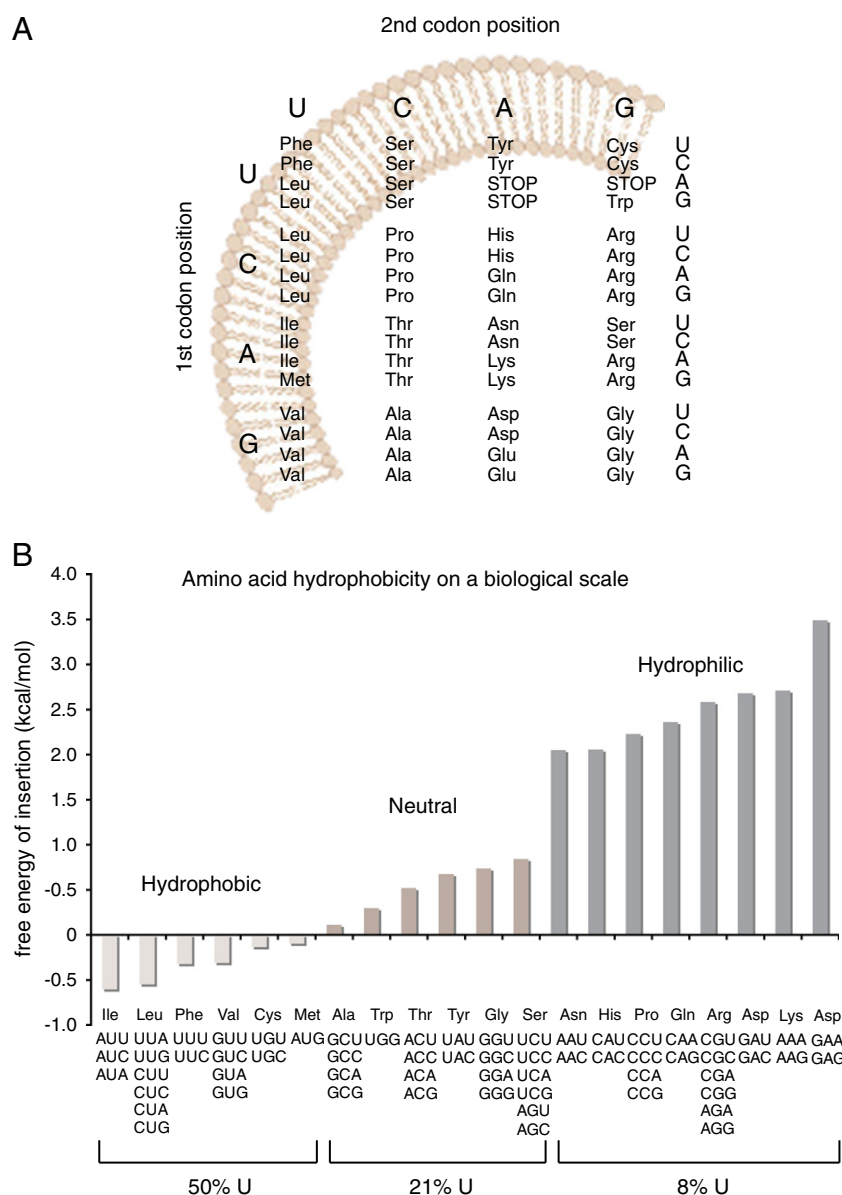
Synonymous codon use is not uniform. Some codons are frequently used whereas others are not; the latter are commonly referred to as *rare codons*. Synonymous codon use also varies between different genes and genomes [2–5], and different indices have been developed to describe this phenomenon (i.e. to distinguish frequent codons from rare codons). For instance, the  $N_c$  scale describes the use of a specific codon relative to the number of synonymous codons in a genome [6]. Alternatively, codon usage can be described by the concentrations of the complementary tRNAs in the cell. Although these two scales correlate well [3,7–9], more refined descriptions, such as the codon bias index (CBI) [10] or the tRNA adaptation index (tAI) [11], can be obtained by combining them. Finally, the codon adaptation index

\* This article is part of a Special Issue entitled: Protein Folding in Membranes.

\* Correspondence to: M.H.H. Nørholm, Plant Biochemistry Laboratory, Department of Plant Biology and Biotechnology, University of Copenhagen and Section for Plant Pathway Discovery, Novo Nordisk Foundation Center for Biosustainability, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark. Tel.: +45 353 33337; fax: +45 353 33300.

\*\* Correspondence to: D.O. Daley, Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91, Sweden. Tel.: +46 8 16 29 10; fax: +46 8 15 36 79.

E-mail addresses: [mhnh@life.ku.dk](mailto:mhnh@life.ku.dk) (M.H.H. Nørholm), [ddaley@dbb.su.se](mailto:ddaley@dbb.su.se) (D.O. Daley).



**Fig. 1.** The genetic code and its relation to hydrophobicity of the encoded amino acids. (A) Typical schematic representation of the genetic code, illustrating how the four different nucleotides (U, C, A and G) encode 20 different amino acids and stop codons. Notably, hydrophobic amino acids such as phenylalanine (Phe), leucine (Leu), isoleucine (Ile), methionine (Met) and valine (Val) that are over-represented in transmembrane protein segments, all contain a uridine nucleotide in the second codon position. Codons with a U in the first position also tend to encode amino acids that are over-represented in membrane proteins. In the figure, amino acids that frequently occur in transmembrane segments have been emphasized with a lipid bilayer in the background. (B) Hydrophobicity of the 20 different amino acids on a biological scale, specified as the free energy of membrane insertion (kcal/mol) when the indicated amino acid is placed in the middle of a 19-residue hydrophobic stretch [29]. Below is shown the codons that encode for the different amino acids, illustrating the U-bias of hydrophobic versus hydrophilic residues.

(CAI) computes statistical codon usage relative to codon usage in highly expressed genes as a prediction of protein expression levels [12]. Recent findings indicate that growth conditions affect codon usage, and the kinetics of recharging tRNAs may also be important for describing codon usage [13,14]. Clearly, it is not straightforward to develop an efficient description of synonymous codon usage.

## 2. Synonymous codon usage can affect membrane protein expression and biogenesis

The availability of synonymous codons means that a single protein can be encoded by a myriad of different DNA sequences. So does it matter which synonymous codon is used? In most situations synonymous codon choice is neutral, however several studies indicate that synonymous codon changes can influence mRNA stability, mRNA structure,

translational initiation, translational elongation and protein folding (reviewed in [15–18]). Thus the genetic code has the capacity to contain deeper layers of information than simply the amino acid sequence.

There are numerous examples of membrane proteins, whose expression levels are sensitive to synonymous codon use. For instance, a single synonymous codon change in FtsH, a membrane-bound protease in *E. coli*, increases the stability of mRNA structure around the ribosome-binding site and inhibits translational initiation. As a result there was a considerable reduction in protein levels [19]. In the human dopamine receptor D2, a synonymous codon change lowered the mRNA stability and caused a reduction in protein levels [20]. Furthermore, synonymous codon changes in the *E. coli* outer membrane protein OmpA resulted in a 10-fold lowering of both mRNA and protein levels [21]. Membrane protein folding can also be sensitive to synonymous codon use. A frequent-to-rare synonymous codon change in the human

P-glycoprotein (an ATP driven efflux pump) resulted in a protein with altered conformation and substrate specificity [22]. In this study it was speculated that the synonymous codon change had affected the timing of translation and the co-translational folding of the protein.

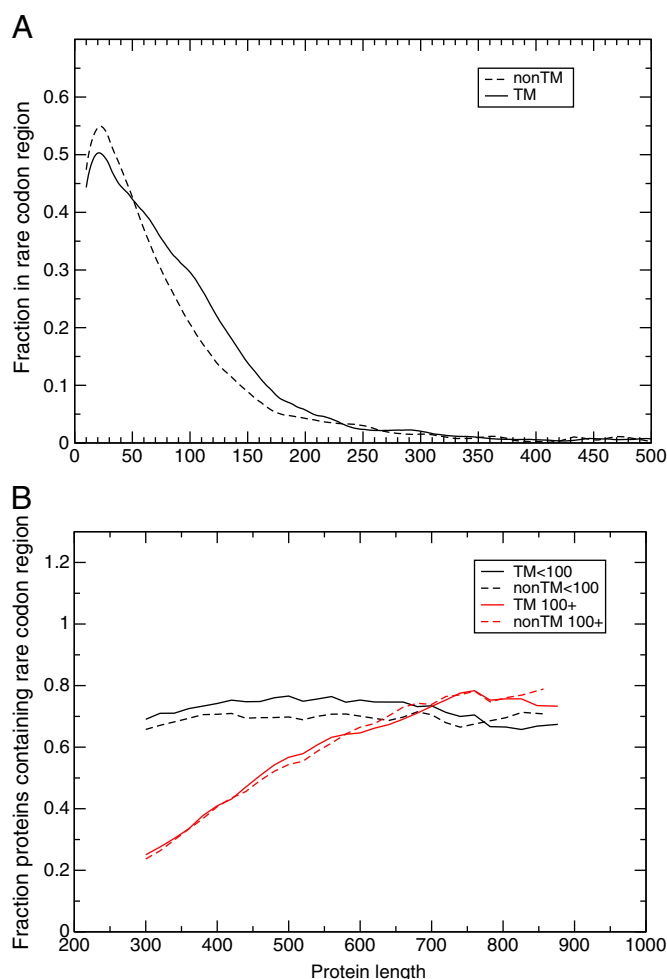
Codon choice can also influence folding of soluble proteins. It has long been recognized that slowly translated regions can be localized downstream of protein domain boundaries [23] and/or secondary structures [24], thus facilitating co-translational folding of protein domains. Clusters of rare codons (in this case defined as codons that are read by less abundant tRNAs) are also predicted to cause translational pausing of domain boundaries in the SufI protein in *E. coli* [25]. When rare codons in these clusters are changed to more frequent synonymous codons, the protein folds incorrectly even though it has the same amino acid sequence. In another study, synonymous codon changes decreased the solubility of a fatty acid binding protein when expressed in *E. coli* [26]. Similar effects of codon usage on protein folding have also been demonstrated *in vitro* (e.g. [27,28]).

### 3. Codon use in membrane protein mRNAs

The effect of codon use on folding of membrane proteins is a poorly understood but important aspect of protein biogenesis, as it has implications for gene design and for understanding single nucleotide polymorphisms in disease states. If we are to effectively manipulate the genetic code for membrane protein production, we must first understand these deeper layers.

Codon use in membrane protein mRNAs differs from that in soluble protein mRNAs. The difference is predominantly a reflection of differences in amino acid usage, as membrane proteins are enriched in hydrophobic amino acids (*i.e.* F, M, I, L, V, C) [29–31]. Intriguingly, the codons for most of these hydrophobic amino acids usually contain a uracil (U) in the second position [32] and a disproportionately higher number of U's compared to codons for other amino acids (Fig. 1). Membrane proteins are also enriched in two hydrophilic amino acids (*i.e.* S and Y) [31], whose codons also contain a disproportionately high number of U's (Fig. 1). As a result, mRNAs encoding membrane proteins contain a high U-bias compared to mRNAs encoding soluble proteins [31]. The U-bias phenomena is more pronounced in bacteria than in eukaryotes and it has been speculated that it may be an evolutionary relic of an mRNA targeting pathway. In support of this hypothesis, it has been shown that mRNAs encoding two *E. coli* membrane proteins (LacY and Bgl) are localized to the inner membrane through the regions encoding the transmembrane helices (*i.e.* the regions that are most U-biased) [33]. Other membrane-protein specific trends, such as GC-richness in the third codon position have been noted [34]. Whilst the U-bias and the 3rd-position-GC-richness phenomena are intriguing, the physiological relevance remains to be determined.

An interesting but poorly understood characteristic of all mRNAs is the presence of rare codon clusters, which can induce ribosomal pausing. Such clusters have been detected in membrane protein mRNAs from *S. cerevisiae* [35], *E. nidulans* [36], *E. coli* and *B. subtilis* [25]. Our analysis of the *E. coli* data set indicates that there is little difference in the occurrence or location of the rare codon clusters between membrane protein mRNAs and soluble protein mRNA's. Approximately 76% of membrane protein mRNAs and 66% of soluble protein mRNA's contain at least one predicted rare codon cluster (an average of 1.54 and 1.37 per mRNA, respectively). As noted in our analysis and in earlier studies the clusters are most often located at the 5' of the mRNA (Fig. 2A and [37–41]). This observation is in agreement with other experimental and bioinformatics studies, which indicate a universally conserved translation speed ramp at the 5' end of genes [42,43]. Such a ramp might serve to minimize ribosome collisions during the early stages of translation and increase overall translational efficiency [44]. Previous studies have also shown that rare codon regions are more common in long proteins,



**Fig. 2.** Patterns of rare codon clusters are similar in membrane and soluble protein mRNAs. Rare codon clusters were identified in protein coding genes from the *E. coli* strain MG1655, using the method of Zhang et al. [25]. Membrane proteins (*i.e.* proteins containing at least one predicted transmembrane helix) were separated from soluble proteins using SCAMPI [68]. Transmembrane helices in the first 40 residues were excluded since these may constitute a signal peptide. (A) Rare codon clusters are more prevalent at the 5' end. For each position in the sequence the fraction of residues present in a rare codon region was calculated. Running averages were calculated with a window size of 21 amino acids. (B) The fraction of proteins with rare codon clusters versus protein length (in amino acids). The solid lines represent transmembrane proteins (TM) and the dotted lines represent the soluble (nonTM) proteins. The red lines represent rare codons at or downstream of the 100th amino acid whilst black lines represent rare codons within the first 100 amino acids.

as maybe expected by chance alone [25]. However, our analysis of the *E. coli* proteome indicated that rare codon regions occurring near the 5' end of mRNAs are present in long and short proteins at similar frequencies (Fig. 2B).

Whilst the bioinformatics analyses indicate that there should be instances of ribosome pausing during translation of many membrane proteins, pausing has to the best of our knowledge only been experimentally demonstrated outside the '5 ramp', viz. for the chloroplast CF<sub>0</sub>-1 subunit of the ATP synthase and the D1 subunit of Photosystem II [45,46]. Given the scarcity of experimental examples it is difficult to fully understand how important pausing might be for membrane protein biogenesis. For the D1 subunit it was hypothesized that the pause was important for co-translational insertion of co-factors and therefore for correct folding of the protein. Furthermore (as mentioned above), a single synonymous codon change was sufficient to alter folding of P-glycoprotein [22]. More experimental work is therefore required to tease out the sequence characteristics (or molecular code) in mRNA that governs control of translation rate, and to

determine the role that ribosomal pauses play in membrane protein folding.

One observation that may guide experimentation is that rare codon clusters are often located 45 or 70 codons downstream of a transmembrane spanning helix in *S. cerevisiae* [35,36]. Since the ribosome exit tunnel can accommodate 30–72 amino acids (depending on secondary structure of the nascent polypeptide) [47,48], it is hypothesized that the pause would often occur as a transmembrane helix is leaving the ribosome exit tunnel or the translocon (Fig. 3). One can speculate that increasing the time spent by a transmembrane helix in the translocon might influence (i) how efficiently it partitions into the surrounding membrane, (ii) how efficiently it interacts with more N-terminally located transmembrane helices, or (iii) how efficiently it is glycosylated on the regions flanking the transmembrane domain. In support of the last point, it has been shown that efficient glycosylation of tyrosinase (a type I membrane glycoprotein) is sensitive to the translation rate [49].

#### 4. Implications for gene design

Determining the structures of biomedically important proteins is central to understanding and modulating function. A first step toward this goal is to obtain milligram amounts of folded proteins for structural studies using X-ray crystallography, NMR and EM, as well as for biochemical and biophysical analysis. This is not a trivial process for membrane proteins as they are difficult to overexpress. A die-hard assumption is that rare codons cause low expression, and that optimizing codon usage will improve production levels. However, a growing number of reports challenge this simplistic view [15,50–52]. In the following section we have tried to make sense of the somewhat confusing reports relating codon use to membrane protein production. What have we learned?

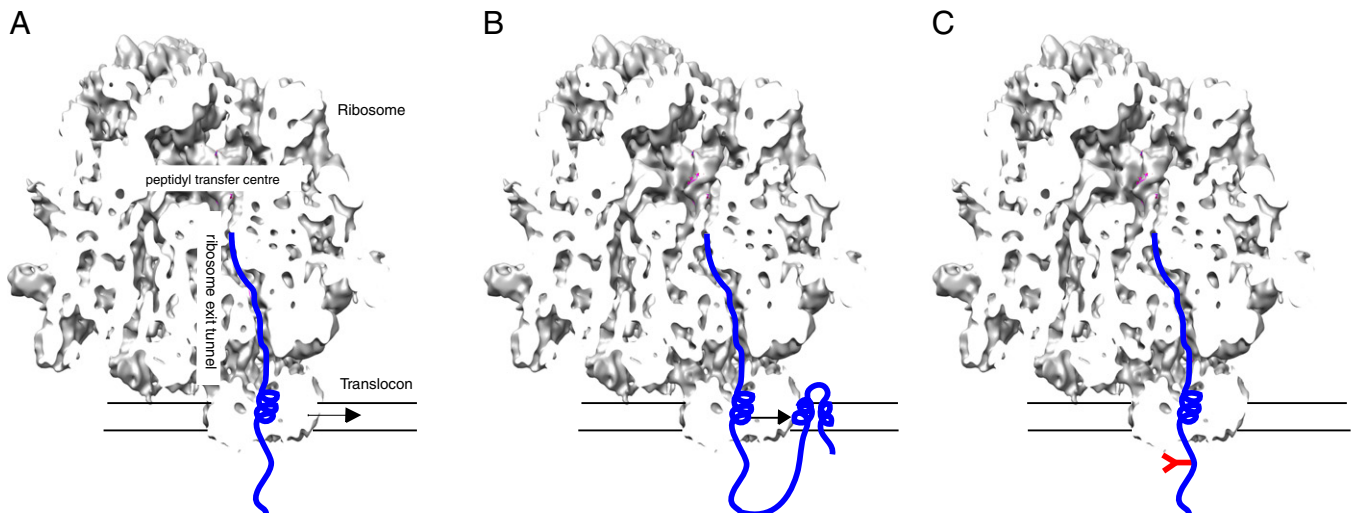
Not surprisingly, there are reports that synonymous codon changes can influence membrane protein overexpression levels. A 6- to 9-fold increase in expression was observed when genes for the GluCl $\alpha$  and GluCl $\beta$  ion channels from *C. elegans* were codon optimized and expressed in E18 rat hippocampal neurons [53]. Likewise, two G-protein coupled receptors were produced after codon-optimization [54,55], although one of them appeared to be misfolded and ended up in inclusion bodies. Moreover, a recent multi-gene study reported an increase in expression success rate (from 39 to 50%) when 28 membrane proteins were

optimized by multi-parameter gene optimization [56]. In the study, codon quality, GC-content, sequence motifs and probability to form stable mRNA secondary structures were all concomitantly optimized. Similar multi-parameter codon optimization algorithms have been described elsewhere [14,52,57,58]. The algorithm of Fath and co-workers [58] was capable of improving the expression of 12 out of 14 membrane proteins, but only by 1–3 fold [58]. An alternative strategy to codon optimization is to supplement the host organism with rare tRNAs [50,59,60].

Whilst there are numerous examples where codon-engineering strategies have been effective for overexpression of membrane proteins [53–56,61,62], there are also examples where they have failed [63,64]. Significantly, analyses of large data sets have failed to find a correlation between codon usage and overexpression levels of membrane proteins in *E. coli* and *S. cerevisiae* [65,66]. This conclusion was corroborated by Kudla et al. who were unable to find a correlation between codon use and overexpression levels in a library of synonymous GFP variants [15]. These examples indicate that there is still much to learn about codon optimization of membrane proteins.

One way to further our understanding is to analyze both successful and failed experiments. Unfortunately the failed experiments are rarely published. In our laboratory, four membrane proteins that had been optimized by different commercial multi-parameter algorithms exhibited little or no improvement in overexpression in *E. coli*. In agreement with this observation, 8 out of 10 codon-optimized variants of a membrane transporter did not express better than the native construct in *S. cerevisiae* (David Drew personal communication). Furthermore, the two that did express aggregated during purification. These observations, although under-represented in the literature, reinforce the point that rare codons may serve important roles and cannot *per se* be regarded as non-optimal (for recent reviews on this particular topic [5,51,52]). The take home message is, that the deeper layers of the RNA code and the species-related differences have not been systematically studied, and are not yet well enough understood, to be effectively exploited for production of membrane proteins.

Adding nucleotide extensions to the gene-of-interest (*i.e.* non-coding leader sequences, protein-coding leader sequences, whole genes) is a generic solution that can improve the mRNA characteristics for membrane protein overexpression. For example, a translational fusion between GFP and a membrane subunit of the ATP synthase stabilized the mRNA, eliminated toxic effects and resulted in high-yield overexpression [67]. Similarly, a 28-codon tag fused to the N-termini



**Fig. 3.** A model for how a rare codon cluster could pause translation and thereby affect the biogenesis of a membrane protein. The spacing of rare codon clusters and transmembrane helices in *S. cerevisiae* membrane proteins suggests that a pause might occur as a transmembrane helix exits the ribosome (see text for details). This event could influence (A) how efficiently the transmembrane helix partitions into the surrounding membrane, (B) how efficiently the transmembrane helix interacts with more N-terminally located transmembrane helices, or (C) how efficiently the regions flanking the transmembrane helix are glycosylated. The image was generated from a cryo-EM structure of the eukaryotic ribosome with bound Sec61 [69]. Figure courtesy of Dr. Shashi Bhushan, Wuerzburg University.



of a library of poorly expressed GFP codon-variants normalized expression to a high level [15]. In our hands the same 28-codon tag was able to stimulate overexpression levels for approximately 30% of the membrane proteins tested in *E. coli* (Nørholm, von Heijne and Daley, *manuscript in preparation*).

## 5. Conclusions

Gene sequences are shaped by evolution and are rarely optimized for translational efficiency. High protein production is a need dictated by biotechnology, whereas nature most likely requires minimization of resources. In this article we have presented examples where manipulation of the genetic code has led to higher production of functionally active membrane proteins in heterologous hosts, and other examples where codon optimization has had no effect, or has led to misfolded and unstable products. Clearly there is still a lot to learn about codon use if we are to manipulate it for protein production. In membrane proteins, the relevance of mRNA structures and rare codon clusters has not been thoroughly explored. Whilst it seems reasonable to suggest that they encode programmed translational pauses, there has been no systematic experimental analysis of their effect on translation rates and protein biogenesis. In the never-ending quest for higher production levels of membrane proteins, these deeper layers need to be kept in mind.

## Acknowledgements

There is a tremendous amount of literature on codon usage, and we would like to apologize to those whose work we could not cite. We would like to thank Shashi Bhushan for help with Fig. 3 and we acknowledge support from the NIH (R01 GM081827-01) and the Swedish Research Council. SL is supported through the Bioinformatics for Life Science platform.

## References

- [1] F.H. Crick, L. Barnett, S. Brenner, R.J. Watts-Tobin, General nature of the genetic code for proteins, *Nature* 192 (1961) 1227–1232.
- [2] M. Gouy, C. Gautier, Codon usage in bacteria: correlation with gene expressivity, *Nucleic Acids Res.* 10 (1982) 7055–7074.
- [3] T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol.* 2 (1985) 13–34.
- [4] D. Chen, D.E. Texada, Low-usage codons and rare codons of *Escherichia coli*, *Gene Ther. Mol. Biol.* 10 (2006) 1–12.
- [5] J.B. Plotkin, G. Kudla, Synonymous but not the same: the causes and consequences of codon bias, *Nat. Rev. Genet.* 12 (2011) 32–42.
- [6] F. Wright, The 'effective number of codons' used in a gene, *Gene* 87 (1990) 23–29.
- [7] T. Ikemura, Measurement of relative amount of *E. coli* tRNAs: codon choice in *E. coli* genes is largely constrained by the concentration of anticodons (author's transl.), *Tanpakushitsu Kakusan Koso* 25 (1980) 668–678.
- [8] T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol.* 151 (1981) 389–409.
- [9] T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes, *J. Mol. Biol.* 146 (1981) 1–21.
- [10] J.L. Bennetzen, B.D. Hall, Codon selection in yeast, *J. Biol. Chem.* 257 (1982) 3026–3031.
- [11] M. dos Reis, R. Savva, L. Wernisch, Solving the riddle of codon usage preferences: a test for translational selection, *Nucleic Acids Res.* 32 (2004) 5036–5044.
- [12] P.M. Sharp, W.H. Li, The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.* 15 (1987) 1281–1295.
- [13] J. Elf, G.W. Li, X.S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell, *Science* 316 (2007) 1191–1194.
- [14] M. Welch, S. Govindarajan, J.E. Ness, A. Villalobos, A. Gurney, J. Minshull, C. Gustafsson, Design parameters to control synthetic gene expression in *Escherichia coli*, *PLoS One* 4 (2009) e7002.
- [15] G. Kudla, A.W. Murray, D. Tollervey, J.B. Plotkin, Coding-sequence determinants of gene expression in *Escherichia coli*, *Science* 324 (2009) 255–258.
- [16] M. Marin, Folding at the rhythm of the rare codon beat, *Biotechnol. J.* 3 (2008) 1047–1057.
- [17] C.J. Tsai, Z.E. Sauna, C. Kimchi-Sarfaty, S.V. Ambudkar, M.M. Gottesman, R. Nussinov, Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima, *J. Mol. Biol.* 383 (2008) 281–291.
- [18] K. Fredrick, M. Ibbá, How the sequence of a gene can tune its translation, *Cell* 141 (2010) 227–229.
- [19] S. Makino, J.N. Qu, K. Uemori, H. Ichikawa, T. Ogura, H. Matsuzawa, A silent mutation in the *ftsH* gene of *Escherichia coli* that affects FtsH protein production and colicin tolerance, *Mol. Gen. Genet.* 254 (1997) 578–583.
- [20] J. Duan, M.S. Wainwright, J.M. Comeron, N. Saitou, A.R. Sanders, J. Gelernter, P.V. Gejman, Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor, *Hum. Mol. Genet.* 12 (2003) 205–216.
- [21] A. Deana, R. Ehrlich, C. Reiss, Silent mutations in the *Escherichia coli* *ompA* leader peptide region strongly affect transcription and translation in vivo, *Nucleic Acids Res.* 26 (1998) 4778–4782.
- [22] C. Kimchi-Sarfaty, J.M. Oh, I.W. Kim, Z.E. Sauna, A.M. Calcagno, S.V. Ambudkar, M.M. Gottesman, A "silent" polymorphism in the MDR1 gene changes substrate specificity, *Science* 315 (2007) 525–528.
- [23] T.A. Thanaraj, P. Argos, Ribosome-mediated translational pause and protein domain organization, *Protein Sci.* 5 (1996) 1594–1612.
- [24] R. Saunders, C.M. Deane, Synonymous codon usage influences the local protein structure observed, *Nucleic Acids Res.* 38 (2010) 6719–6728.
- [25] G. Zhang, M. Hubalewska, Z. Ignatova, Transient ribosomal attenuation coordinates protein synthesis and co-translational folding, *Nat. Struct. Mol. Biol.* 16 (2009) 274–280.
- [26] P. Cortazzo, C. Cervenansky, M. Marin, C. Reiss, R. Ehrlich, A. Deana, Silent mutations affect in vivo protein folding in *Escherichia coli*, *Biochem. Biophys. Res. Commun.* 293 (2002) 537–541.
- [27] A.A. Komar, T. Lesnik, C. Reiss, Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation, *FEBS Lett.* 462 (1999) 387–391.
- [28] V. Ramachandiran, G. Kramer, P.M. Horowitz, B. Hardesty, Single synonymous codon substitution eliminates pausing during chloramphenicol acetyl transferase synthesis on *Escherichia coli* ribosomes in vitro, *FEBS Lett.* 512 (2002) 209–212.
- [29] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, G. von Heijne, Recognition of transmembrane helices by the endoplasmic reticulum translocator, *Nature* 433 (2005) 377–381.
- [30] M.B. Ulmschneider, M.S. Sansom, A. Di Nola, Properties of integral membrane protein structures: derivation of an implicit membrane potential, *Proteins* 59 (2005) 252–265.
- [31] J. Prilusky, E. Bibi, Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs, *Proc. Natl. Acad. Sci. U.S.A.* 106 (2009) 6662–6666.
- [32] R.V. Wolfenden, P.M. Cullis, C.C. Southgate, Water, protein folding, and the genetic code, *Science* 206 (1979) 575–577.
- [33] K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, O. Amster-Choder, Translation-independent localization of mRNA in *E. coli*, *Science* 331 (2011) 1081–1084.
- [34] K. Lin, S.B. Tan, P.R. Kolatkar, R.J. Epstein, Nonrandom intragenic variations in patterns of codon bias implicate a sequential interplay between transitional genetic drift and functional amino acid selection, *J. Mol. Evol.* 57 (2003) 538–545.
- [35] F. Kepes, The "+70 pause": hypothesis of a translational control of membrane protein assembly, *J. Mol. Biol.* 262 (1996) 77–86.
- [36] P. Dessen, F. Kepes, The PAUSE software for analysis of translational control over protein targeting: application to *E. nidulans* membrane proteins, *Gene* 244 (2000) 89–96.
- [37] G.F. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes, *Nucleic Acids Res.* 18 (1990) 1465–1473.
- [38] A. Eyre-Walker, M. Bulmer, Reduced synonymous substitution rate at the start of enterobacterial genes, *Nucleic Acids Res.* 21 (1993) 4599–4603.
- [39] D.A. Phoenix, E. Korotkov, Evidence of rare codon clusters within *Escherichia coli* coding regions, *FEMS Microbiol. Lett.* 155 (1997) 63–66.
- [40] C.M. Stenstrom, H. Jin, L.L. Major, W.P. Tate, L.A. Isaksson, Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*, *Gene* 263 (2001) 273–284.
- [41] T.F. Clarke, P.L. Clark, Rare codons cluster, *PLoS One* 3 (2008) e3412.
- [42] N.T. Ingolia, S. Ghaemmaghami, J.R. Newman, J.S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling, *Science* 324 (2009) 218–223.
- [43] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, An evolutionarily conserved mechanism for controlling the efficiency of protein translation, *Cell* 141 (2010) 344–354.
- [44] N. Mitarai, K. Snieppen, S. Pedersen, Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization, *J. Mol. Biol.* 382 (2008) 236–245.
- [45] J. Kim, P.G. Klein, J.E. Mullet, Ribosomes pause at specific sites during synthesis of membrane-bound chloroplast reaction center protein D1, *J. Biol. Chem.* 266 (1991) 14931–14938.
- [46] N.E. Stollar, J.K. Kim, M.J. Hollingsworth, Ribosomes pause during the expression of the large ATP synthase gene cluster in spinach chloroplasts, *Plant Physiol.* 105 (1994) 1167–1177.
- [47] J. Frank, A. Verschoor, Y. Li, J. Zhu, R.K. Lata, M. Radermacher, P. Penczek, R. Grassucci, R.K. Agrawal, S. Srivastava, A model of the translational apparatus based on a three-dimensional reconstruction of the *Escherichia coli* ribosome, *Biochem. Cell Biol.* 73 (1995) 757–765.

- [48] S. Bhushan, M. Gartmann, M. Halic, J.P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D.N. Wilson, R. Beckmann, Alpha-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel, *Nat. Struct. Mol. Biol.* 17 (2010) 313–317.
- [49] A. Ujvari, R. Aron, T. Eisenhaure, E. Cheng, H.A. Parag, Y. Smicun, R. Halaban, D.N. Hebert, Translation rate of human tyrosinase determines its N-linked glycosylation level, *J. Biol. Chem.* 276 (2001) 5924–5931.
- [50] C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression, *Trends Biotechnol.* 22 (2004) 346–353.
- [51] M. Welch, A. Villalobos, C. Gustafsson, J. Minshull, You're one in a googol: optimizing genes for protein expression, *J. R. Soc. Interface* 6 (Suppl 4) (2009) S467–S476.
- [52] G. Wu, L. Dress, S.J. Freeland, Optimal encoding rules for synthetic genes: the need for a community effort, *Mol. Syst. Biol.* 3 (2007) 134.
- [53] E.M. Slimko, H.A. Lester, Codon optimization of *Caenorhabditis elegans* GluCl ion channel genes for mammalian cells dramatically improves expression levels, *J. Neurosci. Methods* 124 (2003) 75–81.
- [54] J.L. Baneres, A. Martin, P. Hullot, J.P. Girard, J.C. Rossi, J. Parelo, Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1, *J. Mol. Biol.* 329 (2003) 801–814.
- [55] F.F. Hamdan, A. Mousa, P. Ribeiro, Codon optimization improves heterologous expression of a *Schistosoma mansoni* cDNA in HEK293 cells, *Parasitol. Res.* 88 (2002) 583–586.
- [56] B. Maertens, A. Spriestersbach, U. von Groll, U. Roth, J. Kubicek, M. Gerrits, M. Graf, M. Liss, D. Daubert, R. Wagner, F. Schafer, Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*, *Protein Sci.* 19 (2010) 1312–1326.
- [57] M. Allert, J.C. Cox, H.W. Hellinga, Multifactorial determinants of protein expression in prokaryotic open reading frames, *J. Mol. Biol.* 402 (2010) 905–918.
- [58] S. Fath, A.P. Bauer, M. Liss, A. Spriestersbach, B. Maertens, P. Hahn, C. Ludwig, F. Schafer, M. Graf, R. Wagner, Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression, *PLoS One* 6 (2011) e17596.
- [59] U. Brinkmann, R.E. Mattes, P. Buckel, High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product, *Gene* 85 (1989) 109–114.
- [60] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, *Curr. Opin. Biotechnol.* 6 (1995) 494–500.
- [61] S.J. Park, S.K. Lee, B.J. Lee, Effect of tandem rare codon substitution and vector-host combinations on the expression of the EBV gp110 C-terminal domain in *Escherichia coli*, *Protein Expr. Purif.* 24 (2002) 470–480.
- [62] H. Tegel, S. Tourle, J. Ottosson, A. Persson, Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta(DE3), *Protein Expr. Purif.* 69 (2010) 159–167.
- [63] M.F. Alexeyev, H.H. Winkler, Gene synthesis, bacterial expression and purification of the *Rickettsia prowazekii* ATP/ADP translocase, *Biochim. Biophys. Acta* 1419 (1999) 299–306.
- [64] S.E. Bane, J.E. Velasquez, A.S. Robinson, Expression and purification of milligram levels of inactive G-protein coupled receptors in *E. coli*, *Protein Expr. Purif.* 52 (2007) 348–355.
- [65] D.O. Daley, M. Rapp, E. Granseth, K. Melen, D. Drew, G. von Heijne, Global topology analysis of the *Escherichia coli* inner membrane proteome, *Science* 308 (2005) 1321–1323.
- [66] M. Osterberg, H. Kim, J. Warringer, K. Melen, A. Blomberg, G. von Heijne, Phenotypic effects of membrane protein overexpression in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 11148–11153.
- [67] I. Arechaga, B. Miroux, M.J. Runswick, J.E. Walker, Over-expression of *Escherichia coli* F<sub>1</sub>F<sub>0</sub>-ATPase subunit a is inhibited by instability of the *uncB* gene transcript, *FEBS Lett.* 547 (2003) 97–100.
- [68] A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elofsson, Prediction of membrane-protein topology from first principles, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 7177–7181.
- [69] T. Becker, S. Bhushan, A. Jarasch, J.P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E.C. Mandon, R. Beckmann, Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome, *Science* 326 (2009) 1369–1373.