4-15-2010

How the Sequence of a Gene Can Tune Its Translation

Kurt Fredrick
The Ohio State University

Michael Ibba
Chapman University, ibba@chapman.edu

Recommended Citation

Fredrick, K. and Ibba, M. (2010) How the sequence of a gene can tune its translation. *Cell* **141**, 227-229. https://doi.org10.1016/j.cell.2010.03.033

How the Sequence of a Gene Can Tune Its Translation

Comments

Creative Commons License

Copyright

Published in final edited form as:
Cell. 2010 April 16; 141(2): 227–229. doi:10.1016/j.cell.2010.03.033.

How the sequence of a gene can tune its translation

Kurt Fredrick and **Michael Ibba**
Department of Microbiology, Ohio State Biochemistry Program, and Center for RNA Biology, The Ohio State University, Columbus, Ohio 43210, USA

Abstract

Sixty-one codons specify 20 amino acids offering cells many options for encoding a polypeptide sequence. Two new studies (Cannarozzi et al, 2010, Tuller et al., 2010) now foster the idea that patterns of codon usage can control ribosome speed, fine-tuning translation to increase the efficiency of protein synthesis.

Just about every molecular biologist, intentionally or not, has conducted experiments on the roles of codon usage during translation. For example, to prepare a protein of interest, a foreign gene might be expressed in a heterologous host like the bacterium Escherichia coli, but disappointingly either no protein is produced or the resulting product is inactive. There are numerous possible reasons such experiments fail, but one that can usually be excluded from these is that the genetic code differs between the host and the foreign organism. With very few exceptions, the genetic code is universal – no matter how evolutionarily distant two organisms are, they will always use the same combination of nucleotide triplets (codons) to encode the same amino acids. That being said, the universality of the genetic code comes with a few caveats. There are not twenty codons, but up to 61. As a consequence, some amino acids are encoded multiple times within the genetic code whereas others are not. For example, there are six different codons for leucine, but only two for lysine. These codons, with different sequences, but coding for the same amino acid, are termed synonymous codons. The frequency with which different synonymous codons are used for a particular amino acid varies greatly between different organisms – the phenomenon is referred to as codon usage. Understanding differences in codon usage, and making appropriate adjustments, can help to improve yields when trying to express foreign proteins, but the varying success and unpredictability of these approaches suggests that we don't yet understand all the rules guiding translation. One long-standing idea is that the order with which codons are used is far from random. Examining individual mRNA sequences to try and identify patterns in codon choice has yielded relatively little information. Recent studies have instead focused on searching for genome-wide trends in codon choice, with increasingly striking results as clear patterns start to emerge. These include roles for codon selection in protein folding (e.g. Zhang et al., 2009) and, as described in this issue of *Cell*, potential functions in controlling translation elongation (Cannarrozzi et al., 2010; Tuller et al., 2010).

Correspondence to: Dr. Michael Ibba, Dr. Kurt Fredrick, Department of Microbiology, The Ohio State University, 318 West 12th Avenue, Columbus, Ohio 43210-1292, ibba.1@osu.edu, fredrick.5@osu.edu.

Translation speed varies over mRNA length

One of the first questions to arise when considering the possible effects of altering codon usage is how it might impact the rate at which an mRNA is translated. The notion that translation rates can change for different regions within a single mRNA, and in so doing facilitate processes such as frame shifting or the folding of nascent polypeptides, has been discussed in the literature for some time and has recently garnered additional experimental support (Siller et al., 2010; Zhang et al., 2009). To get a more global picture of translation rates, Tuller and colleagues first set out to estimate how efficiently each individual codon is translated in a particular gene, based upon the predicted availability of the corresponding tRNA. Similar approaches have been applied before and averaged over individual genes, but never codon-by-codon over an entire transcriptome. To further refine the analysis, Tuller *et al.* also factored in whether or not translation of a codon required a perfect codon-anticodon match (more efficient) or a codon-anticodon wobble interaction (less efficient). The final adjustment compared to earlier studies was to take advantage of recent technological advances to measure global cellular tRNA pools (Zaborske et al., 2009) and to use these values when calculating tRNA availability. For yeast grown on rich medium, there was a robust correlation between actual tRNA abundance and tRNA gene copy number. Extrapolating from this finding allowed the authors to broaden their studies to organisms other than yeast. One qualification here is that whereas this correlation between copy number and tRNA abundance holds true under some conditions, there are other cellular states where this may not reflect the charging of tRNAs with their appropriate amino acids (Dittmar et al., 2005; Elf et al., 2003). When all of these factors are taken into consideration, conserved patterns start to emerge both within a single organism's transcriptome as well as across species. The most obvious finding is that the speed of translation is predicted to be slow during the first 30–50 codons (the "ramp"), and then to increase to a plateau level for the remainder of the gene (Figure 1A). Support for this model comes from the striking correlation between profiles of ribosome density along mRNAs predicted computationally and those observed experimentally (Ingolia et al., 2009). The clear exception to this rule is for the second codon, which is predicted to be translated at a higher rate than the surrounding codons, thereby promoting rapid release and recycling of the initiator tRNA.

What purpose could this ramp play in translation? Slowing translation elongation immediately after initiation would effectively generate more uniform spacing between ribosomes further down the mRNA, which should prevent ribosome congestion and promote efficient protein synthesis. The ramp is predominantly associated with highly-expressed genes, consistent with a role for the ramp in increasing product yield amid heavy ribosome traffic. Traffic jams in translation, and also in transcription (Ehrenberg et al., 2010), can decrease processivity and increase stalling and termination, providing precedence for the idea that the ramp increases the overall efficiency of protein synthesis. Although this hypothesis is attractive, it remains to be tested experimentally. Another potential role for the ramp involves protein folding. The length of the ramp corresponds remarkably well to the length of polypeptide needed to fill the exit tunnel of the ribosome (Ban et al., 2000), so the nascent peptide chain should emerge from the ribosome as it transitions from the slow (ramp) stage to the fast stage of elongation. This raises the possibility that the ramp might somehow facilitate interactions between the emerging peptide and the chaperone proteins, thereby increasing the fraction of correctly folded product. This idea will also be worth investigating experimentally.

The importance of the ramp may explain why foreign gene expression sometimes fails – the ramp's function would diminish or disappear in heterologous systems due to changes in tRNA availability (Tuller et al., 2010). The need to compensate for the absence of ramps by slowing down translation might contribute both to anecdotal "cures" for increasing foreign protein production (e.g., lowering growth temperature) and to more rational recoding-based approaches (Siller et al., 2010). It also may help to explain why the over-expression of rare

tRNAs, another common strategy to improve foreign protein production, does not always work and can even be counterproductive. In some cases, rare codons can indeed be problematic as the scarcity of the corresponding tRNA may lead to premature termination; in other examples excessive over-expression of a low abundance tRNA could significantly change the translation efficiency of certain codons thereby disrupting the ramp to the detriment of protein synthesis.

Although the effects of the ramp on the early rounds of elongation may indeed help to tune translation, it is important to keep in mind that the rate of product yield for a given mRNA is strictly governed by the rate of initiation. Kudla *et al.* (2009) recently examined the effects of synonymous codon substitutions on the efficiency of translation of a gene encoding green fluorescent protein in *E. coli*. They found that the sequence at the beginning of the gene strongly influenced translation, and that the expression level was inversely correlated with predicted mRNA secondary structure. These data are consistent with other studies pointing to the importance of mRNA structure in controlling translation initiation (de Smit and van Duin, 2003; Studer and Joseph, 2006). As the regions defined by Kudla *et al.* and Tuller *et al.* overlap, the nucleotide sequence in this region may control initiation and/or early elongation, depending on the particular gene.

Finding order from degeneracy

The genetic code is degenerate, that is to say it can use many different combinations of codons to make exactly the same protein. The 61 codons that encode the 20 standard amino acids are not equally abundant in mRNA. Within synonymous sets, some codons are used far more frequently than others, and the limited strategic use of rare codons can sometimes be exploited for regulatory purposes (Chandra and Chater, 2008). The discovery of ramps emphasizes that codon choice is not uniform, but is instead highly selected and broadly conserved (Tuller et al., 2010). In a related vein, Cannarrozzi and colleagues describe a different pattern of codon usage in yeast that appears to reflect differences in translation rate (Cannarrozzi et al., 2010). They find that when an amino acid recurs in a protein (for instance, xLxxL) there is a strong tendency to use the same codon the second time as for the first occurrence of the amino acid. This predisposition towards selecting particular codons rather than arbitrarily choosing one from the synonymous set has important implications for the dynamics of mRNA translation and the protein synthesis machinery.

Cannarrozzi and colleagues focused on groups of codons that encode the same amino acid, and asked whether these synonymous codons were randomly or non-randomly ordered along genes. In so doing, they were able to analyze all consecutive synonymous codon pairs in the yeast genome. What this revealed was that identical codons have a strong tendency to be used again when an amino acid recurs, and if the same codon is not reused, there is instead a bias towards the most closely related synonymous wobble codons. This observed reuse of codons, termed “auto-correlation,” is not simply the result of the most frequent codons accumulating in genes, as rare codons are just as likely to be reused as common ones. These highly conserved patterns suggest that reusing codons may benefit translation in some way, a notion reinforced by the frequent reuse of rare codons in highly expressed genes. This trend extends to larger groups of genes. For example, regulons that are highly upregulated under certain conditions, such as those involved in cell cycle progression and environmental stress responses, showed the strongest correlations in sequential codon usage. The overall model that emerges from these and other analyses performed by the authors is that, simply put, codon reuse may provide an effective mechanism to speed up translation. To put this model to the test, the authors engineered reporter genes in which the synonymous codons were either positionally auto-correlated or anti-correlated (i.e., the two opposite extremes of their model) (Figure 1B). These reporter genes encoded the identical protein and had an identical total codon composition–only the arrangement of the synonymous codons differed. These genes were expressed in yeast

cells in the presence of radiolabeled amino acids, the cells were lysed, and the translation products (partial and complete) were purified from the remainder of the lysate using an N-terminal epitope tag. These translation products were separated using SDS-PAGE, and the size distribution of the nascent chains allowed ribosome density along each mRNA to be compared. It was found that the reporter mRNAs with auto-correlated codons had lower ribosome density than their counterparts with anti-correlated codons. Because ribosome density is inversely proportional to elongation rate, the authors could deduce that translation on auto-correlated mRNA was substantially (~30%) faster than on anti-correlated mRNA.

This significant increase in the speed of translation achieved by the appropriate arrangement of codons has important implications for how the translation machinery functions. To achieve the higher translation speed that comes from correlating codon choices, the availability of the corresponding tRNAs may need to be greater. To assess the availability of tRNAs during high-speed translation, Cannarrozzi *et al.* modeled a number of different scenarios in several eukaryotes, and also took into account the effect of distance between synonymous codons on the degree of auto-correlation. They concluded that the advantage to codon reuse comes from an ability to reuse the corresponding tRNA. This implies that the tRNA molecules exiting the ribosome remain associated with the translational machinery so that they are readily available when the next identical codon comes along. This model necessitates that the enzymes responsible for attaching amino acids to tRNAs, the aminoacyl-tRNA synthetases, are also associated with ribosomes. Extensive evidence exists that the aminoacyl-tRNA synthetases form ordered complexes in eukaryotes (Deutscher, 1984; Mirande et al., 1985), and that these complexes associate with ribosomes (Kaminska et al., 2009). Moreover, it has recently been shown that these complexes can increase translation rates by promoting the “channeling” of charged tRNAs to the ribosome for protein synthesis (Kyriacou and Deutscher, 2008), which is highly consistent with the model of Cannarrozzi and coworkers. One interesting issue that remains to be explored is how these models will play out in bacteria, where translation rates are substantially higher but evidence for channeling is scant at best.

Perspectives

The studies of Cannarrozzi, Tuller and their colleagues provide evidence that the pattern of codon usage modulates the rate of protein synthesis, and suggest how this might be exploited on the genome scale to fine-tune the efficiency with which certain sets of genes are translated. These effects may also be accentuated by translation factors, thereby providing an additional layer of regulation. For example, the translation factor eIF5A is known to increase the efficiency of translation (Saini et al., 2009), and how this and other factors could potentially change the gradient of a ramp or alter the impact of codon correlation is of significant interest. Further modulation of ramps and codon correlation effects may also be provided by isodecoder tRNAs, that is, tRNAs with identical anticodons but otherwise diverse primary sequences. Recent studies have shown that huge numbers of isodecoders exist in eukaryotes. There are approximately 270 isodecoders in the human genome alone (Geslain and Pan, 2010), and variations in the expression of these tRNAs may further heighten the codon-dependent changes in translation efficiency predicted by Tuller, Cannarrozzi and their colleagues (Cannarrozzi et al., 2010; Tuller et al., 2010).

Acknowledgments

This work was supported by grants from the National Institutes of Health (GM 072528 to KF; GM 065183 to MI).

REFERENCES

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. Science 2000;289:905–920. [PubMed: 10937989]

Cannarrozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, Roth AC, Gonnet P, Gonnet G, Barral Y. Cell. 2010 In Press.

Chandra G, Chater K. Antonie van Leeuwenhoek 2008;94:111–126. [PubMed: 18320344]

de Smit MH, van Duin J. J Mol Biol 2003;331:737–743. [PubMed: 12909006]

Deutscher MP. J Cell Biol 1984;99:373–377. [PubMed: 6746733]

Dittmar KA, Sorensen MA, Elf J, Ehrenberg M, Pan T. EMBO Rep 2005;6:151–157. [PubMed: 15678157]

Ehrenberg M, Dennis PP, Bremer H. Biochimie 2010;92:12–20. [PubMed: 19835927]

Elf J, Nilsson D, Tenson T, Ehrenberg M. Science 2003;300:1718–1722. [PubMed: 12805541]

Geslain R, Pan T. J Mol Biol 2010;396:821–831. [PubMed: 20026070]

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Science 2009;324:218–223. [PubMed: 19213877]

Kaminska M, Havrylenko S, Decottignies P, Le Marechal P, Negrutskii B, Mirande M. J Biol Chem 2009;284:13746–13754. [PubMed: 19289464]

Kudla G, Murray AW, Tollervey D, Plotkin JB. Science 2009;324:255–258. [PubMed: 19359587]

Kyriacou SV, Deutscher MP. Mol Cell 2008;29:419–427. [PubMed: 18313381]

Mirande M, Le Corre D, Waller JP. Eur J Biochem 1985;147:281–289. [PubMed: 3971983]

Saini P, Eyler DE, Green R, Dever TE. Nature 2009;459:118–121. [PubMed: 19424157]

Siller E, Dezwaan DC, Anderson JF, Freeman BC, Barral JM. J Mol Biol. 2010 doi:10.1016/j.jmb.2009.12.042.

Studer SM, Joseph S. Mol Cell 2006;22:105–115. [PubMed: 16600874]

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. Cell. 2010 In Press.

Zaborske JM, Narasimhan J, Jiang L, Wek SA, Dittmar KA, Freimoser F, Pan T, Wek RC. J Biol Chem 2009;284:25254–25267. [PubMed: 19546227]

Zhang G, Hubalewska M, Ignatova Z. Nat Struct Mol Biol 2009;16:274–280. [PubMed: 19198590]

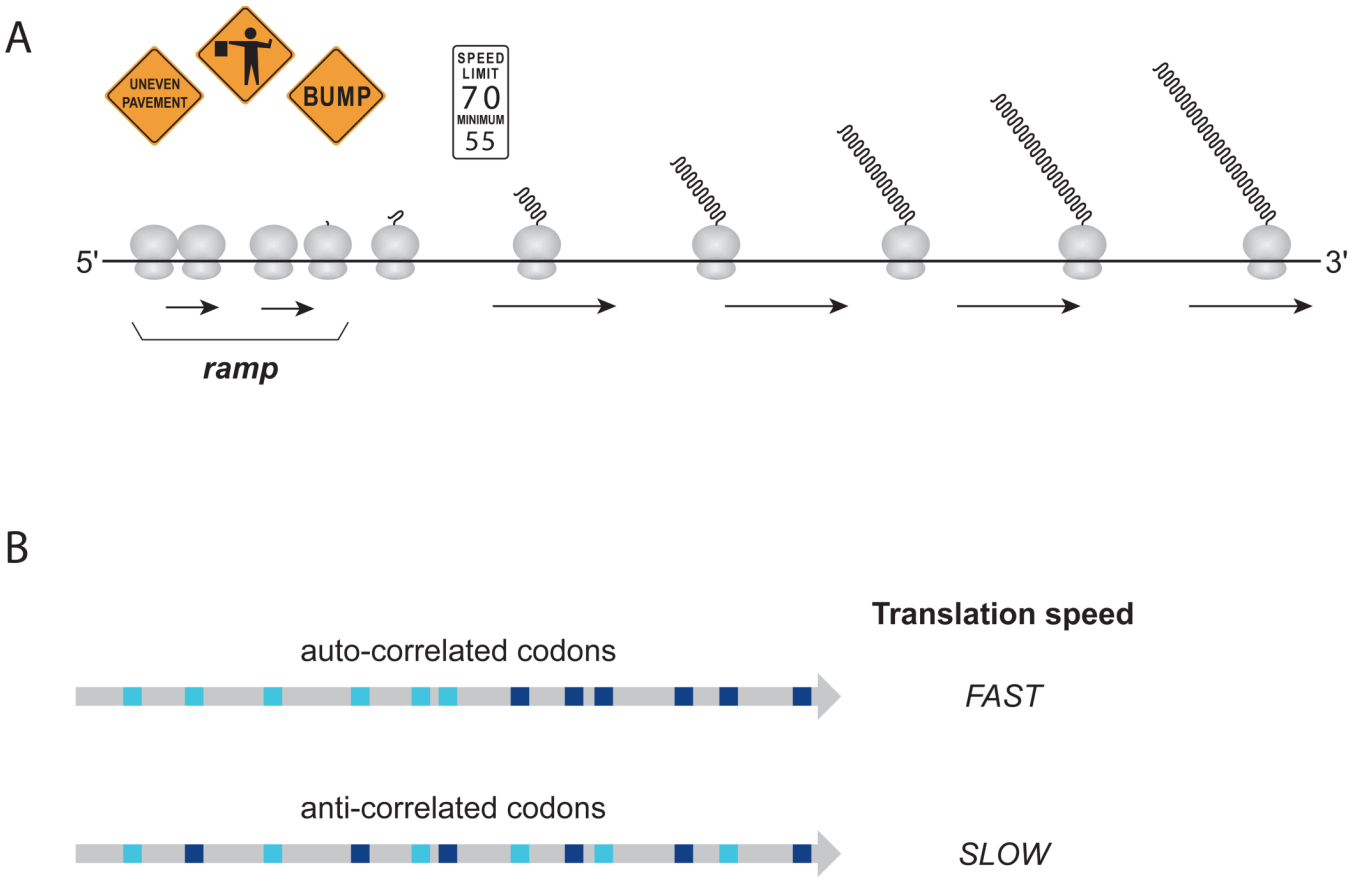

Figure 1. Messenger RNA sequences set the speed limit

There are three times as many codons as amino acids, meaning that a given amino acid can be encoded by several synonymous codons and identical proteins can be made from very different mRNA sequences. Codon choice is not random, but is highly selected across a broad range of organisms to optimize protein production. (**A**) For many genes, codons recognized by low-abundance tRNAs are overrepresented in the first part of the gene. This pattern suggests that ribosomes translate more slowly over the initial 50 codons or so (ramp stage), and then translate the remainder of the mRNA at full speed. The mRNA template itself controls the speed of ribosomes, somewhat analogous to how poor road conditions limit the speed of cars in a construction zone. (**B**) The arrangement of synonymous codons along a gene influences translation speed. Shown is a simple example in which two different codons (represented by different shades of blue) encode the same amino acid. When the identical codons are consecutively arranged along the mRNA (auto-correlated), translation is faster than when they are alternatively arranged (anti-correlated).