molecular systems biology

Efficient translation initiation dictates codon usage at gene start in bacteria

Kajetan Bentele, Paul Saffert, Robert Rauscher, Zoya Ignatova, Nils Blüthgen

Corresponding author: Nils Blüthgen, Charite Universitätsmedizin Berlin

Review timeline:

Submission date:	26 February 2013
Editorial Decision:	02 April 2013
Revision received:	30 April 2013
Accepted:	14 May 2013

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 03 April 2013

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who accepted to evaluate the study. As you will see, the referees find the topic of your study of potential interest and are rather supportive. However, they raise a series of concerns and make suggestions for modifications, which we would ask you to carefully address in a revision of the present work. As suggested by reviewer #2, the definition of rare and abundant codons needs to be carefully considered in order to allow direct comparison of the results to previous studies. Moreover, some of the reviewers' comments refer to the need to clarify and better describe several points throughout the manuscript.

Thank you for submitting this paper to Molecular Systems Biology.

REFEREE REPORTS:

Reviewer #1 (Remarks to the Author):

In this work, Bentele et al. use bioinformatics and statistical tools to explain codon usage at the very beginning of open reading frames in bacteria. In particular, they contrast two opposing hypotheses. According to the first, usage of rare codons at the beginning of ORFs serve the purpose of slowing down initiation of translation to reduce ribosomal traffic jams further down stream of the ORFs. According to the second, the early codon usage serves to minimize mRNA secondary structure formation so as to maximize the speed of initiation of translation. The authors present, to my opinion, convincing data and arguments to show that the second hypothesis is correct. That is, the codon usage at the beginning of bacterial ORFs is not about using rare codons to reduce ribosomal queuing down stream from the initiation site, but to reduce secondary structure formation near the

© European Molecular Biology Organization

initiation codon.
They also provide experiments to demonstrate this very point, i.e. that early, rare codons do not per se favor gene expression. Instead, it is the absence of secondary structure formation that favors gene expression, which is greatly inhibited by stable, secondary mRNA structures.

The only major problem with the manuscript concerns language and exposition of logic. The language is quite convoluted and the description of experiments and theory quite awkward. The paper would gain greatly from language correction and revision, so that its important points become more easily accessible to its readers.

Reviewer #2 (Remarks to the Author):

This is a very interesting manuscript that addresses the evolutionary origin of the "ramp" in codon usage at the beginning of genes, described in the 2010 Cell paper by Tuller et al. The authors distinguish between two plausible hypotheses: either the ramp results from the need to slow down translation, perhaps to avoid ribosome traffic jams, or it results from selection for weak secondary structure, most likely to facilitate translation intiation. The data clearly supports the second hypothesis: the codon ramp appears to be caused by selection for weak RNA folding in bacteria. The experiments performed by the authors complement these theoretical results very well.

Although the analysis is generally convincing, the definitions of rare and abundant codons used by the authors makes it difficult to compare the results to previous studies. In the present manuscript, rare codons are defined as the 15 least abundant codons in the genome, whereas Tuller et al. (2010) defined rare and frequent codons based on their tAI scores. Thus, the meaning of the "codon usage ramp" observed in both studies is somewhat different. "Low-tAI" codons are thought to be translated slowly, whereas little is known about "low-abundance" codons, so the definition based on tAI appears preferable. I would recommend that the authors reanalyse the data using either CAI or tAI scores for splitting codons into slow and fast groups.

Minor comment: p.2, "the amino acids are encoded by two, four or six different codons". Or three codons, in the case of isoleucine.

Reviewer #3 (Remarks to the Author):

In this paper, the authors try to address one of the long-standing debates regarding the evolutionary forces shaping codon usage at start of genes. One of the key features of genome-scale ribosome profiling data is the apparent presence of a high density of ribosomes in the 5' region of genes, which has led to the `ramp hypothesis' suggesting an adaptive role of slow codons at 5' end of genes. This paper challenges the central tenet of this hypothesis by comparing the properties of rare and abundant codons in this region across a wide range of bacteria.

The authors begin by recapturing earlier findings of unusual codon usage in the 5' region of genes. However, across 414 bacterial genomes, they find a consistent bias towards AU rich codons irrespective of the GC content of the genome. Moreover, consistent with their hypothesis, they show that 5' codons are biased towards `slower' codons only if they are AU-rich. This is in direct contradiction to the `ramp hypothesis' which predicts a bias towards slow codons irrespective of their GC content. The authors then go on to verify their hypothesis with wet-lab experiments in \emph{E. coli} by modifying the codon usage and 5' folding energy to two genes - \emph{ypdE} and \emph{pykA}. In my opinion, the manuscript conclusively demonstrates that any selection on the 5' region of genes is due to selection for weaker secondary structure than `slower' codons.

Without doubt, this manuscript makes an important contribution to the wider debate regarding the nature of selection pressures shaping genomic features affecting protein translation, and certainly deserves to be published. I am largely happy with the present manuscript as is. I have only a few specific comments.

1) It is unclear as to why for both genes, a low CA version of the gene shows a higher protein

abundance than higher CA version. The authors need to explain this
discrepancy more clearly in the main text.

2) Also, it in not entirely clear as to what the mRNA expression levels of two transgenes were (say induced versus non-induced). The effects of codon usage and
5' folding energy on protein expression depends strongly on the mRNA abundance. The authors need to more thoroughly explain their results in the context of
ypdE and pykA mRNA abundance levels.

3) The axes on figures need to be made more consistent with each other. For instance, in Fig.~4, (a) shows the fold-change in frequencies but (b) and (c) switch to absolute frequencies. Changing everything to fold-change would make it easier to interpret the plots. Similarly in Fig.~3 (c and d), showing relative frequencies with y-axis going from 0-1 would be helpful.

4) The authors should be aware of and cite these related papers - Keller et. al. (Genome Biol Evol 2012) and Ding et.al. (Genome Biol Evol 2012).

1st Revision - authors' response 30 April 2013

© European Molecular Biology Organization

Dear Dr. Polychronidou,

thank you for giving us the opportunity to revise the paper. We also thank the reviewers for their very supportive comments.

As detailed below, we addressed all referee comments. The most important additions to the manuscript include the use of tAI scores as an alternative measure to distinguish between slow and fast codons. Furthermore, we edited the entire manuscript the manuscript for clarity.

Yours sincerely,

Nils Bluthgen

Reviewer #1 (Remarks to the Author):

In this work, Bentele et al. use bioinformatics and statistical tools to explain codon usage at the very beginning of open reading frames in bacteria. In particular, they contrast two opposing hypotheses. According to the first, usage of rare codons at the beginning of ORFs serve the purpose of slowing down initiation of translation to reduce ribosomal traffic jams further down stream of the ORFs. According to the second, the early codon usage serves to minimize mRNA secondary structure formation so as to maximize the speed of initiation of translation. The authors present, to my opinion, convincing data and arguments to show that the second hypothesis is correct. That is, the codon usage at the beginning of bacterial ORFs is not about using rare codons to reduce ribosomal queuing down stream from the initiation site, but to reduce secondary structure formation near the initiation codon.
They also provide experiments to demonstrate this very point, i.e. that early, rare codons do not per se favor gene expression. Instead, it is the absence of secondary structure formation that favors gene expression, which is greatly inhibited by stable, secondary mRNA structures.

The only major problem with the manuscript concerns language and exposition of logic. The language is quite convoluted and the description of experiments and theory quite awkward. The paper would gain greatly from language correction and revision, so that its important points become more easily accessible to its readers.

We thank the reviewer for his very supportive comments. We have edited our manuscript for clarity, and also consulted a native speaker.

Reviewer #2 (Remarks to the Author):

This is a very interesting manuscript that addresses the evolutionary origin of the "ramp" in codon usage at the beginning of genes, described in the 2010 Cell paper by Tuller et al. The authors distinguish between two plausible hypotheses: either the ramp results from the need to slow down translation, perhaps to avoid ribosome traffic jams, or it results from selection for weak secondary structure, most likely to facilitate translation initiation. The data clearly supports the second hypothesis:

the codon ramp appears to be caused by selection for weak RNA folding in bacteria. The experiments performed by the authors complement these theoretical results very well.

Although the analysis is generally convincing, the definitions of rare and abundant codons used by the authors makes it difficult to compare the results to previous studies. In the present manuscript, rare codons are defined as the 15 least abundant codons in the genome, whereas Tuller et al. (2010) defined rare and frequent codons based on their tAI scores. Thus, the meaning of the "codon usage ramp" observed in both studies is somewhat different. "Low-tAI" codons are thought to be translated slowly, whereas little is known about "low-abundance" codons, so the definition based on tAI appears preferable. I would recommend that the authors reanalyse the data using either CAI or tAI scores for splitting codons into slow and fast groups.

In line with this comment, we have calculated tRNA-copy numbers for all genomes, and thus used tAI scores to define sets of slow and fast codons. Using these sets, we have repeated the analysis and find the same results. The results are described in the respective parts in the main text and shown in supplementary figures S3, S7, S9, S10, S12.

Minor comment: p.2, "the amino acids are encoded by two, four or six different codons". Or three codons, in the case of isoleucine.

Thank you, we have changed it to two to six.

Reviewer #3 (Remarks to the Author):

In this paper, the authors try to address one of the long-standing debates regarding the evolutionary forces shaping codon usage at start of genes. One of the key features of genome-scale ribosome profiling data is the apparent presence of a high density of ribosomes in the 5' region of genes, which has led to the `ramp hypothesis' suggesting an adaptive role of slow codons at 5' end of genes. This paper challenges the central tenet of this hypothesis by comparing the properties of rare and abundant codons in this region across a wide range of bacteria.

The authors begin by recapturing earlier findings of unusual codon usage in the 5' region of genes. However, across 414 bacterial genomes, they find a consistent bias towards AU rich codons irrespective of the GC content of the genome. Moreover, consistent with their hypothesis, they show that 5' codons are biased towards `slower' codons only if they are AU-rich. This is in direct contradiction to the `ramp hypothesis' which predicts a bias towards slow codons irrespective of their GC content. The authors then go on to verify their hypothesis with wet-lab experiments in \emph{E. coli} by modifying the codon usage and 5' folding energy to two genes - \emph{ypdE} and \emph{pykA}. In my opinion, the manuscript conclusively demonstrates that any selection on the 5' region of genes is due to selection for weaker secondary structure than `slower' codons.

Without doubt, this manuscript makes an important contribution to the wider debate regarding the nature of selection pressures shaping genomic features affecting protein translation, and certainly deserves to be published. I am largely happy with the present manuscript as is. I have only a few specific comments.

1) It is unclear as to why for both genes, a low CA version of the gene shows a higher protein abundance than higher CA version. The authors need to explain this discrepancy more clearly in the main text.

The effects of codon adaptation on protein levels are variable: For ypdE, both increased and decreased CA reduce protein expression, and for pykA increase of CA reduces protein expression and decrease of CA increases protein expression. We cannot exclude that this also reflects differences in mRNA structure; Although the overall folding energy remains the same, the base-paring of the mRNA may change and thus the efficiency of ribosome binding. In order to get more insights into the role of codon adaptation, one would need to systematically vary the CA in a larger ensemble of genes, which is beyond the scope of this study.

2) Also, it in not entirely clear as to what the mRNA expression levels of two transgenes were (say induced versus non-induced). The effects of codon usage and 5' folding energy on protein expression depends strongly on the mRNA abundance. The authors need to more thoroughly explain their results in the context of ypdE and pykA mRNA abundance levels.

The expression levels of our constructs are relatively high, and clearly impact on cell physiology. To disentangle the effect of high expression levels and codon adaptation on translation, one would need much more experimental data, in particular about cell physiology, tRNA levels and demand. To avoid over-interpreting our findings, we would like to ratehr carefully interpretate the data in the light of expression levels. In the light of this comment, we changed Fig. Fig S15 such that it shows the non-normalized RT-PCR expression values.

3) The axes on figures need to be made more consistent with each other. For instance, in Fig.~4, (a) shows the fold-change in frequencies but (b) and (c) switch to absolute frequencies. Changing everything to fold-change would make it easier to interpret the plots. Similarly in Fig.~3 (c and d), showing relative frequencies with y-axis going from 0-1 would be helpful.

We agree that this makes the figures easier to interpret, and changed the axis of Fig 4 and 3 accordingly.

4) The authors should be aware of and cite these related papers - Keller et. al. (Genome Biol Evol 2012) and Ding et.al. (Genome Biol Evol 2012).

Thank you, we added these papers in the revised version of the manuscript.

Acceptance letter 14 May 2013

Thank you again for sending us your revised manuscript. We are now satisfied with the modifications
made and I am pleased to inform you that your Article has been accepted for publication.

Thank you very much for submitting your work to Molecular Systems Biology.

REFEREE REPORTS:

Reviewer #2 (Remarks to the Author):

The authors have constructively addressed all my comments.

Reviewer #3 (Remarks to the Author):

The authors have addressed all of my suggestions for improving the manuscript and, apparently, the suggestions of the other referees as well.

The work remains, in my mind, a very important contribution to the debate on the relative importance of initiation versus elongation in shaping protein production. I am certain that this work will help advance the field towards a common understanding of what guides protein translation.

Congratulations on a wonderful joint computationa/experimental study!

© European Molecular Biology Organization