Poor codon optimality is a signal to degrade transcripts with frameshifts

Miquel Àngel Schikora-Tamarit[1], Lucas B. Carey[1]

[1]Systems Bioengineering Program, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Carrer Dr. Aiguader 88, 08003, Barcelona, Spain.

Corresponding authors:

MAST: mikischikora@gmail.com, LBC: lucas.carey@upf.edu

Keywords: nonsense-mediated decay, codon bias, frameshifts, mRNA quality, translation

Abstract:

Frameshifting errors are common and mRNA quality control pathways, such as nonsense-mediated decay (NMD), exist to degrade these aberrant transcripts. Recent work has shown the existence of a genetic link between NMD and codon-usage mediated mRNA decay. Here we present computational evidence that these pathways are synergic for removing frameshifts.

Frameshifting errors in gene expression

All biochemical pathways are intrinsically stochastic processes. Transcription, splicing, and translation are especially error prone, with error rates 4-6 orders of magnitude higher than that of DNA polymerase [1–6]. Such errors can result in single-amino acid substitutions, as well as truncation of the protein due to nonsense mutations or frameshifting errors. The latter can occur due to insertion and deletion events during transcription, splicing errors, and ribosomal slippage during translation **(Figure 1)**.

Frameshifts in protein coding genes are likely to be among the most damaging events, as they result in truncated proteins which may be misfolded or form dominant negative alleles [7,8] **(Figure 1)**. This justifies an evolutionary pressure for cells to contain mRNA surveillance pathways that remove transcripts bearing frameshifts. Suppression of frameshift errors is thought to be one of the major roles of the mRNA quality control machinery [9].

Nonsense-mediated decay for removing frameshifting errors

In eukaryotes, nonsense-mediated decay (NMD) is a conserved mRNA surveillance pathway that is often assumed to fulfill a frameshift-removing role [10]. This follows from the observation that frameshifts generate premature termination codons (PTCs), recognition of which targets the transcript for NMD. However, the quantitative effects of NMD, when measured, are often small [11,12]. In addition, a large fraction native transcripts (between 5%-30% depending on the genome) are targeted by NMD [13]. In the context of mRNA quality control, these are poor evidence for NMD being an effective quality control pathway.

The mechanism of NMD may be species-specific [10,12] and has even been proposed to be a passive result of the degradation of unprotected transcripts [14]. In yeast, NMD is thought to act on long 3'UTRs [15,16], so that transcripts bearing 3'UTRs longer than 250 nucleotides are targeted by NMD **(Figure 1)**. Recent work from our group has shown that this is mostly true and, importantly, the strength of NMD depends linearly on 3'UTR length [11] **(Figure 3B)**. However, native 3'UTR lengths are highly variable, ranging from 0 to 1461 nucleotides [17]. Frameshifts in native transcripts with short 3'UTRs are unlikely to result in efficient NMD.

These data suggest that NMD is both inaccurate and inefficient discretizing "correct" vs "incorrect" transcripts. We propose that an efficient quality control pathway should be better able to distinguish and degrade incorrect transcripts.

Codon bias and mRNA quality control

Recent work from our group [11] provides an unexpected clue towards understanding mRNA quality control. We found that two mechanisms of co-translational regulation, NMD and codon bias-dependent mRNA expression [18,19] **(Figure 2A)** are genetically linked; both pathways are regulated by the DEAD-box RNA helicase Dbp2 and by promoter architecture. A quantitative analysis of the impact of these pathways on mRNA levels gives rise to the hypothesis that they may act in a synergistic manner to remove transcripts with frameshifts. In addition to generating a PTC, frameshifts generate a second signal of "wrong transcript": a run of normally out-of-frame codons between the frameshift and the PTC that are now translated **(Figure 1)**. Below we provide computational support of this hypothesis.

The meaning and role of codon bias

All transcriptomes exhibit imbalances in the synonymous codons used for each amino acid. Not all synonymous codons are equally abundant, a phenomena called “codon bias”[20,21]. Highly expressed genes use codons translated by abundant tRNAs [22] and are coded by optimized codons **(Figure 2)**, leading to efficient protein synthesis. Highly expressed genes with efficient translation initiation but with suboptimal codon usage are deleterious and affect the expression of the rest of the proteome [23].

It was previously noted that use of optimal codons increased not only protein levels, but also mRNA levels [24], suggesting that ribosome speed might regulate mRNA stability. Recently, a pathway that involves the DEAD-box RNA helicase Dhh1 was found to target transcripts with suboptimal codon usage for decay [18,25]**.** Even short stretches of twelve suboptimal codons reduce mRNA levels [19], likely due to slower translation [26].

While most genes do not have highly optimized codon usage, the majority of the yeast transcriptome is populated by highly optimized mRNAs (**Figure 2B**). The top 10% of expressed genes have highly optimized codon usage. In yeast these genes account for 77% of the transcripts in a cell. Translational selection [27] will result in the optimized codon usage of constitutively highly expressed genes but will act less efficiently on genes with lower expression, genes that are rarely expressed, and of course on out-of-frame codons.

Codon optimality for removing frameshifting errors

In addition to producing PTCs, frameshifts are likely to introduce a stretch of non-optimized codons at the 3’end of the ORF **(Figure 1)**. In genes with optimized codons, this will result in a sudden changes in translation efficiency after the frameshift, which will reduce protein synthesis and target the transcript for decay (**Figure 3A**). This reasoning follows the observation that the impact of low codon optimality on translation efficiency and mRNA decay is local and can act over as few as twelve codons [19,26]. The magnitude of the decrease in codon optimality will be highest for transcripts with high codon optimization (most of the mRNAs in the cell **(Figure 2B)**), which correspond to highly expressed genes that likely bear most of the frameshifts (assuming a uniform distribution of errors across

transcripts [1]). Our hypothesis is that frameshift-removing mechanisms are especially relevant for such highly-expressed genes.

To compare the role of NMD and codon bias in mRNA quality control we ran a frameshift-introducing simulation on yeast transcripts. We generated random single-base deletions in native transcripts and calculated codon optimality (tRNA adaptation index, tAI [28]) and 3'UTR length with and without the frameshift. Because errors occur on a per transcript basis, each gene received a number of errors proportional to its mRNA expression level (**Figure 3C**).

We found that almost all frameshifts produce a large decrease in tAI after the mutation (**Figure 3D**). The change in tAI range due to frameshifts decreases mRNA levels [11] **(Figure 3A)**. In contrast, ~50% of errors produce 3'UTRs in the range of native 3'UTR lengths (**Figure 3D**), likely unaffected by NMD [11] **(Figure 3B)**. These findings indicate that selection for codon-optimality (which acts on highly expressed genes) can be a robust way to define "correct transcripts" and thus remove transcripts that contain frameshifts

Conclusions and open questions

Cells needs to remove transcripts with errors; mutants with increased error rates or that are unable to remove transcripts with errors grow slowly [1,29]. Frameshift errors are likely to be deleterious, both by generating deleterious protein isoforms, and because suboptimal codons titrate away both tRNAs and ribosomes [23,30]. However, both the sequence features that cells recognize and the mechanisms by which they do so remain poorly understood. Many open questions remain.

NMD is weak [11,12] and affects 5-20% of the native transcriptome [13], so it may be both inefficient and unspecific for removing errors. Removing transcripts with low codon optimality may be more accurate and efficient. This is consistent with the fact that NMD strength follows a linear relationship with 3'UTR length, while codon optimality has a sigmoidal impact on expression **(Figure 3)**. Small changes in codon optimality can lead to a large decrease in expression.

We observe that ~50% of frameshifts generate 3'UTRs within the range of native transcripts, likely unaffected by NMD. This exemplifies how a model based on a qualitative basis ("NMD removes frameshifts *because* these have longer 3'UTRs") can fail to predict of the quantitative behavior of a system.

Our recent work suggests a genetic link between codon bias and NMD [11]. Here we report a possible explanation of this interaction, but it remains to be seen which is the impact on measured expression levels of both processes. The mechanism of this link also remains to be established.

In frameshifted mRNAs, the quantitative impact of the low-tAI stretches of ORF in expression remains elusive. It will be interesting to see if they can explain more or less quality control than NMD. In addition, the effect of codon bias on expression is expected to impact protein levels [20,23], not only mRNA . This predicts that the impact of codon bias on expression is higher than reported here **(Figure 3A)**, which is not true for NMD. This could explain why we observe a lot of splice isoforms that have PTCs in humans, which may arise from frameshifting splicing errors. NMD does not remove them (as we can detect them), but it is likely that they have lower codon adaptation and reduced protein levels.

Finally, this work suggests and adaptive benefit for imbalanced tRNA repertoires [22], which would confer the ability to degrade transcripts that are not supposed to be highly expressed. It is almost certain that cells avoid selecting the expression of ORFs with a random composition of codons. Frameshifts generate such random stretches, that are likely targeted for decay. Thus, there may be an evolutionary pressure for imbalanced tRNA repertoires to ensure proper mechanisms of mRNA quality control.

Funding details:

This work was supported by the Ministerio de Economía y Competitividad (MINECO) under grant BFU2015-68351-P, by AGAUR under grants 2014SGR0974 & 2017SGR1054, and the Unidad de Excelencia María de Maeztu, funded by the MINECO, under grant MDM-2014-0370.

Figure legends

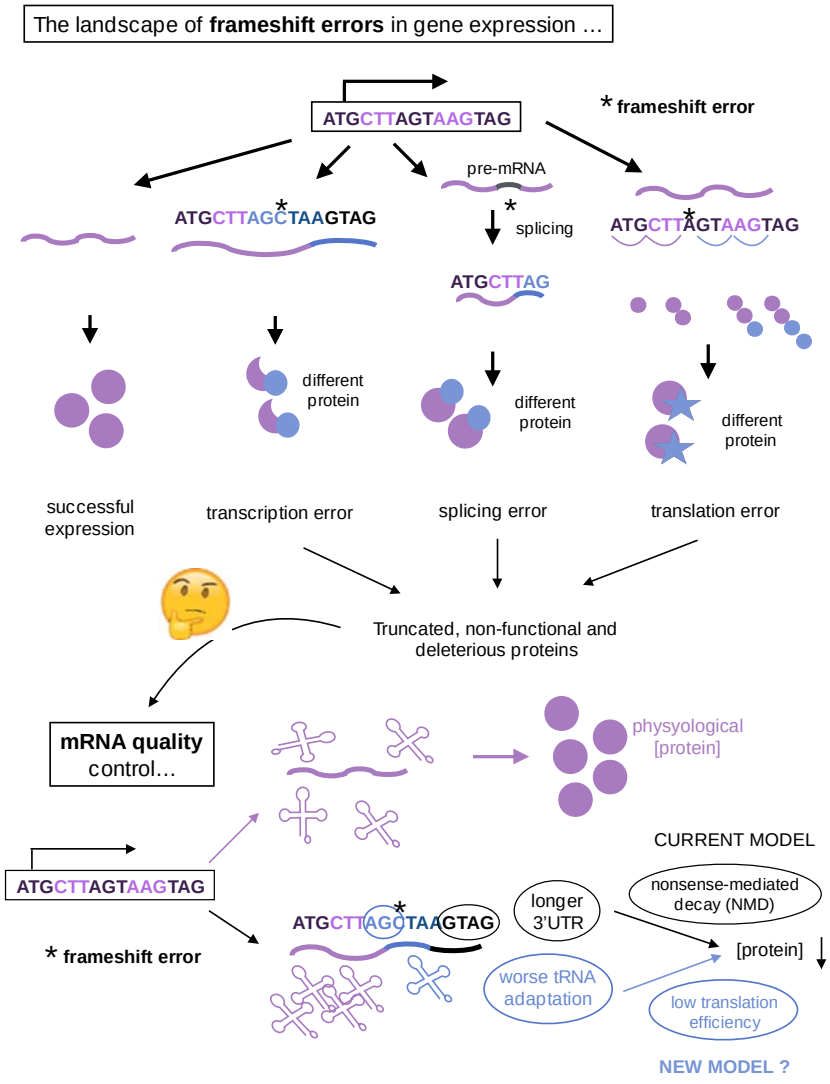

Figure 1: The impact of frameshifting errors in gene expression. Gene expression can result in frameshifting errors (indicated as *) due to transcriptional insertion/deletion epimutations, errors in splicing or ribosomal slippage during translation (top). These processes potentially generate deleterious proteins, which justifies the need of mRNA quality control mechanisms in cells (bottom). In the absence of errors, mRNAs are translated leading to physiological protein levels. The current model indicates that frameshifting errors generate Premature Termination Codons (PTC) that trigger Nonsense-Mediated Decay (NMD) on them, mainly because of the generated long 3'UTR (in yeast). Our hypothesis is that NMD is often nonspecific for errors, so that other quality control mechanisms must exists. We note that another signal of "incorrectness" may appear in transcripts with frameshifts: a stretch of poorly-optimized codons (in blue, indicating worse tRNA adaptation) between the error and the PTC. This should lead to reduced translation efficiency, mRNA decay and lower protein concentrations of the frameshifted transcript.

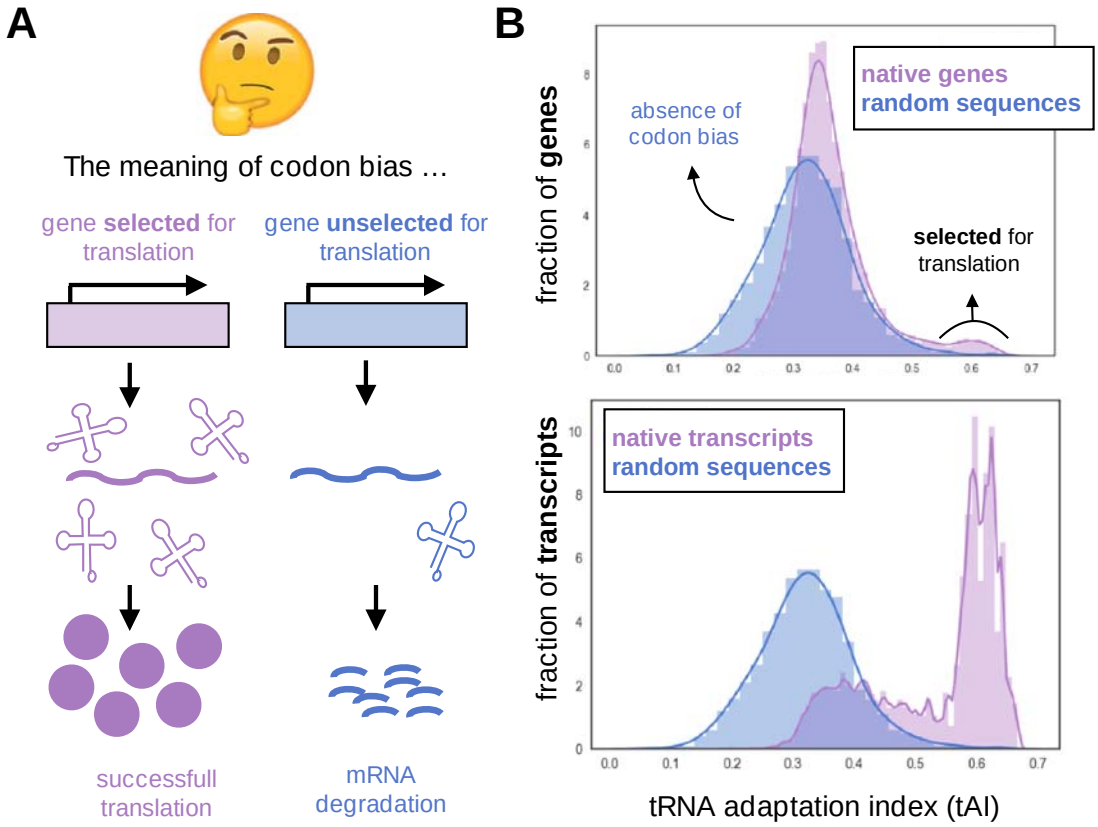

Figure 2: The meaning of codon bias in the transcriptome. (A) Highly expressed genes are often selected to have optimized codons in agreement with the cellular tRNA pool, allowing efficient translation of them (purple). This is known as "translational selection" [20–23]. On the other hand, genes with a poor codon optimization are inefficiently translated and targeted for mRNA decay (blue) [18]. **(B)** Top: in yeast, most native *genes* (purple) exhibit a tRNA Adaptation Index (tAI, as a measure of codon optimality) in the range of ORFs predicted from random transcription throughout the genome (blue). Such random ORFs simulate the absence of codon bias in terms of tRNA adaptation. A small fraction of *genes* have non-random tAI, which corresponds to genes "selected for translation". Bottom: most native *transcripts* (purple) have high tAI, as compared to random ORFs (blue). This histogram was generated weighting each gene by mRNA expression level (which is exponentially distributed), which indicates the per-transcript distribution of tAI.

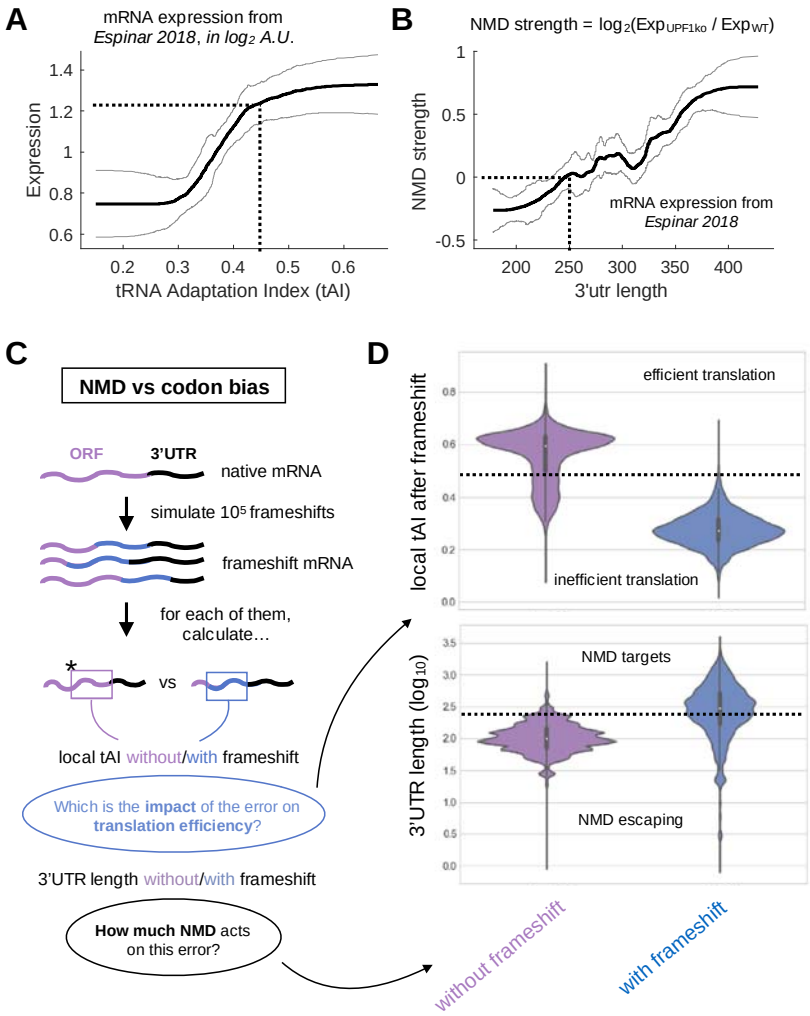

Figure 3: Codon bias can implement quality control of mRNAs with frameshifts. (A) tAI follows a negative sigmoidal relationship with mRNA expression levels. Expression was calculated as the $\log_2$-ratio between mRNA and DNA abundance of a synthetic ORF library of random fragments from the yeast genome, expressed in a plasmid [11]. The dashed line represent a threshold in which decreasing tAI reduces expression. **(B)** NMD strength follows a positive linear relationship with 3'UTR length. NMD was measured as the expression (calculated as in A) $\log_2$-ratio between identical ORF libraries built in a *Δupf1* or a *wt* strain [11]. This ratio indicates the impact of NMD for each sequence in the library (which has variable 3'UTR lengths), as UPF1 is responsible for NMD [10]. The dashed line represent a threshold in which increasing 3'UTR generates NMD (positive values in the Y axis). **(C)** A pipeline for predicting the impact of NMD and codon on frameshift quality control. As an example of frameshift, we simulated 10^5 random single-base deletions on native transcripts. Each gene includes a number of mutations proportional to its expression level. For each error (and corresponding native transcript) we calculated tAI between the frameshift and the PTC (local tAI) and the resulting 3'UTR length. We used these as measures of the impact of error on translation efficiency and/or NMD targeting. **(D)** Transcripts with frameshifts (blue) have lower tAI (top) and longer 3'UTRs (bottom), when compared to native mRNAs (purple). The dashed lines represent the thresholds described in A,B.

References

[1] Gout J-F, Li W, Fritsch C, et al. The landscape of transcription errors in eukaryotic cells. Science Advances. 2017;3:e1701484.

[2] Gout J-F, Thomas WK, Smith Z, et al. Large-scale detection of in vivo transcription errors. Proc. Natl. Acad. Sci. U. S. A. 2013;110:18584–18589.

[3] Carey LB. RNA polymerase errors cause splicing defects and can be regulated by differential expression of RNA polymerase subunits. Elife [Internet]. 2015;4. Available from: http://dx.doi.org/10.7554/eLife.09945.

[4] Milo R, Jorgensen P, Moran U, et al. BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res. 2010;38:D750–D753.

[5] Imashimizu M, Oshima T, Lubkowska L, et al. Direct assessment of transcription fidelity by high-resolution RNA sequencing. Nucleic Acids Res. 2013;41:9090–9104.

[6] Fox-Walsh KL, Hertel KJ. Splice-site pairing is an intrinsically high fidelity process. Proc. Natl. Acad. Sci. U. S. A. 2009;106:1766–1771.

[7] Weterman MAJ, Sorrentino V, Kasher PR, et al. A frameshift mutation in LRSAM1 is responsible for a dominant hereditary polyneuropathy. Hum. Mol. Genet. 2012;21:358–370.

[8] Sadhu MJ, Bloom JS, Day L, et al. Highly parallel genome variant engineering with CRISPR–Cas9. Nat. Genet. 2018;50:510–514.

[9] Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev. 2007;21:1833–3856.

[10] Behm-Ansmant I, Kashima I, Rehwinkel J, et al. mRNA quality control: An ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett. 2007;581:2845–2853.

[11] Espinar L, Schikora Tamarit MÀ, Domingo J, et al. Promoter architecture determines cotranslational regulation of mRNA. Genome Res. [Internet]. 2018; Available from: http://dx.doi.org/10.1101/gr.230458.117.

[12] Lindeboom RGH, Supek F, Lehner B. The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat. Genet. [Internet]. 2016; Available from: http://dx.doi.org/10.1038/ng.3664.

[13] Peccarelli M, Kebaara BW. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryot. Cell. 2014;13:1126–1135.

[14] Brogna S, McLeod T, Petric M. The Meaning of NMD: Translate or Perish. Trends Genet. 2016;32:395–407.

[15] Zhang J, Sun X, Qian Y, et al. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Mol. Cell. Biol. 1998;18:5272–5283.

[16] Amrani N, Ganesan R, Kervestin S, et al. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature. 2004;432:112–118.

[17] Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320:1344–1349.

[18] Radhakrishnan A, Chen Y-H, Martin S, et al. The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell [Internet]. Available from: http://www.sciencedirect.com/science/article/pii/S0092867416311503.

[19] Chen S, Li K, Cao W, et al. Codon-Resolution Analysis Reveals a Direct and Context-Dependent Impact of Individual Synonymous Mutations on mRNA Level. Mol. Biol. Evol. 2017;34:2944–2958.

[20] Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 1985;2:13–34.

[21] Grantham R. Working of the genetic code. Trends Biochem. Sci. 1980;5:327–331.

[22] Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet. 2012;28:574–581.

[23] Frumkin I, Lajoie MJ, Gregg CJ, et al. Codon usage of highly expressed genes affects proteome-wide translation efficiency. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2018; Available from: http://dx.doi.org/10.1073/pnas.1719375115.

[24] Te'o VS, Cziferszky AE, Bergquist PL, et al. Codon optimization of xylanase gene xynB from the thermophilic bacterium Dictyoglomus thermophilum for expression in the filamentous fungus Trichoderma reesei. FEMS Microbiol. Lett. 2000;190:13–19.

[25] Harigaya Y, Parker R. Codon optimality and mRNA decay. Cell Res. 2016;26:1269–1270.

[26] Yu C-H, Dang Y, Zhou Z, et al. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Mol. Cell. 2015;59:744–754.

[27] Akashi H, Eyre-Walker A. Translational selection and molecular evolution. Curr. Opin. Genet. Dev. 1998;8:688–693.

[28] dos Reis M, Wernisch L, Savva R. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Res. 2003;31:6976–6985.

[29] Cherry JM, Hong EL, Amundsen C, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40:D700–D705.

[30] Shah P, Ding Y, Niemczyk M, et al. Rate-limiting steps in yeast protein translation. Cell. 2013;153:1589–1601.