



# Kent Academic Repository

**Chu, Dominique, Thompson, James and von der Haar, Tobias (2014) *Charting the dynamics of translation*. Biosystems, 119 . pp. 1-9. ISSN 1872-8324.**

## Downloaded from

<https://kar.kent.ac.uk/41513/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1016/j.biosystems.2014.02.005>

## This document version

UNSPECIFIED

## DOI for this version

## Licence for this version

UNSPECIFIED

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



# Charting the dynamics of translation



Dominique Chu<sup>a,\*</sup>, James Thompson<sup>a</sup>, Tobias von der Haar<sup>b</sup>

<sup>a</sup> School of Computing, University of Kent, CT2 7NF Canterbury, UK

<sup>b</sup> School of Biosciences, University of Kent, CT2 7NJ Canterbury, UK

## ARTICLE INFO

### Article history:

Received 3 May 2013

Received in revised form 17 February 2014

Accepted 24 February 2014

Available online 12 March 2014

### Keywords:

Translation

Modelling

Codon usage bias

## ABSTRACT

Codon usage bias (CUB) is the well-known phenomenon that the frequency of synonymous codons is unequal. This is presumably the result of adaptive pressures favouring some codons over others. The underlying reason for this pressure is unknown, although a large number of possible driver mechanisms have been proposed. According to one hypothesis, the decoding time could be such a driver. A tacit assumption of this hypothesis is that faster codons lead to a higher translation rate which in turn is more resource efficient. While it is generally assumed that there is such a link, there are no rigorous studies to establish under which conditions the link between translation speed and rate actually exists. Using a computational simulation model and explicitly calculated codon decoding times, this contribution maps the entire range of dynamical regimes of translation. These simulations make it possible to understand precisely under which conditions translation speed and rate are linked.

© 2014 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

The genetic code is highly degenerate. There are 20 amino-acids but 64 codons. An inevitable consequence of this is that each amino acid sequence could be encoded by a very large number of different mRNAs. Large scale analyses of codons have shown that individual species prefer some codons over others. This is commonly referred to as the *codon usage bias* (CUB). While the bare fact of CUB is well established, its underlying biological reasons are not. A number of drivers of the CUB have been proposed, including the abundance of isoacceptor tRNA, pre-mRNA level selection, mRNA concentration (Coghlan and Wolfe, 2000), mRNA secondary structure (Tuller et al., 2011), the efficiency of translation initiation (Sato et al., 2001), GC content (Knight et al., 2001), gene length (Moriyama and Powell, 1998), translation error (Stoletzki and Eyre-Walker, 2007; Shah and Gilchrist, 2010), protein structure (Xie et al., 1998; Mukhopadhyay et al., 2007) and others (Novoa and Pouplana, 2012; Gingold and Pilpel, 2011).

Perhaps one of the more important drivers of the CUB is the decoding time (Shah and Gilchrist, 2011). The current best understanding of the factors determining the decoding time goes back to a model by Gromadski and Rodnina (2004). The central element of the model is that cognate aa-tRNA species compete with near matches (the so-called *near-cognate* aa-tRNA) for access to

the ribosome. The latter are thought to occupy the ribosomal A-site for significant amounts of time before eventually unbinding; while bound they prevent access for the cognate aa-tRNA (Fluitt et al., 2007) thus causing a delay.

For many codons, near cognates are much more abundant than cognates. Even though each near-cognate occupies the ribosome only for a short time, collectively they cause a major bottleneck for translation as a whole (Chu et al., 2011). Consequently, the elongation time depends primarily on the ratio of cognate to near-cognates rather than on the absolute number of cognates. This model of cognate/near-cognate interaction has recently been corroborated experimentally (Chu et al., 2011).

A key prediction of the Gromadski–Rodnina model is that the decoding time may vary strongly even between synonymous codons. For example, in *Saccharomyces cerevisiae* the fastest codon (AGA) is read nearly 44 times faster than the slowest one (CUC). Similarly, among the synonymous codon sequences for a given protein the predicted *translation speed* (i.e. the inverse of the average time to read one codon) of the fastest sequence may be as much as five times lower than that of the slowest. Despite these large differences, the importance of speed for the evolution of CUB is currently unclear. The *prima facie* argument why translation speed should be selected for is as follows (Navon and Pilpel, 2011; Shah and Gilchrist, 2011): higher translation speeds lead to higher achievable *translation rates* (i.e. the number of translation termination events per time unit) given a fixed ribosome pool; hence by decreasing the time required for a ribosome to read a transcript, the cell can reduce the number of ribosomes while keeping the translation rate fixed.

\* Corresponding author. Tel.: +44 1227 827690.  
E-mail address: [d.f.chu@kent.ac.uk](mailto:d.f.chu@kent.ac.uk) (D. Chu).

Given that ribosomes are metabolically costly (Chu et al., 2011), it would seem natural to assume that there is a strong adaptive pressure towards faster mRNAs.

A tacit assumption of this resource argument is that it is actually the case that a faster transcriptome leads to a higher translation rate. This makes intuitively sense, but on further reflection it is not clear that it is always true. One simplified model of translation are *totally asymmetric exclusion processes* (TASEP) (Blythe and Evans, 2007); these systems are known to have three dynamically distinct phases. A low density, high density and a maximal current phase. For the first two, the flux (translation rate) is independent of the transition rate between sites (corresponding to the codon reading times). Whilst real ribosomes do not behave exactly like their TASEP models, many of the results of the theory still provide useful insights.

Direct empirical evidence for the conjectured link between translation rate and decoding time is ambiguous. Using *Escherichia coli* as a host Kudla et al. (2009) measured the translation rates of an extensive library of synonymous sequences with widely varying speeds. The authors reported no correlation between codon adaptedness and translation rate. Similarly, Qian et al. (2012) demonstrated experimentally that the time required to translate an ORF is not a good predictor for the translation rate. Another recent study by Charneski and Hurst (2013) analysed deep sequencing data and found that there is a speed difference between individual codons, but this difference is due to the biophysical characteristics of the nascent polypeptide, rather than bio-chemical parameters of the translation system. Charneski and Hurst concluded that the folding energy of the transcript plays at most a sub-ordinate role for the translation rate.

This partial evidence contrasts with received wisdom in biotechnology where codons of recombinant proteins are engineered routinely to maximise expression (Gustafsson et al., 2004), suggesting that codon choice can indeed impact the translation rate. Theoretically this view is also supported by Tuller et al. (2010) who found a correlation between codon adaptedness and expression level in a genome wide study involving both *Saccharomyces cerevisiae* and *E. coli*. Interestingly, these authors also noted that the folding energy modulates (weakens) the coupling between codon adaptedness and expression level. Further evidence for an important adaptive role of codon speed comes from sequence analysis. Common measures of codon adaptedness such as the CAI (Sharp and Li, 1987) or tAI (dos Reis et al., 2003) are often used as proxies for decoding speed and are able to predict various transcriptomic and proteomic key measures, including expression levels of both mRNA and protein (Gingold and Pilpel, 2011).

There is strong experimental evidence for the Gromadski–Rodnina model. For one, the original authors based their model on careful measurements of the interactions between cognate and near-cognate tRNA. Then, more recently Chu et al. (2011) showed for Firefly Luciferase in a yeast host system that simulations based on the Gromadski–Rodnina model can to a very good degree of accuracy predict the effect of synonymous codon substitutions and changes in the aa-tRNA abundance on the overall expression rate. This corroborates the Gromadski–Rodnina model.

While there is good evidence for the Gromadski–Rodnina model, there still seems to be some confusion as to what it entails about the effects of codon usage on the translation rate. Traditionally, the effect of translation speed (that is the time required to read individual codons) and the translation rate (i.e. the amount of protein produced per time unit) is framed in terms of limitation scenarios. For example, it is claimed frequently that when initiation is limiting, then the codon speed should not impact on the translation rate at all. Similarly, one might be tempted to conclude that the translation speed is irrelevant when ribosome availability is limiting.

While translation as a dynamical system appears to be simple, this simplicity is deceptive. Translation in organisms is highly concurrent and competition for a common ribosome pool introduces interactions that complicate the dynamics considerably. Purely verbal reasoning about this system can be difficult. Hence, formal reasoning tools are required.

In this contribution we will use a computational model of translation (Chu et al., 2012) and generate a comprehensive map of all dynamical regimes relevant to the system. Previously, this model (Chu and von der Haar, 2012) has been applied to model *Saccharomyces cerevisiae*. For this purpose, it was parametrised specifically according to known quantitative details of the yeast system. In this article, we will use the model differently. Instead of committing to a specific parametrisation corresponding to the translation system of a particular species, we will elucidate the dynamics of translation globally. The aim of this is to provide insight into the possible dynamical regimes of the system.

We find that a higher translation speed nearly always entails a higher translation rate, with only two caveats: The first one is the codon position effect. When a transcript is concurrently occupied by a large number of ribosomes, then the translation rate depends on the decoding speed and on how codons are arranged. Secondly, there is no link between translation rate and speed if the ribosome affinity to the 5'-cap structure is sufficiently low to make initiation a major limiting factor of the system. Yet, even if this is the case, we find that mRNA circularisation (whereby ribosomes immediately re-initiate on the same transcript upon termination) can re-establish this link. This means that the widely held belief that in initiation limited systems the codon speed does not impact the translation rate is not necessarily true.

## 2. Simulation model

The computational model we used here has been described in Chu et al. (2012) and is used with the *Saccharomyces cerevisiae* cognate/near-cognate scheme as reported in Chu and von der Haar (2012). The model is agent-based representing explicitly every single mRNA and ribosome. The latter bind to individual transcripts following first order kinetics and then perform a directed random walk with transition rates calculated following the Gromadski–Rodnina model (Gromadski and Rodnina, 2004; Fluitt et al., 2007). Upon termination ribosomes may re-initiate at the same transcript or unbind into the cell volume to rebind to a randomly chosen transcript at a later time again. The model allows the user to set an upper limit to the number of consecutive re-initiation events. Unless stated otherwise, this maximal number was set to 1 in the simulations presented here.

The full simulations presented in Fig. 3 assume 3 million tRNA molecules, 200,000 ribosomes and 15,000 mRNA sequences distributed over 3624 different species. This resulted in average mRNA reading speeds of between 1.6 and 7.8 codons per second for the standard sequence, between 5.9 and 11.8 for the optimised sequence, and 0.65 and 1.65 for the de-optimised sequence.

In all other simulations reported here we used the Firefly Luciferase gene that is frequently used as a reporter gene. In *Saccharomyces cerevisiae* the Firefly Luciferase sequence StaFLuc is of medium speed and it can be experimentally (de-)optimised by appropriate synonymous codon substitutions. The speed-optimised version – MaxFLuc – is obtained from the standard sequence by exchanging all codons for the fastest available synonym. Analogously, the de-optimised MinFLuc is obtained by replacing all codons by the slowest synonym. On sparsely populated transcripts the average reading times per codon for MinFLuc, StaFLuc and MaxFLuc are 0.53, 0.25 and 0.126s respectively. This

means that StaFLuc is read roughly twice as fast as MinFLuc and MaxFLuc is again read at about twice the speed.

In the simulations reported in Figs. 1, 2, and 6 we created mixed sequences by concatenating the head of one sequence with the tail of another sequence at a given breakpoint. For example, in Fig. 6 the sequence MIN–MAX with breakpoint 211 is identical to MaxFLuc for the first 211 codons; the remainder of the sequence is identical to MinFLuc.

### 3. Results

#### 3.1. Translation rate and codon speed

If elongation is a limiting factor for the translation rate, then an increase of the elongation rate entails an increase of the number of proteins that are produced per time unit from each transcript. Synonymous codon substitutions are one way to change the elongation rate. Hence, increasing the average codon speed of an ORF will always lead to an increased rate of translation as long as elongation is limiting. To better understand how codon substitutions actually impact on the translation rate, we implemented a reduced model of translation (see Section 2 for a description of the model). It differs from the full model that has been described previously in that it has only a small number of (up to 100) mRNAs and ribosomes. This simplification allows extensive parameter sweeps across all dynamical regimes in a way that would not be possible with a full model. In particular the simplified model makes it possible to explore systematically all possible limitation scenarios that could apply to translation. This will allow us to understand under which conditions there is a link between codon speed and the translation rate in very general terms.

We first assumed that the affinity of ribosomes for the mRNA sequence is sufficiently high so that ribosomes bind as soon as the initiation site is free. If one then also assumes that there is only a single ribosome and mRNA then the behaviour of the system is easy to understand. Due to the high ribosome affinity the time between termination and re-initiation can be ignored. Consequently, this system is well described by a circular motion. A faster average decoding time then simply means that within a given period of time more circular events can be completed, i.e. the translation rate is higher.

The same argument can be made for more than one ribosome as long as ribosomes are not interfering with one another, i.e. there are no traffic jams on the transcript. Hence, for a low number of ribosomes one would, within this toy set-up, expect that the translation rate scales with the average decoding speed of the sequence. This is precisely what we find in our simulations. In Fig. 1a we vary the average translation rate ( $x$ -axis) and the number of ribosomes ( $y$ -axis) and observe a steady increase of the translation rate as the transcript speed increases at the lower end of the graph.

For a higher number of ribosomes the qualitative behaviour changes. The transition from slow to fast sequences becomes much more abrupt. It is still the case that the fastest sequence on the right hand side of the graph always shows a higher translation rate than the slowest sequence on the left hand side; however the average codon speed no longer has the proportional effect on the translation rate that can be observed for low ribosome numbers. Ribosome–ribosome interactions or traffic jams are an additional complication in this regime, so that the average reading time of the sequence becomes secondary to the overall order of codons on the sequence.

The steady relationship between the sequence speed and the translation rate can be restored by increasing the number of mRNAs. To show this in simulation, we fixed the number of ribosomes to 35 and varied the number of mRNAs from 1 to 65. Fig. 1b

summarises the results. Its bottom line corresponds exactly to the top line in Fig. 1a displaying substantial traffic jams. As the number of mRNA increases, the ribosomes distribute equally across all the transcripts and the effects of traffic jams become less important. This restores the smooth increase of the translation rate with the increase of the average decoding time of the sequence. At the top of the figure the system is extremely initiation limited with roughly 2 mRNAs competing for each ribosome. In the most extreme case with 60 transcripts in the system competing for 35 ribosomes, every mRNA will be unoccupied half of the time with associated long waiting times between subsequent initiation events. Still, our simulations show that over the whole range of parameters, the speed optimised sequence is translated at a substantially higher (by a factor >4) rate than the de-optimised sequence.

The dependence of the translation rate on the decoding time is broken when elongation ceases to be rate limiting. In real cells this could be the case when the affinity of the ribosome to the transcript is very low. Formally, this corresponds to making the first step of elongation very slow compared to all subsequent elongation steps. In this case then the speed of the other elongation steps is irrelevant for the overall translation rate as long as they are much faster than the first elongation step. Newly terminated ribosomes cannot rapidly re-initiate in this scenario even if there are free transcripts available. Faster codons do not lead to higher translation rates (see Fig. 2b).

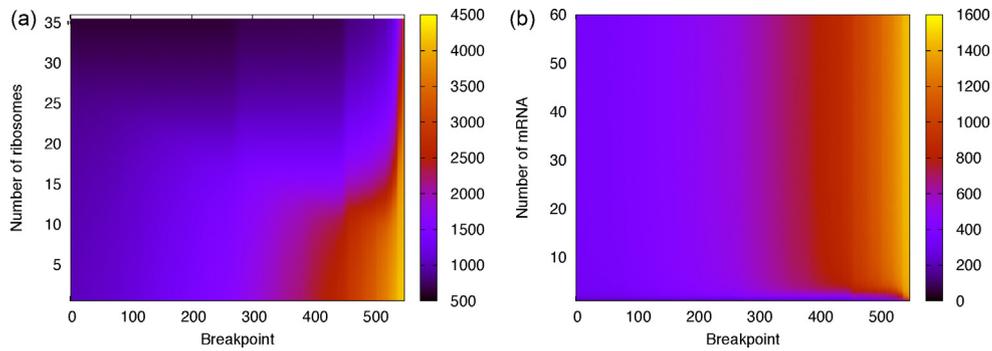
In this scenario translation is initiation-limited in the sense that initiation events are rare. Another way to implement initiation limitation is to make ribosomes a limiting resource (while restoring the ribosome affinity to high values, as discussed above). *Prima facie* those two ways of realising initiation limitation have the same immediate consequence (few initiation events), but dynamically they are still very different. If only ribosome availability is limiting, then the link between codon speed and translation rate is preserved; if ribosome affinity is limiting then the translation rate becomes independent of the codon composition.

#### 3.2. Local and global effects

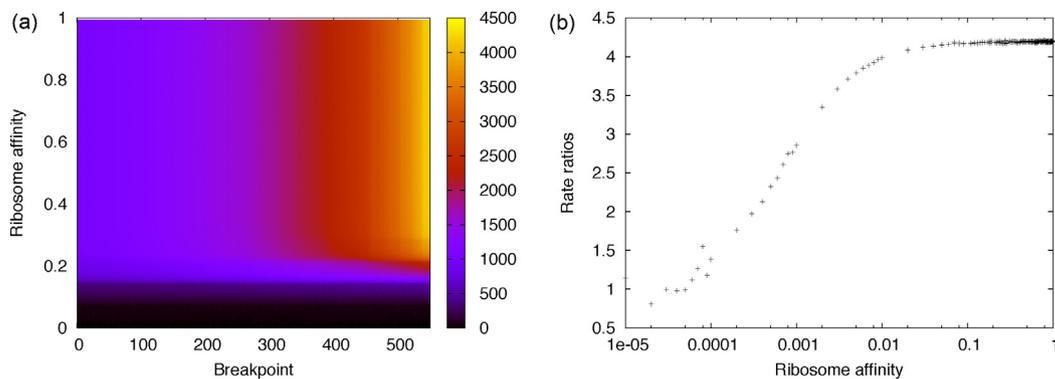
The codon usage literature sometimes distinguishes between purely local and global changes of the CUB. The former refers to changes that leave the system-wide conditions quasi-constant. Dynamically, local effects correspond to situations where the number of free ribosomes is, for all practical purposes, infinite (while having a finite concentration). This decouples the dynamics of one transcript from that of all others and the initiation rate becomes a constant.

Yeast transcripts are not globally speed optimised. This is not surprising because translation speed is unlikely to be the only selective force acting on codon selection. A number of other genomic parameters including GC-contents (Knight et al., 2001), mRNA folding (Tuller et al., 2011; Bentele et al., 2013), co-translation protein folding (Zhang et al., 2009; Xie et al., 1998; Mukhopadhyay et al., 2007) are also relevant. Moreover, even if translation speed were the only adaptive force, it would be very difficult to maintain a globally optimised genome in the face of continuous mutational pressure.

Notwithstanding this, in simulation, one can check what would happen if they were optimised, or indeed de-optimised. To understand this, we implemented a realistically sized model of translation in *Saccharomyces cerevisiae* using the best known parameters for the system with all mRNA sequences speed-optimised/de-optimised. This model is based on best available information about the quantitative properties of yeast cells. For further details see Section 2 and Chu and von der Haar (2012). As expected from the above results we found the global codon optimisation/de-optimisation to lead to a global increase/decrease

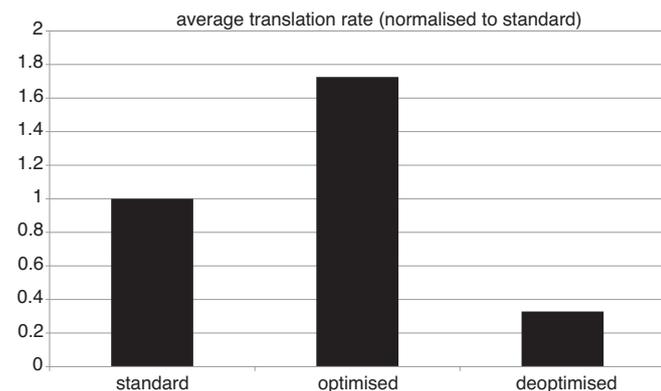


**Fig. 1.** A heatmap showing the change of protein expression as a function of ribosome availability in a simplified model. Along the x-axis the average decoding speed increases. At the far left end the sequence is MinFLuc; the right end of the graph shows results for MaxFLuc. In-between these two extremes the sequences are concatenation of the two with the head corresponding to MaxFLuc and the tail to MinFLuc. The breakpoint indicates the codon number where the concatenation was made. The ribosome affinity factor was set to 1, which means that ribosomes initiate immediately upon termination if the initiation site is free. The colour of the heatmap indicates the number of translation terminations within 1500 s. (a) Varying the number of ribosomes and changing the sequence. There is exactly one mRNA molecule in the simulation. Along the y-axis the number of ribosomes are increased. (b) The same as (a), but the number of ribosomes is kept fixed at 35 and the number of mRNA molecules is varied instead. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)



**Fig. 2.** (a) Same as Fig. 1, but the ribosome affinity is varied instead. The simulation consists of a single mRNA and a single ribosome. For very low ribosome affinities a faster sequence does not imply a higher translation rate. (b) Using the data from (a) it shows the ratio of the translation rates of MaxFLuc and MinFLuc for different affinities. To improve readability the x-axis is represented in log-scale. A ratio  $>1$  indicates that the fast sequences are translated at a higher rate; if the ratio is smaller than 1, then this means that the fast sequences are translated at a lower rate. Finally, a value of exactly 1 means that there is no difference.

of the translation rates respectively (see Fig. 3). Indeed, a globally optimised transcriptome is also locally more efficient. To illustrate this we compared the translation rate of a speed-optimised version of YFR055W in a standard background with that of the same ORF in an overall optimised transcriptome. We found that in a standard

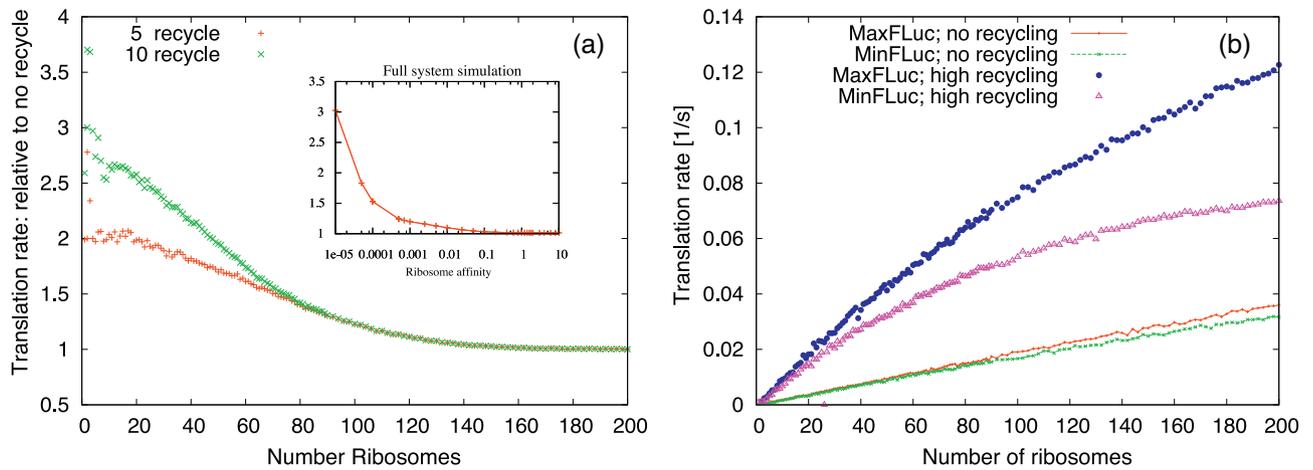


**Fig. 3.** Comparing the translation rate in a model of translation of *Saccharomyces cerevisiae*. The bars indicate global translation rates (relative to a standard model) obtained from system wide simulations of a yeast transcriptome. The parameters for the model are taken from Chu and von der Haar (2012). Optimising all sequences leads to an overall increase of the translation rate relative to the actual codon usage patterns. Similarly, when all sequences are de-optimised then the translation rate decreases.

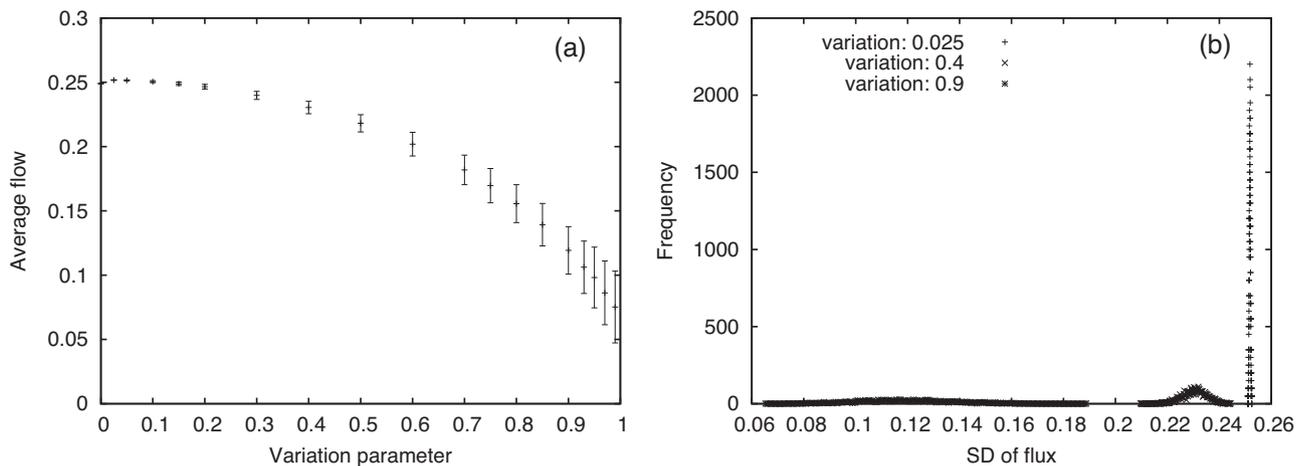
background, that is when we use the yeast transcriptome then the optimised version of YFR055W translates  $\approx 1.8$  times more protein than the standard version (the average reading speed is increased by a factor of  $>2$ ). This indicates that using the best known parameters for yeast suggests that the system is in a regime where local codon substitutions are effective at increasing the translation rate. If placed within a globally optimised background the expression rate of the optimised transcript increased by a further 7%. By the same token, globally de-optimising codons leads to a decrease of the translation rate both locally and globally.

### 3.3. Transcript circularisation

It has been shown for yeast that eIF4E/eIF4G/Pab1p complexes can circularize capped polyadenylated mRNA, suggesting that this could lead to ribosome recycling (Kopeina et al., 2008; Wells et al., 1998) (i.e. upon termination the ribosome immediately re-initiates on the same transcript). This could increase the (local) translation rate of an ORF if its 5'-end is not conducive to ribosome initiation, i.e. ribosomes have a low affinity for it. Ribosome recycling can also increase the global translation rate in that it reduces the “dead-time” of ribosomes between termination and initiation at the next transcript. Fig. 4a summarises the results of simulations of a single StaFLuc ORF with a very low affinity for ribosomes. For low numbers of ribosomes the simulations predict an increased translation rate by a factor of 3. The same increase can be achieved globally (see Fig. 4a, inset).



**Fig. 4.** The average expression rate as a function of the number of ribosomes for different values of ribosome recycling. (a) A simulation of StaFLuc in isolation shows that the effect of ribosome recycling is dynamically more important when ribosomes are more scarce. The inset shows global translation rates for system wide simulations. Again, ribosome recycling becomes more important as ribosomes become scarce. (b) This graph assumes a very low ribosome affinity (value of 0.00001 in the model). Without ribosome circularisation the translation rate is quasi independent of the codon speed. MinFLuc and MaxFLuc translate at about the same rate. The curves labelled “high recycling” show simulations where the ribosomes are recycled up to ten times. In this case, the translation rate depends again on the speed. MaxFLuc translates at a higher rate than MinFLuc.



**Fig. 5.** A simple TASEP varying the parameters. (a) The average flux as a function of the variation of the hopping rates in a TASEP model where the transition rates at each site were allowed to vary by up to the variation parameter. (b) A histogram showing the distribution of fluxes for 3 different variation parameters. The three histograms correspond to points in (a).

Ribosome recycling also has an effect on speed. When the link between translation rate and ribosome affinity is broken at low ribosome affinities, then circularisation can re-establish the link. Fig. 4a compares simulations of MinFLuc and MaxFLuc assuming a very low ribosome affinity. When there is no ribosome recycling, then the translation rates for the two sequences are nearly the same, but they differ substantially in the presence of recycling.

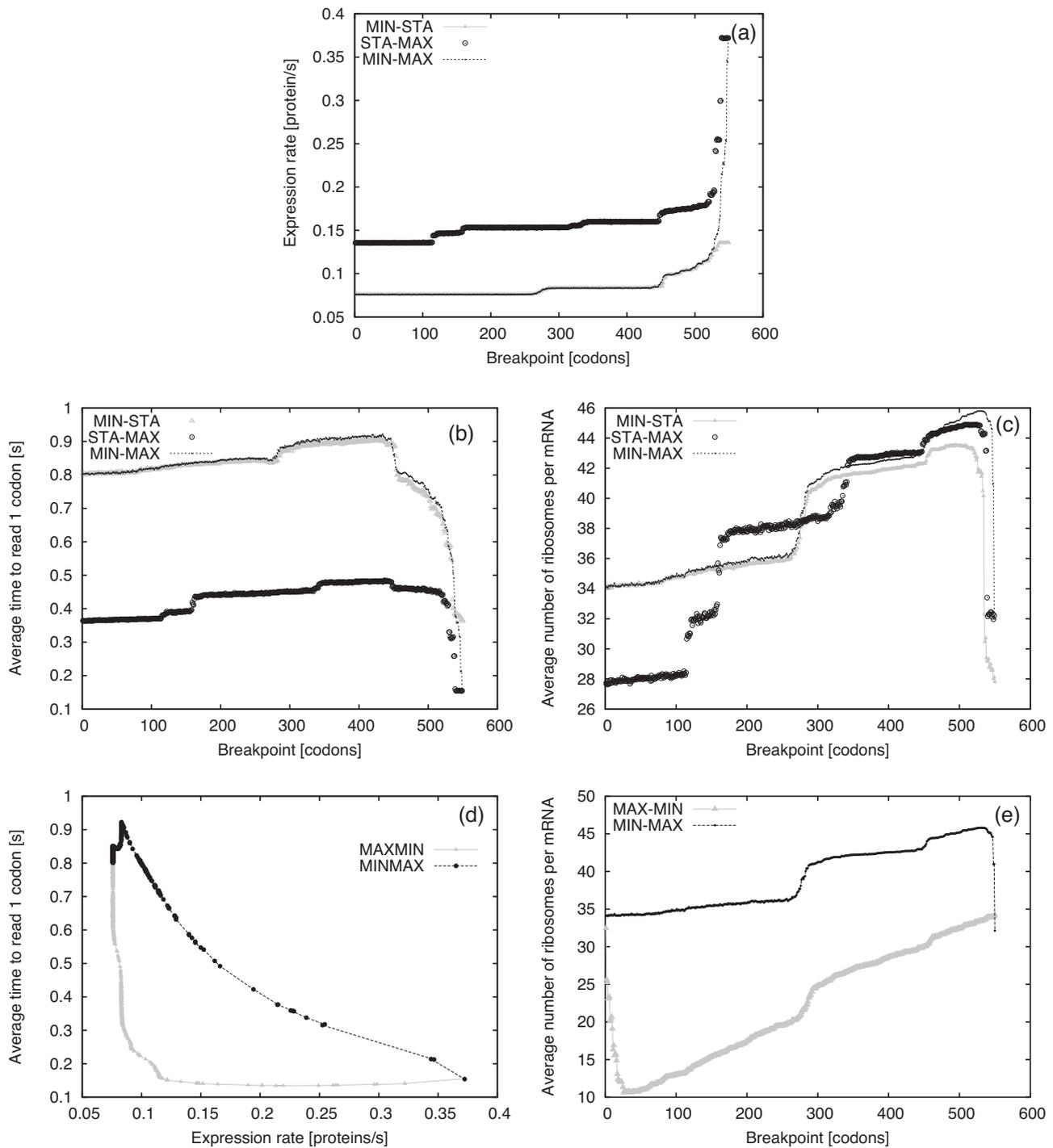
### 3.4. Dependence on codon order

It is frequently conjectured that the order of codons on a transcript is biologically relevant. Possible mechanisms for this include local aa-tRNA depletion, effects of codon usage on mRNA structure, variation of the decoding speeds to aid protein folding or slow codons at the beginning of sequences (Novoa and Pouplana, 2012). There are also purely dynamical position effects caused by the ribosome-ribosome interactions on the transcript (i.e. “traffic jams”). Various aspects of those have been widely studied using various simplified model systems (most notably TASEP (Blythe and Evans, 2007; Greulich

et al., 2012)) and simulations (Ciandrini et al., 2010; Tuller et al., 2011).

TASEP assumes all hopping rates to be the same (equal to 1). In this case one can show analytically (Blythe and Evans, 2007) that the maximal flux of particles ( $\approx$  translation rate) is 0.25. If one allows each of the transition rates to deviate from 1 (while keeping the mean transition rate at 1) then the average maximal flux (i.e. translation rate) goes down while the spread of actual fluxes around the mean flux increases (see Fig. 5a). This spread is a simple example of a positional effect, where the flux depends on how transition rates (i.e. codons) are arranged on the linear sequence.

Naively, one may conclude from this that in the presence of traffic jams more homogeneous sequences tend to be translated at a higher rate than less homogeneous ones. However, this theoretical insight is difficult to apply to mRNA sequences because synonymous codon substitutions will not normally conserve the average total decoding time. To better understand positional effects we considered again the transition from MinFLuc/StaFLuc to MaxFLuc (see Fig. 6) now assuming a non-limiting initiation rate which leads to substantial traffic jams. In the simulations we replaced



**Fig. 6.** Characterising the behaviour of mixed sequences at high ribosome availability. The  $x$ -axis is as in Fig. 1. The key “MIN-MAX” in the legend indicates that the left-most sequence is MinFLuc and that its head is replaced by MaxFLuc at the given breakpoint. The other keys have an analogous meaning. All simulations assume a ribosome affinity of 1, a single mRNA and 300 ribosomes. In practice this means that that ribosomes constantly attempt to initiate at the mRNA. (a) Translation rate as a function of sequence composition. The average sequence speed increases from left to right, but the translation rate does not always increase. (b) The average time to read 1 codon, i.e. the dwell time. A faster underlying sequence does not necessarily mean a lower dwell time. (c) The average number of ribosomes (i.e. ribosome sequestration) depends strongly on the sequence composition. We use this as a measure of traffic jams. For sequences with a fast head and a slow tail ribosome usage increases significantly. Interestingly, StaFLuc which is not the fastest sequence has a much lower tendency for traffic jams than both MinFLuc and MaxFLuc which are half and twice as fast respectively. (d) The average reading time as a function of the expression rate. The two curves considered are MAX-MIN and MIN-MAX. (e) Ribosome sequestration for mixed sequences with fast heads and slow heads. Clearly, a slow head can substantially reduce the number of ribosomes on the transcript.

the slow head of a MinFLuc sequence with the fast head of the MaxFLuc sequence. In Fig. 6 the size of the head increases from left to right (see Section 2); at the far right end of the graph the sequence is pure MaxFLuc. We also compared these results with the transition from MaxFLuc to MinFLuc where the fast

MaxFLuc head is replaced by the slow MinFLuc head (Fig. 6d and e).

We use the average number of ribosomes on the mRNA sequence as a measure for the propensity of a sequence for ribosome-ribosome interactions. The more ribosomes there are on

the sequence, the more interactions there will be between ribosomes. These simulations predict that mixed sequences with a fast head and a slow tail tend to have a high propensity for jams compared to pure sequences. When saturated with ribosome MinFLuc and MaxFLuc each carry on average fewer than 35 ribosomes. In the mixed sequences with a fast head and a slow tail this number can go up to 46 (see Fig. 6c). Likewise, a slow head followed by a fast tail dramatically reduces traffic jams and ribosome numbers can go as low as ten (Fig. 6d and e). Interestingly, with an average occupancy of about 28 codons StaFLuc (which is the WT version of the gene) is in its pure form much less prone to traffic jams than either MinFLuc or MaxFLuc.

These quite dramatic effects of codon choice on traffic jams for the mixed sequences are reflected in the average actual time required to read a codon (the dwell time, as opposed to the underlying reading time of a codon in absence of ribosome–ribosome interactions). Traffic jams increase the dwell time because the translocation step of ribosomes may be held up by other ribosomes ahead on the sequence. This may have interesting consequences: in the presence of traffic jams there may be synonymous codon changes that decrease the average reading time of codons but increase the average dwell time. The transition in Fig. 6b has several examples of this. From left to right the sequences get faster but the dwell time sometimes increases. This is a strictly position dependent effect. Making all sequences faster will always reduce the dwell time, whereas making some sequences faster will only sometimes reduce the dwell time.

For the sequences considered here an increased dwell time caused by a faster codon does not reduce the translation rate. Chimeras of slow heads and fast tails still tend to be translated at a higher rate (albeit modestly so) than the all slow sequences (Fig. 6a). However, large increases of the translation rate only take place at the far right hand side of the graph, when the sequences approach their pure states. This indicates that the translation rate may be strongly influenced by relatively short sections of the sequence.

#### 4. Discussion

Our results indicate that quite independent of any limitation scenarios the translation rate is determined by the translation speed. An increase of the latter can often lead to an increase of the former. This is also true when translation is extremely limited by ribosome availability. Similarly, traffic jams do not break the relation between translation speed and rate, although they do complicate it. When ribosome–ribosome interactions on the transcript dominate the dynamics of translation then the translation rate no longer depends on the average speed of the transcript alone, but it also starts to matter how codons are arranged on the mRNA. This is demonstrated in Fig. 6d and e which compare the average dwell times and ribosome sequestration for slow heads followed by fast tails and fast tails followed by slow tails. In real sequences, which will normally have fast and slow sections interspersed, the detailed dependence on codon order will be more complicated than in the idealised cases considered here. However, the basic insights gained from our simulations transfer to real cases: When there are traffic jams, then synonymous codon substitutions will often not have an effect on the translation rate.

Another limitation scenario that breaks the link between translation speed and translation rate is limitation by ribosome affinity. Biologically, this scenario could be realised, for example, when the initiation sequence of the message is strongly folded, preventing access for the ribosome. However, our simulations also show that even for low initiation rates transcript circularisation can re-establish the link between translation speed and rate.

There is good evidence that mRNA is circularised allowing ribosomes to re-initiate upon termination. The effect of this is that even when there is a strong limitation by affinity, faster codons would still lead to a higher translation overall compared to slower codons.

Within the field of translation researchers often distinguish between so-called *local* and *global* codon substitutions. The latter are large scale changes of the codon usage patterns across a high number of genes. Such global changes of codon usage have the potential to affect key variables of the system resulting in more (or less) efficient translation at a system wide scale. For example, if one could somehow decrease the reading time for *all* codons by a factor of two then this would increase the overall translation rate of every single message by a factor of two as well. A local change would be to make codon substitutions on one type of mRNA only.

*Prima facie* the codon substitutions in our simulations are all global in the sense that in our model there is only one type of transcript. However, our model can still be used to understand the effects of local codon substitutions by concentrating on the relevant regimes as we shall discuss below. Within the field of translation research it is widely believed that local codon substitutions cannot affect translation rates. The reasoning behind this assumption seems to be as follows: Within the cell each transcript “experiences” a constant background of free ribosomes. Since there is a large number of other transcripts, codon substitutions on individual transcripts do not affect this pool of free ribosomes. Consequently, the dynamics of translation of different transcripts effectively decouples in that a local change in one transcript does not affect the conditions of the others. This means that ribosome initiation can be described by a single constant rate. If one now also takes into account that for each initiating ribosome exactly one protein is made, then it is easy to see, so the reasoning, that the speed with which the ribosomes are dispatched over the transcript must be irrelevant. Hence, according to the argument, local changes of the translation speed have no effect on the translation rate and there is no case to be made for the local optimisation of codon usage. This reasoning is valid if ribosome affinity is the limiting factor. In this case ribosomes will have some substantial waiting time between a termination event and a subsequent initiation event. In our simulations this local regime can be located at the lower parts of Fig. 2a where, indeed, the translation rate is independent of the translation speed.

Interestingly, while based on this or similar arguments, the relevance of local codon substitutions is often doubted, it is also generally assumed that translation is under a selection pressure for global optimisation. It is unclear to us how a denial of local codon usage optimisation can be made consistent with this assumption of a global optimisation. It is hard to imagine that any selection for such a global optimised state is achieved by anything other than a series of local changes of codon usage. Hence, if the genome evolved into a globally optimised state, then local changes must have at least some noticeable effects as well.

Furthermore, and more importantly, the above argument against the effectiveness of local codon usage has to assume that ribosome affinity is the limiting factor for translation which may not be true. If it is the case that the transcript is circularised then even under conditions of low affinity the translation rate will depend on the translation speed. In our simulations this regime corresponds to Fig. 4; the circularisation partially undoes the effect of affinity limitation and re-introduces the dependence of the (local) translation rate on the (local) translation speed. Finally, another regime is when the rate of initiating ribosomes is high. This would correspond to the upper regions in Fig. 1a and lead to traffic jams. In this scenario, as discussed above, the translation rate

would again depend locally on translation speed (as well as codon order).

The precise nature of the dynamical regime of translation in real organisms remains unclear and may in any case be species dependent. However, at least some understanding of the dynamical regime in real organisms is now forthcoming. The assumption that translation initiation is the main limiting factor (Aitken, 2012) seems to be corroborated by Kudla et al. (2009) who found that for *E. coli* the ribosome affinity to be the main determinant of gene expression, whereas local codon substitutions have a small effect only. This is interesting because an affinity limited regime would be inefficient in that it would leave ribosomes, which are expensive to make, unused. A subsequent re-analysis of the data by Tuller et al. (2010) led to a more refined view, where the ribosome affinity modulates the dependence of the expression rate on the codon composition. In our model, this effect can also be seen for some parameter, and is illustrated in Fig. 2b. Tuller's findings would imply that ribosome affinities are limiting, but only weakly so. There has now also been direct experimental evidence (Chu et al., 2013) that decreasing the ribosome affinity also decreases the expression rate differences between synonymous codon variants. Again, this suggests that ribosome affinity is not the only limiting factor *in vivo* and that local codon optimisation can lead to increased protein production. At the same time, a study by Shah et al. (2013) suggested that ribosome availability is an important limiting factor of translation. Based on a detailed dynamical model of yeast the same conclusion had been reached by Chu and von der Haar (2012).

Still unresolved is the question of traffic jams. While direct observations of traffic jams are technically challenging, analyses of footprinting data (Siwiak and Zielenkiewicz, 2010; Ingolia et al., 2009) can provide some insights. From this it appears that traffic jams play a minor role in the dynamics of yeast translation. Whether or not the same is true in other species or whether entirely different dynamical regimes apply in those, remains to be seen.

Altogether, it seems at present uncertain in which regime translation is. It should be noted that several limiting factors can operate concurrently. For example, translation could be limited by ribosomes and also by mRNA availability. The following picture now emerges: If the translation speed is codon dependent, then there are dynamical regimes where both the global and the local translation rates depend on the translation speed. Only when the affinity is very low will the translation rate become independent of the translation speed.

This relationship between the translation speed and rate is a very generic property of translation system and cannot be avoided as long as the basic underlying model, i.e. the Gromadski–Rodnina model, is correct. Should it be found experimentally that the translation rate is not sensitive to the codon speed and should it not be possible to explain this by low ribosome affinities, then this would require re-assessing very basic mechanistic assumptions about translation.

## References

- Aitken, J.L.C., 2012. A mechanistic overview of translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* 6, 568–576. <http://dx.doi.org/10.1038/nsmb.2303> <http://www.nature.com/nsmb/journal/v19/n6/full/nsmb.2303.html>
- Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., Bluthgen, N., 2013. Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675. <http://dx.doi.org/10.1038/msb.2013.32>
- Blythe, R., Evans, M., 2007. Nonequilibrium steady states of matrix-product form: a solver's guide. *J. Phys. A: Math. Theor.* 40 (46), R333. <http://stacks.iop.org/1751-8121/40/i=46/a=R01>
- Charneski, C., Hurst, L., 2013. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11 (3), e1001508. <http://dx.doi.org/10.1371/journal.pbio.1001508>
- Chu, D., von der Haar, T., 2012. The architecture of eukaryotic translation. *Nucleic Acids Res.* 40 (20), 10098–10106. <http://dx.doi.org/10.1093/nar/gks825>
- Chu, D., Barnes, D., von der Haar, T., 2011. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 39 (15), 6705–6714. <http://dx.doi.org/10.1093/nar/gkr300> <http://nar.oxfordjournals.org/content/39/15/6705.abstract>
- Chu, D., Zabet, N., von der Haar, T., 2012. A novel and versatile computational tool to model translation. *Bioinformatics* 28 (2), 292–293. <http://dx.doi.org/10.1093/bioinformatics/btr650> <http://bioinformatics.oxfordjournals.org/content/28/2/292.abstract>
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M., von der Haar, T., 2013. Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 33 (1), 21–34. <http://dx.doi.org/10.1002/embj.201385651>, arXiv:<http://emboj.embopress.org/content/33/1/21.full.pdf>, <http://emboj.embopress.org/content/33/1/21>
- Ciandrini, L., Stansfield, I., Romano, M., 2010. Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mRNA translation. *Phys. Rev. E* 81 (5 Pt 1), 051904.
- Coghlan, A., Wolfe, K.H., 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast* 16 (12), 1131–1145. <http://dx.doi.org/10.1080/0898/2-F>
- dos Reis, M., Wernisch, L., Savva, R., 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* 31 (23), 6976–6985.
- Fluitt, A., Pienaar, E., Viljoen, H., 2007. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* 31 (5–6), 335–346. <http://dx.doi.org/10.1016/j.compbiolchem.2007.07.003>
- Gingold, H., Pilpel, Y., 2011. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* 7, 481. <http://dx.doi.org/10.1038/msb.2011.14>
- Greulich, P., Ciandrini, L.L., Allen, R., Romano, M., 2012. Mixed population of competing totally asymmetric simple exclusion processes with a shared reservoir of particles. *Phys. Rev. E* 85, 011142. <http://dx.doi.org/10.1103/PhysRevE.85.011142>
- Gromadski, K., Rodnina, M., 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell* 13 (2), 191–200. [http://dx.doi.org/10.1016/S1097-2765\(04\)00005-X](http://dx.doi.org/10.1016/S1097-2765(04)00005-X)
- Gustafsson, C., Govindarajan, S., Minshull, J., 2004. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22 (7), 346–353. <http://dx.doi.org/10.1016/j.tibtech.2004.04.006>
- Ingolia, N., Ghaemmaghami, S., Newman, J., Weissman, J., 2009. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324 (5924), 218–223. <http://dx.doi.org/10.1126/science.1168978>
- Knight, R.D., Freeland, S.J., Landweber, L.F., 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.* 2 (4).
- Kopeina, G., Afonina, Z., Gromova, K., Shirokov, V., Vasiliev, V., Spirin, A., 2008. Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. *Nucleic Acids Res.* 36 (8), 2476–2488. <http://dx.doi.org/10.1093/nar/gkm1177>
- Kudla, G., Murray, A., Tollervey, D., Plotkin, J., 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324 (5924), 255–258. <http://dx.doi.org/10.1126/science.1170160>
- Moriyama, E.N., Powell, J.R., 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* 26 (13), 3188–3193.
- Mukhopadhyay, P., Basak, S., Ghosh, T., 2007. Synonymous codon usage in different protein secondary structural classes of human genes: implication for increased non-randomness of gc3 rich genes towards protein stability. *J. Biosci.* 32 (5), 947–963.
- Navon, S., Pilpel, Y., 2011. The role of codon selection in regulation of translation efficiency deduced from synthetic libraries. *Genome Biol.* 12 (2), R12. <http://dx.doi.org/10.1186/gb-2011-12-2-r12>
- Novoa, E., Poupiana, L., 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28 (11), 574–581. <http://dx.doi.org/10.1016/j.tig.2012.07.006>
- Qian, W., Yang, J., Pearson, N., Maclean, C., Zhang, J., 2012. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 8 (3), e1002603. <http://dx.doi.org/10.1371/journal.pgen.1002603>
- Sato, T., Terabe, M., Watanabe, H., Gojobori, T., Hori-Takemoto, C., Miura, K., 2001. Codon and base biases after the initiation codon of the open reading frames in the *Escherichia coli* genome and their influence on the translation efficiency. *J. Biochem.* 129 (6), 851–860.
- Shah, P., Gilchrist, M., 2010. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet.* 6 (9), <http://dx.doi.org/10.1371/journal.pgen.1001128>
- Shah, P., Gilchrist, M., 2011. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. U.S.A.* 108 (25), 10231–10236. <http://dx.doi.org/10.1073/pnas.1016719108>
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., Plotkin, J., 2013. Rate-limiting steps in yeast protein translation. *Cell* 153 (7), 1589–1601. <http://dx.doi.org/10.1016/j.cell.2013.05.049>
- Sharp, P.M., Li, W.H., 1987. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15 (3), 1281–1295.
- Siwiak, M., Zielenkiewicz, P., 2010. A comprehensive, quantitative, and genome-wide model of translation. *PLoS Comput. Biol.* 6 (7), e1000865. <http://dx.doi.org/10.1371/journal.pcbi.1000865>

- Stoletzki, N., Eyre-Walker, A., 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* 24 (2), 374–381, <http://dx.doi.org/10.1093/molbev/msl166>.
- Tuller, T., Waldman, Y., Kupiec, M., Rupp, E., 2010. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.* 107 (8), 3645–3650, <http://dx.doi.org/10.1073/pnas.0909910107>.
- Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Rupp, E., Ziv-Ukelson, M., 2011. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.* 12 (11), R110, <http://dx.doi.org/10.1186/gb-2011-12-11-r110>.
- Wells, S., Hillner, P., Vale, R., Sachs, A., 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* 2 (1), 135–140.
- Xie, T., Ding, D., Tao, X., Dafu, D., 1998. The relationship between synonymous codon usage and protein structure. *FEBS Lett.* 434 (1–2), 93–96.
- Zhang, G., Hubalewska, M., Ignatova, Z., 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Biol.* 16 (3), 274–280, <http://dx.doi.org/10.1038/nsmb.1554>.