Accurate measures of translation efficiency and traffic using ribosome profiling

Juraj Szavits-Nossan*[1] and Luca Ciandrini†[2,3]

[1]SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom
[2]CBS, Université de Montpellier, CNRS and INSERM, Montpellier, France
[3]Laboratoire Charles Coulomb (L2C), Université de Montpellier and CNRS, Montpellier, France

August 2019

Abstract

Ribosome profiles are often interpreted with underlying naive models that are unable to grasp the complexity of the biological process. We develop a Non-Equilibrium Analysis of Ribo-seq (NEAR), and provide unprecedented estimates of translation initiation and elongation efficiencies, emphasising the importance of rating these two stages of translation separately. When examining ribosome profiling in yeast, we observe that translation initiation and elongation are close to their optima, and traffic is minimised at the beginning of the transcript to favour ribosome recruitment. However, we find many sites of congestion along the mRNAs where the probability of ribosome interference can reach 50%. Our analysis excludes unreliable datapoints and discusses biases of the experimental technique; as a result, we conclude that current analyses and definitions of translation efficiency are not adapt to quantitatively asses ribosome interference.

1 Introduction

Understanding the rationale behind codon usage bias and the role of synonymous codons in regulating protein synthesis are among the main open questions in molecular biology. Despite the fact that mRNA translation is one of pivotal stages of gene expression, its sequence determinants are in fact still largely elusive (Gorochowski and Ellis, 2018). Recent technical advances, such as ribosome profiling, allow probing translation at codon resolution, making quantitative studies of transcript efficiency more accessible.

Ribosome profiling (or Ribo-seq or ribosome footprinting, as it is often called), is an experimental technique delivering a snapshot of ribosome positions along all transcripts in the cell at a given condition. Its archetypal version has been developed at the end of the 1960s to study translation initiation (Steitz, 1969; Kozak, 1981), then it has been extended in the 1980s to investigate the role

*jszavits@staffmail.ed.ac.uk
†luca.ciandrini@umontpellier.fr

of slow codons and ribosome pausing (Wolin and Walter, 1988). Recently, Ingolia et al. (2009) revamped this technique in the light of the Next Generation Sequencing era, and since then it is considered to be the state-of-the-art technique for studying gene expression at the level of translation.

In short, the method consists in isolating mRNA fragments covered by a ribosome engaged in translation ($\sim$ 30 nucleotides), which are then sequenced and aligned to build histograms of ribosome occupancy at codon resolution on each transcript. This technique has granted unprecedented view on translation leading to many new discoveries (Brar and Weissman, 2015). Examples include detecting novel translation initiation sites (Fritsch et al., 2012), identifying actively translated open reading frames (Calviello et al., 2016), quantifying the extent of stop codon readthrough (Dunn et al., 2013) and elucidating the translation of long non-coding RNAs (Guttman et al., 2013).

While this technique has significantly advanced our understanding of translation, there exist related issues that concern both the experimental side and the analysis of the raw data. The former includes for instance artefacts induced by the use of cycloheximide, an elongation inhibitor present in the original protocol to *freeze* ribosomes in their *in vivo* configuration, see for instance Weinberg et al. (2016); Duncan and Mata (2017). Besides, the processing of raw reads in order to determine the ribosome's A-site position is a complex task that requires advanced statistical methods (Ahmed et al., 2019), and it could be susceptible to known biases during the bioinformatic analysis, including the normalisation in order to obtain absolute values of the ribosomal densities (Bartholomäus et al., 2016). Although ribosome profiling has been very successful in studying mRNA translation at the genomic scale, the analysis of Ribo-seq still lacks a general and common scheme. The rate of mRNA translation, which is also called translation efficiency (TE), is commonly assumed to be proportional to the ribosomal density: the more ribosomes on a transcript, the more efficient is protein synthesis. On the one hand, this definition implicitly assumes that ribosome interference is absent. On the other hand, the extent to which ribosome queuing affects translation has been frequently debated (Li, 2015; Dao Duc et al., 2018; Dao Duc and Song, 2018; Diament et al., 2018; Riba et al., 2019).

Our work aims to establish a rigorous procedure for interpreting ribosome profiling data based on kinetic modelling, which we name NEAR (Non-Equilibrium Analysis of Ribo-seq). NEAR is able to estimate the balance between initiation and elongation rates at codon resolution by means of a non-equilibrium model of translation. Additionally, we propose a new gene-specific index for Translation Initiation Efficiency (TIE) that can be used to estimate the relative rate of initiation and elongation, and a codon-specific index for Translation Elongation Efficiency (TEE) that can detect ribosome queuing.

Codon elongation rates are usually estimated from Ribo-seq data assuming that the ribosome density at codon i is proportional to the time spent by the ribosome on that particular codon (Li, 2015; Weinberg et al., 2016). However, translation is intrinsically a multi-ribosome process and thus ribosome interference (traffic) is supposed to give rise to correlations between ribosome positions. The crux of the matter is that it is difficult to distinguish from the ribosome profiling data whether the ribosome spent long time on a particular codon because of the long decoding time or because it had to wait for the downstream ribosome to move away. Consequently, the relationship between codon elongation rates and ribosome density profile is potentially much more complex compared to the naive estimate that ignores correlations between ribosome positions. A few recent works have

attempted to investigate ribosome profiling using the same model as we do (Sharma et al., 2019; Dao Duc et al., 2018), but the aforementioned effect of ribosome interference was otherwise largely ignored, and the conclusions in the literature are contradictory.

In this work we show how to account for ribosome interference based on the mathematical solution of the exclusion process that we recently developed (Szavits-Nossan et al., 2018a,b). We also emphasise that it is not possible to quantitatively estimate *absolute* elongation rates from ribosome profiling data alone. Instead, we show how with NEAR it is only possible to infer elongation rates of an mRNA *relative* to the initiation rate of that transcript.

In the rest of the paper we first introduce the model underlying our method to infer elongation determinants, show that we cannot estimate absolute elongation rates from ribosome densities alone, and then explain the principles of our framework. Then we apply NEAR to a Ribo-seq dataset to identify the efficiency of translation initiation (compared to elongation), show that codon elongation rates are highly variable and context dependent, and identify codons at which the model is inconsistent with the ribosome profiling data. Moreover, we propose a new codon-specific method to quantify the elongation efficiency, and show evidence of local queuing *in vivo*.

2 Results

We base our method on a well-established stochastic model for mRNA translation known in the literature as the *totally asymmetric simple exclusion process* (TASEP), which has been introduced more than fifty years ago (MacDonald et al., 1968; MacDonald and Gibbs, 1969). Since then the model has been refined in many ways (von der Haar, 2012; Zur and Tuller, 2016) and has been explicitly and implicitly used in state-of-the-art modelling approaches of transcription and translation (Klumpp and Hwa, 2008; Mitarai et al., 2008; Reuveni et al., 2011; Ciandrini et al., 2013; Dao Duc and Song, 2018; Cholewa-Waclaw et al., 2019).

2.1 The model

In the TASEP, particles (ribosomes) move on a discrete one-dimensional lattice (the mRNA) with only steric interactions, i.e. particles cannot overtake each other. Each lattice site corresponds to a codon, and particles cover $\ell = 10$ sites, as the ribosome footprint covers ~ 30 nt or equivalently ~ 10 codons. A ribosome initiates translation with a rate α and it is then positioned with its A-site at codon 2; this happens only if the beginning of the lattice is not occupied by another ribosome. The ribosome then advances from codon i to codon $i+1$ with a rate k_i, provided that the next codon is not covered by the downstream ribosome (see top right drawing of the model in Fig. 1); eventually, when the A-site of the ribosome is at the STOP codon (the L'th site), it detaches with rate $k_L = \beta$. A coding sequence is then defined by a set of L rates $\alpha, \{k_i\}$, with $i = 2, \dots L$. Further information about the model can be found in the Material and Methods section.

Normally, the knowledge of all those rates allows one to compute simulated ribosome density profiles and protein production rates that can be compared to experimental outcomes; this is how this class of models has been exploited in the literature so far. However, there is an open debate regarding the estimates of the initiation and elongation rates, and no direct experimental method to

measure these rates exists. For example, codon-specific translation elongation rates k_i are often assumed to be proportional to the tRNA gene copy number or to the local tAI (Shah et al., 2013; Rudorf and Lipowsky, 2015; Gorgoni et al., 2016). Instead, we use this modelling framework in order to quantitatively determine codon elongation rates from ribosome profiling data.

This is an *inverse problem* in which we need to optimise the parameters $\{k_i\}$ in order to match the experimental Ribo-seq data. At first sight this may seem a straightforward problem, but there are several subtle complications. First of all, the parameter space is extremely vast: a typical protein consists of a few hundreds of amino acids, meaning that one would generally need to optimise a comparable number of parameters. Secondly, a change in a single k_i could in principle affect large

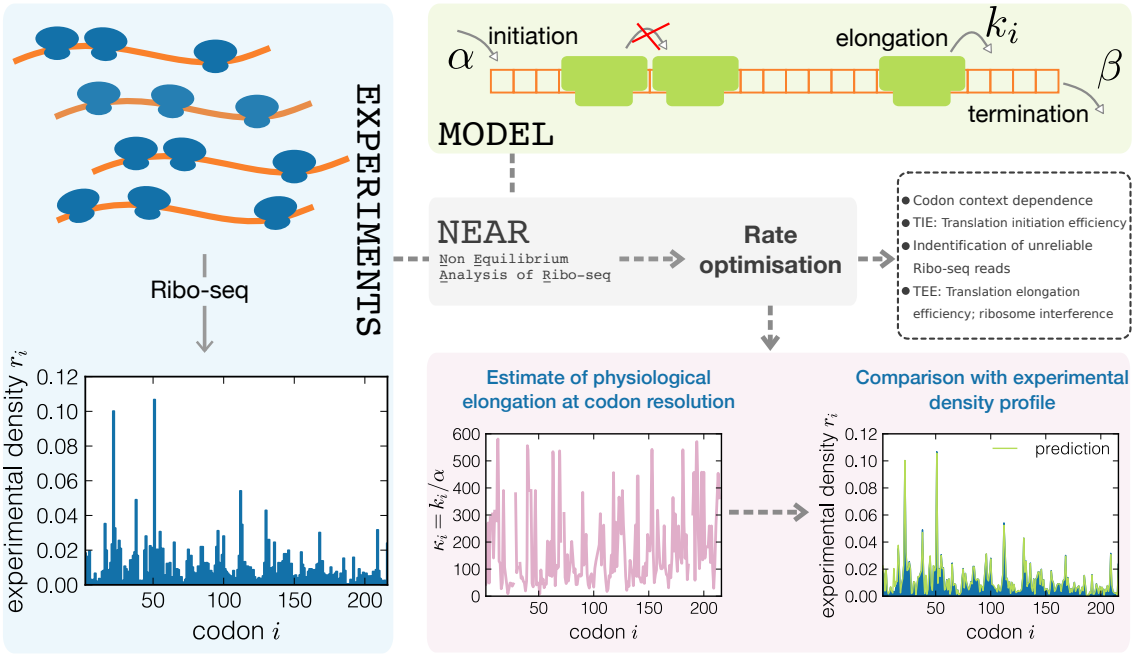

Figure 1: Sketch of the framework proposed. Experimental Ribo-seq profiles are first normalised and then processed using the stochastic model. In this figure we randomly chose a gene (YAL007C) to represent the work-flow. Its normalised density profile $\{r_i\}$ is represented in the bottom left panel. The model is shown in the top right green box: ribosomes covering ℓ sites are injected in the lattice with an initiation rate α, provided that there are ℓ empty sites between the A-site of the injected ribosome and the following one. Ribosomes then move from site i to site $i+1$ with rate k_i provided that the A-site of the ribosome downstream is more than ℓ sites away. Eventually, ribosomes leave the lattice with rate β ($k_L = \beta$) when their A-site is on the last site. In this drawing $\ell = 3$ for clarity, while in our analysis we used $\ell = 10$. The NEAR procedure then informs the model with the experimental data, and after optimising the parameters of the model in order to match the experimental ribosome profile for each gene, it estimates elongation rates (relative to initiation rate) at codon resolution. By further analysing the outcomes we define the translation initiation and elongation efficiencies (TIE and TEE), highlight the strong context dependence of elongation rates, and identify problematic transcript regions in which Ribo-seq are not consistent with the model.

part of the density profile or, in other words, there is a complex non-linear relation between the set of rates and the number of reads of each codon. Finally, from ribosome profiling data we have only access to measurements corresponding to the number of reads for each one of the $L-1$ codons in the coding sequence (excluding thus initiation). However, counting initiation, there are L rates to fix for each profile, meaning that it is not possible to estimate *absolute* rates without integrating more information. We emphasise this non-trivial point in the following section.

2.2 Ribosome profiles alone cannot estimate absolute elongation rates

We first want to show that ribosome density profiles alone are not sufficient to estimate absolute elongation rates at codon resolution. In the Supplementary Material we show that it is only possible to extract codon elongation rates *relative to the initiation rate of the gene of interest*. In essence, it is possible to write equations for the ribosome density at codon i involving all other densities, the initiation rate α, and the elongation/termination rates k_i ($i = 2, \ldots, L$). From these equations we observe that the ribosome density at codon i depends only on the ratio between the elongation/termination rates $\{k_i\}$ and the initiation rate α. This is expected since ribosome density at codon i measures the probability to find the ribosome at that particular codon and thus bears no information on the absolute timescales.

In other words and without further assumptions, for a gene n it will only be possible to estimate the elongation-to-initiation ratio $\kappa_i^{(n)} := k_i^{(n)}/\alpha^{(n)}$, where $k_i^{(n)}$ is the elongation rate of codon i of the gene n, and $\alpha^{(n)}$ is the initiation rate of the same gene. However, since the initiation rate is severely gene-dependent, it will not be possible to compare the absolute elongation rates $k_i^{(n)}$ from different genes without the knowledge of $\alpha^{(n)}$ of each gene.

Despite that, many works in the literature fix a unique timescale shared by all mRNAs (for instance the average ribosome speed of 5.6 aa/s in yeast (Ingolia et al., 2011) to compute absolute elongation rates. This typical timescale could however be highly variable from transcript to transcript, which in turn would introduce a bias when obtaining absolute rates and when comparing the outcome of different genes.

To make this point clear, we show a practical example by means of a simulation mimicking the outcome of an *in silico* Ribo-seq experiment. We fix the coding sequence or, being more precise, we fix the elongation speed profile $\{k_i^{(n)}\}$ of the isolated ribosomes along the coding sequence. We then simulate the dynamics of ribosomes according to the exclusion process at several ribosome recruitment rates (initiation) $\alpha^{(n)}$'s. Unsurprisingly, this leads to different profiles shown in Fig. 2(a).

On the other hand, identical density profiles may emerge from different sets of absolute elongation rates $k_i^{(n)}$, but with the same elongation-to-initiation profile $\{\kappa_i^{(n)}\} = \{k_i^{(n)}/\alpha^{(n)}\}$. We thus emphasise that ribosome profiles depend on the elongation-to-initiation ratios $\{\kappa_i\}$ only, and thus that indistinguishable density profiles can be obtained with different absolute elongation rates $\{k_i\}$. This is evident from the analytical solution of the exclusion process (and the qualitative argument given above). To further demonstrate this point, in Fig. 2(b) we show the outcome of simulations of translation having different absolute rates $\{k_i\}$ but same relative speed profile $\{\kappa_i\}$.

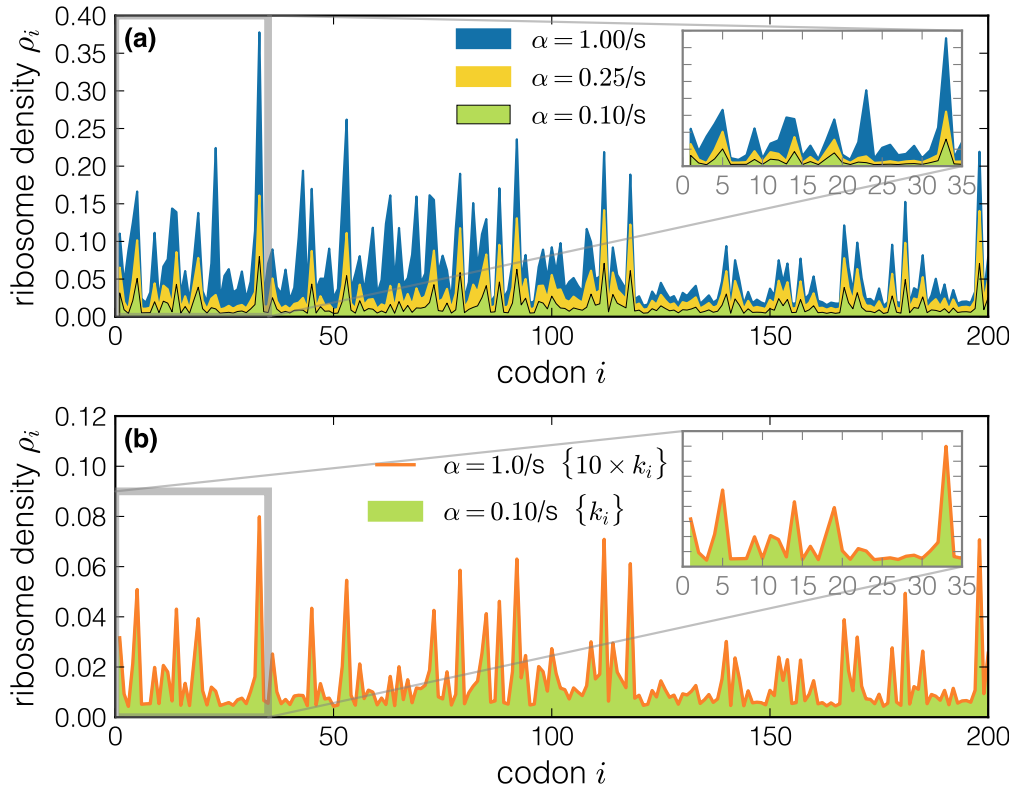

Figure 2: Density profiles obtained by stochastic simulations. In panel (a) we fix the speed profile $\{k_i\}$ of the *in silico* transcript for three different values of the initiation rate α. As expected, by increasing the initiation rate we obtain different profiles with increasing density and traffic effects. The green (filled) profile in panel (b) corresponds to the density obtained with the same speed profile of panel (a) for $\alpha = 0.1$/s. The yellow line corresponds to the profile of a transcript simulated with a tenfold larger initiation rate, but keeping $\{\kappa_i\}$ constant (i.e. also increasing the elongation rates by a factor 10). This shows that densities obtained with the same elongation-to-initiation ratios $\{\kappa_i\}$ are indistinguishable.

The fact that Ribo-seq profiles are determined by the ratio between elongation and initiation rates, see Fig. 2(b), rather than absolute elongation rates alone is crucial to interpret Ribo-seq data, and this is an issue so far neglected in the literature.

For the sake of simplicity, in the remaining of the text we drop the index (n) of the gene in all the equations.

2.3 Non-Equilibrium Analysis of Ribo-seq (NEAR)

In order to improve existing estimates of translation determinants we develop NEAR, a Non-Equilibrium Analysis of Ribo-seq data. It consists of a model-driven interpretation of ribosome density profiles based on the exclusion process. Our method is able to formally extract information on

the determinants of elongation at codon resolution from experimental ribosome footprints.

NEAR infers the relative elongation profiles $\{\kappa_i\}$ with an optimisation procedure that aims to find a theoretical density profile $\{\rho_i\}$ which is the closest as possible to the experimental one $\{r_i\}$, and which is predicted by the exclusion process explained in the Model Section and further in the Supplementary Information. To this end we developed an *inverse method* that returns the elongation-to-initiation ratio κ_i of each codon i from the knowledge of the experimental ribosome density at each codon. This has been possible since we have recently found a mathematical expression for the ribosome density profile in terms of translation initiation, elongation and termination rates. This expression was obtained under the assumption of a limiting initiation rate α (Szavits-Nossan et al., 2018a,b), which is supposed to hold for most of the mRNAs in physiological conditions. The mathematical details of the method are summarised in the Supplementary Information.

In Fig. 1 we outline the procedure: single-gene experimental ribosome profiles are integrated with the stochastic model of translation and, after optimising the parameters of the model, NEAR finds the set of elongation rates (relative to the initiation of the gene of interest) that generate a theoretical density profile which is the closest as possible to the one obtained experimentally. As a proof of concept, we apply our method to extract information from the the dataset experimentally obtained by Weinberg et al. (2016) and mapped to A-site positions by Dao Duc et al. (2018) (yeast). In more detail:

1. We first normalise the number of footprint reads on each codon by the total number of reads. This number is then multiplied by the absolute ribosome density for that particular gene obtained from sucrose gradient experiment. The end result is a normalised ribosome density profile $\{r_i\}$ that reveals how likely is to find a ribosome at codon i. When applying NEAR in yeast, we normalise the Ribo-seq profiles from Dao Duc et al. (2018) using absolute (gene-dependent) ribosome densities found in MacKay et al. (2004).

2. Next we run the optimisation of $\{\kappa_i\}$ based on our theory (details in the Supplementary Information).

3. We then compute the density profile $\{\rho_i^{\mathrm{sim}}\}$ from stochastic simulations based on the estimated $\{\kappa_i\}$.

4. We perform a quality check on each κ_i, and verify that:

   (a) the theoretical approximation holds by comparing analytical density profiles and simulated ones. This step is necessary because our solution of the model is approximate, see Refs. (Szavits-Nossan et al., 2018a,b) and also the Supplementary Information.

   (b) the estimated κ_i reproduces the experimental density r_i (within 5% tolerance).

These are the main steps of the proposed method, and we provide further mathematical details in the Supplementary Information. We stress that NEAR is also able to reject codons whose κ_i cannot be trusted, and identify why the inference of elongation rates for those codons is problematic (quality check - point 4). We are in fact able to say whether our theoretical solution is satisfied or not, or if the problem is due to the model being inconsistent with the experimental data. This particular point will be developed later on in this work.

As an example, we tested NEAR on a mock sequence $\{k_i\}$, and show that it can accurately predicts the elongation rates of the original sequence; we also remark that the quality check allows to push the analysis to relatively high elongation rates (Figure S3).

2.3.1 Using NEAR to study translation of individual genes

Before applying our method genome-wide to investigate the elongation-to-initiation ratio of each codon type, we first show that its implementation successfully works at the individual gene. We describe NEAR on a concrete example, and we randomly choose a gene (here YLR301W) to highlight differences with previous methods and to point out the excellence of the results obtained.

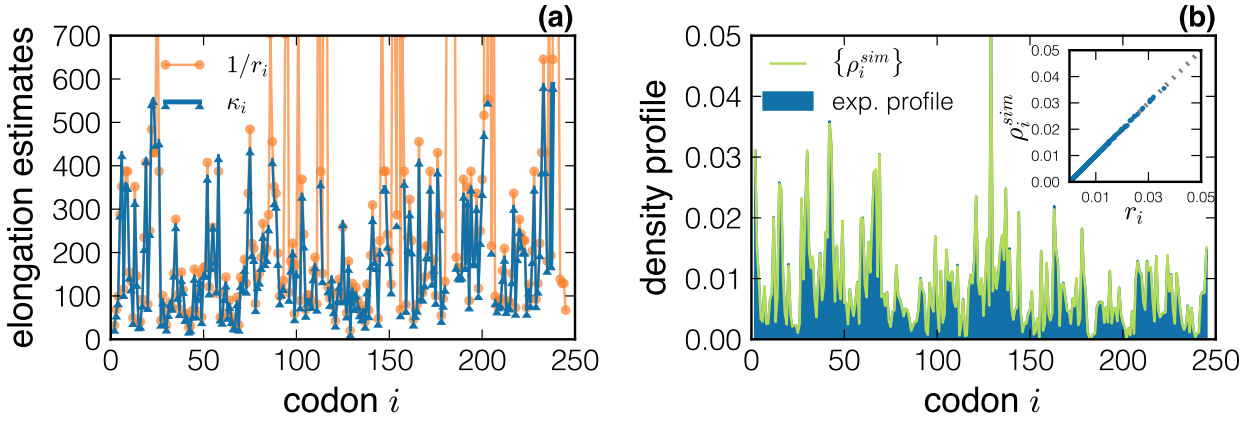

Figure 3: As an example we apply NEAR to the YLR301W gene. (a) The NEAR profile $\{\kappa_i\}$ is plotted (blue line, triangle markers) as a function of the codon position i, and compared to the naive estimates $\{1/r_i\}$ (orange line, round marker). In panel (b) we compare the theoretical density profile obtained using the inferred $\{\kappa_i\}$ (green line) with the experimental normalised profile $\{r_i\}$. The inset shows the scatter plot between the two densities (for each codon i) demonstrating an excellent agreement between theory and experiments.

We first compute the normalised experimental density profile $\{r_i\}$ using the experimental total density r from MacKay et al. (2004) (in units of ribosomes/codon). This profile is then analysed following the method explained in the previous section. A set of elongation-to-initiation ratios $\{\kappa_i\}$ is drawn by optimising the match between the theoretical density profile and the experimental one. Each value κ_i is then examined to see whether it provides a good prediction for that particular codon and to check for inconsistencies in the method as previously explained. As a result, a few values of κ_i that do not pass this quality check are rejected and are not included in the final analysis. In Fig. 3(a) we plot the final $\{\kappa_i\}$ profile (blue line, triangle markers) obtained for a randomly selected yeast gene (YLR301W). In the same figure we plot the naive estimate $1/r_i$ for each codon i (orange line, round marker) that ignores ribosome interference but it is usually judged as a good estimator of the elongation rate. As it is evident from the figure, the two profiles $\{\kappa_i\}$ and $\{1/r_i\}$, although they share a common trend, quantitatively differ from each other. Moreover, we are able to identify controversial values of κ_i that we do not consider as reliable, while this cannot be done when using

$\{1/r_i\}$ as a proxy for elongation determinants.

The result of a stochastic simulation of ribosome dynamics performed with the first set of optimised elongation weights $\{\kappa_i\}$ is shown in Fig. 3(b). The agreement between the simulated density profile $\{\rho_i^{\mathrm{sim}}\}$ (green line) and the experimental one (in blue) is excellent for most of the genes. The inset shows the scatter plot between the values (for each codon) of the simulated and experimental ribosome density.

2.4 Estimate of elongation-to-initiation ratios at codon resolution in yeast

We gather the NEAR elongation-to-initiation ratios $\{\kappa_i\}$ for each gene and compute their distribution for each codon type. We thus aim to understand if each codon type has a characteristic elongation timescale and verify or reject the well accepted hypothesis that elongation rates are determined by the availability of aminoacyl-tRNAs, or if elongation is context dependent. However, since κ_i is by definition the ratio k_i/α between the elongation rate of codon i and the initiation rate α of the gene, we cannot easily compare elongation-to-initiation ratios of the same codon types from different genes. The distribution of all κ belonging to a given codon type shows large variations (see Supplementary Figure S4) and we are not able to determine if the source of the broad distribution resides in the elongation rates of the codon or if it is generated by the broad distribution of the initiation rates.

However, we observe that STOP codons are the codons with less variability of the elongation-to-initiation ratio κ (Supplementary Figure S3). This result has been previously found by Dao Duc and Song (2018), and it is consistent with the expectation of a context-independent termination. Thanks to this observation, we then compute the elongation-to-termination ratio $\kappa_i/\kappa_L = k_i/\beta$, i.e. the elongation rate of codon i with respect to the termination of the gene under investigation. This quantity does not depend on the initiation rate α that is supposed to be context-dependent and largely varies from gene to gene. We can thus better compare elongation of the same codon type extracted from different genes.

The result of the NEAR analysis of *Saccharomyces cerevisiae* ribosome profiles from Weinberg et al. (2016) is shown in Fig. 4. In Fig. 4(a) we plot the boxplot of the distribution of the elongation-to-termination ratios κ/κ_L for each codon type. We find a large variability in the median values of κ/κ_L for different codon types. In particular there is a $\approx$5-fold difference in the median value of κ/κ_L between the slowest (CGG) and the fastest (GUU) codon, in accordance with a common view that translation elongation is a non-uniform process. In Fig. 4(b) we compared median values of κ/κ_L against two common measures of tRNA availability, a codon-dependent rate of translation based on the tRNA gene copy number (GCN) from Ciandrini et al. (2013) and the tRNA adaptation index (tAI) (Reis et al., 2004). We find a moderate correlation between the median of the κ/κ_L distributions and both measures of tRNA availability ($k(GCN)$: Spearman $\rho = 0.53$, $p = 1.4 \times 10^{-5}$; tAI: $\rho = 0.46$, $p = 3 \times 10^{-4}$).

In our dataset we also find a large variability in the values of κ/κ_L belonging to the same codon type. This variability, together with a moderate correlation between κ/κ_L and tRNA availability, suggests that the elongation speed of individual codons is only partially determined by their codon type.

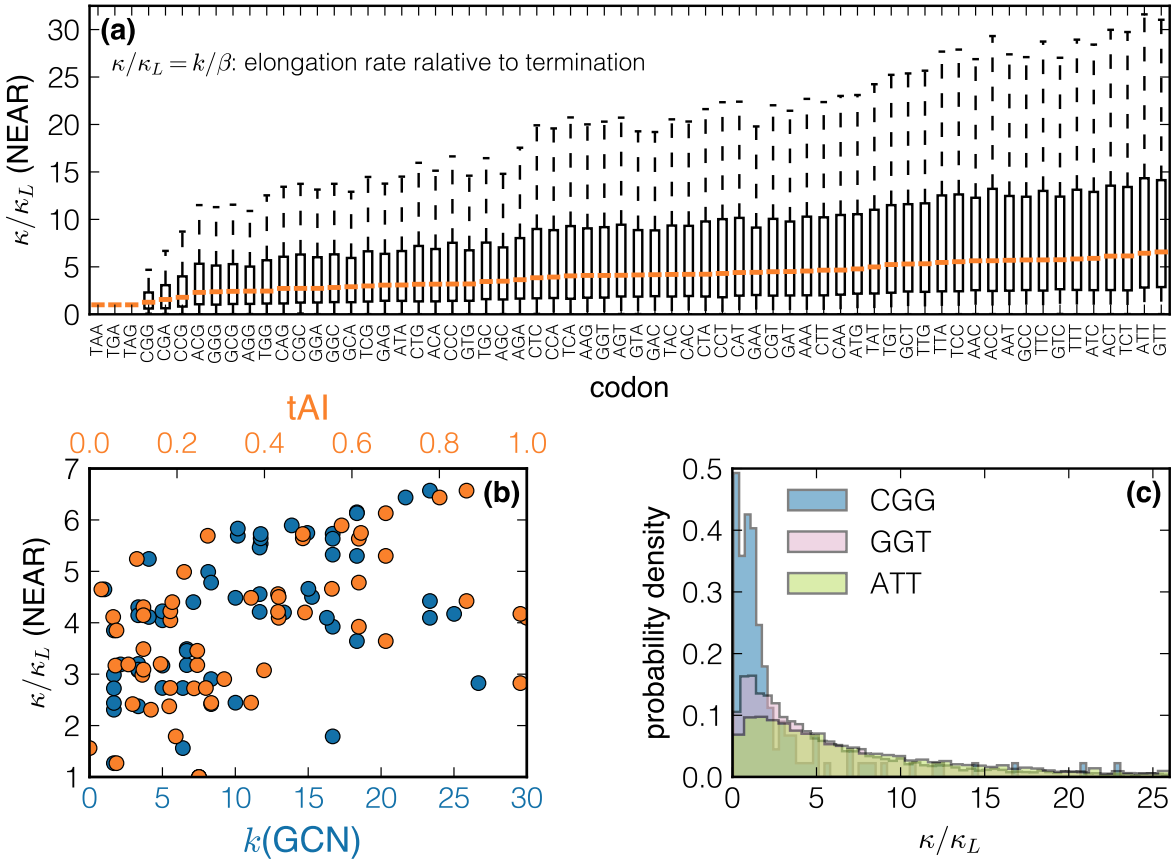

Figure 4: Result of the NEAR procedure. Panel (a) shows the distributions of the estimated κ_i/κ_L passing the quality check for all the codon types (x-axis). The boxplots show the quartiles $Q1$ and $Q3$ ($\Delta Q = Q3 - Q1$) of the distribution, the median is the orange horizontal line; the whiskers extend to the most extreme, non-outlier data points (i.e. to the last data point less than $Q3 + 1.5\Delta Q$ and greater than $Q1 - 1.5\Delta Q$). The comparison between the NEAR estimates (median values of the κ/κ_L distributions) and $k(GCN)$ (blue) or the tAI (orange) for each codon type is shown in panel (b). In panel (c) we show the distribution of the κ_i/κ_L for three different codon types (slow-medium-high median).

Furthermore we find that the degree of variability in the elongation speed differs among different codon types. Typically we find a narrow distribution of κ/κ_L for slow codons such as CGG and a wide distribution for fast codons such as AUU, see Fig. 4(c).

A distinctive feature of the exclusion process is that it takes into account ribosome interference that may increase the time ribosome spends on a particular codon. If the ribosome movement was never hindered by other ribosomes, we would expect to see $\kappa_i \approx 1/r_i$ corresponding to the usual estimate of the rates (Li, 2015). However, as it is already clear from the example shown in Fig. 3, the outcomes of NEAR deviate from the naive estimates. In Fig. 5(a) we plot the scatter plot of κ_i versus the corresponding inverse ribosome densities $1/r_i$ of each codon analysed. We find that there are

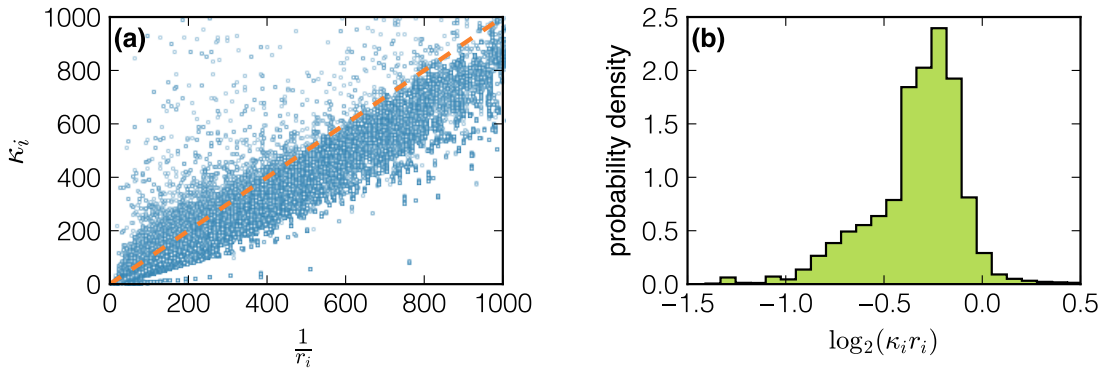

Figure 5: Elongation-to-termination ratio κ_i compared to the inverse ribosome density $1/r_i$ of each codon analysed using NEAR (a). Dashed line corresponds to the naive estimate $\kappa_i = 1/r_i$ which is expected to hold if ribosome movement is not hindered by other ribosomes on the same mRNA. In panel (b) we show the $\log_2$ of the fold change $\kappa_i/(1/r_i)$ between the NEAR and the naive estimate.

many inferred κ_i that are not accurately described by the inverse ribosome density and, as it can be seen in Fig. 5(a), in extreme cases κ_i could be 50% smaller than the corresponding value of $1/r_i$. In panel (b) we also plot the $\log_2$ of the fold change between the NEAR estimate κ_i and $1/r_i$.

Our findings in Fig. 5 thus suggest that the effect of ribosome interference is not negligible. In order to quantify its impact more precisely, we devise two new measures of translation efficiency: the translation initiation efficiency (TIE) and the translation elongation efficiency (TEE), which we describe in detail below.

2.5 Translation Initiation Efficiency (TIE)

From the simulation of translation with the inferred rates we can measure the ribosomal current J, and extract J/α which is a quantity dependent on $\{\kappa_i\}$ only. The current J can be used as a proxy for protein synthesis rate (at the level of translation).

In the biological literature it is often assumed that $J = \alpha$, thus identifying translation initiation with protein synthesis rate. However this is true only if initiation is much slower than elongation so that substantially only one ribosome is translating a transcript at a given time. Yet, this approximation is too crude to quantitatively describe translation (Szavits-Nossan et al., 2018a). Instead, when more than one ribosome is engaged in translation, J becomes a function of α and the elongation rates $\{k_i\}$; the current J can be thought of as the "bare" initiation rate α multiplied by the probability that the first codons of the mRNA are not occupied by another ribosome (which would otherwise obstruct initiation).

Therefore we propose to use J/α as a measure of the Translation Initiation Efficiency (TIE), which takes values between 0 and 1. The TIE would be equal to 1 in the optimal case in which initiation is not hindered by ribosome traffic ($J = \alpha$, hence TIE = 1). Otherwise, the TIE gives the probability that the first codons, potentially interfering with ribosome recruitment and initiation, are unoccupied. A TIE smaller than 0.5 means that more than half of the times a new ribosome tries to initiate, it fails because of another ribosome whose A-site is located within the first ten codons. In

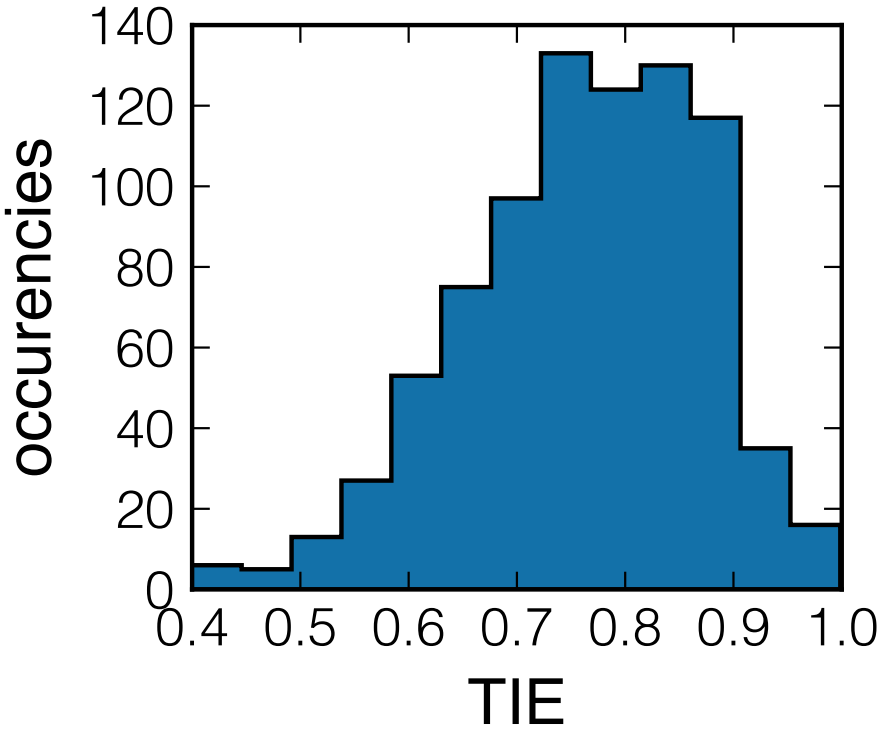

Figure 6: Histogram of estimated Translation Initiation Efficiency (TIE) for all *Saccharomyces cerevisiae* genes included in our study. The TIE gives the probability that the first codons are unoccupied.

Fig. 6 we plotted the histogram of TIE for all the genes included in our study. We find that almost all genes show TIE $>$ 0.5 with a median and mean values around 0.77 and 0.76, respectively. These values suggest that the first codons are mainly free from ribosomes already engaged in translation. Our previous theoretical work on the exclusion process showed that if translation is rate-limited by initiation, then TIE predominately depends on the κ_i of the first $\ell \approx 10$ codons, which is the ribosome footprint on the mRNA (Szavits-Nossan et al., 2018a). Based on that prediction, TIE $>$ 0.5 indicates that the first codons of *Saccharomyces cerevisiae* genes have been selected for fast elongation (relative to initiation) in order to maximise protein production rate.

2.6 *Translation Elongation Efficiency (TEE) shows congestion of ribosomes in vivo*

Analogously to the TIE, we now define an efficiency index for translation elongation aiming to identify local ribosome interference along the transcript and not only around the initiation region. We remind that the ribosomal current is given by $J = k_i \rho_i P(\ell \text{ free codons } | i)$, where $P(\ell \text{ free codons } | i)$ is the probability that the $i+1 \dots i+\ell$ codons are not occupied, given that a ribosome's A-site is at site i. We name this probability Translation Elongation Efficiency (TEE), which is a measure of the local mRNA congestion seen by a ribosome translating the codon i. When the ribosome does not

encounter interference then TEE=1, otherwise $0 <$ TEE < 1. In the extreme case of a transcript completely jammed with ribosome one would get TEE ≈ 0, i.e. the probability that a ribosome advances is negligible. Based on the definition of the TIE given in the previous section, we estimate the elongation efficiency as $\text{TEE}_i = \text{TIE}/(\kappa_i \rho_i)$. More details are given in the Supplementary Information. We also stress that the TEE is a function of $\{\kappa_i\}$ only, meaning that the balance between initiation and elongation rates is what determines ribosome interference.

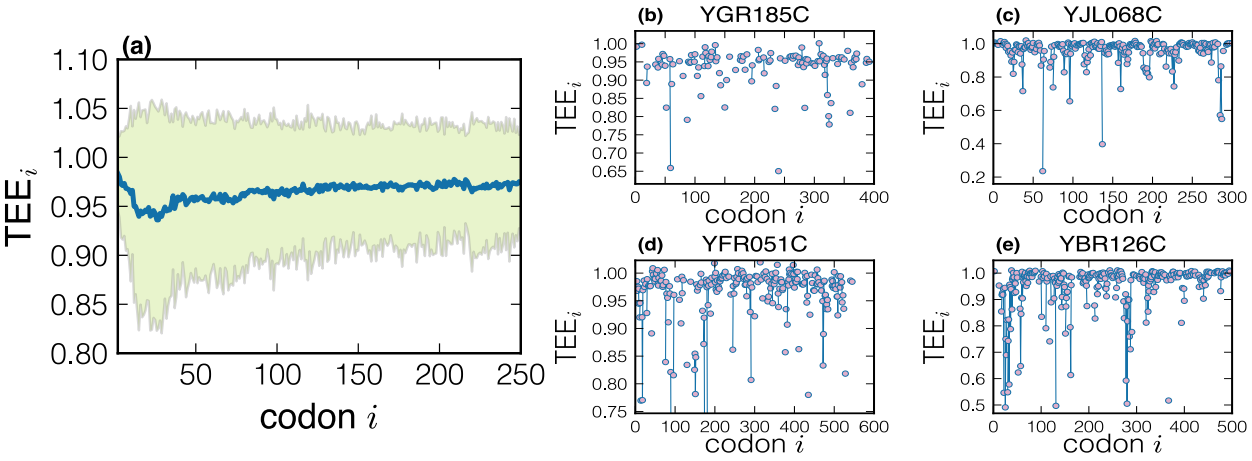

Figure 7: (a) Translation Elongation Efficiency profile averaged among the entire set of genes included in our study. The full line is the averaged TEE profile, while the shadow (green) area corresponds to the region included between the averaged TEE profile and $\pm$ the standard deviation of the distribution. Panels (b)-(c)-(d)-(e) are single-gene TEE profiles.

We first evaluate the TEE profiles of all the yeast genes previously analysed, and then compute an averaged metagene TEE profile by aligning the genes at their START codon; the result is plotted in Figure 7(a). This genome-averaged profile shows that TEE is close to its optimum value, and that the first 10 sites seem to have a larger elongation efficiency. This result is consistent with the large TIE previously observed, and it confirms the importance of the beginning of the coding sequence in controlling translation.

From Fig. 7(a), since TEE is close to 1 one could conclude that ribosome interference is not relevant in translation. However, as it can be remarked from the standard deviation of its distribution, the TEE for each fixed codon position i varies substantially. When looking more carefully at the profiles of individual genes, although the TEE is typically close to 1 indicating that traffic is negligible for most of codons, we are able to identify codons where ribosome interference is significant, with ribosome congestion (defined as $1 - TEE$, i.e. the probability of finding a ribosome obstructing the movement) reaching values close to 50%. In order to show that, we randomly selected four genes of the analysed dataset and plotted the TEE profile in Fig. 7(b)-(c)-(d)-(e). We thus stress that the method developed can locate, at codon resolution and excluding unreliable estimates, particular regions on the transcript that are affected by ribosome interference.

3 Discussion

In this work we introduce NEAR, a Non-Equilibrium Analysis of Ribo-seq, which is based on a well-studied model borrowed from statistical physics. The model tracks individual ribosomes engaged in translation and predicts their spatial distribution on the mRNA using initiation, elongation and termination rates as input parameters. Normally one has to makes certain assumptions in order to estimate the input parameters. Here we do the opposite: we develop a method that infers elongation-to-initiation ratios at codon resolution directly from ribosome profiling experiments. Thanks to this method we can inspect ribosome profiling data to reveal "slow" elongation regions of each transcript, quantify ribosome interference and the interplay between initiation and elongation with unprecedented detail.

We first emphasise that Ribo-seq profiles, being an averaged snapshot of the translatome, do not contain information on the timescale of the process, and that thus it is possible to estimate relative rates only. NEAR infers the elongation-to-initiation ratios κ_i, which are verified and reproduce the experimental data. We can then discard problematic estimates that are not consistent with the model nor the experiments.

A source of inconsistency between the model and the data is possibly hidden in the nature of the ribosome profiling technique. Queuing ribosomes generate large footprints (Guydosh and Green, 2014) that are usually discarded in the experimental pipeline. Thus, one would expect ribosome profiling not to be sensitive to ribosome interference. However, NEAR finds evidence of local jamming despite the experimental bias that discards jammed ribosomes. We first show this by demonstrating a difference between the estimated elongation-to-initiation ratios and the naive estimates $1/r_i$ computed by neglecting ribosome interference (see Fig. 5). Additionally, the Translation Elongation Efficiency index TEE better highlights the codons at which ribosomes are dwelling because their movement is hindered by the presence of downstream ribosomes. When analysing the TEE of individual profiles we observe that the probability of finding ribosome interference can go as up as $\sim 50\%$. The TEE profile is however generally flat with the value of TEE close to 1, which suggests that along most of the transcript the ribosome traffic is negligible.

When averaging the TEE profiles among all the genes we observe slightly higher values at the beginning of the coding sequence. This could mean that elongation is optimised in this region to avoid queuing and allow faster ribosome recruitment at the start codon. This result is consistent with two recent studies that stressed the importance of the first 10 codons in regulating protein production (Chu et al., 2014; Szavits-Nossan et al., 2018a). Furthermore, the first codons are also experimentally recognised to be critical in determining protein synthesis (Kudla et al., 2009; Bentele et al., 2014; Cambray et al., 2018; Verma et al., 2019). Finally, this result is also consistent with the "slow ramp" hypothesis (Tuller et al., 2010): a ribosome translating the first (slower) codons will be less prone to interfere with other ribosomes downstream the sequence. By averaging the elongation-to-initiation profiles of different genes we also observe signatures of a bottleneck at the beginning of the transcript (see Supplementary Fig. S5).

These are all signatures that translation elongation is largely optimised although one can locally observe high levels of ribosome interference. Having a higher TEE at the beginning of the transcript is also a way to optimise the Translation Initiation Efficiency (TIE), which we define as the probability that a ribosome attempting to initiate translation is not obstructed by another ribosome on the mRNA.

Figure 6 clearly shows that the coding sequence potentially affecting ribosome recruitment is most likely under evolutionary pressure (Tuller et al., 2010)

Despite all the advances in the past ten years, we still do not have enough knowledge on the translation process to formally identify all its sequence determinants. In this work we present a method for interpreting ribosome profiling data and show its strengths when applied to a particular dataset. We can affirm that there is a correlation between used indices of codon optimality, such as the local tAI, and the estimated elongation-to-initiation ratios. However, the variability of the estimated rates of each individual codon type is such that using those indices for protein synthesis optimisation or other synthetic applications will probably not lead to the expected results. Instead, our findings indicate that codon context in the sequence seems to be relevant more than the particular codon used, and further studies should focus on the discovery of mechanisms giving rise to the codon context dependence. For instance secondary structures might be relevant, particularly around the initiation region (Kudla et al., 2009; Bentele et al., 2014; Cambray et al., 2018) or the amino-acid charge at the beginning of the coding sequence (Dao Duc et al., 2018). What we can formally state so far is that elongation is largely optimised (with respect to initiation) for endogenous sequences; since we observe a TEE close to 1, elongation rates seem to be selected to avoid ribosome interference. This also holds for the initiation step, for which we find TIE > 0.7. Instead of finding a separate mechanism for efficient initiation and elongation, we highlight the importance of the balance and the relative role of these two stages of translation. Consequently, we believe that ordinary measures of translation efficiency cannot capture in detail the complexity of this biological process.

Despite the fact that our results show signatures of efficient initiation and elongation, we find evidence that local congestion happens *in vivo* in the genes we analysed (yeast). Moreover, we stress once more that Ribo-seq are intrinsically biased towards genes that do not present ribosome clustering and, consistently, small TEE can be observed in genes where many estimates are not considered reliable and then discarded (see Supplementary Fig. S6). Future developments of NEAR will include a formal study of this bias. Another key question that quantitative studies using ribosome profiling should address in the future is the role of density normalisation in order to better compare the outcome of different genes and of different organisms.

4 Methods and Materials

4.1 Notations

In this section we summarise the notations used in the paper. The main symbols for densities, rates, and rates relative to initiation are given in Table 1. When the quantity is codon specific we use the suffix $i = 2, \ldots, L$ to identify the codon number (the first codon after the START codon is at $i = 2$, the STOP codon is at $i = L$). Brackets $\{\cdot\}$ indicate a set of values: for instance $\{a_i\}$ is the set of all the values a_i for $i = 2, \ldots, L$.

We emphasise that we use *normalised* densities, in units of ribosomes (A-sites) per codon. The total density is thus the averaged ribosome profile $r = \sum_{i=2}^{L} r_i/(L-1)$, and the number of ribosomes translating an mRNA is $N = r(L-1)$.

Symbol	Meaning
L	length of the mRNA (in codons, including START)
ℓ	length of the ribosome (in codons)
α	initiation rate [s^{-1}]
k_i	elongation rate [s^{-1}] of codon i, with $i = 2,\ldots,L-1$
k_L (or β)	termination rate [s^{-1}]
$\{k_i\}$	speed profile (elongation) of a given transcript
$\kappa_i = k_i/\alpha$	relative (to initiation) elongation rate of codon i, or *elongation-to-initiation ratio*
$\{\kappa_i\}$	relative (to initiation) elongation profile
r_i	experimental (normalised) density of codon i
$\{r_i\}$	experimental (normalised) density profile
$r = \sum_{i=2}^{L} r_i/(L-1)$	mean density of a given gene
ρ_i	theoretical (normalised) density of codon i
$\{\rho_i\}$	theoretical (normalised) density profile
ρ_i^{sim}	simulated (normalised) density of codon i
$\{\rho_i^{\mathrm{sim}}\}$	simulated (normalised) density profile

Table 1: Summary of the symbols used and their meaning.

4.2 The Exclusion Process

We model translation by a stochastic process called the totally asymmetric simple exclusion process (TASEP) introduced by MacDonald et al. (1968); MacDonald and Gibbs (1969). Messenger RNAs are represented by a one-dimensional lattice consisting of L codons labelled by $1,\ldots,L$, whereby each ribosome takes $\ell = 10$ lattice sites. Ribosomes on the lattice are tracked according to the position of their A-site. A codon i that is occupied by the A-site of the ribosome is labelled by A_i and is otherwise labelled by $\emptyset_i$. The model accounts for translation initiation, elongation and termination of individual ribosomes described by the following first-order kinetic reactions:

$$\text{initiation:} \quad \emptyset_2\emptyset_3\ldots\emptyset_{\ell+1} \xrightarrow{\alpha} A_2\emptyset_3\ldots\emptyset_{\ell+1} \tag{1}$$

$$\text{elongation:} \quad A_i\emptyset_{i+1}\ldots\emptyset_{i+\ell+1} \xrightarrow{k_i} \emptyset_i A_{i+1}\emptyset_{i+2}\ldots\emptyset_{i+\ell+1} \tag{2}$$

$$\text{termination:} \quad A_L \xrightarrow{k_L} \emptyset_L. \tag{3}$$

The process is described by the probability density $P(C,t)$ to find a configuration C of ribosomes on an mRNA at a particular time t. By *configuration* we mean a particular arrangement of ribosomes described by the positions $\{A_i\}$ of their A-sites. The time evolution of $P(C,t)$ is governed by the master equation

$$\frac{\mathrm{d}P(C,t)}{\mathrm{d}t} = \sum_{C'\neq C} W(C'\to C)P(C') - \sum_{C'\neq C} W(C\to C')P(C), \tag{4}$$

where $W(C \to C')$ is the rate of transition from C to C'. We assume that translation takes place in the stationary limit so that $\mathrm{d}P(C,t)/\mathrm{d}t = 0$ in which case Eq. (4) becomes a system of linear equations,

$$0 = \sum_{C' \neq C} W(C' \to C)P(C') - \sum_{C' \neq C} W(C \to C')P(C). \tag{5}$$

The three main quantities of interest are the rate of translation J, which measures the amount of proteins produced per unit time, the local ribosome densities ρ_i, which measure the probability of detecting a ribosome at codon i, and total ribosome density ρ, which measures the average number of ribosomes per unit length of the transcript. In the stationary TASEP J, ρ_i and ρ are defined as

$$J = k_L \langle \tau_L \rangle \tag{6}$$

$$\rho_i = \langle \tau_i \rangle \tag{7}$$

$$\rho = \frac{1}{L-1} \sum_{i=2}^{L} \rho_i, \tag{8}$$

where averaging is performed with respect to the steady-state probability $P(C)$ and τ_i is an occupation number whose value is equal to 1 if codon i is occupied by the A-site of the ribosome and is 0 otherwise. Computing these quantities requires an exact knowledge of $P(C)$, which is known only in the biologically unrealistic case of $\ell = 1$ (Derrida et al., 1993). Instead, we compute J, ρ_i and ρ using two approximation methods: the mean-field approximation developed in MacDonald et al. (1968); MacDonald and Gibbs (1969) and initiation-limited approximation developed in Szavits-Nossan et al. (2018a,b).

4.3 Dataset used

We analysed the reads of the mapped A-site positions from Dao Duc et al. (2018), which are based on the experimental results by Weinberg et al. (2016) in *Saccharomyces cerevisiae* (flash freeze). Our method successfully optimised 838 of the total 839 genes from the dataset of Dao Duc et al. (2018) for which the experimental ribosome density necessary for normalisation was known from MacKay et al. (2004) (for one gene out of 839 genes the normalisation was not possible because it resulted in ribosome density larger than 1). In Supplementary Table S1 we summarise the features extracted from the data by NEAR, and the results for each individual gene at the codon detail can be found in the Supplementary files.

5 Acknowledgments

LC would like to thanks G. Cambray for useful discussions. JSN was supported by the Leverhulme Trust Early Career Fellowship under grant number ECF-2016-768. This project has been initially supported by the Défi InPhyNiTi (exploratory project funded by the CNRS).

References

Ahmed N, Sormanni P, Ciryam P, Vendruscolo M, Dobson CM, OBrien EP. Identifying A- and P-site locations on ribosome-protected mRNA fragments using Integer Programming. Scientific Reports 2019 9:1. 2019 4; 9(1):6256. 10.1038/s41598-019-42348-xdoi: 10.1038/s41598-019-42348-x.

Bartholomäus A, Del Campo C, Ignatova Z. Mapping the non-standardized biases of ribosome profiling. Biological Chemistry. 2016 1; 397(1):23–35. 10.1515/hsz-2015-0197doi: 10.1515/hsz-2015-0197.

Bentele K, Saffert P, Rauscher R, Ignatova Z, Bluthgen N. Efficient translation initiation dictates codon usage at gene start. Molecular Systems Biology. 2014 4; 9(1):675–675. 10.1038/msb.2013.32doi: 10.1038/msb.2013.32.

Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nature Reviews Molecular Cell Biology. 2015 11; 16(11):651–664. 10.1038/nrm4069doi: 10.1038/nrm4069.

Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U. Detecting actively translated open reading frames in ribosome profiling data. Nature Methods. 2016 2; 13(2):165–170. 10.1038/nmeth.3688doi: 10.1038/nmeth.3688.

Cambray G, Guimaraes JC, Arkin AP. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. Nature Biotechnology. 2018 11; 36(10):1005–1015. 10.1038/nbt.4238doi: 10.1038/nbt.4238.

Cholewa-Waclaw J, Shah R, Webb S, Chhatbar K, Ramsahoye B, Pusch O, Yu M, Greulich P, Waclaw B, Bird AP. Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. Proceedings of the National Academy of Sciences of the United States of America. 2019 7; p. 201903549. 10.1073/pnas.1903549116doi: 10.1073/pnas.1903549116.

Chu D, Kazana E, Bellanger N, Singh T, Tuite MF, Von Der Haar T. Translation elongation can control translation initiation on eukaryotic mRNAs. EMBO Journal. 2014; 10.1002/embj.201385651doi: 10.1002/embj.201385651.

Ciandrini L, Stansfield I, Romano MC. Ribosome Traffic on mRNAs Maps to Gene Ontology: Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation. PLoS Comput Biol. 2013 1; 9(1).

Dao Duc K, Saleem ZH, Song YS. Theoretical analysis of the distribution of isolated particles in totally asymmetric exclusion processes: Application to mRNA translation rate estimation. Physical Review E. 2018; 97(1). 10.1103/PhysRevE.97.012106doi: 10.1103/PhysRevE.97.012106.

Dao Duc K, Song YS. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. PLoS Genetics. 2018; 14(1):1–32. 10.1371/journal.pgen.1007166doi: 10.1371/journal.pgen.1007166.

Derrida B, Evans MR, Hakim V, Pasquier V. Exact solution of a 1D asymmetric exclusion model using a matrix formulation. Journal of Physics A: Mathematical and General. 1993 4; 26(7):1493–1517. 10.1088/0305-4470/26/7/011doi: 10.1088/0305-4470/26/7/011.

Diament A, Feldman A, Schochet E, Kupiec M, Arava Y, Tuller T. The extent of ribosome queuing in budding yeast. PLOS Computational Biology. 2018 1; 14(1):e1005951. 10.1371/journal.pcbi.1005951doi: 10.1371/journal.pcbi.1005951.

Duncan CDS, Mata J. Effects of cycloheximide on the interpretation of ribosome profiling experiments in Schizosaccharomyces pombe. Scientific Reports. 2017 12; 7(1):10331. 10.1038/s41598-017-10650-1doi: 10.1038/s41598-017-10650-1.

Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife. 2013 12; 2. 10.7554/eLife.01179doi: 10.7554/eLife.01179.

Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome research. 2012 11; 22(11):2208–18. 10.1101/gr.139568.112doi: 10.1101/gr.139568.112.

Gorgoni B, Ciandrini L, McFarland MR, Romano MC, Stansfield I. Identification of the mRNA targets of tRNA-specific regulation using genome-wide simulation of translation. Nucleic Acids Research. 2016 11; 44(19):9231–9244.

Gorochowski TE, Ellis T. Designing efficient translation. Nature Biotechnology. 2018 11; 36(10):934–935. 10.1038/nbt.4257doi: 10.1038/nbt.4257.

Guttman M, Russell P, Ingolia N, Weissman J, Lander E. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell. 2013 7; 154(1):240–251. 10.1016/J.CELL.2013.06.009doi: 10.1016/J.CELL.2013.06.009.

Guydosh N, Green R. Dom34 Rescues Ribosomes in 3' Untranslated Regions. Cell. 2014 2; 156(5):950–962. 10.1016/J.CELL.2014.02.006doi: 10.1016/J.CELL.2014.02.006.

von der Haar T. MATHEMATICAL AND COMPUTATIONAL MODELLING OF RIBOSOMAL MOVEMENT AND PROTEIN SYNTHESIS: AN OVERVIEW. Computational and Structural Biotechnology Journal. 2012 4; 1(1):e201204002. 10.5936/CSBJ.201204002doi: 10.5936/CSBJ.201204002.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science (New York, NY). 2009 4; 324(5924):218–23. 10.1126/science.1168978doi: 10.1126/science.1168978.

Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011; 147(4):789–802. 10.1016/j.cell.2011.10.002doi: 10.1016/j.cell.2011.10.002.

Klumpp S, Hwa T. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. Proceedings of the National Academy of Sciences. 2008 11; 105(47):18159–18164. 10.1073/pnas.0806084105doi: 10.1073/pnas.0806084105.

Kozak M. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Research. 1981 10; 9(20):5233–5252. 10.1093/nar/9.20.5233doi: 10.1093/nar/9.20.5233.

Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science. 2009; 324(5924):255–258. 10.1126/science.1170160doi: 10.1126/science.1170160.

Li GW. How do bacteria tune translation efficiency? Current Opinion in Microbiology. 2015 4; 24:66–71. 10.1016/j.mib.2015.01.001doi: 10.1016/j.mib.2015.01.001.

MacDonald CT, Gibbs JH. Concerning the kinetics of polypeptide synthesis on polyribosomes. Biopolymers. 1969 5; 7(5):707–725. 10.1002/bip.1969.360070508doi: 10.1002/bip.1969.360070508.

MacDonald CT, Gibbs JH, Pipkin AC. Kinetics of biopolymerization on nucleic acid templates. Biopolymers. 1968 1; 6(1):1–25. 10.1002/bip.1968.360060102doi: 10.1002/bip.1968.360060102.

MacKay VL, Li X, Flory MR, Turcott E, Law GL, Serikawa KA, Xu XL, Lee H, Goodlett DR, Aebersold R, Zhao LP, Morris DR. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. Molecular & cellular proteomics : MCP. 2004 5; 3(5):478–89. 10.1074/mcp.M300129-MCP200doi: 10.1074/mcp.M300129-MCP200.

Mitarai N, Sneppen K, Pedersen S. Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. J Mol Bio. 2008 9; 382(1):236–245. 10.1016/J.JMB.2008.06.068doi: 10.1016/J.JMB.2008.06.068.

Reis Md, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Research. 2004 9; 32(17):5036–5044. 10.1093/nar/gkh834doi: 10.1093/nar/gkh834.

Reuveni S, Meilijson I, Kupiec M, Ruppin E, Tuller T. Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model. PLoS Comput Biol. 2011 9; 7(9):e1002127. 10.1371/journal.pcbi.1002127doi: 10.1371/journal.pcbi.1002127.

Riba A, Di Nanni N, Mittal N, Arhné E, Schmidt A, Zavolan M. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. Proceedings of the National Academy of Sciences of the United States of America. 2019 7; p. 201817299. 10.1073/pnas.1817299116doi: 10.1073/pnas.1817299116.

Rudorf S, Lipowsky R. Protein synthesis in E. coli: Dependence of codon-specific elongation on tRNA concentration and codon usage. PLoS ONE. 2015; 10(8):1–22. 10.1371/journal.pone.0134994doi: 10.1371/journal.pone.0134994.

Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-limiting steps in yeast protein translation. Cell. 2013; 153(7):1589–1601. 10.1016/j.cell.2013.05.049doi: 10.1016/j.cell.2013.05.049.

Sharma AK, Sormanni P, Ahmed N, Ciryam P, Friedrich UA, Kramer G, OBrien EP. A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. PLOS Computational Biology. 2019 5; 15(5):e1007070. 10.1371/journal.pcbi.1007070doi: 10.1371/journal.pcbi.1007070.

Steitz JA. Polypeptide Chain Initiation: Nucleotide Sequences of the Three Ribosomal Binding Sites in Bacteriophage R17 RNA. Nature. 1969 12; 224(5223):957–964. 10.1038/224957a0doi: 10.1038/224957a0.

Szavits-Nossan J, Ciandrini L, Romano MC. Deciphering mRNA Sequence Determinants of Protein Production Rate. Phys Rev Lett. 2018; 120:128101.

Szavits-Nossan J, Romano MC, Ciandrini L. Power series solution of the inhomogeneous exclusion process. Physical Review E. 2018; 97(5):1–13. 10.1103/PhysRevE.97.052139doi: 10.1103/PhysRevE.97.052139.

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 2010; 141(2):344–354. 10.1016/j.cell.2010.03.031doi: 10.1016/j.cell.2010.03.031.

Verma M, Choi J, Cottrell KA, Lavagnino Z, Thomas EN, Pavlovic-Djuranovic S, Szczesny P, Piston DW, Zaher H, Puglisi JD, Djuranovic S. Short translational ramp determines efficiency of protein synthesis. bioRxiv. 2019 3; p. 571059. 10.1101/571059doi: 10.1101/571059.

Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. Cell Reports. 2016; 14(7):1787–1799. 10.1016/j.celrep.2016.01.043doi: 10.1016/j.celrep.2016.01.043.

Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. The EMBO Journal. 1988 11; 7(11):3559–3569. 10.1002/j.1460-2075.1988.tb03233.xdoi: 10.1002/j.1460-2075.1988.tb03233.x.

Zur H, Tuller T. Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. Nucleic Acids Research. 2016 9; 44(19):gkw764. 10.1093/nar/gkw764doi: 10.1093/nar/gkw764.