The effect of tRNA levels on decoding times of mRNA codons

Alexandra Dana and Tamir Tuller[*]

Department of Biomedical Engineering, Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv 69978, Israel

ABSTRACT

The possible effect of transfer ribonucleic acid (tRNA) concentrations on codons decoding time is a fundamental biomedical research question; however, due to a large number of variables affecting this process and the non-direct relation between them, a conclusive answer to this question has eluded so far researchers in the field. In this study, we perform a novel analysis of the ribosome profiling data of four organisms which enables ranking the decoding times of different codons while filtering translational phenomena such as experimental biases, extreme ribosomal pauses and ribosome traffic jams. Based on this filtering, we show for the first time that there is a significant correlation between tRNA concentrations and the codons estimated decoding time both in prokaryotes and in eukaryotes in natural conditions (-0.38 to -0.66, all P values <0.006); in addition, we show that when considering tRNA concentrations, codons decoding times are not correlated with aminoacyl-tRNA levels. The reported results support the conjecture that translation efficiency is directly influenced by the tRNA levels in the cell. Thus, they should help to understand the evolution of synonymous aspects of coding sequences via the adaptation of their codons to the tRNA pool.

INTRODUCTION

The way in which intracellular transfer ribonucleic acid (tRNA) levels affect messenger RNA (mRNA) decoding times is still debatable, due to the complexity of quantifying these effects (1–11). First, gene expression is affected by a large number of factors; for example, gene translation efficiency is determined by various features of the transcript (e.g. mRNA folding (2,12), context of the start codon (13,14), length of the different parts of the transcripts (15), charge of the amino acids (16,17), intracellular concentrations of mRNA molecules (18), ribosomes (19), tRNA molecules (20–23), aminoacyl-tRNA synthetases (aaRS)

(24) and the intracellular concentrations of dozens of initiation and elongation factors (24,25)); thus it is impossible to completely control for non-causal relations between these two variables (1–3,6,7). Second, heterologous gene expression, which is often used to study such relations, may not reflect the decoding time of endogenous transcripts since they tend to violate the natural intracellular regimes (1,5,7). Third, although most large-scale measurements of the different stages of gene expression do not directly measure translation elongation rates, they are nevertheless used as a proxy of this variable (e.g. protein levels are often used as a proxy for translation rates (1–3,5,8,26), neglecting additional levels of control that govern the synthesis and degradation of mRNAs and proteins).

The current cutting-edge methodology for studying mRNA translation is ribosome profiling (Ribo-seq), which is based on deep sequencing of ribosome-protected mRNA fragments and produces a detailed account of ribosome occupancy on specific mRNAs under endogenous conditions (27). Recently, several studies (5,10,11) using ribosome profiling data found insignificant correlations between tRNA levels and codons decoding times, inconsistent with previous studies based on other methodologies and data sources (1,4,8,9,28–30).

In this study, we develop a novel statistical approach specifically tailored for analyzing ribosome profiling data of both prokaryotes and eukaryotes. The new approach enables a better understanding of the different variables that contribute to the codon decoding time. We show for the first time that when filtering out rare events such as long pauses in translation elongation, the correlation between codon decoding times and tRNA levels is significant for endogenous transcripts in all analyzed organisms. This relationship is not only fundamental for human health (31–33) but also affects biotechnology (7,8) and disciplines such as molecular evolution (1,5,30,34,35) and functional genomics (6,9,26,28,29,36).

---

[*]To whom correspondence should be addressed. Tel: +972 507 601260; Fax: +972 3 6409357; Email: tamirtul@post.tau.ac.il

MATERIALS AND METHODS

Reconstruction of the Open Reading Frame (ORFs) ribosomal profiles of the analyzed organisms

Saccharomyces cerevisiae ribosomal profiles were reconstructed using the data published in the GEO database, accession number GSE13750 (GSM346111, GSM346114) (27). *Caenorhabditis elegans* ribosomal profiles of genes expressed in the L4 larval stage were built from Illumina sequencing results (NCBI Sequence Read Archive, accession number SRR52883) (37). *Escherichia coli* and *Bacillus subtilis* profiles were built from the published Illumina sequencing results (GEO database, accession number GSE35641) (11). Full details regarding the alignment method appear in the Supplementary text.

Calculating the normalized footprint count (NFC) - data normalization

To avoid analyzing ribosomal profiles of genes with many missing read counts (RCs) that may result in a non-reliable estimation of the local ribosome density, only genes with a median RC above 1 were included in the analysis. Previous studies indicated an increase of RCs at the beginning of the ORF (10,38) and for some organisms at the end of ORF (11), therefore the first and last 20 codons were excluded when determining these thresholds or when calculating the average RCs per ORF. The exact number of genes included in the analysis after applying this filter is depicted in Supplementary Table S1.

To enable comparison and analysis of RCs of codons of the same type originating from different genes, RCs of each codon were normalized by the average RCs of each gene; this normalization controls for possible different mRNA levels and initiation rates of different genes and has been performed in previous studies (5,11). To prevent biasing the average with codons containing less than one RC, those were excluded from the analysis (a similar procedure has been performed in a previous study (11)). Therefore this normalization enables measuring the relative time a ribosome spends translating each codon in a specific gene relative to other codons in it, while considering the total number of codons in the gene, resulting in its normalized footprint count (NFC):

$$\mathrm{NFC}_j = \frac{\mathrm{RC}_j}{\frac{1}{J-40}(\mathrm{RC}_{21} + \mathrm{RC}_{22} + \cdots + \mathrm{RC}_{J-20})},$$

$$j = 21..J - 20$$

where J is the number of codons in the gene and j is the index of a codon; see also Figure 1B.

Calculating the NFC distributions

To study the translation time properties of different codons, for each codon type we generated a vector consisting of NFC values originating from all analyzed genes. These vectors were used to generate for each codon type a histogram reflecting the probability of observing each NFC value in the expressed genes (the number of times each NFC value occurs in the data normalized by the total number of
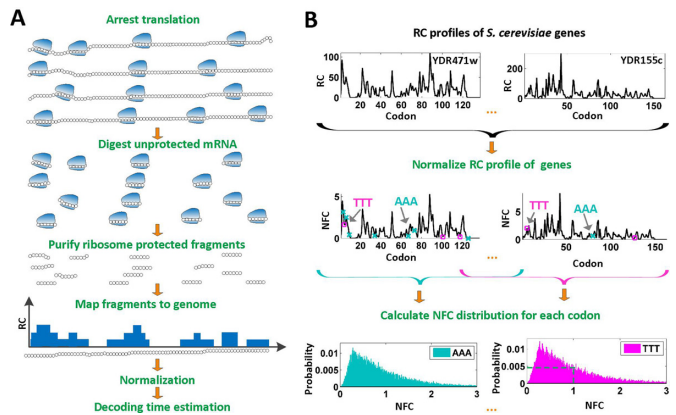
Figure 1. Schematic description of the ribosome profiling method and generating the NFC distributions. (**A**) Translation of codons on mRNAs (black circles) by ribosomes (blue shapes) is arrested, then exposed mRNA is digested. Protected mRNA footprints are then sequenced and mapped to the genome, creating for each gene its read count (RC) profile. (**B**). Illustration of the codons NFC distributions inference from a set of RC profiles: to control for different mRNA levels and translation initiation rates, RC profiles of *S. cerevisiae* (here only YDR471w and YDR155c RC profiles are illustrated) are normalized per gene by their mean RC value, resulting in NFC profiles. Then, NFC values of each specific codon type (NFC values of codons of type 'AAA' are marked by teal 'x' while NFC values of codons of type 'TTT' are marked by magenta squares) are collected from all analyzed genes and presented using a histogram, where the x-axis represents the NFC values and the y-axis represents the fraction of the time (probability) each NFC value appears in the analyzed genes, thus creating the NFC distribution of a codon. For example, the codon 'TTT' appears with an NFC value that equals 1 in the analyzed genes in 0.48% of the times (probability of 0.0048; dashed green lines in the 'TTT' histogram).

times the codon appears in the data) which was named the 'NFC distribution' of the codon (see also Figure 1B). The black histograms in Supplementary Figures S1–S4 depict the NFC distributions of each type of codon, for all analyzed organisms. As summarized in Supplementary Table S2 the mean variance of NFC values of the codons differs between different organisms, such that prokaryotes have a higher NFC standard deviation range, which is associated with more prominent translational pauses (11).

The skewness of a distribution

Skewness measures the lack of symmetry of a distribution. Given a data set of samples $Y_1, Y_2, ..., Y_N$ their skewness is defined as

$$\mathrm{skewness}(Y_1, Y_2, ..., Y_N) = \frac{\sum\limits_{i=1}^{N} (Y_i - \bar{Y})^3}{(N-1)\sigma^3},$$

where $\bar{Y}$ symbolizes the mean, σ the standard deviation and N the number of points.

Inferring the two components of the NFC distributions

Based on the characteristics of the NFC distributions, we suggest that their topology could result from a superposition of two distributions/components: the first one describes the 'typical' decoding time of the ribosomes, which was modeled by a normal distribution characterized by its

mean μ and variance σ^2 with a probability density function $f_x(x; \mu, \sigma)$ (for a random variable X) of

$$f_x(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The second component describes relatively rare translational pauses and ribosomal interactions such as traffic jams due to the codons' different translation efficiency and was modeled by a random variable with an exponentially distribution, characterized by one parameter λ with a probability density function, $f_z(y; \lambda)$, (for a random variable Y) of

$$f_y(y; \lambda) = \{ \begin{array}{l} \lambda e^{-\lambda y}, \ y \geq 0 \\ 0, \ y < 0 \end{array}.$$

It is known that the distribution of a random variable $w(t)$ that is a sum of two independent random variables $f(t)$ and $g(t)$ (i.e. $w(t) = f(t) + g(t)$) is calculated as convolution between the two distributions (39):

$$w(t) = f(t) * g(t)$$

$$= \int_{-\infty}^{t} f(\tau)g(t-\tau)d\tau \quad \forall f, g : [-\infty, \infty) \rightarrow \mathbb{R}.$$

Thus, the summation of two independent normal and exponential random variables corresponding to the distributions mentioned above results in a distribution that is named 'exponentially modified Gaussian' (EMG) and is a convolution of a normal and exponential distribution; formally, the EMG distribution function, $f_z(y; \mu, \sigma, \lambda)$, of a random variable Z (where $Z = X + Y$)(40) is

$$f_z(z; \mu, \sigma, \lambda) = \frac{\lambda}{2} e^{\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2z)} erfc(\frac{\mu + \lambda\sigma^2 - z}{\sqrt{2}\sigma}),$$

where

$$erfc(x) = 1 - erf(x) = \int_{x}^{\infty} e^{-t^2} dt.$$

The parameters μ, σ, λ were estimated by fitting the measured NFC distributions to the EMG distribution, under the log-likelihood criterion. Estimated μ, σ, λ values appear in Supplementary Table S11. It should be mentioned that the EMG distribution does not necessarily define an equal weight for the Gaussian and exponential distributions. Rather, the optimization of the μ, σ, λ parameters defines their optimal weight. For additional insights regarding the expected relation between the μ and λ parameters see the Supplementary section.

In addition, we fit the NFC data using an EMG model with an identical exponential distribution parameter for all codons; in this model, the λ parameter was optimized for all codons under the log-likelihood criterion. Moreover, the NFC distribution was fit also using the exponential and Gaussian distributions separately. The goodness of these fittings was compared and analyzed by using the Akaike information criterion (AIC) score (more details below).

Evaluating the goodness of models using the AIC score

Comparison between the different fitting models was evaluated using the AIC (41). This measure considers both the number of parameters in the model and the fit (log-likelihood) of the data to the model (i.e. it penalizes more complicated models). Lower AIC score is related to a model that is better fitted to the data.

Calculating codons adaptiveness value to tRNA levels w_i

The tRNA adaptation index (tAI) (42) uses the adaptiveness of the codons of a gene to the tRNA pool. Let us mark the adaptiveness value of codon of type i with W_i. Let tCGN$_{ij}$ be the copy number of the j-th anti-codon that recognizes the i-th codon and let S_{ij} be the selective constraint of the codon–anti-codon coupling efficiency. The s vector (42,43) $[s_{I:U}, s_{G:C}, s_{U:A}, s_{C:G}, s_{G:U}, s_{I:C}, s_{I:A}, s_{U:G}, s_{L:A}]$ was defined for prokaryotes as [0, 0, 0, 0, 0.41, 0.63, 0.9749, 0.68, 0.95] and for eukaryotes as [0, 0, 0, 0, 0.561, 0.28, 0.9999, 0.68, 0.89]. Then, the absolute adaptiveness value of a codon is defined by

$$W_i = \sum_{j=1}^{n_i} (1 - S_{ij}) \text{tCGN}_{ij}.$$

Let us mark the relative adaptiveness value of codon i with w_i by normalizing each W_i with the maximal W_i value among the 61 W_i values. A codon typical decoding time CDT$_i$ used for the Totally Asymmetric Simple Exclusion Process (TASEP) simulations was defined as

$$\text{CDT}_i = \frac{1}{w_i},$$

where codons with low adaptiveness values to the tRNA pool will be more slowly translated.

Simulating ribosome density profiles using the TASEP model

Ribosome density profiles were simulated using the TASEP biophysical translation model, previously used in different studies (44,45). In this model, the mRNA was simulated using a lattice of N sites, where N represents the number of codons of the ORF. Each ribosome was defined to cover 11 codons and its A site was located at the sixth codon. During translation, any codon could be covered at a time by a single ribosome at most. In each step of the simulation, a single ribosome was allowed to attach itself to the lattice/advance to the next codon if the first/next six codons were not occupied. The time between initiation attempts was set to be exponentially distributed with a constant rate ψ. Similarly, the time between jump attempts from site i to site $i + 1$ was assumed to be exponentially distributed with rate ψ_i.

The time between events (initiation or jumping between sites) is therefore exponentially distributed (minimum of exponentially distributed random variables) with rate

$$\mu(n_i) = \psi + \sum_{i=1}^{N} n_i \psi_i,$$

where i describes the site (codon) number on the lattice and $n_i = 1$ if codon i is being translated, otherwise $n_i = 0$.

Therefore the initiation probability is given by $\psi/\mu(n_i)$ and the probability of a ribosome to jump from site i to site $i+1$ is given by $n_i\psi_i/\mu(n_i)$.

In this study *S. cerevisiae* genes that contained a sufficient number of RCs (see the criterion in Supplementary Table S1) and with a least 50 codons were simulated. Codons' decoding rate ψ_i was set according to their adaptiveness to the tRNA pool (specific codon decoding times (CDT) values used in the simulation are presented in Supplementary Table S3), where

$$\psi_i = w_i.$$

The initiation rate ψ was set to be as the decoding rate of the fastest codon. To reach a steady-state distribution on the mRNA, the simulation of each mRNA copy was run until 200 ribosomes finished translating the gene. At this point of the simulation, codons covered by ribosomes were referred as ribosomal protected fragments. For each simulated gene the TASEP model was run for m times, where m represents the mRNA level of the gene. This parameter was set to be the maximal measured RC value of each gene. The protected fragments of each one of the m simulations of a gene were used to create its ribosomal profile as previously described for *S. cerevisiae*.

To simulate the effect of translational pauses on the NFC distributions, codons of *S. cerevisiae* with NFC higher than 3-fold of the mean NFC of the ORF (based on the real measurements, when excluding codons with zero RCs (11), i.e. codons with NFC higher than 3) were defined as locations of translational pauses. The decoding time of these codons was set to be proportional to the NFC of the measured pauses in the real ribosomal profiles. Specifically, we defined the simulated translational pauses translation time to be proportional to the mean simulated mean RCs, as in the real data, i.e.

$$\frac{\text{simulated\_}TT\text{\_}TP(k)}{\text{simulated\_mean\_}TT(\text{gene})} = \frac{RC\text{\_}TP(k)}{\text{mean\_}RC(\text{gene})},$$

where k is the index of the codon; $\text{simulated\_}TT\text{\_}TP(k)$ is the simulated translational time of codon k; $\text{simulated\_mean\_}TT(\text{gene})$ is the mean simulated translational time of all codons in the gene not defined as translational pauses; $RC\text{\_}TP(k)$ is the real RC of the translational pause of the gene at codon k; and $\text{mean\_}RC(\text{gene})$ is the real measured mean RC of all codons in the gene not defined as translational pauses.

Protein abundance and mRNA level measurements of *S. cerevisiae* and *E. coli*

S. cerevisiae protein abundance (PA) measures were averaged from four quantitative large-scale measurements: two large-scale measurements in two conditions (46) and a large-scale PA measurement from two sources (47,48). mRNA levels were determined by averaging large-scale measurements of mRNA levels (27,49). *E. coli* PA measurements were downloaded from the PRIDE database (50) and mRNA level measurements were taken from another source (51). Aminoacyl-tRNA synthetase protein levels and quantities of *S. cerevisiae* (52) and *E. coli* (53) were used from (52) and (53), respectively.

Inferring the weighted typical decoding times and tRNA levels per amino acid

To enable comparison between aaRS levels, tRNA levels and typical decoding times, we defined the representative tRNA level of each amino acid as the sum of tRNA levels corresponding to the amino acid weighted by the frequency of codons in the ORFs that translate it. Let f_i denote the frequency of codon i and tRNA$_i$ its corresponding tRNA level. Thus, the weighted tRNA level of an amino acid equals

$$\text{tRNA}_{\text{weighted}} = \sum_{i=1}^{P} f_i \text{tRNA}_i.$$

Similarly, the typical amino acid decoding time μ_{aa} also took into consideration the codon frequencies of the codons coding it, thus was defined as

$$1/\mu_{\text{aa}} = \sum_{i=1}^{P} f_i \frac{1}{\mu_i},$$

where P depicts the number of codons coding to a specific amino acid and μ_i depicts the typical codon decoding time.

Estimating the typical amino acid decoding time using a linear regressor of the weighted tRNA levels and aaRS levels

To estimate to what extent the aaRS levels improve the prediction of the amino acid decoding time (relative to tRNA levels only), we have modeled the relationship between amino acid decoding times and the two aforementioned possible explanatory variables by using a linear regressor (54)

$$\mu_{\text{aa}} = c + w_1 \text{aaRS} + w_2 \text{tRNA}_{\text{weighted}},$$

where the coefficients c, w_1 and w_2 are estimated as to minimize the mean square error difference between the prediction vector $\tilde{\mu}_{\text{aa}}$ and the real measurements μ_{aa}. Absolute Spearman correlation between μ_{aa} and $\tilde{\mu}_{\text{aa}}$ was compared to Spearman correlation between μ_{aa} and tRNA$_{\text{weighted}}$ to determine if the linear regressor based on aaRS levels improves the correlation. The correlation coefficient between μ_{aa} and $\tilde{\mu}_{\text{aa}}$ was adjusted due to an increased number of explanatory variables using the correction equation (55)

$$r^2 = 1 - (1 - r^2)\frac{n-1}{n-p-1},$$

where r depicts the Spearman correlation coefficient, n depicts the sample size (in this case $n = 20$) and p depicts the total number of regressors (in this case $p = 2$). For each of the coefficients c, w_1 and w_2 their confidence interval (56) was calculated to determine the reliability of the estimates (at a 95% confidence interval). A coefficient's confidence interval that does not contain zeros implies that it significantly contributes to the regression, i.e. the coefficient is not zero (56).

RESULTS

### Estimating the codons relative decoding times from ribosome profiling data

We began our analysis by reconstructing ribosome profiles for *E. coli* (11), *B. subtilis* (11), *C. elegans* (37) and *S. cerevisiae* (27) expressed genes. The ribosome profiling method produces ribosome footprint counts that are proportional to the time spent in decoding each codon of all translated transcripts in a genome, at single nucleotide resolution (see Figure 1A and Supplementary Methods). To enable a comparison between footprint counts of codons from different genes we applied to them a conventional normalization (11) (see the Materials and Methods section; Figure 1B) and named this value the Normalized Footprint Count (NFC). To gain an initial understanding of the NFC values in natural conditions, we calculated a histogram reflecting the distribution of the NFC values of each codon in the expressed genes for each codon in each organism; we call this histogram the 'NFC distribution' of the codon. Each NFC distribution (see Figure 1B and Supplementary Figures S1–S4) describes the probability (y-axis) of observing each of the codon's NFC values (x-axis) in the ORFs of the analyzed organism.

The mean codons decoding time is extremely sensitive to extreme values caused by translational pauses

The intuitive way for representing the decoding time of a codon is by calculating the average of all NFC values for each codon (we name this estimation mean NFC), which was implemented in several previous studies (5,11) with some variation. However, as can be seen in Figure 2A the NFC histograms show a right-skewed distribution with 'right tails', reflecting possible interactions between ribosomes such as ribosomal jams, rare but extreme events such as translational pauses (10,11) and/or experimental biases (10,57,58). Therefore, straightforward mean NFC estimations (see also Supplementary Tables S4 and S5) may have been biased by these factors, thus did not accurately represent the typical codon decoding time.

To illustrate the statistical problem and evaluate the suggested approach in a controlled environment where the codons decoding times are known, we started our analysis by first creating different TASEP simulations of the ribosome profiling experiment (see the Materials and Methods section and Figure 2B). In the first simulation, all codons decoding times were set to be exponentially distributed with the same parameter $\psi = 1$ and with a low initiation rate of 0.3 (thus generally no ribosomal jamming was created; no translational pauses were added). The resulting NFC distribution of the codons (Supplementary Figure S05) shows that when no interactions between codons/biases/pauses are present, the expected NFC distribution is similar to a Normal distribution (comparison of codon NFC distribution to Normal distribution under Kolmogorov–Smirnov (KS) test: mean *P*-value = 0.27).

However, deviations from the theoretic Normal distribution are expected due to ribosomal jamming caused by different codons decoding efficiency, translational pauses and experimental biases such as a low number of mRNA
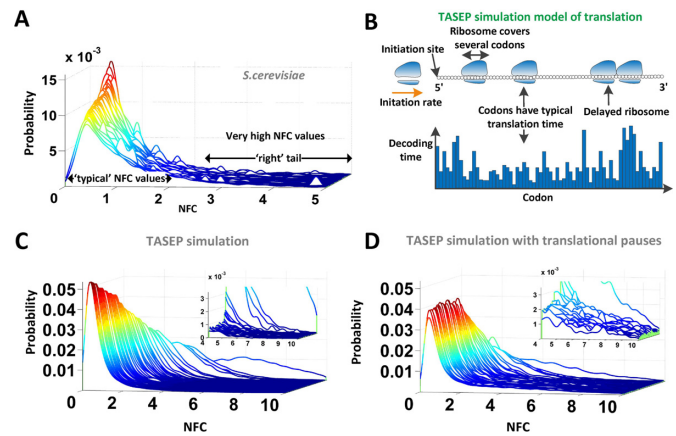
Figure 2. Visualization of the NFC distributions in real and simulated data. (**A**) NFC distributions (e.g. *S. cerevisiae*) consist of two components: one that is centered around the most frequent (typical) NFC values in the distribution and another that consists mainly of high NFC values, associated with extremely slow translation events, e.g. translation pauses. The graph shows the NFC distributions of all *S. cerevisiae* codons. (**B**) Description of the TASEP simulation model: each gene has its own specific initiation rate; ribosomes cover multiple codons and can continue translating if the adjacent downstream codons are not covered by ribosomes. Each codon is associated with its typical decoding time. (**C**) NFC distributions of all codons based on the *S. cerevisiae* genes simulated with the TASEP model without translation pauses. In the absence of translational pauses the 'right tail' region is negligible. (**D**) NFC distributions of all codons based on the simulated *S. cerevisiae* genes when 4% of the codons are altered to cause translation pauses. The additional translational pauses are reflected as increased peaks in the 'right tail' region of the distribution.

copies per gene. To demonstrate this, we created an additional TASEP simulation of the ribosome profiling experiment where each codon was assigned a unique decoding time. According to the TASEP model, a queue of ribosomes (or ribosomal jamming) is expected upstream of a slow codon. These jams can result from slower codons (reflected by high RCs) or from translational pauses (reflected by very high RCs). Therefore the effective decoding time of a codon c_i can be higher than its nominal decoding time if a slower codon (or a translational pause) appears downstream of it, causing a traffic jam that eventually affects codon c_i. As a result, the NFC distribution of a codon type can be skewed from the expected symmetric 'Normal like' distribution in the TASEP simulation where no pauses are present and all codons have equal translation decoding times. Thus, overall it is expected that codons with faster translation time will be more influenced by ribosomal jams and translational pauses, leading to an NFC distribution with a thicker 'right tail' and higher 'skewness' (i.e. asymmetricity; see the Materials and Methods section). As expected, in the presence of ribosomal jamming, the simulated NFC distributions deviated from expected Normal distribution, creating a right-skew/right tail (see Figure 2C and Supplementary Figure S06). Specifically, the mean/median/max skewness was 0.078/0.086/0.13, respectively. However, even when the codons decoding time differed, still a highly significant correlation was observed between the mean NFC values and simulated decoding times (Spearman correlation: $r = 0.99$; $P = 3.5 \times 10^{-52}$; Figure 3B).
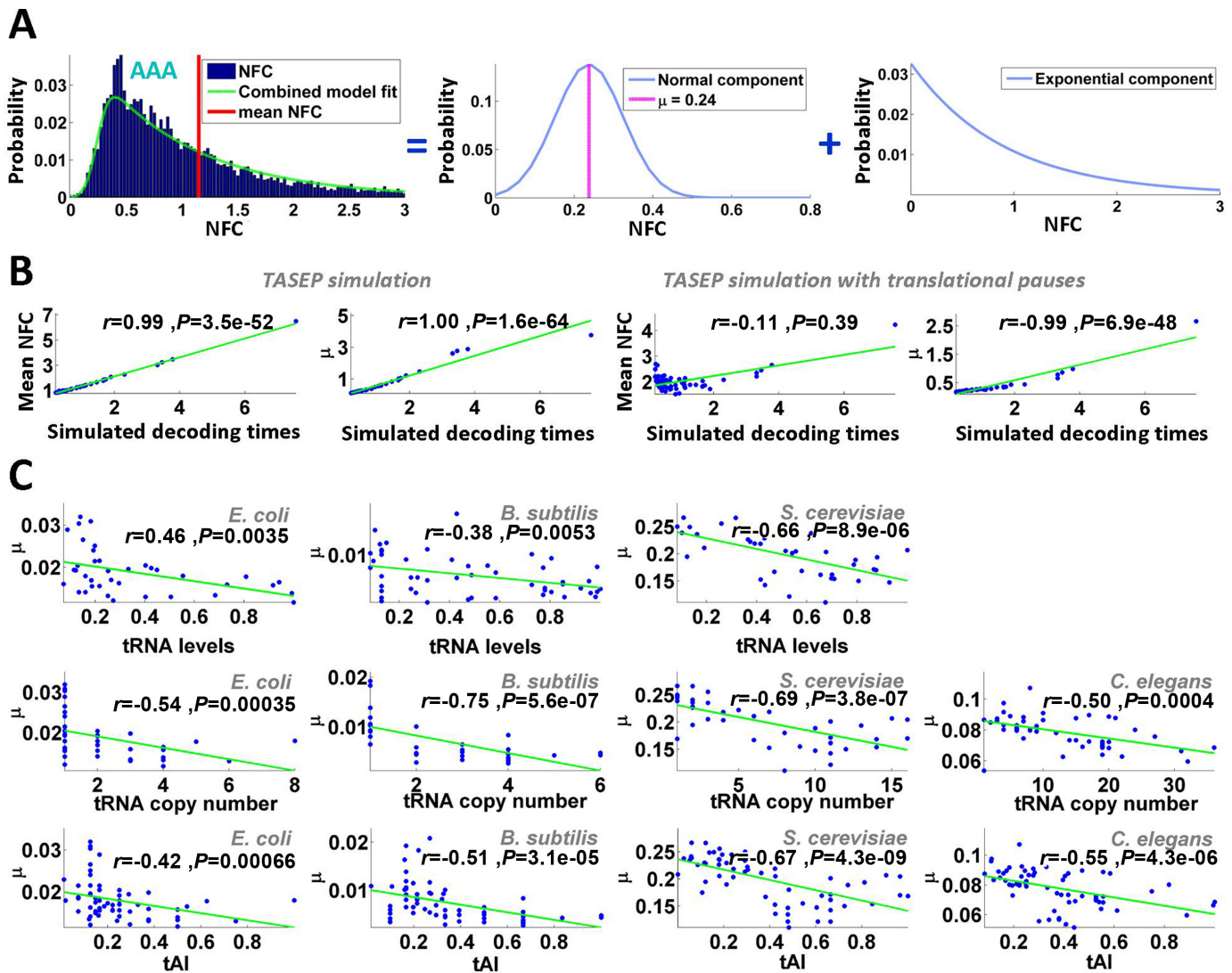
Figure 3. tRNA levels are correlated with the estimated μ in real and simulated data. (**A**) NFC distribution of codon AAA of *S. cerevisiae* is depicted as a histogram. The combined normal/exponential model fitting is plotted as a curve. The position of the mean NFC value is presented with a vertical line. The NFC distribution can be decomposed into a normal (which includes the parameter μ) and an exponential component using a log-likelihood fitting. (**B**) Estimated mean NFC/μ values of simulative ribosome profiles versus simulated decoding times in the absence and presence of translation pauses. (**C**) Estimated μ values of the real ribosome profiles plotted against tRNA levels (first row), tRNA copy numbers (second row) and tAI values (third row) for the various organisms analyzed. Fitting lines are also shown in all cases.

To demonstrate that the mean codon decoding time is very sensitive to rare and extreme pauses, we added to the previous TASEP simulation translational pauses in only 4% of the positions of the ORFs (see details in the Materials and Methods section). In the presence of translational pauses, the skewness of the NFC distributions (see Figure 2D and Supplementary Figure S07) increased (mean/median/max of skewness = 0.18/0.19/0.23, respectively) and dramatically reduced the correlation between mean NFC and simulated decoding times (Spearman correlation: $r = -0.11$; $P = 0.39$; Figure 3B). Therefore, this result suggests that the mean estimator is very sensitive to phenomena such as extreme values caused by translational pauses.

The estimated typical codon decoding times are significantly correlated with tRNA levels

To estimate the typical decoding time of each codon based on NFC distributions, we developed a novel statistical model, which takes into consideration the skewed nature of the NFC distribution that includes three parameters. In this model, the typical codon decoding time was described by a Normal distribution with two parameters: mean (μ) and standard deviation; the μ parameter represents the location of the mean of the theoretical Gaussian component that should be obtained if there are no phenomena such as pauses/biases/ribosomal traffic jams (see Supplementary Figure S5); σ represents the width of the Gaussian component, while λ represents the skewness of the NFC distribution. Rare and extreme translation pauses and translational

features such as ribosomal jamming caused by codons with different decoding times were described by an exponential single-parameter distribution (see illustration in Figure 3A and more details in the Materials and Methods section). We used the conventional maximum likelihood criterion to estimate these three parameters for each codon and organism by fitting the suggested model to the NFC distribution (see the Materials and Methods section and Supplementary Figures S1–S4). When employing this approach on the simulated data generated in the presence of translation pauses as mentioned above, we obtained an extremely high correlation between the simulated decoding times and the estimated decoding time values (μ) over the entire set of codons (Spearman correlation: $r = 0.99$; $P = 6.9 \times 10^{-48}$; Figure 3B); this result demonstrates the high accuracy of the new approach.

Next, we applied our approach on real ribosome profiling data. First, we fit the NFC data to the EMG distribution. For comparison purposes the AIC score (see the Materials and Methods section) was calculated for this model and we compared the goodness of fitting when using only a Gaussian or an exponential fitting (see Supplementary Table S6). The results indicate that in all organisms the EMG exhibits a lower AIC score, supporting the conjecture that EMG distribution better describes the NFC distribution than either the normal or the exponential distributions alone.

In addition, we also considered an EMG model with an identical exponential distribution parameter for all codons (see the Materials and Methods section) and show that it also decreases the data fit (see Supplementary Table S7). This further showed that the EMG model with three free parameters better fits the NFC data.

If tRNA levels affect codons translation efficiency, we would expect them to correlate with the estimated typical codon decoding times (μ). Indeed, we found such a significant correlation between μ and tRNA levels in all organisms with available tRNA level measurements (*E. coli* (59), *B. subtilis* (60) and *S. cerevisiae* (28); Spearman correlation was between -0.38 and -0.66, all *P*-values <0.006; see Figure 3C). Significant Spearman correlations were also obtained when tRNA copy numbers were used as a proxy for absolute tRNA levels (28,60) ($r = -0.5$ to -0.75, all *P*-values <0.0004; Figure 3C). A similar observation was made when using the tAI (2,42) of the different codons (see the Materials and Methods section), which considers not only codon–anti-codon interactions but also wobble interactions with other tRNA molecules ($r = -0.42$ to $r = -0.67$; all *P*-values <0.0007; Figure 3C). To control for possible biases resulting from codons which are more abundant and thus for example are known to be recognized by more abundant tRNA species (20,21,23,45,61), the analysis was repeated when calculating the NFC distributions based on an equal amount of RCs for all codons sampled from the real ribosomal profiling data; the results in this case were similar (Supplementary Figure S8). This indicates that our conclusions are not dependent on the amount of RCs per codon type, therefore eliminating a possible bias caused by this factor. Finally, we obtained similar significant correlation also when we considered an EMG model with an identical exponential distribution parameter for all codons (Supplementary Table S8).

The estimated typical codon decoding times are not correlated with aaRS levels given the tRNA levels

In the previous section, we showed that the estimated codon decoding times are significantly correlated with tRNA levels. However, the decoding time of codons could be affected by additional factors such as aaRS concentrations. In this section, we estimate the effect of aaRS concentrations on the typical decoding times and compare it to the effect of tRNA levels on the typical decoding times.

Each amino acid (or its precursor) has only one corresponding aaRS which catalyzes its esterification to one of all its compatible cognate tRNAs; thus in this section, we analyze the effect of aaRS PA and mean tRNA levels on the typical amino acid decoding times μ_{aa} (see details in the Materials and Methods section). Specifically, the amino acid decoding times μ_{aa} and their tRNA levels $tRNA_{weighted}$ were computed via a weighted average that considers the relative frequencies of all codons coding for each specific amino acid (we analyzed two organisms with reliable aaRS protein levels, *E. coli* and *S. cerevisiae*).

As expected, in both organisms aaRS protein levels negatively correlate with amino acid decoding times, i.e. higher protein levels correspond to shorter decoding times; only in *S. cerevisiae* this correlation was significant ($r = -0.32$, -0.66, $P = 0.16/0.0018$ for *E. coli* and *S. cerevisiae*, respectively). However, since tRNA levels and aaRS concentrations are expected to undergo co-evolution to optimize translation cost, they are expected to positively correlate, i.e. amino acids with higher aaRS protein levels tend to have higher levels of tRNA corresponding to their codons ($r = 0.49/0.66$; $P = 0.031/0.0016/$ for *E. coli* and *S. cerevisiae*, respectively) and thus are expected to also affect codon decoding time in the same direction.

To study the distinct statistical contribution of tRNA levels/aaRS protein levels to the measured μ_{aa} we calculated partial Spearman correlation between μ_{aa} and $tRNA_{weighted}$ given the aaRS protein levels and also partial Spearman correlation between μ_{aa} aaRS protein levels given $tRNA_{weighted}$. We found that the partial correlation corr(μ_{aa}, $tRNA_{weighted}$| aaRS protein levels) is very high and significant ($r = -0.67/-0.63$; $P = 0.00017/0.0036$ for *E. coli* and *S. cerevisiae,* respectively) while the partial correlation corr(μ_{aa}, aaRS protein levels | $tRNA_{weighted}$) is not significant ($r = 0.04/-0.20$; $P = 0.88/0.42$ for *E. coli* and *S. cerevisiae*, respectively); these results support the conjecture that in the analyzed organisms weighted tRNA levels rather than aaRS levels explain the variance in μ_{aa} indicating that μ_{aa} is more likely to be determined by the tRNA levels in the cell.

Similar results were obtained based on a linear regression analysis: when considering the number of explaining variables in the regressor (i.e. adjusting the correlation to the number of variables in the regression; see the Materials and Methods section), in both analyzed organisms the correlation between μ_{aa} and its regressor based on aaRS protein levels and the weighted tRNA levels per amino acid (see the Materials and Methods section) was not higher than the absolute value of the correlation between tRNA levels and μ_{aa}($r = 0.58/0.77$; $P = 0.0029/0.0001$ versus $r = -0.71/-0.81$; $P = 0.00063/1.8 \times 10^{-5}$ for *E. coli* and *S. cere-*

visiae, respectively); in addition, in both analyzed organisms the confidence interval of the regressor's aaRS protein levels coefficient was not significantly different than zero (see the Materials and Methods section), suggesting the insignificant contribution of aaRS protein levels to the prediction of μ_{aa}.

These results indicate that in the analyzed organisms aaRS protein levels do not significantly improve the explained variance of the typical codon decoding times given the tRNA levels; thus, our analyses support the conjecture that typically tRNA levels and not aaRS protein levels are a rate limiting factor during translation elongation.
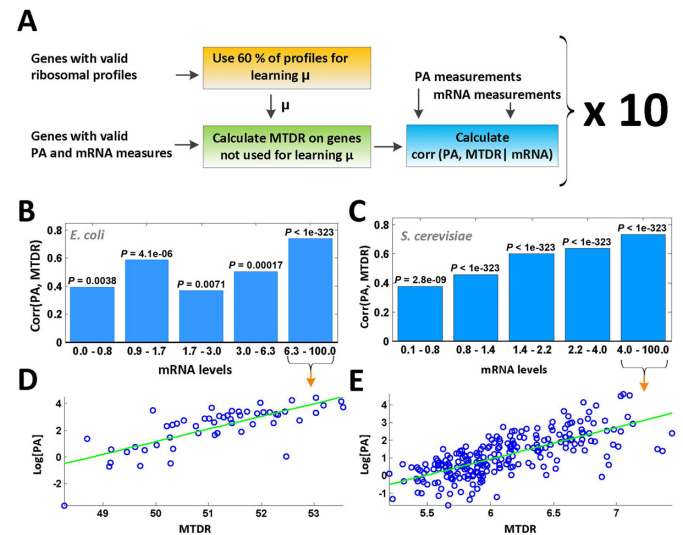
The mean-estimated typical codons decoding times of genes significantly correlate with their number of proteins per mRNA molecules

To evaluate the influence of the estimated typical decoding times on the genes' translation efficiency (i.e. the protein levels of a gene when controlling its mRNA levels or the number of proteins per mRNA molecule), we defined for each gene a translation efficiency measure. This measure was defined as the geometric Mean of the Typical Decoding Rates $(1/\mu)$ of its codons (MTDR), in a similar manner that previous related codon bias usage indexes have been defined (2,62).

Two organisms (*E. coli* and *S. cerevisiae*) with available large-scale measurements of protein and mRNA levels were analyzed. To prevent overfitting the results, the following procedure was performed: for each organism, codons decoding times were estimated based on only 60% of the genes (train set) with sufficient RCs (see details in the Materials and Methods section); then, for the rest of the genes (test set) with existing mRNA and protein abundance levels (not participating in the learning phase) we computed the MTDR based on the learned μ values.

Finally, partial Spearman correlation between PA levels and MTDR when controlling mRNA levels corr(PA, MTDR | mRNA) was computed for the test set; the procedure was repeated 10 times (see also a flow diagram in Figure 4A), obtaining similar results: the correlation coefficient for *E. coli* was between 0.49 and 0.58 (all $P < 3.1 \times 10^{-16}$) and for *S. cerevisiae* the obtained correlation coefficient was between 0.5 and 0.57 (all $P < 7.9 \times 10^{-76}$). Results of a single representative run of the procedure are presented in Figure 4B–E. When controlling an equal amount of RCs for all codons, the partial correlation obtained for *E. coli* was between 0.44 and 0.61 (all $P < 1.2 \times 10^{-10}$) and for *S. cerevisiae* was between 0.56 and 0.58 (all $P < 2.6 \times 10^{-109}$).

Significant results were also obtained when computing the correlations between MTDR and PA/mRNA (the number of proteins per mRNA molecule) on the test set: correlation coefficient for *E. coli* was between 0.41 and 0.52 (all $P < 3.1 \times 10^{-11}$) and for *S. cerevisiae* the obtained correlation coefficient was between 0.25 and 0.29 (all $P < 9.1 \times 10^{-19}$). Similarly, when constraining an equal amount of RCs for all codons, the correlation obtained for *E. coli* was between 0.41 and 0.55 (all $P < 1.3 \times 10^{-8}$) and for *S. cerevisiae* between 0.29 and 0.31 (all $P < 1 \times 10^{-323}$). These results indicate that the translation efficiency of genes can be partially explained by the codons typical decoding times estimated

Figure 4. Translation efficiency index based on the estimated typical codons decoding time significantly explains the translation efficiency in *E. coli* and *S. cerevisiae*. (**A**) Diagram showing the process of the calculating Spearman partial correlation between protein abundance levels of *S. cerevisiae* and genes translation efficiency index (MDTR) when controlling mRNA levels (corr(PA, MTDR | mRNA)). (**B**) Calculating the correlation between protein abundance and MDTR measure (corr(PA, MTDR)) on *E. coli* genes for five equal-sized bins with similar mRNA level values, for a single representative run (Spearman partial correlation is 0.53, $P < 1.2 \times 10^{-20}$). The specific P-value calculated for each bin apart appears above it. mRNA level values for this presentation were normalized to range [0, 100]. (**C**) The same presentation as in (B) calculated for *S. cerevisiae*. The partial Spearman correlation for this run is 0.54 ($P < 1.26 \times 10^{-91}$). (**D**) Dot-plot of the log protein abundance (**PA**) values (normalized to range [0, 100]) versus MTDR values for genes in the bin obtaining the highest correlation presented in (B), marked with an arrow. (**E**) The same presentation as in (D) for *S. cerevisiae*.

using the newly suggested method, which in turn are significantly influenced by the abundance of tRNA molecules in the cell. Finally, a similar analysis based on the mean NFC estimator yields no significant correlation or very low correlation to protein abundance: the correlation for *S. cerevisiae* was between 0.10 and 0.2 ($P < 0.0008$) and for *E. coli* the correlation was between 0.02 and 0.11 ($P < 0.77$).

It should be mentioned that the mean NFC estimator did not result in a significant correlation with PA as it is very sensitive to extreme translational pauses and to other phenomena such as traffic jams and experimental biases which are non-codon specific and blur the actual effect of the nominal codon decoding time. The MDTR based on our approach estimates the typical/nominal translation speed of each codon, therefore considers the direct contribution of the codons to the elongation speed. Thus a correlation between PA and MDTR was observed, but not with the geometrical mean (over the codons in the ORF) of the mean codon NFC values.

It is important to explain that the fact that there is correlation between MTDR and PA per mRNA does not contradict the fact that initiation usually tends to be more rate limiting than elongation. Many previous studies have demonstrated that even though initiation has higher correlation with protein per mRNA (or other measures of translation rate), elongation can also significantly explain

non-negligible portion of the variance in translation rate (2,3,8,20,21,23,63).

As translation efficiency is influenced not only by the elongation rate but also by initiation rates and other unknown additional factors, our correlation is, of course, significant but not perfect, i.e. $r = 0.5$–0.6. Thus, MTDR can explain around 25–36% of the variance in PA per mRNA levels. Therefore additional major factors (such as initiation) are needed to explain the variance in translation efficiency.

DISCUSSION

The analyses performed in this study support the conjecture that both in eukaryotes and prokaryotes tRNA concentrations affect codons decoding times. The decoding time of codons recognized by tRNA species with lower intracellular concentrations tends to be longer. It was suggested before that there is a relation between the adaptation of an ORF to the intracellular tRNA concentrations and its translation rate (1–4,6,7,20,22,23,26,28,29,31,34,36,47,64); however, this is the first time that a direct relation between these two variables is shown. We believe that due to noise and biases in the data (e.g. possibility to accurately determining the exact P-site position related to the read) the reported results are only lower bounds of the actual relations. Nevertheless our findings do not contradict the possible effect of other factors on translation elongation speed (2,11,45,65); for example, it was recently shown that the anti-Shine–Dalgarno-like sequences (11), charge of amino acids (45,65) and folding strength of the mRNA molecule (45,66) affect elongation rate and translation efficiency. Several different factors can affect this process and it is not always trivial to isolate the effect of each variable on it; furthermore in different organisms, conditions and/or genes, some factors may be more dominant than others. Thus, the results reported in this study emphasize the complexity of the translation elongation process and the way its efficiency is encoded in the ORF. Naturally, the reported relation between codon decoding times and tRNA concentrations has various fundamental ramifications.

First, we show that translation elongation speed is affected by the adaptation of codons to the tRNA pool, explaining a mechanism by which the codon content of the ORF can affect the organismal fitness. Thus, the reported results should help to understand the evolution of synonymous aspects of coding sequences via the adaptation of their codons to the tRNA pool (1,7,28,30,33–35,66).

Second, it was shown that the folding of proteins partially occurs during translation elongation (67–71). In addition, several studies also suggested that the speed of translation elongation affects co-translational folding of the nascent peptide chain (34,70,72–75); specifically, it was suggested that codon bias usage can affect the folding of proteins, among others due to the two points above. The results reported here suggest that the link between codons bias usage and protein co-translational folding is partially related to the effect of the adaptation of codons to the tRNA pool on their decoding speed. Thus, our results support previous studies that have suggested that codons adaptation to the tRNA pool may affect protein folding/function (2,4,26,29,34,36) and thus even cause human disease (31,32,65).

Third, various studies have suggested the existence of co-evolution between transcriptomic codon frequencies and the intra-cellular levels of the tRNA molecules recognizing them (5,61,76,77): codons with higher frequencies are recognized by tRNA molecules with higher intra-cellular levels; evolutionary changes that affect tRNA levels corresponding to a certain codon are followed by evolutionary changes that affect the codon frequency in transcripts in the same direction and vice versa. The results reported in this study suggest that this co-evolutionary process may at least partially be mediated by the effect of tRNA species on codon decoding times; specifically, increasing the frequency of codons recognized by highly concentrated tRNA species should improve elongation speed and thus the organismal fitness.

It was also suggested that since eventually there is a balance between demand (codon frequency) and supply (tRNA levels), the codon decoding times of all codon should be identical (5). Our study shows that this co-evolutionary process does not produce a 'perfect' balance as there is still a high correlation between codon decoding times and intra-cellular levels of tRNA levels (the decoding time of codon recognized by more abundant tRNA molecules is shorter). This scenario may be possible due to the fact that there may be a delay between evolutionary changes in the intracellular tRNA repertoire and the time it takes for evolutionary forces to shape codon frequencies to match this repertoire. It is also possible that there is a selection for non-prefect balance between tRNA levels and codon frequency; the possible functional advantage of non-perfect balance may be related to the fact that different codon decoding time may regulate translation and protein folding in various ways as mentioned above (7,26,28,29,34,36,70,72–75).

Finally, here we studied the effect of tRNA and aaRS levels on the typical codons decoding time; however, it should be mentioned that the typical codon decoding time is affected by several additional factors that the ribosome profiling method cannot directly and/or separately measure. Among others, these factors include the amino acid–tRNA accommodation time, peptidyl transfer time and the time for the whole EF-G-driven translocation process. It is not clear if the effect of these factors on elongation rate differs among different codons and amino acids; we also do not know to what extent these factors can explain the non-prefect correlation between codon decoding time and tRNA levels. The understanding of the contribution of these factors to codon decoding time and specific answers to these questions are deferred to future studies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENT

REFERENCES

1. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science*, **324**, 255–258.
2. Tuller,T., Waldman,Y.Y., Kupiec,M. and Ruppin,E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 3645–3650.
3. Supek,F. and Smuc,T. (2010) On relevance of codon usage to expression of synthetic and natural genes in Escherichia coli. *Genetics*, **185**, 1129–1134.
4. Lithwick,G. and Margalit,H. (2003) Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res.*, **13**, 2665–2673.
5. Qian,W., Yang,J.R., Pearson,N.M., Maclean,C. and Zhang,J. (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.*, **8**, e1002603.
6. Novoa,E.M. and Ribas de Pouplana,L. (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.*, **28**, 574–581.
7. Plotkin,J.B. and Kudla,G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
8. Gustafsson,C., Govindarajan,S. and Minshull,J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.*, **22**, 346–353.
9. Cannarozzi,G., Schraudolph,N.N., Faty,M., von Rohr,P., Friberg,M.T., Roth,A.C., Gonnet,P., Gonnet,G. and Barral,Y. (2010) A role for codon order in translation dynamics. *Cell*, **141**, 355–367.
10. Ingolia,N.T., Lareau,L.F. and Weissman,J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
11. Li,G.W., Oh,E. and Weissman,J.S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, **484**, 538–541.
12. Zur,H. and Tuller,T. (2012) Strong association between mRNA folding strength and protein abundance in S. cerevisiae. *EMBO Rep.*, **13**, 272–277.
13. Kozak,M. (1984) Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature*, **308**, 241–246.
14. Zur,H. and Tuller,T. (2013) New universal rules of eukaryotic translation initiation fidelity. *PLoS Comput. Biol.*, **9**, e1003136.
15. Tanguay,R.L. and Gallie,D.R. (1996) Translational efficiency is regulated by the length of the 3' untranslated region. *Mol. Cell. Biol.*, **16**, 146–156.
16. Trylska,J., Konecny,R., Tama,F., Brooks,C.L. III and McCammon,J.A. (2004) Ribosome motions modulate electrostatic properties. *Biopolymers*, **74**, 423–431.
17. Lu,J. and Deutsch,C. (2008) Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.*, **384**, 73–86.
18. Schwanhausser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
19. Arava,Y., Wang,Y., Storey,J.D., Liu,C.L., Brown,P.O. and Herschlag,D. (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. U.S.A.*, **100**, 3889–3894.
20. Ikemura,T. (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.*, **146**, 1–21.
21. Ikemura,T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
22. Dittmar,K.A., Goodenbour,J.M. and Pan,T. (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet.*, **2**, e221.
23. Duret,L. (2000) tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.*, **16**, 287–289.
24. Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K. and Walter,P. (2002) *Molecular Biology of the Cell.* Garland Science, New York.
25. Jackson,R.J., Hellen,C.U. and Pestova,T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, **11**, 113–127.
26. Xu,Y., Ma,P., Shah,P., Rokas,A., Liu,Y. and Johnson,C.H. (2013) Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature*, **495**, 116–120.
27. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
28. Tuller,T., Carmi,A., Vestsigian,K., Navon,S., Dorfan,Y., Zaborske,J., Pan,T., Dahan,O., Furman,I. and Pilpel,Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, **141**, 344–354.
29. Frenkel-Morgenstern,M., Danon,T., Christian,T., Igarashi,T., Cohen,L., Hou,Y.M. and Jensen,L.J. (2012) Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.*, **8**, 572.
30. Novoa,E.M., Pavon-Eternod,M., Pan,T. and Ribas de Pouplana,L. (2012) A role for tRNA modifications in genome structure and codon usage. *Cell*, **149**, 202–213.
31. Kimchi-Sarfaty,C., Oh,J.M., Kim,I.W., Sauna,Z.E., Calcagno,A.M., Ambudkar,S.V. and Gottesman,M.M. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, **315**, 525–528.
32. Sauna,Z.E. and Kimchi-Sarfaty,C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.*, **12**, 683–691.
33. Chamary,J.V., Parmley,J.L. and Hurst,L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.*, **7**, 98–108.
34. Pechmann,S. and Frydman,J. (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.*, **20**, 237–243.
35. Drummond,D.A. and Wilke,C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, **134**, 341–352.
36. Zhou,M., Guo,J., Cha,J., Chae,M., Chen,S., Barral,J.M., Sachs,M.S. and Liu,Y. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, **495**, 111–115.
37. Stadler,M., Artiles,K., Pak,J. and Fire,A. (2012) Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets. *Genome Res.*, **22**, 2418–2426.
38. Ingolia,N.T. (2010) Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol.*, **470**, 119–142.
39. Damelin,S.B. and Miller,W. Jr (2011) *The Mathematics of Signal Processing*. Cambridge University Press, New York.
40. Grushka,E. (1972) Characterization of exponentially modified Gaussian peaks in chromatography. *Anal. Chem.*, **44**, 1733–1738.
41. Akaike,H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Trans.*, **19**, 716–723.
42. dos Reis,M., Savva,R. and Wernisch,L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.
43. Sabi,R. and Tuller,T. (2014) Modeling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res.*, in press.
44. Shaw,L.B., Zia,R.K. and Lee,K.H. (2003) Totally asymmetric exclusion process with extended objects: a model for protein synthesis. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **68**, 021910.
45. Tuller,T., Veksler-Lublinsky,I., Gazit,N., Kupiec,M., Ruppin,E. and Ziv-Ukelson,M. (2011) Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.*, **12**, R110.
46. Newman,J.R., Ghaemmaghami,S., Ihmels,J., Breslow,D.K., Noble,M., DeRisi,J.L. and Weissman,J.S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, **441**, 840–846.
47. Ghaemmaghami,S., Huh,W.K., Bower,K., Howson,R.W., Belle,A., Dephoure,N., O'Shea,E.K. and Weissman,J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.

48. Lee,M.V., Topper,S.E., Hubler,S.L., Hose,J., Wenger,C.D., Coon,J.J. and Gasch,A.P. (2011) A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol. Syst. Biol.*, **7**, 514.

49. Wang,Y., Liu,C.L., Storey,J.D., Tibshirani,R.J., Herschlag,D. and Brown,P.O. (2002) Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci. U.S.A.*, **99**, 5860–5865.

50. Vizcaino,J.A., Cote,R.G., Csordas,A., Dianes,J.A., Fabregat,A., Foster,J.M., Griss,J., Alpi,E., Birim,M., Contell,J. *et al.* (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.*, **41**, D1063–D1069.

51. Taniguchi,Y., Choi,P.J., Li,G.W., Chen,H., Babu,M., Hearn,J., Emili,A. and Xie,X.S. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533–538.

52. Craig,R., Cortens,J., Fenyo,D. and Beavis,R. (2006) Using annotated peptide mass spectrum libraries for protein identification. *J. Proteome Res.*, **5**, 1843–1849.

53. Jakubowski,H. and Goldman,E. (1984) Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J. Bacteriol.*, **158**, 769–776.

54. Seber,G.A. and Lee,A.J. (2012) *Linear Regression Analysis*. John Wiley & Sons, Hoboken, New Jersey.

55. Glantz,S.A. and Slinker,B.K. (1990) *Primer of Applied Regression and Analysis of Variance*. McGraw-Hill, New York, NY.

56. Kendall,M. and Stuart,A. (1979) The advanced theory of statistics. *Inference and Relationship*, 4th edn. **2**, Griffin, London, pp. 1.

57. Dana,A. and Tuller,T. (2012) Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.*, **8**, e1002755.

58. O'Connor,P.B., Li,G.W., Weissman,J.S., Atkins,J.F. and Baranov,P.V. (2013) rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. *Bioinformatics*, **29**, 1488–1491.

59. Dong,H., Nilsson,L. and Kurland,C.G. (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.*, **260**, 649–663.

60. Kanaya,S., Yamada,Y., Kudo,Y. and Ikemura,T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, **238**, 143–155.

61. Rocha,E.P. (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.*, **14**, 2279–2286.

62. Sharp,P.M. and Li,W.H. (1987) The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.

63. Chu,D., Kazana,E., Bellanger,N., Singh,T., Tuite,M.F. and Haar,T. (2014) Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.*, **33**, 21–34.

64. Gustafsson,A., Baverud,V., Gunnarsson,A., Pringle,J. and Franklin,A. (2004) Study of faecal shedding of Clostridium difficile in horses treated with penicillin. *Equine Vet. J.*, **36**, 180–182.

65. Charneski,C.A. and Hurst,L.D. (2013) Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.*, **11**, e1001508.

66. Shabalina,S.A., Spiridonov,N.A. and Kashina,A. (2013) Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Res.*, **41**, 2073–2094.

67. Fedorov,A.N. and Baldwin,T.O. (1997) Cotranslational protein folding. *J. Biol. Chem.*, **272**, 32715–32718.

68. Kramer,G., Boehringer,D., Ban,N. and Bukau,B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.*, **16**, 589–597.

69. Komar,A.A. (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.*, **34**, 16–24.

70. Zhang,G. and Ignatova,Z. (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Curr. Opin. Struct. Biol.*, **21**, 25–31.

71. Hoffmann,A., Becker,A.H., Zachmann-Brand,B., Deuerling,E., Bukau,B. and Kramer,G. (2012) Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell*, **48**, 63–74.

72. Komar,A.A. and Jaenicke,R. (1995) Kinetics of translation of gamma B crystallin and its circularly permutated variant in an in vitro cell-free system: possible relations to codon distribution and protein folding. *FEBS Lett.*, **376**, 195–198.

73. Zhang,G., Hubalewska,M. and Ignatova,Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.*, **16**, 274–280.

74. O'Brien,E.P., Vendruscolo,M. and Dobson,C.M. (2012) Prediction of variable translation rate effects on cotranslational protein folding. *Nat. Commun.*, **3**, 868.

75. O'Brien,E.P., Vendruscolo,M. and Dobson,C.M. (2014) Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat. Commun.*, **5**, 2988.

76. Ikemura,T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. *J. Mol. Biol.*, **158**, 573–597.

77. Bulmer,M. (1987) Coevolution of codon usage and transfer RNA abundance. *Nature*, **325**, 728–730.