THE
EMBO
JOURNAL

Manuscript EMBO-2013-85651

Translation elongation can control translation initiation on eukaryotic mRNAs

Dominique Chu, Eleanna Kazana, Noémie Bellanger, Tarun Singh, Mick F Tuite and Tobias von der Haar

Corresponding author: Tobias von der Haar, University of Kent

Review timeline:		
	Submission date:	13 May 2013
	Editorial Decision:	20 June 2013
	Revision received:	02 September 2013
	Editorial Decision:	18 September 2013
	Revision received:	09 October 2013
	Accepted:	15 October 2013

Editor: Anne Nielsen

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   20 June 2013

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below. As you will see from the reports all three referees express great interest in your findings, but also raise a number of concerns that you will have to address before they can support publication of your manuscript.

Given these positive recommendations, I would like to invite you to submit a revised version of the manuscript. In addition to addressing the numerous minor concerns raised by all three referees, I would like to particularly emphasize the need for you to:

-> strengthen the statistical analysis (ref#1)
-> address variations in mRNA and protein expression levels (ref #1, 2 and 3)
-> address the consequence for slow codon usage on polysome association (ref#1 and 3)
-> clarify text and terminology used (ref#1 and 3)

I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance or rejection of your manuscript will therefore depend on the completeness of your responses in this revised version.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: http://www.nature.com/emboj/about/process.html

---

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

REFEREE REPORTS

Referee #1

Von der Haar et al.

Neither the lines nor the pages are numbered, making it rather cumbersome to refer to specific points or passages in the text. Nor are the actual figures.

MAJOR POINTS

1. Throughout, how have the authors corrected for any differences in the levels of the different mRNAs they are using? This must be a key parameter when trying to assess translational efficiency.

2. Introduction: I did not find the second paragraph under 'Translation initiation and elongation rates in yeast' very clear. Also, what is meant in terms of what we know about the mechanisms of translation by the expression 'the inherent ribosome affinity of an mRNA' (used again as 'implicit ribosome affinity' at the start of the Discussion)? An mRNA contains many diverse sequences and structural features, and engages with the ribosome indirectly through the intervention of translation factor, so what meaning can this expression have? It is also used elsewhere in the text; perhaps, what the authors mean is the efficiency of initiation, which is not really anything to do with this idea of ribosomal affinity. Also, initiation is a dynamic process, not a static one, so the concept of affinity is less relevant than the frequency with which ribosomes (strictly 40S-based initiation complexes) arrive at the start codon and the rate at which the 60S subunit is able to associate following start codon recognition.

3. In the section about 'start codon liberation rates limit translation initiation...', it needs to be made clear that the authors' discussion of initiation rates is not based on any direct experimental measurement of those rates.

4. Fig. 4. How was the passage time calculated? It is not clear to this reviewer how these figures were arrived at.

5. Related to Fig. 5C: are the authors suggesting that 'slow' codons near the 3'-end of a coding sequence cause a build-up of ribosomes all along the mRNA which may block escape of newly-initiated ribosomes from the start codon? Is there evidence that mRNAs with slow codons accumulate in (slowly translating) polysomes, which are large compared to those for mRNAs of the same ORF length but containing fast codons? This is what one would expect from this idea.

6. The use of statistics in the manuscript is inconsistent and when done the conclusions drawn are inappropriate for the statistical test used. What is the rationale for only performing statistical tests for the data presented in figure 4 and figure 7C? Were tests performed on other data (for example, but not only, the data in Figs. 4 and 5C)? If not, why not? If they were, why are they not reported? There does not appear to be any rationale given for the use of statistics in some circumstances and not others.
Furthermore, the author's note in the results and figure captions that they performed an ANOVA and then draw conclusions regarding a comparison between one particular manipulation versus others. There are several issues with this: 1) The F-statistic for an ANOVA along with the appropriate

degrees of freedom are not reported as they should be and 2) A statistically significant one-way ANOVA only means that there is some difference in the dependent variable due to the experimental manipulation, but does not provide information regarding the significance of differences observed between specific groups. To do this the author's would need to follow up the significant ANOVA with a post-hoc test (e.g, a Bonferonni t-test, Tukey etc.) in which all pairwise comparisons can be made whilst correcting the p-value for multiple tests.

MINOR POINTS

1. Abstract, line 6: the authors are presumably referring to 'ribosome movement immediately after the start codon'.

2. Introduction, third paragraph, last sentence - a more recent mathematical and experimental analysis from the McCarthy lab suggest elongation (strictly, elongation factor levels) may be limiting - this paper (Firczuk et al.) is cited later but should also be mentioned here.

3. The sentence starting: 'Movement near the start codon cannot be fully decoupled ...' is not clear to me; it should be rephrased to explain better what is meant.

4. There are some grammatical errors or misuse of words which need to be corrected.

5. Discussion: the idea of 'slow codon ramps' is not universally consonant, e.g., with ribosome profiling data and therefore not fully accepted. Also, would the authors expect such ramps only to be quite short, following the start codon, or more extensive? How do their expectations fit with the available data?

Referee #2

Recent studies reveal the role of codon usage in shaping translation efficiency. Tobias von der Haar et al. investigated the interplay between codon usage and ribosome affinity in shaping protein expression levels. Specifically, the authors suggest that efficient translation requires both high ribosome affinity and high elongation speed, where the last is specifically needed for fast liberation of the start codon by initiating ribosomes.

Overall, I found the paper of high interest and worth publication. Yet, in order to allow the reader to fully understand and evaluate the impact of the choice between synonymous codon on translation efficiency, the following changes have to be made:

Major issues

1. The authors analyzed the effect of codon usage on protein expression levels. Yet, the most appropriate measure of protein synthesis rate is not the protein expression level, but the ratio between protein abundance to mRNA levels. Since the authors specifically discuss the effect of codon usage on translation efficiency, the explicit association between the calculated decoding speed and protein-to-mRNA ratio must be shown (especially in the light of the high correspondence between expression levels and mRNA steady-state levels, as can be seen in figure 2). In that context, the authors should supply mRNA levels for the variants introduced in figures 3 & 4, and normalize to such quantities.

2. In many cases, the expression levels of the CFLuc codon variant are given in terms of luciferase activity. One major question is whether such measurement captures the actual changes in protein expression level, or does it at least partially reflecting effect of the fraction of properly folded proteins? This issue is particularly bothering since 'slow codons' are defined here by the ratio between near-cognate and cognate tRNAs, and thus are, by definition, subjected to high probabilities of translational errors. Direct investigation of protein level (e.g. by comasi staining, as done by Kudla et al in their GFP work (Science 2008), could tell whether effects are on protein level. This aspect should also be taken into consideration while discussing the results.

3. Start codon clearance control can be used to probe codon decoding times in vivo - the experiment

nicely illustrates that different codons are decoded with different speeds in vivo. However, I found the association between the results to the "start codon clearance" less convincing - what if the maxCFLuc would be modified in non-consecutive ten positions along the transcript? Do the authors expect less striking differences between GAA- and GAG-related variants?

Minor issues

1. The terminology of near-cognate and non-cognate tRNAs should be clearly defined at the beginning of the paper.
2. Codon decoding time is partial predictor pf protein expression levels - it is recommended to shortly describe to what extent the variants differ from each other (i.e, the total number of substitutions); in addition, is the secondary structure defined in terms of ΔG? It is also important for the reader to understand what are the relations between the secondary structure of the different variants - which one is more tight or loose?
3. Start codon liberation rates limit translation initiation on transcripts with high ribosome affinity - for the clarity of reading, it is highly recommended to re-write this section. In addition - It is of interest to know what is the fraction of the five single-gene encoded tRNAs in the min346maxCFLuc variant.
4. Start codon clearance control can be used to probe codon decoding times in vivo - the median decoding time of all the codons should be denoted (first paragraph).
5. Since the definition of translation speed in terms of the ratio between near cognate to cognate tRNAs is less established compared to measures of translation efficiency in terms of tRNA availability, it would be of great interest to add to figure 2 measurement of predicted translation efficiency by one of the most common indexes. However, this last comment is only recommendation and it is completely subjected the authors judgment.

Referee #3

This is an interesting manuscript that links elongation to the control of protein synthesis while most previous studies had focused on either initiation or mRNA turnover as the regulated steps. One item that is a little unclear to this reviewer which should be made more clearly is whether every codon change is for the same amino acid, just one used more or less frequently that the starting one and second the actual methods used. The Chu et al., 2011 paper actually cites additional manuscripts for the methods and it would be worthwhile to have a full paragraph that described how the studies were performed. Second, the figures with the amino acid sequences are difficult to analyze for codon change. The authors might consider a designation of the following type: GUG81GUA (+2.3-fold) which would indicate the codon changed and the increase or decrease in codon usage relative to the standard. Obviously for several of the mutants, this would be a list of 50 to 100 changes so the supplemental figures would not get any smaller.

Specific concerns

1. Much of the translation literature has focused on getting the first ribosome on the mRNA (i.e. the 80S pathway) and not considered the kinetics of translation in polysomes. One expects this to be different for several reasons including the prior removal of proteins from the 5' UTR and the "intramolecular" affect of the "circular polysome" where the terminating ribosome has a competitive advantage in being the next ribosome to initiate translation. If the preference for the terminating ribosome is say 10-fold greater than adding a free ribosome, then one would expect the timing of ribosome addition to be set or limited by termination which if rapid, is actually set by elongation. Thus, one should be able to use the transit time and ribosome density (number of ribosomes per mRNA) to predict how frequently an mRNA is reinitiated.

2. Thus, in Figure 1, it would be helpful to know if the mRNA (min, sta and max-CFluc) is in small, medium or large polysomes (i.e. generally high levels of expression would predict large polysomes; however, slowly elongating mRNAs might also be in large polysomes). Secondly, it would be good to determine the half transit time for the mRNAs to see if a linear relationship exists between transit time and level of expression (expected).

3. Figure 2 would be more clear if there were 5 bar graph panels with the first being protein

expression. Secondly, there is also the correction for mRNA levels. Is there any fixed ratio for the reporter protein to mRNA or does this vary depending on the type and placement of the variable codons?

4. The math - Generally, most researchers have assumed the "average" mRNA in eukaryotes is elongated at a rate of 4 to 5 amino acids per second (12-15 nucleotides) which is about 4 to 5-fold slower than in bacteria. This would tend to predict that it would take at least 2 seconds to obtain clearance for the next initiation complex even if the 40S subunit might be bound to the 5' UTR before that. If the average mRNA has a coding region of 500 amino acids or 1500 nucleotides, then ribosome loading at every 2 seconds (or 30 nucleotides) would result in the mRNA having about 50 ribosome bound to it. Yet from most polysome profiles, the average size appears to be more like 8 to 12. So how does this match to your numbers (recruit a ribosome on average every 0.8 seconds).

5. "These constructs were termed "slow" CFLucs as they reduce the rate with which ribosomes arrive at the CFLuc start codon." This reviewer wonders whether it is the rate at which the ribosomes arrive at the start codon or the number of ribosomes that arrive at the start codon. The rate statement would seem to only apply if it is known that all the ribosomes that bind to the mRNA will in fact sooner or later arrive at the correct start codon and not be released from the mRNA.

6. Figure 4 - is passage time the same as transit time (the time required for the ribosome to go from the start codon to the terminating codon?

7. Figure 6 - The use of the switch from GAA to GAG raises the concern over the possible formation of G quadraplex structures which the authors need to rule out as an alternate explanation. Alternatively, one could use different codons that bring in either U or C to avoid this problem.

8. Figure 7 - given the considerable difference in His3 expression and the relatively small change in growth rate (about 2-fold or less), it would appear that the absence of histidine is not exclusively limiting and thus this makes growth rate changes a relatively poor measure of expression.

Minor concerns
1. Per journal style, the authors should not use terms such as "for the first time", "this is the first reported mechanism", etc. It is assumed that all manuscripts that will be published represent original findings.
2. The authors need to be consistent with the use of abbreviations (i.e. kcal, not kcals) and the term mRNA should be used instead of message.

1st Revision - authors' response   02 September 2013

Responses to Comments

Comments highlighted by the editors:

Point 1: Strengthen the statistical analyses

We have now conducted rigorous statistical tests on all data, in response to the concerns expressed by referee 1. The relevant procedures are described in a new Materials and Methods section "Statistical Analyses". Specifically: **1)** we have used standard T-tests for comparisons between two samples, and have reported the resulting p-values (figure 1 D and E, figure 2, figure 5C and figure 6B). **2)** we have used one-way ANOVA for comparisons between more than two samples and report all F-statistics and p-values in the figure legends. We have identified individual samples different from the reference sample by post-hoc analysis based on Tukey's honestly significant difference criterion (figure 1B and C, figures 2-5, figure 7) . **3)** we have labelled significant differences from the reference sample in all figures using a consistent labelling scheme throughout the manuscript (* for $p<0.05$, ** for $p < 0.01$, *** for $p<0.001$). **4)** we noticed that in the original manuscript, error bars denoted standard deviation in some figures and standard error of the mean in others. We have now unified this to the standard error of the mean in all figures and this is stated in the figure legends.

Point 2: Address variations in mRNA and protein expression levels
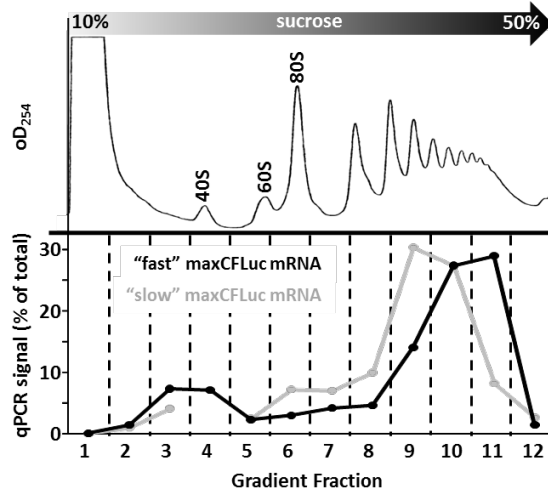
We have now generated qRT-PCR-based mRNA abundance information for all constructs, in response to comments made by all three referees. Due to the fact that the different codon variants differ in sequence and thus cannot all be amplified with a single primer pair, this required design of a specific primer strategy. Primer sequences and details on how the different primers were used to compare the various sequences are now presented in detail in the supplemental information. Resulting from this work, all luciferase activity data from figure 3 onwards are now accompanied by corresponding mRNA data, as are the HIS3 expression data in figure 7. Together with the existing mRNA data in figure 2 and figure 6, all protein/activity level measurements are now presented with accompanying mRNA level data. Reference to these data has been introduced into the relevant passages in the text.

As part of the analysis of the new qPCR data, we also re-analysed the existing data. While doing so, we noticed a mistake in the mRNA value displayed for maxHIS3 in figure 2. This has now been corrected to display the actually measured value.

Point 3: Address the consequence for slow codon usage on polysome association

Referee 1 and 3 raise questions about the relationship between codon usage, protein expression levels, ribosome density, and known features of polysomal gradients. In general, the relationship between ribosome speed and protein expression levels follows the rules we identify in this paper: if ribosome movement around the start codon is slow, then translation initiation rates are limited. In contrast, ribosome density has no simple correlation with speed. Instead, density depends on the distribution of slow and fast codons along the entire ORF. Because of the complex interactions that occur for example through traffic jams, this relationship is non-trivial and non-linear. In consequence, there is no *general* predictable relationship between codon usage-dependent translational control and ribosome density.

Our own CFLuc constructs have relatively uniform speed profiles. Since our model predicts that on these sequences ribosomes are loaded as soon as the preceding ribosome has liberated the start codon, and since the uniform speed profile means that ribosome interactions (traffic jams) should not occur, they should have high ribosome density throughout (this is true for minCFLuc, staCFLuc and maxCFLuc to similar extent). We can address this experimentally by analysing the location of CFLuc mRNAs in polysomal gradients: The figure shows a representative gradient centrifugation experiment conducted with the maxCFLuc construct. The top panel shows the polysome trace of a strain expressing this construct for reference. The gradient was fractionated, RNA isolated from each fraction, and maxCFLuc mRNA detected by qPCR. The second panel shows the resulting qPCR signals: for maxCFLuc (black line), the signal peaks at the very end of the polysomal profile, consistent with a high ribosome density of this mRNA. This contrasts with the "slow initiation" maxCFLuc construct (grey line), where ribosome loading is limited by the initiation frequency but ribosomes move through the ORF quickly once initiation is completed: when a strain containing this variant was analysed in the same way as discussed above, the peak for this sequence was located earlier in the gradient than for the "fast initiation" maxCFLuc.

We show this selected experiment here for the benefit of the referees and to address their specific concerns. However, as the space required to address the reviewers comments on this issue demonstrates, a detailed and thorough investigation of ribosome density on our constructs is beyond the scope of our study and would in any case only reveal features specific to our particular sequences, rather than revealing general rules related to translational control by codon usage, we have decided to leave these results out of the current manuscript.

Specific response to the comment on this topic by referee 1: We are suggesting that slow codons near the 3'-end of a coding sequence cause slow movement of ribosomes in regions of the mRNA before the collision site, and that this blocks access of newly-initiating ribosomes to the start codon. We are not aware of any experimental evidence that directly supports this notion, however, ribosome interactions on crowded message have been studied extensively in theory (this is at the heart of approaches like TASEP of which we cite several studies), and this scenario is expected in that field. On the comment that if ribosome queuing really occurs on the max346min construct this should have higher ribosome density than the non-collision sequences: The max346min construct should have high ribosome density before the fast/slow boundary, possibly slightly higher than the non-collision variants (which all have high ribosome density, see above). On the other hand the ribosomes would clear the remaining stretch of sequence quickly and so density on the 3'-fast stretch of sequence would be lower. The average ribosome density in the collision construct would likely change only marginally compared to maxCFLuc, but should in fact be slightly lower.

In response to the comments on this topic made by referee 3: This referee makes several related comments, which we can best address individually.

'the "average" mRNA in eukaryotes is elongated at a rate of 4 to 5 amino acids per second': By way of comparison, the median elongation rates on our luciferase constructs according to our computer models are 2.3 codons per second (minCFLuc), 6.6 codons per second (staCFLuc) and 8.3 codons per second (maxCFLuc). Our predicted rates thus agree excellently with the experimentally observed ones.

'If the average mRNA has a coding region of 500 amino acids or 1500 nucleotides, then ribosome loading at every 2 seconds (or 30 nucleotides) would result in the mRNA having about 50 ribosomes bound to it': This is a similar argument to the one we use above, that if ribosomes are loaded onto an mRNA as soon as the previous ribosome has liberated the start codon, then this mRNA should be densely loaded with ribosomes. For the 1500 nucleotide luciferase mRNA, this should indeed approach 50 ribosomes per mRNA. The experiment shown above shows that ribosome density on maxCFLuc is indeed high (much higher than for most yeast mRNAs), although an absolute quantification in terms of numbers of ribosomes in this region of the gradient is not possible.

'how does this match to your numbers': We think the numbers match reasonable well. We anticipate that only a subset of mRNAs is under elongation control, and so only a subset of mRNAs should be densely ribosome loaded. An *average* mRNA load of 8 to 12 ribosomes per message would thus not contradict our model (BTW the average yeast mRNA has 300 codons and is thus shorter than luciferase – if all mRNAs were loaded to maximum, the average mRNA load in yeast would be 30 ribosomes). Moreover, it is unclear how an apparent average ribosome load of 8 to 12 ribosomes per mRNA on polysomal gradients relates to the real ribosome load in the cell. For example, there are older papers (eg PMID 1260053) which say that a substantial proportion of ribosomes is bound to various cellular membranes. These would be cleared from the extract and invisible in standard polysomal gradients, which could strongly skew the data. Consistent with this notion, if one sums up all ribosome densities from genome-wide density assays such as those done by Arava et al (PMID 15860778), one gets significant underestimates of ribosome numbers (between 30% and 60% of ribosome numbers determined by other assays, depending on whether the total mRNA concentration of a yeast cell is closer to 30,000 or 60,000 as different studies report). In short, the numbers as they appear on polysomal gradients may well not be representative of *in vivo* numbers.

The numbers cited by the referee indicate that ribosomes take on average two seconds to liberate the start codon, whereas our data indicate that initiation events are on average 0.8 seconds apart. Given that both are mean values of parameters that have a considerable spread in reality, and that both are based on imperfect measurements, we would hold that the referee's data support our conclusion made in the manuscript, that physiological elongation and initiation rates are in a range where they can potentially interfere with each other. Our remaining data show (we hope convincingly) that they do indeed interfere with each other given the right combination of leader sequence and ORF sequence.

Point 4: Clarify text and terminology used

We have clarified the text and terminology used as requested, which has involved editing of a significant proportion of the manuscript. We have endeavoured to simplify the text without diluting the essential messages, and have removed all statements identified as problematic by the referees (such as "inherent ribosome affinity").

Comments made by Referee 1

We have now introduced line and page numbers.

Major point 1: "*…how have the authors corrected for any differences in the levels of the different mRNAs…?*"

mRNA level data are now shown in all figures from figure 2 onwards. Please see response to point 2 under "Comments highlighted by the editors" above.

Major point 2: *"inherent ribosome affinity" and other issues regarding clarity*
In response to comments made by all three referees, we have edited a significant portion of the manuscript text.

Regarding specifically the use of the term "inherent/implicit ribosome affinity": We agree that we used this without providing a good definition of the term. We had wanted to highlight that there is a difference between a ribosome-free mRNA, which would recruit ribosomes with an apparent affinity (the "inherent" affinity), and an mRNA already containing ribosomes, which might have a reduced affinity compared to the "inherent" affinity if the already bound ribosomes modify the initiation process. We have now completely removed this term, and re-written the text using more general terminology (which we hope will be clearer and more accessible to non-specialist readers).

Major point 3: *"discussion of initiation rates is not based on any direct experimental measurement"*

We have now explicitly stated the nature of the data used (ie protein levels, protein turnover and mRNA levels) and that initiation rates were calculated from these data.

Major point 4: *"How was the passage time calculated?"*

This is simply the sum of the mean decoding times for each codon, which gives the time required by one ribosome in isolation (ie in the absence of any traffic jams) to translate an ORF. This is now stated in the legend to figure 4.

Major point 5: *ribosome density on different constructs*

Please see under "comments highlighted by the editors", point 3, above.

Major point 6: *Use of statistics*

There was some confusion caused by our sloppy use of technical terms. All the comments made by the referee are of course correct. The statement that the min8max construct was significantly different from max in figure 4 was actually based on a post-hoc test (we used Tukey) as the referee suggests. We have now statistically analysed all data shown in the manuscript and included the results in the figures and figure legends as stated in detail in the section "Comments highlighted by the editors" above.

Minor point 1: '*ribosome movement immediately after the start codon*'

This sentence has been altered as suggested by the referee.

Minor point 2: *Reference to Firczuk* et al.

The earlier reference to this publication has been introduced as suggested.

Minor point 3: '*Movement near the start codon cannot be fully decoupled ...*'

We have reworded this sentence to make it clearer, and have added a reference to an earlier paper where we discuss the relationship between 'traffic jams' and average ribosome speed in more detail.

Minor point 4: *grammatical errors and misuse of words*

We have thoroughly re-read the manuscript and removed such errors where we have found them.

Minor point 5: *discussion of slow codon ramps*

Indeed, consistent with the referee's comments on a lack of universal acceptance for the ramp model, a paper published while this manuscript was under review (Shah *et al* 2013, *Cell*, Pubmed ID   23791185) shows that this model is not consistent with ribosome footprinting data. In light of the emerging contradictive data in the literature, and since our model of elongation control is not dependent on the existence of slow ramps, we have decided to remove this section from the text. Regarding the length of the ramp we would expect: Our data in figure 4 show that slow regions or ramps can be quite short yet exert an effect, but as an evolved feature of mRNAs, the length of such ramps (if they exist) would be subject to a complicated balance of selective forces and what length they would have in that case can only be guessed.

Comments made by referee 2

Major point 1: *ratio of protein abundance to mRNA levels*

We have now provided mRNA data for all constructs (including the data in figures 3 and 4 which the referee requests). For more details see point 1 of the response to comments highlighted by the editors above. In light of the referee's comment that protein-to-mRNA ratios must be shown: we deemed it important to show separate protein and mRNA level data, which convey more information than processed protein/mRNA ratios. We tried including protein/mRNA ratios as a third series of bars but the data then got very messy and difficult to understand. We hope that showing the data separately addresses the concerns of this referee sufficiently.

Major point 2: *do luciferase measurements capture the actual changes in protein expression level, or do they at least partially reflect an effect of the fraction of properly folded proteins?*

In figure 1, we provide western blot evidence for changes in protein levels, as well as activity evidence (bar graphs) for firefly luciferase, Renilla luciferase and the mCherry constructs. In all cases protein levels and protein activities vary with very similar ratios, indicating that activity reductions arising from codon-dependent misfolding are below our detection limit. For firefly luciferase, similar data have already been shown and discussed in our earlier publication (Chu et al 2011 NAR).

Major point 3: *probing of decoding times* in vivo

The constructs with the initial 10-Glu codon runs were constructed with a specific purpose in mind. The combination of 10 slow codons (in the case of GAG) with a following luciferase sequence that contains only fast codons means that, once the slow GAG barrier has been passed, ribosomes move away quickly through the rest of the transcript. Because 10 codons is also the reported physical extension of one ribosome, this means that the next ribosome can initiate exactly when the 10 codon sequence has been passed, and protein synthesis rates from this construct are reliable indicators of the time required to translate 10 GAG codons. We have re-phrased this section to make our thinking clearer.

If the construct is modified with GAG codons in non-consecutive positions, we would indeed expect that the effect is quantitatively different. How exactly it would be different could be predicted for any particular sequence based on our computational models, although there are no intuitively understandable or generally applicable rules for this.

While our manuscript was under review, we became aware of a study from the Stanfield lab which describes constructs where ten glutamine codons precede an unmodified luciferase ORF. The reported results further confirm our approach (although the fact that Kemp et al used an unmodified Luc sequence which has a lower average speed than our optimised one makes their construct less widely applicable as a codon speed reporter). We have now introduced a reference to this paper in acknowledgement of their parallel work.

Minor point 1: *"The terminology of near-cognate and non-cognate tRNAs should be clearly defined at the beginning of the paper."*

We have now introduced a sentence in the fourth paragraph explaining the difference between near-cognate and cognate tRNAs.

Minor point 2: *Describe the difference between variants, explain secondary structure better*

The substitutions made to generate the variants are now clearly identifiable in the colour-coded alignments which we introduced in the supplemental material.

We have now introduced an additional sentence into the legend to figure 2: "The calculated secondary structure content was defined in terms of ΔG and then normalised to each sta construct. Higher values indicate more stable secondary structure".

Minor point 3: *Re-write "Start codon liberation rates limit translation initiation on transcripts with high ribosome affinity"*

This section has been re-written extensively, shortening sentences and hopefully clarifying all points. The fraction of single-gene tRNA decoded codons in min346max as well as minCFLuc has been calculated and is now stated (20% and 25%, respectively).

Minor point 4: *"Start codon clearance control can be used to probe codon decoding times in vivo - the median decoding time of all the codons should be denoted"*

We assume that the referee asks for the median decoding time of all the maxCFLuc codons, to be compared to the GAA and GAG decoding times? The median decoding time of maxCFLuc codons is 0.12 seconds, compared to 0.07 and 0.82 seconds for GAA and GAG codons, respectively, and this is now stated in the text.

Minor point 5: *"add to figure 2 measurement of predicted translation efficiency by one of the most common indexes"*

We would really like to get away from a description of codon usage in terms of statistical usage properties like CAI, and towards a description founded on mechanistic understanding such as decoding times. Since the referee gives us the choice in this matter, we would prefer to not add these data. However, for the benefit of the referee, the CAI indices of the codon usage variants described in figures 1 and 2 (calculated using codonO) are: 0.409, 0.620, 0.990 for min, sta and maxCFLuc; 0.446, 0.587, 0.660 for min, sta and and max*HIS3* ; 0.473 and 0.687 for min and staRLuc; and 0.820 and 0.939 for mCHerry v3 and v4, respectively.

Comments made by referee 3

Comments made in the initial paragraph:

"Is every codon change for the same amino acid, just one used more or less frequently than the starting one?"

Every codon change is for the same amino acid. It was an oversight that we did not state this explicitly, this has now been done.

"The Chu et al., 2011 paper actually cites additional manuscripts for the methods and it would be worthwhile to have a full paragraph that described how the studies were performed."

To give the full description of the method again here would considerably lengthen the manuscript, and given that this information is available, we would argue should not be necessary here. However, to address the referee's comment, we have introduced a detailed paragraph that summarises the modelling logic on which the software is based into this section of materials and methods.

"the figures with the amino acid sequences are difficult to analyze for codon change. The authors might consider a designation of the following type: GUG81GUA (+2.3-fold) which would indicate the codon changed and the increase or decrease in codon usage relative to the standard."

We have tried to implement the format suggested by the referee, but found that this was rather illegible to human readers. We can imagine two uses for such a representation: a) for machine parsing of the speed changes for additional analyses. In order to enable this, we have now introduced a new table into the supplemental material and methods (table 1) that gives the actual decoding times coming out of our model for all sense codons. Together with the sequences themselves, this should enable efficient parsing. b) for a human reader. As we stated above, we do not think that a long list of numbers and letters is really suitable for this purpose. Instead, we have generated alignments of the sequences in which the speed change is colour coded, and these are now available in the supplemental material information. We hope that together, this additional material addresses the referee's comment in a satisfactory manner.

Major point 1: *Comment on intramolecular traffic in circular polysomes*

The referee is of course correct in saying that the issue of circular ribosome movement complicates the analysis of polysome behaviour. However, from the point of view of the problem we discuss here, it is not important whether a ribosome that attempts to initiate at an mRNA originates from a termination event on that same mRNA, or from the soluble ribosome pool. What is important is whether it finds the start codon already occupied by another ribosome, or not. That probability is given by the ratio of the number of ribosomes trying to access the start codon over the rate of movement away from the start codon. The additional statement by the referee, that *"one should be able to use the transit time and ribosome density (number of ribosomes per mRNA) to predict how frequently an mRNA is reinitiated"* is also true, and has indeed been used by us and others to derive initiation rates for the different yeast mRNAs from ribosome footprinting data. However, this statement is again true for both circular and linear translation models.

Major point 2: *"in Figure 1, it would be helpful to know if the mRNA (min, sta and max-CFluc) is in small, medium or large polysomes"*

All three mRNAs are expected to be in large polysomes, as is discussed in more detail in our response to major point 5 of referee 1.

Major point 3: *Organisation of figure 2 and correction for RNA levels*

We have tried to reformat the figure as suggested by the referee and showed the two versions to a couple of test readers, all of whom found the existing version easier to understand. We also note that other publications presenting this kind of analysis use the same format as we do (cf Figure 1A in Qian *et al* 2012 *PLoS Genet* **8(3)**: e1002603).

Regarding the question of how mRNA levels relate to our expression data: in order to address this point, we have now measured and show detailed data on mRNA levels for all our constructs. In answer to the referee's question: there is no fixed RNA: protein ratio, this ratio varies as we would expect if the major point of control is at the translational level. For further details, please see the mRNA data that have now been added to figures 3, 4, 5 and 7 (in addition to data shown in the original version in figures 2 and 6).

Major point 4: *Comments on the maths of elongation speed and polysome densities*

Please see under "comments highlighted by the editors", point 3, above.

Major point 5: *rate vs number of ribosomes arriving at the start codon*

Strictly speaking, whether an mRNA is under elongation control is determined by the ratio of rates of ribosomes arriving at the start codon over the rate of ribosome liberating the start codon. If this rate is higher than one, not all ribosomes arriving can immediately initiate translation – they may wait around until they can access the start codon or fall off and give another ribosome a chance to access the start, we do not know for sure which. If an mRNA is under elongation control, our model predicts that it can be reverted to initiation control by reducing the rate of ribosomes arriving at the start codon. The important point here is only that we achieve a relative reduction in the start codon arrival rate, as our experimental data indicate we did. Whether scanning is processive or not (ie whether ribosomes sometimes fall off the mRNA) is in our opinion not relevant in this context.

In response to this comments as well as comments made by other referees, we have extensively edited this section in what we hope is clearer and less ambiguous language.

Major point 6: *"is passage time the same as transit time (the time required for the ribosome to go from the start codon to the terminating codon?"*

Yes it is. We have replaced the word "passage time" with "transit time" throughout the manuscript, as this seems a better descriptor of what we mean.

Major point 7: *The role of G-quadruplexes in the GAA/GAG constructs*

We thank the referee for drawing our attention to this issue, of which we were not aware. We have found no feasible experimental test to exclude formation of such quadruplexes at the RNA levels. Indirectly, since as far as we understand G quadruplexes can form at the level of RNA and DNA, we would argue that since we see no problems with transcription (see mRNA levels in this figure) this at least rules out such structures at the DNA level.

The suggestion by the referee of comparing codons with a lower G content seems to be the best strategy for ruling out G quadruplexes as the source of reduced translation on the GAG construct. Fortunately in this respect, a study by Kemp et al which was published just before we submitted our manuscript (but of which we had been unaware at the time of submission) compared constructs with Glutamine runs (CAA and CAG), and found results that strictly mirror our data. This study has now been cited in support of our data.

Major point 8: *Relationship between* HIS3 *expression levels and growth rates*

We never meant to use growth in medium lacking histidine as a means of assessing *HIS3* expression levels – expression levels have been unambiguously determined by western blotting. The growth experiments in figure 7C were intended to ask whether elongation control of His3 protein levels has phenotypic consequences. The fact that we can show growth advantages of a *HIS3^{sta}* strain over other codon variants qualitatively says that it does. We have now altered the last sentence of the results section to indicate this more clearly.

Speaking more generally, we would not expect a linear relationship between His3 levels and growth. The relationship between His3 protein and histidine levels is dependent on enzyme kinetics in multiple competing amino acid biosynthesis pathways and therefore strongly non-linear. The question what histidine levels are optimal for growth is an extremely complex optimisation problem (since all amino acids have to be provided in balance with usage). In consequence of these complex relationships, we cannot interpret the relationship between expression levels and growth rates in any finer detail than we have done.

Minor point 1: *"the authors should not use terms such as "for the first time", "this is the first reported mechanism", etc."*
We have removed all but one reference to novelty as requested by the reviewer. We have only left this in the very last sentence of the discussion, where we draw attention to a specific aspect of novelty ie making a connecting between two fields which have so far operated with little contact although they both deal with matters surrounding protein synthesis (translational control and sequence evolution).
Minor point 2: *"The authors need to be consistent with the use of abbreviations"*

We have proof-read the manuscript in this respect and used consistent abbreviations throughout, as well as replacing the term "message" with "mRNA" throughout the manuscript.

2nd Editorial Decision   18 September 2013

I am contacting you to let you know that we have now heard back from two of the original referees for your manuscript (comments included below) and they both support publication of the revised version, pending minor textual revisions and clarifications as pointed out by referee #3 (old numbering). With regards to the remarks made about the manuscript title, we do agree with the referee that a modulation along the lines of 'translation elongation influences translation initiation' would give a more accurate presentation of the current work.

Furthermore, we now encourage the publication of source data, particularly for electrophoretic gels and blots, with the aim of making primary data more accessible and transparent to the reader. Would you be willing to provide files comprising the original, uncropped and unprocessed scans of all gels used in the figures? We would need 1 file per figure (which can be a composite of source data from several panels) in jpg, gif or PDF format, uploaded as "Source data files". The gels should be labelled with the appropriate figure/panel number, and should have molecular weight markers; further annotation would clearly be useful but is not essential. These files will be published online with the article as a supplementary "Source Data". Please let me know if you have any questions about this policy.

Given the referees' positive recommendations, I would like to invite you to submit a revised version of the manuscript, addressing the above-mentioned concerns as well as the comments from the reviewers.

Thank you for the opportunity to consider your work for The EMBO Journal. I look forward to your revision.

REFEREE REPORTS

Referee #1

In their revised manuscript, the authors have dealt with my concerns in a satisfactory way and I am happy that this version is now acceptable for publication.

Referee #3

This is a much improved manuscript and provides some excellent food for thought on the influence of elongation as relates to perhaps being the limiting factor for initiation due to the enhanced initiation efficiency of terminating ribosomes relative to free 40S and 60S subunits. There are however, a few concerns that should be addressed editorially prior to publication.

1. page 5, line 33 - Are "expression variants 3 and 4" equivalent to sta and max?

2. Figure 2 - are the rel expression levels for protein corrected or normalized for mRNA content?

3. Figure 4 - The affect of introducing "slow codons" does not seem to kick in until after 53 slow

codons (i.e. activity for this construct is about 60% relative to 0 slow codons (ratio of activity to amount of mRNA)). This would appear to be well removed from the start codon (by 150 nucleotides) and thus should not influence start codon availability. Is there some other explanation for why this does not occur earlier?

4. For Figure 5, is there a calculated transit time (as in Figure 4)? This would be a help.

5. Figure 6 "predicted decoding time" should probably be codon/sec, not sec.

6. On the following pages, one of two corrections should be made: 1) message should be mRNA and 2) et al should be et al. (period after al):page 9, 10, 11 and 12.

7. This reviewer feels the title too strongly implies elongation controls initiation. A more suitable title might be something like "Influence of elongation rates on translation initiation". This concept would be re-enforced by citations where changes in elongation rates independently have been implicated in changes in elongation rates as in many instances, these seem to be coordinated (i.e. reduction in both initiation and elongation through phosphorylation/dephosphorylation of initiation and elongation factors).

2nd Revision - authors' response   09 October 2013

Response to comments highlighted by the editor

1)   "*change the title to more accurately represent the described work*"
We agree with the referee and the editor that the original title could suggest that elongation *always* controls initiation, which is not the case. However, we feel that the title suggested by the referee (Elongation *influences* initiation) is also not accurate since we show that if elongation is below a critical threshold, it stringently controls initiation. In our view, "Translation elongation *can* control translation initiation on eukaryotic mRNAs" is an accurate summary of our results, which does not suggest that elongation is always in control.

2)   "*supply source data for the gels used in the manuscript*"
Two source data files for gels shown in figure 1 and figure 7 have been prepared and uploaded with this manuscript. During the course of the study we had run multiple blots for each protein. For figures 1B and 1E, were no longer able to unambiguously determine from which blot we had taken the relevant blot sections, and we therefore replaced the CFLuc and mCherry blot sections shown in figure 1B and 1E to ensure that they correspond exactly to the full blots shown in the source data files.

Response to Comments made by referee 3

Comment 1: "*Are "expression variants 3 and 4" equivalent to sta and max?*"

No they are not. The mCherry variants were constructed by another lab, and we have kept the original designations for the variants since we did not want to create the impression that these arose from our work. We tested the two highest expressing variants since it was not clear from that study whether their variants behaved as our theory would predict, but our data show that they do. The nearest designation of v3 and v4 in our terms would be "nearmax" and "max". There is a relevant section discussing the origin of these variants in the text from page 5, line 27 to page 6, line 2.

Comment 2: "*Figure 2 - are the rel expression levels for protein corrected or normalized for mRNA content?*"

No, the data shown are raw protein levels as is indicated in the legend to this figure: "Protein expression levels (grey bars) are the same as in figure 1." In this figure, we compared the raw protein expression levels against various parameters that also vary with codon usage including mRNA content, so it would not have made sense to correct these values for mRNA content.

14

Comment 3: *comment on introduction of slow codon runs*

The effect of introduction of slow codons is significant from the min8max construct, as we state in the text (page 8, lines 31-33). The introduction of further slow codons up to codon 346 gradually reduces expression levels further. This effect can be understood if the overall architecture of the minCFLuc ORF is taken into account: there is a stretch of particularly slow codons near the middle of this ORF, and thus ribosome speed becomes reduced as more and more slow codons are introduced up to this particularly slow point. Again, this is discussed in detail in the text (page 10, lines 9-16).

Comment 4: "*is there a calculated transit time for figure 5?*"

We have now introduced these numbers as requested by the reviewer, with an explanatory sentence in the legend to figure 5.

Comment 5: *comment on units used (sec vs codons/sec)*

No the unit used is correct – it takes 0.072 seconds in our model to decode a GAA codon, but 0.82 seconds to decode a GAG codon.

Comment 6: *typographical errors*

We have replaced all instances of "message" with "mRNA" as requested. The stops behind "et al" were removed by the reference management program for unknown reasons, this has now been corrected.

Comment 7: *change of title and examples of co-regulation of initiation and elongation*

The title has been changed as explained above under the response to comment highlighted by the editor.

Regarding the citing of examples where initiation and elongation factors are co-regulated: We are aware of studies that have identified phosphorylation events on elongation factors, and overview articles relating to this work are cited (Browne and Proud 2002, and Stark 2010, page 12 line 36 to page 13 line 1). In addition, in response to the reviewer's comments we have introduced a reference to the observation of co-regulation of initiation and elongation in mammalian cells via eEF2-kinase during stress in mammalian cells (Patel et al, page 13 line 4).