Genomics 111 (2019) 1292–1297

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods Paper

The optimization of mRNA expression level by its intrinsic properties—Insights from codon usage pattern and structural stability of mRNA

Manish P. Victor[a], Debarun Acharya[a,*], Tina Begum[b], Tapash C. Ghosh[a,*]

[a] Bioinformatics Centre, Bose Institute, Kolkata, West Bengal, India
[b] Department of Ecology and Evolution, University of Lausanne, Switzerland

ARTICLE INFO

Keywords:
Codon usage bias (CUB)
Codon adaptation index (CAI)
Gene expression
mRNA stability
Folding free energy

ABSTRACT

Codon usage bias (CUB) and mRNA structural stability are important intrinsic features of mRNA that correlate positively with mRNA expression level. However, it remains unclear whether the mRNA expression level can be regulated by adjusting these two parameters, influencing the mRNAs' structure. Here we explored the influence of CUB and mRNA structural stability on mRNA expression levels in *Saccharomyces cerevisiae*, using both wild type and computationally mutated mRNAs. Although in wild type, both CUB and mRNA stability positively regulate the mRNA expression level, any deviation from natural situation breaks such equilibrium. The naturally occurring codon composition is responsible for optimizing the mRNA expression, and under such composition, the mRNA structure having highest stability is selected by nature.

1. Introduction

A codon is the combination of three DNA or RNA nucleotides that encode an amino acid and supply it to the polypeptide chain or stop protein synthesis during translation. The codons sum to a total of 61 in number that encodes 20 amino acids. With Tryptophan and Methionine as exceptions, all other amino acids are coded by two to six codons. Such a twofold to sixfold degeneracy in genetic code enables mRNA to carry extra information without altering the encoded amino-acids. All such degenerate codons responsible for coding the same amino acid are referred to as synonymous codons [17]. Although a lot of coding sequences are theoretically possible for every protein, only one of them is selected by nature [41]. Such a deviation from the uniform usage of codons is known as codon usage bias (CUB) that refers to the differences in the frequency of occurrence of synonymous codons in coding DNA [17].

Codon usage bias has its influence at genomic levels like gene length [5,13], GC-composition [3,7,25,40], Mutational patterns [38,39] etc. and is guided by the available tRNA pool of that species [2,21,24,36]. This ultimately impacts the translational efficiency and gene expression level [1,18]. Previous studies on codon usage in Enterobacteria, *Escherichia coli* and Yeast revealed a predominant use of optimal codons in highly expressed genes, compared to the uniform use of all synonymous codons in weakly expressed genes [11,17,34]. So, a small tRNAs pool is presented by the highly expressed class of genes, which may reduce the screening procedure and facilitate rapid expression with fidelity, explaining the selection pressure for translational accuracy [6,20]. It has also been shown by the translational selection hypothesis that the close association of codon usage pattern and the tRNA pool prevents translation machinery from incorporating incorrect amino acid(s), facilitating the formation of desired mRNA sequence responsible for its secondary structure [6,20]. The mRNA secondary structures, in turn, facilitate the appropriate abundance and binding of ribosomes [27]. Therefore, in addition to fidelity and the rate of translation, codon usage bias also influences the mRNA secondary structure and regulates its stability owing to such structural integrity. mRNA stability arising out of its innate secondary structures plays a major role in influencing transcript splicing [30], ribosome abundance and its binding [27], protein abundance [44] and regulation of gene expression level and its accuracy [16,27,37]. It is also responsible for halts in protein synthesis [28,43] and is a major determinant of protein secondary structure [8–10,23]. Studies have also shown the preference of codons pairing to high-abundance tRNAs in the regions having high mRNA secondary structure content, while codons read by significantly less abundant

Abbreviations: CUB, Codon usage bias; CAI, Codon adaptation index; miRNA, micro RNA; ENC/Nc, Effective Number of Codons

* Corresponding authors at: Bioinformatics Centre, Bose Institute, Centenary Campus, P- 1/12, C.I.T. Scheme VIIM, Kankurgachhi, Kolkata 700054, West Bengal, India.

E-mail addresses: debarun@jcbose.ac.in (D. Acharya), tapash@jcbose.ac.in (T.C. Ghosh).

https://doi.org/10.1016/j.ygeno.2018.08.009
Received 27 June 2018; Received in revised form 14 August 2018; Accepted 24 August 2018
Available online 01 September 2018

tRNAs are situated in lowly structured regions of the mRNA [14]. This suggests that the influence of transcriptional forces on codon usage bias improves gene expression by optimizing mRNA secondary structure [41].

In this study, we investigated the influence of codon usage pattern and mRNA stability (folding free energy) in regulating the mRNA abundance in the budding yeast *Saccharomyces cerevisiae*. Our results indicate that both mRNA stability and codon usage pattern fine-tunes the expression level of *S. cerevisiae* mRNAs. This influence of codon usage pattern and mRNA structural stability on mRNA expression levels is not only limited to the transcriptional domain, but also impacts translation as both transcription and translation are articulated biological processes [41].

2. Materials and methods

2.1. Sequence and expression data

We downloaded the *Saccharomyces cerevisiae* ORFs from *Saccharomyces* Genome Database (https://www.yeastgenome.org/) [12]. We removed all erroneous sequences by CodonW (version 1.4.4, https://sourceforge.net/projects/codonw/) [31] and obtained a total of 5894 mRNA sequences. The mRNA molecules per cell were collected from genome-wide expression analysis data from Holstege et al. [19]. We also used another up-to-date large-scale expression dataset for validation purpose [32]. We obtained protein abundance data from Arava et al. [4]. Both datasets hold a strong and significant positive correlation with each other (Table S1 in Supplementary File 1). We averaged across the two datasets of mRNA abundance after normalizing each data set and considered as mRNA abundance in this study.

2.2. Randomization of native mRNA sequence

In this study, we did two types of codon randomization. Firstly, we randomly replaced the codons in the mRNA structure with synonymous codons, making them to use different types of codon composition (different extents of codon usage bias) while keeping the polypeptide sequences unchanged (Fig. 1). In this way, we were able to computationally generate mRNA sequences ranging from high to low ENC values (ENC~20, ENC~30, ENC~40, ENC~50, and ENC~61) with varying AT/GC content.

In another analysis, we generated 100 sequences for each native mRNA sequence by randomly shuffling the synonymous codons within the mRNA, keeping the polypeptide sequence and ENC same as the native form (Fig. 2). In-house PERL scripts were written for both the randomization processes. The PERL scripts for both the randomization processes are provided in Supplementary file 2.

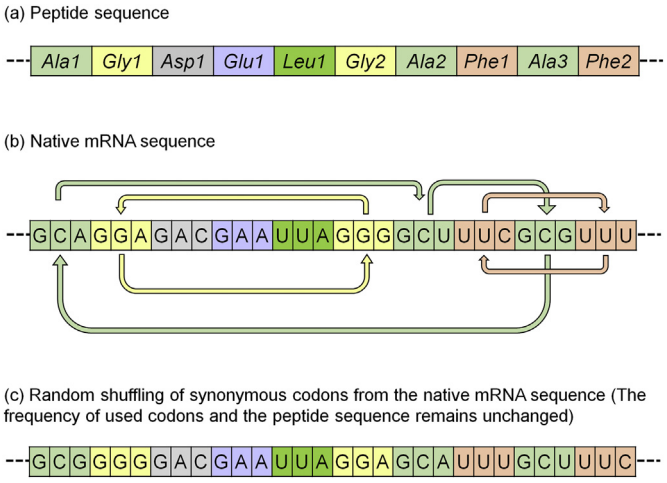


Fig. 2. Schematic representation of creating random mRNA sequences by swapping the synonymous codons within each mRNA sequence. Here, the protein sequences, ENC and CAI remain unchanged, the mRNA folding free energies of randomized sequences show changes.

2.3. CAI and ENC calculation

Codon adaptation index (CAI) and effective number of codons (ENC) for each randomized and native *Saccharomyces cerevisiae* ORFs were determined using the program CodonW [31]. CAI compares the codon composition of a gene with respect to a reference set of highly expressed genes, and is used to assess the expression of the gene solely based on its codon composition. [33]. ENC is a measure of codon usage bias. It measures the deviation from a uniform codon usage based on null hypothesis [42]. A low ENC value (ENC ranges as $20 \geq ENC \geq 61$) denotes a high codon usage bias, and a high CAI value indicates high expression level.

2.4. Folding free energy and statistical calculation

The folding free energy for NATIVE and all the randomized mRNA sequences were calculated using the RNAfold attribute from Vienna

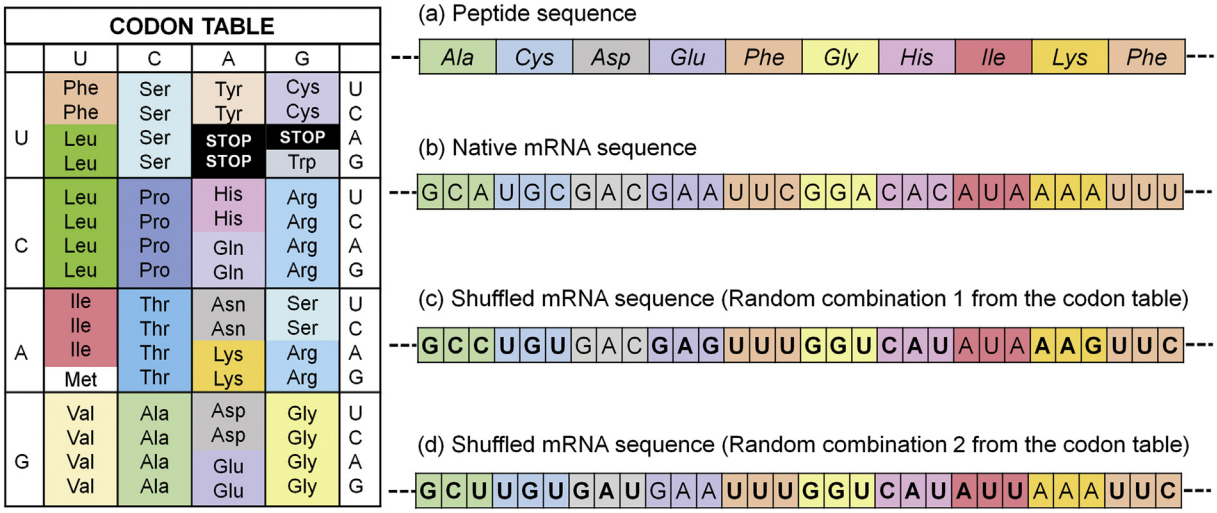


Fig. 1. Schematic representation of creating random mRNA sequences by replacing the existing codons with synonymous codons from the codon table. Here, the protein sequences remain unchanged, ENC, CAI and mRNA folding free energies show changes in the randomized sequences.

software package 2 [26]. The folding free energy was calculated with the window size of 40 nucleotides in the step size of 1 nucleotide, a similar process used by Park et al. [29]. The predicted local folding free energies were averaged to obtain the folding free energy for the entire mRNA sequence.

The statistical calculations were carried out using IBM SPSS 22. All the correlations reported in this study are Spearman correlation. In all statistical analysis, we used 95% level of confidence as a measure of significance.

3. Results

3.1. The relationship between ENC and mRNA abundance in Saccharomyces cerevisiae

Effective number of codons (ENC) represents the state of codon usage biases (CUB) in genes and genomes and ranges from 20 (extreme bias using only one codon per amino acid) to 61 (no bias towards any codon) [42]. In *S. cerevisiae* genome, the distribution of ENC is not uniform, and most of the genes show an unbiased use of codons, as defined by their high ENC values (~63% genes possess ENC values 50–61) (Fig. 3). However, consistent with earlier studies [17], we obtained a negative correlation between ENC and mRNA abundance in *S. cerevisiae* (Spearman $\rho = -0.446$; $P = 6.366 \times 10^{-222}$; $n = 4562$), confirming that genes expressing high codon usage bias show elevated mRNA expression levels, and in turn protein abundance, as these two are positively correlated with each other (Spearman $\rho = 0.663$; $P < 0.01$; $n = 4152$). In addition to mRNA abundance data, when we have used codon adaptation index (CAI) as a proxy of mRNA expression level, we also obtained a negative correlation of ENC and CAI (Spearman $\rho = -0.651$; $P < 0.01$; $n = 5859$).

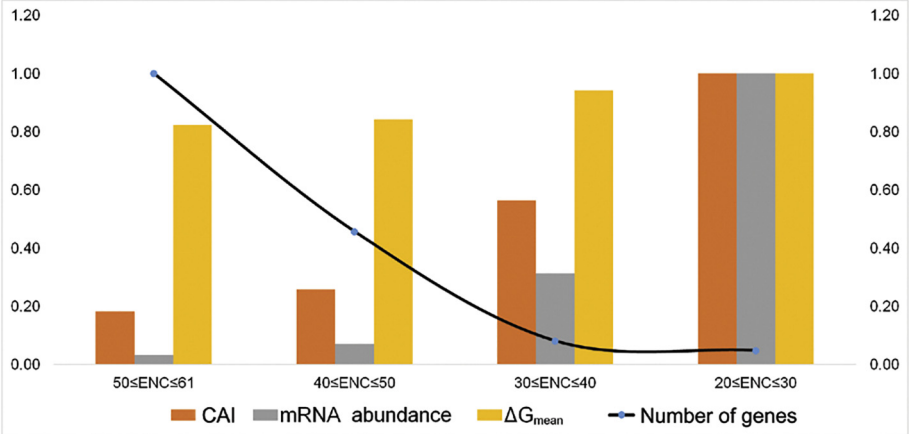
3.2. The relationship between mRNA stability and mRNA abundance in Saccharomyces cerevisiae

mRNA stability in this study is considered as the structural stability of mRNA owing to its secondary structures and not to be confused with mRNA stability indicated by its half-life. The structural stability of mRNA is measured by the folding free energy (ΔG), where a lower ΔG value indicates higher mRNA stability. For native mRNA sequences, there is a significant negative correlation between mRNA abundance and folding free energy (Spearman $\rho = -0.344$; $P = 1.050 \times 10^{-126}$; $n = 4562$). A similar trend is observed between CAI for native sequences and folding free energy (Spearman $\rho = -0.160$; $P = 4.788 \times 10^{-35}$; $n = 4562$). The results indicate that the more stable an mRNA becomes, its expression value increases. To further investigate the tendency of the mRNA abundance with varying mRNA stability, genes were divided into five categories, from highest to lowest mRNA stability. We find that with a nominal drop in the mRNA stability its abundance has a considerable dip. Such a drop in mRNA abundance might not be due to the exclusive effect of mRNA stability (Table 1). (Category1: $\Delta G = -6.412$; mRNA abundance = 9.328) and (Category5: $\Delta G = -5.369$; mRNA abundance = 0.839), instead may be the cumulative effect of it along with ENC, which also holds a positive correlation with folding free energy ($\rho = 0.108$; $P = 1.405 \times 10^{-16}$; $n = 5859$), indicating that genes using more biased codons will possess higher structural stability in their transcript.

Thus, to understand the cumulative impact of codon usage pattern (ENC) and mRNA structural stability (ΔG), we used two randomization

Table 1
The mRNA abundance and protein abundance (NATIVE) show a decrease in the value with increasing folding free energy (ΔG_{mRNA}), (i.e. stability is decreasing from category 1 → 5). The dataset represents 5140 genes for which mRNA stability and protein abundance data are available. The categorisation has been done on the decreasing stability of the mRNA secondary structure.

Category	ΔG_{mRNA} mean	mRNA Abundance $_{mean}$	Protein abundance (protein molecules per second) mean	ENC $_{mean}$	Dataset (n)
1	−6.142	9.328	1.259	47.45	1028
2	−5.778	1.779	0.2	49.32	1028
3	−5.699	1.058	0.099	50.54	1028
4	−5.445	0.752	0.049	51.42	1028
5	−5.369	0.839	0.017	51.43	1028



ENC range	Number of genes	Mean CAI	Mean mRNA abundance	Mean ΔG
50≤ENC≤61	3695	0.132	0.966	-5.573
40≤ENC≤50	1688	0.187	2.135	-5.703
40≤ENC≤50	299	0.408	9.542	-6.383
20≤ENC≤30	177	0.725	30.452	-6.772

Fig. 3. The bar-plot showing the assortment of genes under different intervals of ENC values with average CAI, mRNA abundance and folding free energy of mRNA for NATIVE sequences. The values shown here are normalized from 0 to 1 and are representative of relative quantities (0 = lowest, 1 = highest).

studies, by using codon manipulation by our in-house PERL script (Supplementary file 2). We used CAI as a proxy of mRNA expression levels, as realistic mRNA abundance data cannot be generated computationally, and CAI calculates the relative adaptiveness of the codon usage of a gene with respect to that of highly expressed genes, useful for the estimation of mRNA abundance and holds a strong and significant positive correlation with the latter [22] (Spearman $\rho = 0.584$; $P < 0.01$; $n = 4562$).

3.3. Random shuffling of codon composition and its effect on mRNA expression level in Saccharomyces cerevisiae

Due to the degeneracy in the genetic code, there are numerous theoretically possible combinations of a gene or it's mRNA that may ultimately lead to the production of the same protein. However, all such combinations are not equivalent and may lead to differences in their gene expression level [15,33,35]. Here, we randomly replaced the native mRNA structure of all *Saccharomyces cerevisiae* mRNAs by using synonymous codons in codon degenerate family (Fig. 1), thereby restricting their ENC values to different combinations of codon usage bias (see, Materials and methods). This also influences the mRNA secondary structure stability, as it changes the mRNA sequence responsible for that structure. We computed the CAI values for all cases and observed that the native structure of mRNAs show highest CAI values (Table S2 in Supplementary File 1), suggesting that although highly expressed genes have a high codon-usage bias, increasing the bias of all genes does not result in their elevated mRNA expression. Instead, the native CAI value is always greater than any class of randomly replaced sequences (Fig. 1, Table S2). The results suggest that under strong bias (ENC = 20.96), CAI (0.09) is never approximately equal to the native CAI (0.18). Rather, at a higher ENC value which is closer to native ENC (ENC = 41.60 the CAI =0.14) approaches very near to the $CAI_{NATIVE} = 0.18$. Moreover, these randomly replaced sequences for all the possible codon compositions show a negative correlation between mRNA stability and effective number of codons, as well as for codon adaptation index (Table S2). This is an anomaly to the natural behaviour of native sequences and clearly suggests that the existing codon composition is the key for achieving the optimal mRNA expression in yeast.

3.4. Random shuffling within mRNA sequences and its effect on mRNA expression level in Saccharomyces cerevisiae

For a gene, there are numerous theoretically possible combinations of mRNA sequences for the existing codon composition. If some codons are signatures for elevated gene expression, their presence will lead to higher expression of that gene (regardless of their positions). But the mere presence of those codons for high gene expression needs to be examined. To decipher the effect of mRNA structural stability on mRNA abundance with the same codon composition (same ENC), we did another randomization study. We randomly shuffled the codons within each native mRNA sequences of 5859 *Saccharomyces cerevisiae* genes and generated one hundred random sequences per mRNA that have exactly same codon composition, ENC and CAI values (Fig. 2). However, the stability of the mRNA is expected to vary in different random combinations, as with the change in the position of trinucleotides, structural alterations to this mRNA are expected to happen. We calculated the stability of one hundred such randomly generated combinations for each mRNA (see Materials and methods). We found that 5558 genes among 5859 showed the highest stability for native sequences (Supplementary table S3 in Supplementary file 3), i.e. 94.86% of total genes in a genome (Z score = 97.128; $p < 0.0001$). This result indicates that under the native codon composition, the mRNA structural stability also plays a very important role in obtaining the apt mRNA abundance. Taken together, our results suggest that the native ENC values of the mRNAs help to optimize the mRNA expression level, and under such codon composition, the mRNA structure with the highest stability represents the native structure and is responsible for optimization of mRNA expression level.

4. Discussion

Codons are tri-nucleotide sequences situated in the mRNA and are considered as the key to the gene expression signature of an mRNA. In this study, we explored the association of transcript structural features with its expression level in *Saccharomyces cerevisiae*. Firstly, we focused on the mRNA sequence to observe codon composition and its effect on mRNA expression. We observed a significant negative correlation between the effective numbers of codons (ENC) with mRNA expression levels, indicating a higher mRNA expression levels for genes having biased codon composition, a trend similar to that observed in previous studies [44]. This further signifies that the degeneracy of the genetic code enables mRNA to carry additional information without altering the encoded amino-acid sequence [17]. This is mostly due to its influence on the mRNA secondary structure, as the latter depends on the sequence of mRNA. We observed that the stability of mRNA secondary structure holds a positive correlation with mRNA abundance (Table S1). This suggests that a biased usage of codons results in mRNAs with higher structural stability, and the stable mRNA with high abundance favours greater protein abundance [44]. The greater stability (due to the abundant folded regions in secondary structures) of the mRNA is supported by the high fidelity and translational efficiency [14,16,27,37].

In the next part, we explored whether the codon usage bias is enough to guide the mRNA expression level in *Saccharomyces cerevisiae*. As every gene has its own ENC values reflecting its codon composition, we computationally modified the codon composition to allow each gene to use different levels of codon usage patterns from low to high range of codon usage bias (Fig. 1). We hypothesized that if the codon usage bias is responsible for elevated mRNA expression levels, the randomized group with highest codon usage bias (lowest ENC) should show the highest mRNA expression (Highest CAI value). However, we found that this is not true, as the native condition shows the highest mean CAI values (CAI = 0.18) than any of the random combinations, be it both high as well as low codon usage bias (Fig. 4). Interestingly, among all the randomized conditions, the CAI value becomes highest

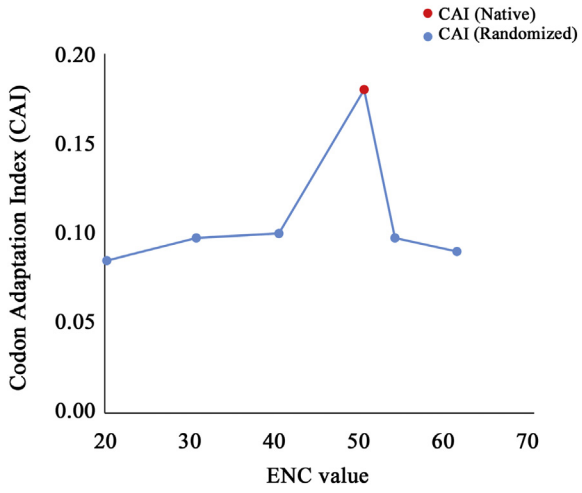


Fig. 4. Average Codon adaptation indices (CAI) of *Saccharomyces cerevisiae* mRNAs in wild-type versus the computationally randomized sequences at different levels of codon compositions. The red dot represents the CAI and ENC in the wild type. Blue dots represent corresponding ENC and CAI values in randomized sequences with different levels of codon composition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(CAI = 0.14) very close to that of genome averaged ENC values (Table S2). This suggests that although in native condition, genes having high codon usage bias have high mRNA expression levels, but genes with low codon usage preferences do not show elevated mRNA expression under computationally generated strong codon usage bias, indicating that the native codon composition is responsible for the optimum expression of *S. cerevisiae* mRNAs. However, if only the codon composition is responsible for the elevated mRNA expression, the mere presence of some codons will give the highest expression level of *S. cerevisiae* mRNAs. Conversely, as mRNA structural stability has its influence on mRNA abundance, the position of these codons in an mRNA may be crucial for optimal expression of the mRNA. Thus, we took the native codon composition of each of *S. cerevisiae* mRNAs and randomly shuffled the degenerate codons within each mRNA sequence, keeping the codon composition and the encoded peptide sequence unchanged, but altering the synonymous codons' positions within each mRNA (Fig. 2). We made 100 such randomly shuffled combinations and calculated the mRNA secondary structure stability (folding free energy, ΔG) and compared them with that of the native sequences. We observed a higher mRNA stability for native sequences than the randomly generated combinations in 94.863% genes in *S. cerevisiae* genome (Z score = 97.128; $p < 0.0001$) (Supplementary table S3 in Supplementary File 3). This indicates that along with the native codon composition, the positions of these codons in the mRNA structure play a major role in achieving the most stable mRNA secondary structure required for the optimal mRNA expression levels of *S. cerevisiae*.

5. Conclusion

Our study suggests that the existing codon composition and the position of such codons are essential for the optimum expression levels of *Saccharomyces cerevisiae* mRNAs. If any deviation is introduced to the native codon composition the mRNA expression level is perturbed. Although highly expressed genes prefer biased usage of codons, such narrow choice of codon pool does not promote the elevation of mRNA expression level globally. Rather, the existing codon composition is responsible for optimal expression of *S. cerevisiae* mRNAs. Additionally, our study also encompasses the positional priority of codons. It is evident from the analysis of random shuffling that under native codon usage pattern there is a strong preference for the most stable secondary structure. In other words, we can say both mRNA stability and meticulous codon usage fine-tunes the gene expression level.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data were obtained from publicly available databases (mentioned in the 'Materials and methods' section) and are freely available online. The datasets used and/or analysed during the current study are also available from the corresponding author on reasonable request.

Conflict of interest

None of the authors have any competing interests.

Funding

This work was financially supported by Centre of Excellence in Bioinformatics, Bose Institute, Kolkata, West Bengal, India.

Authors' contributions

MPV and TB conceived and performed the work. MPV and DA wrote the manuscript, DA, TB and TCG helped in planning and designing of the work and drafting the manuscript. All authors reviewed and approved the manuscript.

Acknowledgements

The authors thank Mr. Sanjib Kumar Gupta, Dr. Sandip Chakraborty and Dr. Jyotirmoy Das for their technical advice.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.08.009.

References

[1] H. Akashi, Translational selection and yeast proteome evolution, Genetics 164 (2003) 1291–1303.
[2] S.G. Andersson, C.G. Kurland, Codon preferences in free-living microorganisms, Microbiol. Rev. 54 (1990) 198–210.
[3] M.A. Antezana, M. Kreitman, The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences, J. Mol. Evol. 49 (1999) 36–43.
[4] Y. Arava, Y. Wang, J.D. Storey, C.L. Liu, P.O. Brown, D. Herschlag, Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3889–3894.
[5] W. Bains, Codon distribution in vertebrate genes may be used to predict gene length, J. Mol. Biol. 197 (1987) 379–388.
[6] S.K. Behura, D.W. Severson, Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes, Biol. Rev. Camb. Philos. Soc. 88 (2013) 49–61.
[7] G. Bernardi, G. Bernardi, Compositional constraints and genome evolution, J. Mol. Evol. 24 (1986) 1–11.
[8] J. Biro, Nucleic acid chaperons: a theory of an RNA-assisted protein folding, Theor. Biol. Med. Model. 2 (2005).
[9] J. Biro, Indications that "codon boundaries" are physico-chemically defined and that protein-folding information is contained in the redundant exon bases, Theor. Biol. Med. Model. 3 (2006).
[10] F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M.V. Rodnina, A.A. Komar, Synonymous codons direct cotranslational folding toward different protein conformations, Mol. Cell 61 (2016) 341–351.
[11] M. Bulmer, The selection-mutation-drift theory of synonymous codon usage, Genetics 129 (1991) 897–907.
[12] J.M. Cherry, E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K.R. Christie, M.C. Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, K. Karra, C.J. Krieger, S.R. Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simison, S. Weng, E.D. Wong, Saccharomyces genome database: the genomics resource of budding yeast, Nucleic Acids Res. 40 (2012) D700–D705.
[13] A. Eyre-Walker, Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? Mol. Biol. Evol. 13 (1996) 864–872.
[14] T.E. Gorochowski, Z. Ignatova, R.A. Bovenberg, J.A. Roubos, Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate, Nucleic Acids Res. 43 (2015) 3022–3032.
[15] M. Gouy, C. Gautier, Codon usage in bacteria: correlation with gene expressivity, Nucleic Acids Res. 10 (1982) 7055–7074.
[16] M. Grunberg-Manago, Messenger RNA stability and its role in control of gene expression in bacteria and phages, Annu. Rev. Genet. 33 (1999) 193–227.
[17] R. Hershberg, D. Petrov, Selection on Codon Bias, Annu. Rev. Genet. 42 (2008) 287–299.
[18] Y. Hiraoka, K. Kawamata, T. Haraguchi, Y. Chikashige, Codon usage bias is correlated with gene expression levels in the fission yeast Schizosaccharomyces pombe, Genes Cells 14 (2009) 499–509.
[19] F.C. Holstege, E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, R.A. Young, Dissecting the regulatory circuitry of a eukaryotic genome, Cell 95 (1998) 717–728.
[20] T. Ikemura, Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs, J. Mol. Biol. 158 (1982) 573–597.
[21] T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms, Mol. Biol. Evol. 2 (1985) 13–34.
[22] M. Jia, Y. Li, The relationship among gene expression, folding free energy and codon usage bias in *Escherichia coli*, FEBS Lett. 579 (2005) 5333–5337.
[23] M. Jia, L. Luo, C. Liu, Statistical correlation between protein secondary structure and messenger RNA stem-loop structure, Biopolymers 73 (2004) 16–26.
[24] S. Kanaya, Y. Yamada, M. Kinouchi, Y. Kudo, T. Ikemura, Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency

and with CG-dinucleotide usage as assessed by multivariate analysis, J. Mol. Evol. 53 (2001) 290–298.
[25] S. Karlin, J. Mrázek, What drives codon choices in human genes? J. Mol. Biol. 262 (1996) 459–472.
[26] R. Lorenz, S.H. Bernhart, C. Höner Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algoritm. Mol. Biol. 6 (2011) 26.
[27] Y. Mao, H. Liu, Y. Liu, S. Tao, Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae, Nucleic Acids Res. 42 (2014) 4813–4822.
[28] K. Mita, S. Ichimura, M. Zama, T.C. James, Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA, J. Mol. Biol. 203 (1988) 917–925.
[29] C. Park, X. Chen, J.R. Yang, J. Zhang, Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E678–E686.
[30] D.J. Patterson, K. Yasuhara, W.L. Ruzzo, Pre-mRNA secondary structure prediction aids splice site prediction, Pac. Symp. Biocomput. (2002) 223–234.
[31] J.F. Peden, Analysis of Codon Usage, Dept of Genetics, University of Nottingham, 1999.
[32] V. Pelechano, S. Chávez, J.E. Pérez-Ortín, A complete set of nascent transcription rates for yeast genes, PLoS One 5 (2010) e15442.
[33] P. Sharp, W. Li, The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications, Nucleic Acids Res. 15 (1987) 1281–1295.
[34] P.M. Sharp, L.R. Emery, K. Zeng, Forces that influence the evolution of codon bias, Phil. Trans. R. Soc. B 365 (2010) 1203–1212.
[35] P.M. Sharp, W.H. Li, An evolutionary perspective on synonymous codon usage in unicellular organisms, J. Mol. Evol. 24 (1986) 28–38.
[36] P.M. Sharp, M. Stenico, J.F. Peden, A.T. Lloyd, Codon Usage: Mutational Bias, Translational Selection, or both? Portland Press Limited, 1993.
[37] E.G. Shpaer, The secondary structure of mRNAs from Escherichia coli: its possible role in increasing the accuracy of translation, Nucleic Acids Res. 13 (1985) 275–288.
[38] N. Sueoka, Directional mutation pressure, selective constraints, and genetic equilibria, J. Mol. Evol. 34 (1992) 95–114.
[39] N. Sueoka, Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C, J. Mol. Evol. 49 (1999) 49–62.
[40] N. Sueoka, Y. Kawanishi, DNA G + C content of the third codon position and codon usage biases of human genes, Gene 261 (2000) 53–62.
[41] E. Trotta, Selection on codon bias in yeast: a transcriptional hypothesis, Nucleic Acids Res. 41 (2013) 9382–9395.
[42] F. Wright, The 'effective number of codons' used in a gene, Gene 87 (1990) 23–29.
[43] M. Zama, Correlation between mRNA structure of the coding region and translational pauses, Nucleic Acids Symp. Ser. (1999) 81–82.
[44] H. Zur, T. Tuller, Strong association between mRNA folding strength and protein abundance in S. cerevisiae, EMBO Rep. 13 (2012) 272–277.