

FORMULATION OF THE PROTEIN SYNTHESIS RATE WITH SEQUENCE INFORMATION

WENJUN XIA

Faculty of Science, Jiangsu University
Zhenjiang, Jiangsu 212013, China

JINZHI LEI*

Zhou Pei-Yuan Center for Applied Mathematics
Tsinghua University
Beijing 100084, China

(Communicated by Qing Nie)

ABSTRACT. Translation is a central biological process by which proteins are synthesized from genetic information contained within mRNAs. Here, we investigate the kinetics of translation at the molecular level by a stochastic simulation model. The model explicitly includes RNA sequences, ribosome dynamics, the tRNA pool and biochemical reactions involved in the translation elongation. The results show that the translation efficiency is mainly limited by the available ribosome number, translation initiation and the translation elongation time. The elongation time is a log-normal distribution, with the mean and variance determined by the codon saturation and the process of aa-tRNA selection at each codon binding site. Moreover, our simulations show that the translation accuracy exponentially decreases with the sequence length. These results suggest that aa-tRNA competition is crucial for both translation elongation, translation efficiency and the accuracy, which in turn determined the effective protein production rate of correct proteins. Our results improve the dynamical equation of protein production with a delay differential equation that is dependent on sequence information through both the effective production rate and the distribution of elongation time.

1. Introduction. Translation is a central biological process by which genetic information contained within mRNAs is interpreted to generate proteins. Ribosomes provide the environment for all activities involved in the translation process, including the formation of the initiation complex, the elongation of the translation involving ribosome movement along the mRNA sequence, and the dissociation of the ribosome from the mRNA. Protein synthesis is principally regulated at the initiation stage, and hence, the protein production rate is mainly limited by the availability of free ribosomes [12, 21]. During translations, the ribosome selects matching aminoacylated tRNAs (aa-tRNA) to the mRNA codons from a bulk of non-matching tRNAs. The reaction rate constants can show 350-fold differences in

2010 *Mathematics Subject Classification.* 92C45.

Key words and phrases. RNA translation, efficiency, delay differential equation, noncoding RNA.

This work is supported by National Natural Science Foundation of China (91430101 and 11272169).

* Corresponding author: Jinzhi Lei.

the stability of cognate and near-cognate codon-anticodon complexes [8]. Hence, the translation efficiency is affected by the mRNA sequence and the competition between cognate and near-cognate tRNAs in addition to the initiation stage [6, 18, 27]. The search of global understanding of how the ribosome number, mRNA sequence, and tRNA pool combine to control the translation kinetics has become an interesting topic in recent years due to its potential impact on biogenesis and synthesis biology [9, 15, 17, 21].

Computational models have been developed to investigate details of translation kinetics and to explore the main factors that affect translation efficiency, such as codon bias, tRNA and ribosome competition, ribosome queuing, and codon order [2, 3, 6, 16, 21, 22, 23]. In these models, the statuses of all the ribosomes and tRNAs along an mRNA are tracked in a continuous timeframe. Translation initiation and the availability of free ribosomes were highlighted in previous studies [3, 21, 23]. A model simulation found that variations in translation efficiency were caused by very short times of translation initiation [23]. Using a model that tracked all ribosomes, tRNAs and mRNAs in a cell, the authors concluded that the protein production in healthy yeast cells was typically limited by the availability of free ribosomes; however, protein production under stress was rescued by reducing the initiation or elongation rates [21]. Codon bias of an mRNA sequence is an important factor that may affect translation efficiency due to competitions for tRNAs [2, 3, 6]. A study of the *S. cerevisiae* genome suggested that tRNA diffusion away from the ribosome was slower than translation, and hence, codon correlation in a sequence could accelerate translation because the same tRNA could be used by nearby codons [2]. A cognate, near-cognate, or non-cognate tRNA may attempt to bind to the A site of a ribosome during the elongation process. A study based on a computational model that contains the detailed tRNA pool composition showed that the competition between near-cognate and cognate tRNAs was a key factor that determined the translation rate [6]. Another study using a mean-field model of translation in *S. cerevisiae* showed that the competition for ribosomes rather than tRNAs limited global translation [3]. A model of the stochastic translation process using *E. coli lacZ* mRNA as a traffic problem demonstrated that ribosome collisions can also reduce the translation efficiency [16]. The mechanism for controlling the efficiency of protein translation is evolutionarily conserved based on a calculation of the adaptation between coding sequences and the tRNA pool [26]. Moreover, a nested model of protein translation and population genetics in the genome of *S. cerevisiae* suggested that the codon usage bias of genes could be explained by evolution due to the selection for efficient ribosomal usage, genetic drift, and biased mutation; thus, the selection for efficient ribosome usage is a central force in shaping codon usage at the genomic scale [22].

Despite extensive studies, many of the details underlying the control of translation by mRNA sequences and the cellular environment remain elusive. Both the number of available free ribosomes and the codon order are important for translation efficiency; however, the mechanism by which various factors combine to determine the translation efficiency has not been clearly formulated. Because a codon can be bound by a near-cognate tRNA, proteins with mismatched amino acids can be produced during translations. Hence, the translation accuracy may depend on the codon usage of the sequence and the composition of tRNAs. However, to the best of our knowledge, little is known about this dependence. The relationship between the timing of the ribosome elongation stage, the sequence and the tRNA pool is

closely related to the modeling of genetic network dynamics in which the elongation time is associated with the time delay in dynamical equations [25, 30, 31], but how the elongation time is formulated remains a mystery.

In this paper, translation kinetics are evaluated using a stochastic computation model with detailed reactions of ribosome dynamics and the composition of the tRNA pools. We further investigated how translation efficiency, accuracy, and elongation time are determined through model simulation. Moreover, the translation dynamics of various mRNA sequences (yeast and human, coding and non-coding mRNAs) were studied to clarify whether the sequence is important for translation efficiency and accuracy. Our results show that translation efficiency is mainly limited by the number of available ribosomes, translation initiation and the elongation time of translation. We demonstrate that the elongation time is a log-normal distribution, with the mean and variance of the logarithm of the elongation time dependent on the sequence due to aa-tRNA usages. Moreover, the translation accuracy exponentially decreases with the sequence length. These results provide a more detailed understanding of the translation process and can improve the mathematical modeling of protein production in gene regulation network dynamics.

2. Model and methods.

2.1. Model description. We referred the model of ribosome kinetics during translation established in [6] (Fig. 1). We summarize the model below and refer to [6] for details¹.

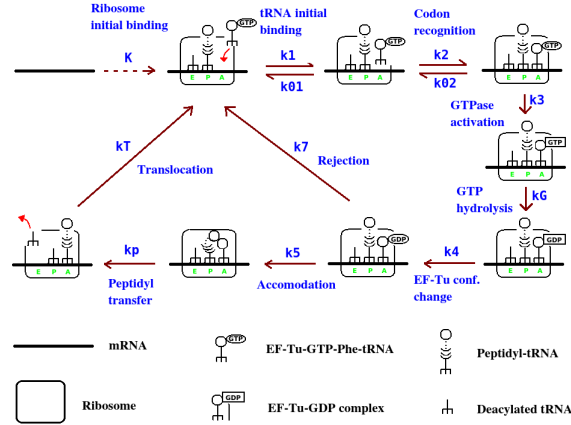


FIGURE 1. Kinetic scheme of RNA translation. Re-drawn from [6].

Protein translation begins from the initiation stage when the start codon (AUG site) of the mRNA sequence is occupied by a ribosome. The peptide between the first two amino acids is formed, with corresponding aa-tRNAs binding to the E and P sites of the ribosome, respectively. Each movement of the ribosome during elongation includes 9 steps, as shown in Fig. 1: initial binding of the aa-tRNA, codon recognition, GTPase activation, GTP hydrolysis, EF-Tu conformation change, rejection, accomodation, peptidyl transfer, and translocation. For each codon on

¹See http://v.youku.com/v_show/id_XNzMxNzEwNjg0.html for an animation of translation. Kindly provided by Prof. Ada Yonath.

TABLE 1. Values of kinetic rate constants (s^{-1}) (refer to [6])

Parameters	Values	Cognate	Near-cognate	Non-cognate
K	0.03	-	-	-
k1	-	140	140	2000
k01	-	85	85	-
k2	-	190	190	-
k02	-	0.23	80	-
k3	-	260	0.4	-
kG	-	1000	1000	-
k4	-	1000	1000	-
k5	-	1000	60	-
k7	-	60	1000	-
kp	-	200	200	-
kT	-	20	20	-

TABLE 2. tRNA pool composition (refer to [5, 6]). Also refer to [6] for the anti-codons for the tRNAs.

tRNA	Molecules/cell	tRNA	Molecules/cell	tRNA	Molecules/cell
Ala1	3250	His	639	Pro3	581
Ala2	617	Ile1	1737	Sec	219
Arg2	4752	Ile2	1737	Ser1	1296
Arg3	639	Leu1	4470	Ser2	344
Arg4	867	Leu2	943	Ser3	1408
Arg5	420	Leu3	666	Ser5	764
Asn	1193	Leu4	1913	Thr1	104
Asp1	2396	Leu5	1031	Thr2	541
Cys	1587	Lys	1924	Thr3	1095
Gln1	764	Met f1	1211	Thr4	916
Gln2	881	Met f2	715	Trp	943
Glu2	4717	Met m	706	Tyr1	769
Gly1	1068	Phe	1037	Tyr2	1261
Gly2	1068	Pro1	900	Val1	3840
Gly3	4359	Pro2	720	Val2A	630
				Val2B	635

the mRNA sequence, tRNAs in the tRNA pool are divided into three types: cognate, near-cognate, and non-cognate (as listed in [6]). All aa-tRNAs can attempt to bind to the A site of the ribosome according to the match between the codon and anticodon [8]; however, only cognate and near-cognate aa-tRNAs can proceed through the step of peptide formation, while non-cognate aa-tRNAs are rejected by codon recognition. Cognate aa-tRNAs yield the correct amino acid following the genetic code, while near-cognate tRNAs often bring incorrect amino acids and yield a defective protein. The reaction rates differ for cognate and near-cognate tRNAs, as reported in [8, 20] and demonstrated in our simulations (Table 1). We note that near-cognate aa-tRNAs are more likely to be rejected at both steps of codon recognition and rejection. Therefore, the competition between cognate and near-cognate tRNAs may be crucial for both the fidelity of peptide synthesis and translation efficiency [6, 8]. After peptidyl transfer, the E site aa-tRNA is released and the ribosome moves forward a codon, leaving the A site free and waiting for the next move. Translation of a polypeptide stops when the ribosome reaches a stop codon (UAG/UAA/UGA), resulting in the release of the polypeptide and the dropping off of the ribosome from the mRNA. One ribosome can synthesize only one polypeptide at a time, and each mRNA can be translated simultaneously by multiple ribosomes. The multiple ribosomes form a queue along the mRNA, with a safe distance of at least 10 codons between two ribosomes [16, 21].

2.2. Numerical scheme. The translation process with multiple ribosomes was modeled with the stochastic simulation algorithm (SSA) [7], which includes the following reactions:

1. binding of a ribosome to the start codon if the first 10 codons are not occupied by ribosomes;
2. binding of an aa-tRNA from the tRNA pool to the A site of an unoccupied ribosome;
3. reactions of codon recognition, energy transformation, and peptide formation;
4. releasing of the tRNA from the E site of a ribosome;
5. translocation of the ribosome to the next codons if the safety condition is satisfied;
6. dropping off of the ribosome once the stop codon is reached.

The kinetic parameters are provided in Table 1, which refer to [6]. The tRNA pool compositions refer to the total number of each tRNA in a yeast cell from [5, 6] and are given in Table 2. To mimic the effects of available tRNAs for each single mRNA translation in our simulations, we used a factor F ($0 < F \leq 1$) for all tRNA numbers to adjust the changes in the numbers of available tRNAs. For the anti-codon of each tRNA and the cognate, near-cognate, and non-cognate for each codon, refer to [6] for details.

The availability of free ribosomes has been shown to be an important limitation for translation efficiency [21]. Here, we introduced a parameter R for the maximum number of available ribosomes that can be used for translation of a single sequence. We note that a ribosome can be re-used after it is released from the stop codon.

An example of translation kinetics obtained from our simulation is provided in Fig. 2. This example shows that the ribosomes sequencing along the mRNA and the amount of protein production increase linearly with the translation time. The average translation rate (amino acids per second) in our simulation is on the order of 10, which is in good agreement with the experimental observations [19]. These results suggest a well-defined translation efficiency, elongation time, and translation accuracy, as discussed given below.

2.3. Translation efficiency, elongation time, and accuracy. To quantify the translation process, we considered the translation efficiency for the protein production rate, the elongation time for the movement kinetics of each individual ribosome, and the translation accuracy for the fidelity of translation. The *translation efficiency* (TE) is defined as the average slope of the increase in the protein production number with the translation time. The *elongation time* of each ribosome is given by the time period from the binding of a ribosome to the start codon to its dropping off from a stop codon. The *elongation time per codon* (ETC ; the average time for a ribosome to move one codon) is often used to describe the translation kinetics. The elongation time is given by $ETC \times L/3$, where L is the length (in *nt*) of an mRNA sequence. Because a protein product may contain mismatched amino acids due to the binding of near-cognate aa-tRNAs with the mRNA, it is possible to have incorrect protein products in the translation. Hence, the ratio of correct proteins in all protein products gives the *translation accuracy*.

3. Results.

3.1. Translation elongation time is a log-normal distribution and sequence dependent. The elongation time measures how long it takes a ribosome to finish

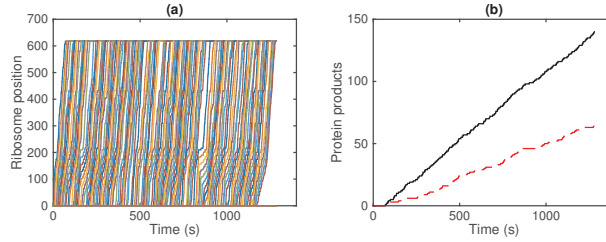


FIGURE 2. Translation kinetics of a single mRNA sequence. (a) Positions of each ribosome on the sequence. (b) Numbers of protein products. The black solid line represents all protein products, and the red dashed line represents the correctly translated proteins (no incorrect amino acid added by near-cognate aa-tRNAs). Here, the sample sequence is the gene YAL003W from the SGD yeast coding sequence, with a sequence length $L = 621nt$. The simulation time, on Mac Pro with 2×3.06 GHz 6-Core Intel Xeon and 16 GB memory, was about 3 min. Parameters are $R = 20$ and $F = 0.03$. For other parameters refer to Table 1.

the translation of a protein, which corresponds to the delay of translation in modeling the dynamics of gene regulation networks through delay differential equations [30, 31]. Protein production can be described by translation efficiency α and mRNA number $M(t)$ through a delay differential equation of form

$$\frac{dP}{dt} = \alpha \int_0^{+\infty} M(t - \tau) \rho(\tau) d\tau, \quad (1)$$

where τ represents the elongation time with a distribution density $\rho(\tau)$. Here we note that the elongation time τ is taken from 0 to $+\infty$ for the mathematical simplicity. Biologically, it is not likely to have infinite elongation time due to the limit life span, however prolonged elongation is still possible by the unexpected traffic jams in translation. Additionally, the finite elongation time can be represented by a density function $\rho(\tau)$ that is non-zero only in a finite subset, and hence Eq. 1 is still valid for a more realistic situation.

Log-normal distribution is often used in biological science for the skewed distributions [14, 24]. In this study, the elongation time came from the accumulation of waiting times of biochemical reactions to complete the translation process, which can be explained with log-normal distribution [14]. To obtain the formulation of the distribution density $\rho(\tau)$, we calculated the elongation time per codon (ETC) during the translation of YAL003W (here we note that $\tau = ETC \times L/3$). The distribution density is shown in Fig 3. The density function was well fitted by the log-normal distribution

$$\ln \mathcal{N}(\mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0. \quad (2)$$

Here, the shape parameters μ and σ are defined so that the logarithm of ETC has mean μ and variance σ^2 . We notice that the fit misses the left tail of the histogram. This may due to the rare effects of fast translation, and remains further investigation in the future.

Let $n = L/3$ be the number of amino acids in a protein product. The density function $\rho(\tau)$ of the elongation time is given from the log-normal distribution Eq. 2 as

$$\rho(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}} e^{-\frac{(\ln \tau - \ln n - \mu)^2}{2\sigma^2}}, \quad (3)$$

and the average elongation time is

$$\bar{\tau} = \int_0^{+\infty} \tau \rho(\tau) d\tau = n e^{\mu + \sigma^2/2}. \quad (4)$$

Here we note that the log-normal distribution Eq. 3 implies a non-zero probability for any large value τ . This is possible following the numerical scheme of stochastic simulation, however not realistic for real biological systems because of the limit life span of a cell. Nevertheless, the probability is very small so that we can neglect it, and the density function Eq. 3 is convenient in terms of mathematical formulation. In the next section, we show that the translation efficiency is dependent on the average elongation time, which enables us to refine our dynamical equation for protein production.

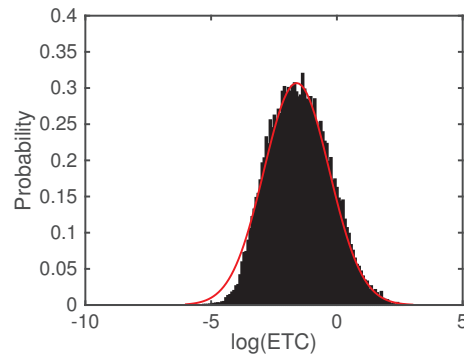


FIGURE 3. Distribution of the elongation time per codon during the translation of YAL003W. All parameters are the same as described in Fig. 2. The red curve is the fit with the normal distribution $\ln \mathcal{N}(-1.6, 1.69)$.

Each move of a ribosome consists of several chemical reactions (shown in Fig. 1), including selections of cognate or near-cognate aa-tRNA from the tRNA pool and a step forward if the safety condition is satisfied. To investigate how the *ETC* depends on the mRNA sequence and translation kinetics, we examined the translation of a set of 1000 sequences from yeast coding genes with length L that vary from 51 to 1995 *nt* (17 to 665 codons). To measure the tRNA usage of each sequence, we calculated the average fraction of cognate, near-cognate, and non-cognate tRNA along the sequence, which are defined as

$$F_\nu = \frac{1}{L/3} \sum_{i=1}^{L/3} \frac{n_{i,\nu}}{\text{Total tRNA number}}, \quad \nu = \text{cog, near, non}, \quad (5)$$

where F_ν ($\nu = \text{cog, near, non}$) measures the average tRNA usage of cognate, near-cognate, and non-cognate tRNAs, respectively. The summation is taken over all codons, and $n_{i,\nu}$ is the number of tRNAs of type ν for codon i along the mRNA sequence.

Fig. 4 shows the dependence of the mean (μ) and variance (σ^2) of the logarithm of *ETC* on tRNA usages. The results suggest that the mean decreases with the cognate tRNA usage F_{cog} , increases with the near-cognate tRNA usage F_{near} , and has no correlation with the non-cognate tRNA usage F_{non} . In contrast, the variance is not dependent on either F_{cog} or F_{near} but weakly decreases with F_{non} . These results suggest that the competition of near-cognate tRNAs tends to increase the elongation time, while the competition of non-cognate tRNAs has only a slight effect on the elongation time. Moreover, Fig. 4 suggests that typical parameters for the distribution of the *ETC* of the yeast coding gene translation are $\mu \approx -1.5$ and $\sigma^2 \approx 1.4$ (refer to Eq. 2). Our simulations suggest no obvious dependence of *ETC* with sequence length L (data not shown).

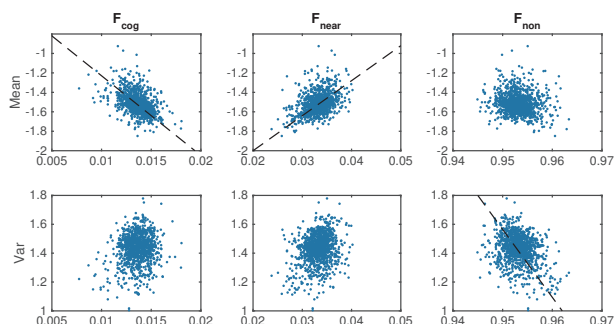


FIGURE 4. Dependence of the *ETC* of yeast coding sequences on tRNA usage. Dots represent the mean (upper panel) and variance (bottom panel) of the logarithm of *ETC* with cognate tRNA usage F_{cog} , near-cognate tRNA usage F_{near} and non-cognate tRNA usage F_{non} . Dashed lines show the linear fitting. Simulations of 1000 yeast coding sequences are shown; each dot corresponds to one sequence. All parameters are the same as described in Fig. 2.

To investigate how the available ribosome number R affects the elongation time, we changed the value R to calculate the dependence of *ETC*. The results showed that both the mean and variance of the logarithm of *ETC* were nonlinearly dependent on R . Indeed, they were mostly independent of R when R was either small or large and demonstrated an obviously increasing dependence when R was an intermediate value (Fig. 5a). A possible reason for the increase of the *ETC* was the traffic jam due to codon occupation. Fig. 5b shows that the average ribosome distance obviously decreased with R in the intermediate region and approached a minimum distance (the safe distance of 10 codons) when R was large. These results reveal that the increasing dependence of the elongation time with ribosome number R ($10 < R < 30$) is due to the increase of traffic jam in the translation kinetics.

The total number of tRNAs was fixed in the above calculations. To further examine how the number of total tRNAs affected the elongation time, we varied factor F from 0.03 to 1 to calculate the dependence of *ETC*. The results showed that both the mean (μ) and variance (σ^2) of the logarithm of *ETC* decreased with F for small value of F and remained nearly unchanged when $F > 0.5$ (Fig. 6). Biologically, these dependencies are obvious because it takes longer to select a cognate or near-cognate tRNA when the tRNA pool is small.

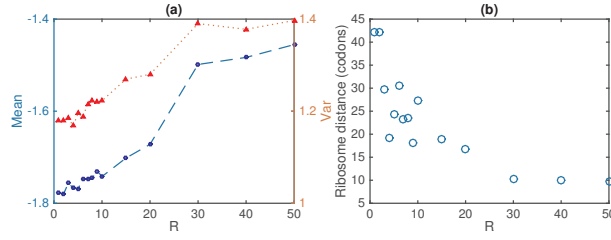


FIGURE 5. Dependence of the elongation time on the available ribosome number R . (a) Average ETC versus R . (b) Ribosome distance (in codons) versus R . The sequence and parameters are the same as described in Fig. 2.

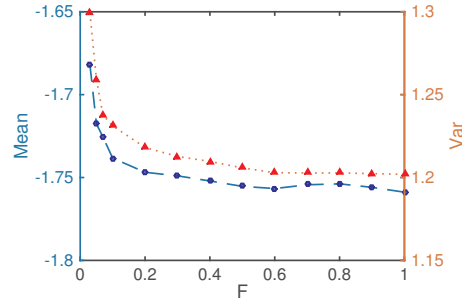


FIGURE 6. Dependence of the ETC on total tRNA number represented by the factor F . The mean (left hand ordinate, blue circles connected with a dashed line) and variance (right hand ordinate, red triangles connected with a dotted line) of the logarithm of ETC are shown as a function of the factor F . The sequence and parameters are the same as described in Fig. 2.

3.2. Translation efficiency is mainly dependent on the elongation time and available ribosome number. Studies in [21, 23] showed that variations in translation efficiency were caused by translation initiation and that availability of free ribosomes was a typical rate limiting step for translation. To investigate how the translation efficiency depended on the translation kinetics and mRNA sequences, we constructed a model to track the dynamics of available ribosomes.

Consider an mRNA with n codons ($n = L/3$). Let R be the number of available ribosomes, $x_i(t)$ ($i = 1, \dots, n$) the number of ribosomes at the i^{th} codon at time t , and $x_0(t)$ the number of free ribosomes. The kinetics of a ribosome during translation is a combination of initiation at a rate K , elongation per codon at a rate c and termination at a rate K_T . Therefore, the dynamics of x_i can be expressed by the following differential equations model

$$\begin{cases} \frac{dx_0}{dt} &= K_T x_n - K x_0 \\ \frac{dx_1}{dt} &= K x_0 - c x_1 \\ \frac{dx_i}{dt} &= c(x_{i-1} - x_i) \quad i = 2, 3, \dots, n-1 \\ \frac{dx_n}{dt} &= c x_{n-1} - K_T x_n. \end{cases} \quad (6)$$

The protein production rate is proportional to x_n . Here, we note

$$0 \leq x_0 \leq R, \quad 0 \leq x_i \leq 1 \quad (i = 1, 2, \dots, n), \quad (7)$$

and

$$\sum_{i=0}^n x_i = R. \quad (8)$$

When $K_T > c$ and R is small, Eq. 6 has a stable equilibrium state that gives

$$x_n = \frac{R}{K_T \frac{n-1}{c} + 1 + \frac{K_T}{K}}. \quad (9)$$

Hence, let $\bar{\tau} = \frac{n-1}{c}$ be approximate to the elongation time (here, we note that $1/c$ corresponds to the average of *ETC*). Then, the translation efficiency satisfies

$$TE \propto \frac{RK}{K\bar{\tau} + 1 + K/K_T}. \quad (10)$$

When R is sufficiently large that all codons are occupied, the translation efficiency is mainly determined by the elongation time so that $TE \propto 1/\bar{\tau}$. Hence, taking Eq. 10 into account, the translation efficiency can be approximated as

$$TE \propto \frac{K \min\{R, R_{\max}\}}{K\bar{\tau} + 1 + K/K_T}, \quad (11)$$

where R_{\max} is the number of available ribosomes required to saturate all codons. We take $R_{\max} = n/10$ in our simulations, which is consistent with Fig. 5. We note that $\bar{\tau}$ is dependent on R according to the discussions above; hence, the relationship in Eq. 11 suggests the following dependence of translation efficiency on the ribosome number R : linear increase when R is small, independent of R when R is large, and nonlinear dependence through the elongation time $\bar{\tau}$ when R takes intermediate values. These results are in agreement with our numerical simulations (Fig. 7).

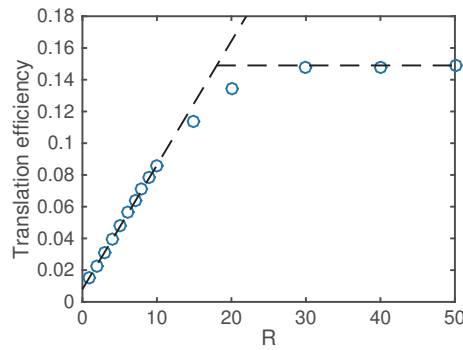


FIGURE 7. Dependence of translation efficiency on the maximum number of available ribosomes R . The dashed lines represent show the two-phase dependence following Eq. 11. The sequence and parameters are the same as described in Fig. 2.

The result of Eq. 11 supports the previous findings that translation initiation and ribosome number are the rate limiting steps of protein production. Moreover, the

translation efficiency decreases with the elongation time, thereby demonstrating the dependence of protein production on the mRNA sequence through the elongation dynamics.

Because the average elongation time $\bar{\tau}$ is proportional to the protein length n , Eq. 11 suggests that the translation efficiency is dependent on the protein length n through a Michaelis-Menten function. Fig. 8a shows translation efficiency versus sequence length for yeast coding sequences with different lengths. The translation efficiency is well fitted by a Michaelis-Menten function, in agreement with our theoretical conclusion based on Eq. 11.

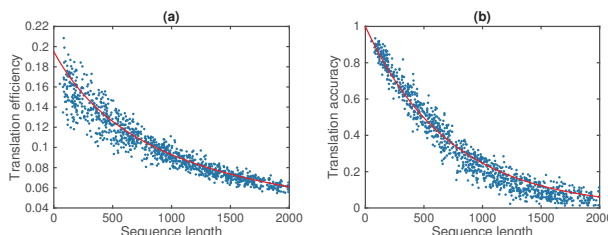


FIGURE 8. Translation kinetics. (a) Translation efficiency versus sequence length for 1000 yeast coding sequences. Red line shows the fitting with $TE = \frac{0.195}{1 + 0.0033n}$. (b) Translation accuracy versus sequence length for 1000 yeast coding genes. Red line shows the fitting with $e^{-0.0042n}$. Here, $n = L/3$ represents the protein chain length. Data were obtained from the simulation shown in Fig. 4.

To further investigate the sensitivity of translation efficiency with the changes in parameters, we increased or decreased each of the parameters in Table 1 and examined the resulting changes in the translation efficiency. The results showed that the translation efficiency was sensitive to changes in k_{02} (or ke_{02} for near-cognate tRNA), k_{01} , k_1 , and k_2 , which corresponded to the process of aa-tRNA selection (Fig. 9). The translation initiation K was also important for the translation efficiency, as we have seen from Eq. 11. Changes in other parameters led to minor changes in the translation efficiency. These results indicate that the steps of aa-tRNA selection are crucial for the translation efficiency through their effects on the elongation time, while changes in the peptide formation steps have minor effects on the translation efficiency.

3.3. Translation accuracy decreases exponentially with sequence length.

During translation, protein products may contain mismatched amino acids when a near-cognate aa-tRNA is selected and successfully forms a peptide. Hence, the translation accuracy (fraction of correct protein products) should exponentially decay with the chain length. The decay rate is associated with the probability of selecting a near-cognate aa-tRNA at each step. Fig. 8b shows the translation accuracy versus sequence length, which is well-fitted with an exponential function.

In living cells, abnormal proteins are usually degraded quickly so that the intracellular amino acids can be recycled efficiently. Hence, only correctly translated proteins are relevant in modeling the dynamics of gene regulation networks. This yields a factor for translation accuracy in the production rate of normal proteins in equation Eq. 1.

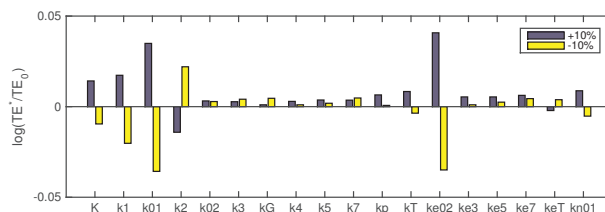


FIGURE 9. Sensitivity analysis of translation efficiency. Bars show changes in the logarithm of translation efficiencies induced by changes in a single parameter $\ln(TE^*/TE_0)$, where TE^* and TE_0 represent the TE for modified and default parameters, respectively. Blue bars correspond to the increase of a parameter by 10%, and yellow bars correspond to the decrease of a parameter by 10%. For parameters refer to Table 1, the parameters ke02, ke3, ke5, ke7, and keT for values of k02, k3, k5, k7, and kT of near-cognate tRNAs (second column in Table 1), respectively, and kn01 for the parameter k01 of the non-cognate tRNAs (third column in Table 1). The sequence and default parameters are the same as described in Fig. 2.

The above discussions suggest a more refined equation for effective protein production using Eq. 1, with $\rho(\tau)$ given by the log-normal distribution Eq. 3 and the effective translation efficiency α given by

$$\alpha = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}}, \quad (12)$$

where parameters a, b depend on the available ribosome numbers, translation initiation and termination and c is related to the composition of the tRNA pool. A crucial refinement of Eq. 12 includes the dependence of the protein chain length n . Other parameters are somehow universal for differential proteins, with the exception of the weak dependence of μ and σ^2 on the sequences shown in Fig. 4 under certain cellular conditions.

3.4. Translation kinetics with sequence dependence. One motivation for this study was to attempt to examine whether there are distinct dynamics for coding and non-coding RNA sequences during translation. We have shown that the translation efficiency depends on the mRNA sequence through the elongation time and that the mean and variance of the elongation time per codon are dependent on the sequence through the aa-tRNA usage. A study of ribosome occupancy showed that many large noncoding RNAs are bound by ribosomes and can hence be translated into proteins [10, 11]. To investigate the translation kinetics of coding and noncoding RNAs, we applied the model simulation to yeast coding RNA, yeast noncoding RNA, human coding RNA and human noncoding RNA. In each sample, we selected 500 sequences with lengths between 200nt and 1000nt; however, most of the noncoding RNAs possess reading frames with lengths less than 300nt, which is in agreement with the observations in [4]. The simulations showed that the previous results can be qualitatively applied to different samples. Fig. 10 shows the distributions of the mean (μ) and variance (σ^2) of the logarithm of the ETC for each set of the simulations (the average of μ and σ^2 for each sample are provided in the table), which are crucial parameters in the density function of Eq. 2. From Fig.

10, we have the following observations: coding RNAs (both yeast and human) have similar distributions in μ and σ^2 , while noncoding RNAs have smaller variances σ^2 compared with the corresponding coding RNAs. These results reveal distinct translation kinetic between coding and noncoding RNAs and will be interesting foundations for studies into the biological significance in the future.

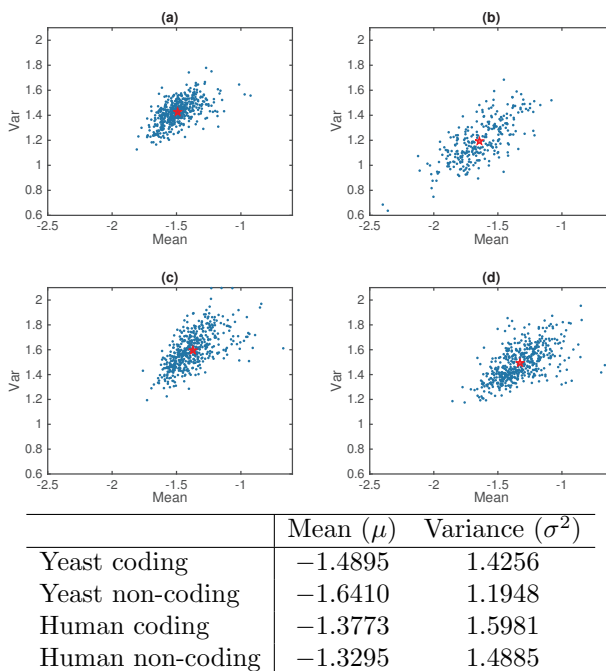


FIGURE 10. *ETC* of the translation for different samples. Distributions of the mean and variance of the logarithm of *ETC* for yeast coding RNAs (a), yeast noncoding RNAs (b), human coding RNAs (c) and human noncoding RNAs (d). Here, the results of 500 random sequences with lengths of $200nt < L < 1000nt$ for each sample are shown. Red stars show the average values for each sample; the values are provided in the table. The parameters are $R = 20, F = 0.03$; for other parameters, refer to Table 1.

4. **Discussion.** We applied a stochastic simulation to study translation kinetics at the molecular level. RNA sequences, ribosome dynamics, the tRNA pool and the biochemical reactions that occur during the elongation step were included in the model. The simulations showed that the *ETC* satisfied a log-normal distribution during translation (Fig. 3) and was mainly determined by both codon saturation (Fig. 5) and the steps of aa-tRNA selection (Fig. 9). During tRNA selection, the relative numbers of near-cognate to cognate aa-tRNAs are crucial for the elongation of a ribosome. Hence, the mean value of the logarithm of *ETC* in this study was dependent on the tRNA usage as defined by F_{cog} and F_{near} (Fig. 4). In the log-normal distribution Eq. 2, the mean μ and variance σ^2 were important for the density function of the elongation time. We showed that these two parameters were slightly different for coding and noncoding RNAs for both yeast and human

samples (Fig. 10). On average, noncoding RNAs have smaller variance than coding RNAs. A simple model of ribosome dynamics revealed that the translation efficiency was mainly determined by the number of available ribosomes, translation initiation and the elongation time; indeed, the translation efficiency was dependent on the elongation time through a Michaelis-Menten function. The translation efficiency increased with the available ribosome numbers when the numbers were small; however, it was insensitive to the ribosome number when the number was sufficiently large to saturate all codons. These results were further confirmed by our simulations. Moreover, the translation accuracy decreased exponentially with the sequence length. These results suggest an improvement for effective protein production when gene expression is modeled in gene regulation networks.

When modeling gene expression, protein production is described by a delay differential equation in the form of Eq. 1 that is dependent on the translation efficiency α and the distribution $\rho(\tau)$ of the elongation time. This study showed that the effective production of correct proteins can be expressed as

$$\alpha = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}}, \quad (13)$$

where n is the protein chain length (number of amino acids), μ and σ^2 are the mean and variance of the logarithm of the *ETC*, respectively, parameters a and b are dependent on the available ribosome number, translation initiation and termination, and c is related to the composition of the tRNA pool. The distribution of the elongation time is formulated as

$$\rho(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}} e^{-\frac{(\ln \tau - \ln n - \mu)^2}{2\sigma^2}}. \quad (14)$$

Hence, the protein production in equation Eq. 1 can be rewritten as

$$\frac{dP}{dt} = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}} \int_0^{+\infty} M(t - \tau) \frac{1}{\tau\sigma\sqrt{2\pi}} e^{-\frac{(\ln \tau - \ln n - \mu)^2}{2\sigma^2}} d\tau, \quad (15)$$

where $M(t)$ is the number of mRNAs at time t . In this equation, the protein chain length n is explicitly included. Moreover, the sequence information is implicitly included in the parameter τ and σ^2 , which are mainly determined by the process of aa-tRNA selection in each step of ribosome movement. However, other parameters are somehow universal under given cellular conditions. With the new techniques of single-molecule approach in cell biology [13, 29], we expect more accurate estimations for these model parameters, which are helpful to improve the quantitative analysis of the kinetics of the translation process. A direct conclusion from Eq. 15 is the extremely low effective production rates due to the long proteins and low translation accuracy for these long chain molecules. This is consistent with biological observations that many transcription factors are small proteins with high production rates (many of them also have high degradation rates) [1], while many structural proteins (*e.g.*, fibers) and transport proteins (*e.g.*, membrane proteins) are large proteins with low production rates (these proteins are mostly very stable) [28]. Hence, this study provides insightful details for known observations and is a valuable resource for further works in whole cell modeling.

Data resources. RNA sequences were downloaded from available databases:

- Yeast coding RNAs from SGD (<http://downloads.yeastgenome.org/>).
- Yeast noncoding RNAs from SGD (<http://downloads.yeastgenome.org/>).

- Human coding RNAs from Ensembl Genome Browser (<http://useast.ensembl.org/>).
- Human noncoding RNAs from Genecode19 (<ftp://ftp.sanger.ac.uk/>).

Acknowledgments. This work was supported by the National Natural Science Foundation of China (91430101 and 11272169). We thank Prof. Zhi Lu in Tsinghua University and his lab members for valuable discussions.

REFERENCES

- [1] M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein and S. A. Teichmann, [Structure and evolution of transcriptional regulatory networks](#), *Curr. Opin. Struct. Biol.*, **14** (2004), 283–291.
- [2] G. Cannarozzi, N. N. Schraudolph, M. Faty, P. von Rohr, M. T. Friberg, A. C. Roth, P. Gonnet, G. Gonnet and Y. Barral, [A role for codon order in translation dynamics](#), *Cell*, **141** (2010), 355–367.
- [3] D. Chu, D. J. Barnes and T. von der Haar, [The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *saccharomyces cerevisiae*](#), *Nucleic. Acids. Res.*, **39** (2011), 6705–6714.
- [4] L. J. Core, A. L. Martins, C. G. Danko, C. T. Waters, A. Siepel and J. T. Lis, [Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers](#), *Nat. Genet.*, **46** (2014), 1311–1320.
- [5] H. Dong, L. Nilsson and C. G. Kurland, [Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates](#), *J. Mol. Biol.*, **260** (1996), 649–663.
- [6] A. Fluitt, E. Pienaar and H. Viljoen, [Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis](#), *Comput. Biol. Chem.*, **31** (2007), 335–346.
- [7] D. T. Gilliespie, [Exact stochastic simulation of coupled chemical reactions](#), *J. Phys. Chem.*, **81** (1977), 2340–2361.
- [8] K. B. Gromadski and M. V. Rodnina, [Kinetic determinants of high-fidelity tRNA discrimination on the ribosome](#), *Mol. Cell*, **13** (2004), 191–200.
- [9] M. Guttman, P. Russell, N. T. Ingolia, J. S. Weissman and E. S. Lander, [Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins](#), *Cell*, **154** (2013), 240–251.
- [10] N. T. Ingolia, S. Ghaemmahami, J. R. Newman and J. S. Weissman, [Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling](#), *Science*, **324** (2009), 218–223.
- [11] N. T. Ingolia, L. F. Lareau and J. S. Weissman, [Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes](#), *Cell*, **147** (2011), 789–802.
- [12] R. J. Jackson, C. U. Hellen and T. V. Pestova, [The mechanism of eukaryotic translation initiation and principles of its regulation](#), *Nat. Rev. Mol. Cell Biol.*, **11** (2010), 113–127.
- [13] G.-W. Li and X. S. Xie, [Central dogma at the single-molecule level in living cells](#), *Nature*, **475** (2011), 308–315.
- [14] E. Limpert, W. Stahel and M. Abbt, [Log-normal distributions across the sciences: Keys and clues](#), *BioScience*, **51** (2001), 341–352.
- [15] Y. Mao, H. Liu, Y. Liu and S. Tao, [Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *saccharomyces cerevisiae*](#), *Nucleic. Acids. Res.*, **42** (2014), 4813–4822.
- [16] N. Mitarai, K. Sneppen and S. Pedersen, [Ribosome collisions and translation efficiency: Optimization by codon usage and mRNA destabilization](#), *J. Mol. Biol.*, **382** (2008), 236–245.
- [17] J. Ninio, [Ribosomal kinetics and accuracy: sequence engineering to the rescue](#), *J. Mol. Biol.*, **422** (2012), 325–327.
- [18] J. B. Plotkin and G. Kudla, [Synonymous but not the same: The causes and consequences of codon bias](#), *Nat. Rev. Genet.*, **12** (2010), 32–42.
- [19] S. Proshkin, A. R. Rahmouni, A. Mironov and E. Nudler, [Cooperation between translating ribosomes and RNA polymerase in transcription elongation](#), *Science*, **328** (2010), 504–508.

- [20] A. Savelsbergh, V. Katunin, D. Mohr, F. Peske, M. Rodnina and W. Wintermeyer, [An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation](#), *Mol. Cell*, **11** (2003), 1517–1523.
- [21] P. Shah, Y. Ding, M. Niemczyk, G. Kudla and J. B. Plotkin, [Rate-limiting steps in yeast protein translation](#), *Cell*, **153** (2013), 1589–1601.
- [22] P. Shah and M. A. Gilchrist, [Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift](#), *Proc. Natl. Acad. Sci. USA*, **108** (2011), 10231–10236.
- [23] M. Siwiak and P. Zielenkiewicz, [A comprehensive, quantitative, and genome-wide model of translation](#), *PLoS Comput. Biol.*, **6** (2010), e1000865.
- [24] S. S. Sommer and N. A. Rin, [The lognormal distribution fits the decay profile of eukaryotic mRNA](#), *Biochem Biophys Res Commun*, **90** (1979), 135–141.
- [25] T. Tian, K. Burrage, P. M. Burrage and M. Carletti, [Stochastic delay differential equations for genetic regulatory networks](#), *J. Comput. Appl. Math.*, **205** (2007), 696–707.
- [26] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman and Y. Pilpel, [An evolutionarily conserved mechanism for controlling the efficiency of protein translation](#), *Cell*, **141** (2010), 344–354.
- [27] T. Tuller, Y. Y. Waldman, M. Kupiec and E. Ruppin, [Translation efficiency is determined by both codon bias and folding energy](#), *Proc. Natl. Acad. Sci. USA*, **107** (2010), 3645–3650.
- [28] G. von Heijne, [Membrane-protein topology](#), *Nat. Rev. Mol. Cell Biol.*, **7** (2006), 909–918.
- [29] X. S. Xie, P. J. Choi, G.-W. Li, N. K. Lee and G. Lia, [Single-molecule approach to molecular biology in living bacterial cells](#), *Annual review of biophysics*, **37** (2008), 417–444.
- [30] L. M. y Terán-Romero, M. Silber and V. Hatzimanikatis, [The origins of time-delay in template biopolymerization processes](#), *PLoS Comput. Biol.*, **6** (2010), e1000726, 15pp.
- [31] E. Zavala and T. T. Marquez-Lago, [Delays induce novel stochastic effects in negative feedback gene circuits](#), *Biophys. J.*, **106** (2014), 467–478.

Received June 04, 2016; Accepted March 16, 2017.

E-mail address: wjqx@ujs.edu.cn

E-mail address: jzlei@tsinghua.edu.cn