BioSystems 119 (2014) 1–9

Contents lists available at ScienceDirect

BioSystems

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

Charting the dynamics of translation

Dominique Chu[a,*], James Thompson[a], Tobias von der Haar[b]

[a] School of Computing, University of Kent, CT2 7NF Canterbury, UK
[b] 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.

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 (D. Chu).

http://dx.doi.org/10.1016/j.biosystems.2014.02.005
0303-2647/

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. Cherneski 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 fasted 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