STOCHASTIC MODEL OF TRANSLATION ELONGATION
BASED ON CONTINUOUS TIME MONTE CARLO METHOD

Vladimirov N.V.*1, Likhoshvai V.A.1,2

1 Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; 2 Ugra Research Institute of
Information Technologies, Khanty-Mansyisk, Russia
* Corresponding author: e-mail: nikita@bionet.nsc.ru

Keywords: translation, ribosome, codon, secondary structure, elongation, stochastic
modeling, model of translation, Monte Carlo method

Summary

Motivation: Protein synthesis is one of the most energy consuming processes in cell. As a result,
there are strong evidences of optimization at different stages of mRNA translation in many
organisms. In particular, for individual mRNAs their codon composition and local secondary
structures are biased, resulting in high optimization of elongation. To reveal fine patterns of
ribosome traffic along mRNA, development of mathematical model of mRNA translation is required.

Results: A stochastic model of elongation based on continuous time Monte Carlo method has been
developed. The model simulates initiation of translation, ribosome movement along mRNA, tacking
into account the current codon in A-site of ribosome, rate of cognate tRNA binding to the codon,
local secondary structures and ribosome size. A new effect of periodical localization of non-optimal
codons in mRNA has been predicted using the model. This effect has been shown to exist in natural
E. coli genes. The length of the period is equal to the ribosome size obtained in experimental studies,
confirming the model validity and its applicability for solving biological problems.

Introduction

The rate of protein synthesis is one of the most crucial characteristics of cell ontogenesis. The
translation machinery can consume up to 50-70 % of total cell energy (for example, E. coli in
exponential growth phase). Therefore, enhancement in translation efficiency is a significant factor
of evolution. In a number of works the usage of synonymous codons with different rates of
translation was shown to be a widespread mechanism of translation optimization (Ikemura, 1985;
Li, Lou, 1996). The optimal codons with high translation rates are preferred in highly expressed
genes. Such codons correspond to the most abundant cognate tRNAs. It has been shown that
for a number of organisms (E. coli, S. cerevisiae and others) the expression level of genes may be
estimated using the frequency of optimal codons usage in their mRNAs (Li, Lou, 1996). Local
complementary mRNA structures which slow down the ribosome movement may also influence
the rate of translation (Likhoshvai, Matushkin, 2002).

In this work we have developed a stochastic model of translation which takes into account the
rate of initiation, successive movement of ribosome along the mRNA, translation rate of current
codon in its A-site, ribosome pausing at local mRNA hairpins, and steric size of ribosome. The
model has been implemented in a program on the base of continuous time Monte Carlo method.
The program estimates the basic characteristics of translation for given mRNAs: rate of protein
synthesis, level of optimization of codon composition, average number of ribosomes moving
along mRNA, fraction of ribosomes standing in queue, average distance between ribosomes in
polysome. It also allows to obtain elongation speed profiles and variances of the estimates.

Recently it has been shown that modeling of translation within deterministic approach (solving
systems of differential equations) can be carried out only in very crude approximation (Likhoshvai,
Matushkin, 2000). We assume that stochastic methods nowadays are the only approach for detailed
modeling of translation.
The model has been used for studying the patterns of elongation profiles of *E. coli* mRNAs. We have confirmed the well-known effect of frequent usage of rare codons in 5′-proximal coding regions of mRNAs (Bulmer, 1991). Besides, we have predicted one previously unknown effect of periodical localization of non-optimal codons due to the queuing of ribosomes. This effect was shown to exist in real *E. coli* genes. The results are consistent with experimental studies about ribosome steric size (Kozak, 1983).

**Methods and Algorithms**

Coding sequences of *Escherichia coli* K12 genes have been extracted from GenBank database. The rate constants $k_{\text{bind}}^j$ of cognate tRNA binding to the $j$-th codon ($j=1,..,61$) exposed in A-site of ribosome, which determine the rate of codon translation, were calculated using the algorithm described in (Likhoshvai, Matushkin, 2002). The algorithm for $k_{\text{bind}}^j$ calculation uses relative frequencies of codon usage in a sample of genes, which is automatically composed by certain rules.

Elementary events in the model are the following: a) binding of the $i$-th ribosome to the start codon; b) binding of cognate aa-tRNA to the codon exposed in A-site of the $i$-th ribosome; c) translocation of ribosome to the next codon (if next ribosome does not block the way); d) termination of translation.

In the implementation of the model the continuous time Monte Carlo method with the algorithm described in (Gibson, 2000) has been used.

The influence of local mRNA hairpins is considered through increasing the time point of ribosome translocation $t_{\text{trans}}^i$ on small value $\Delta t_{\text{LCI}}$, which has a physical meaning of ribosome pausing for untwining the hairpin. The value $\Delta t_{\text{LCI}}$ is a random variable with probability density:

$$
P_{\text{LCI}} = \frac{LCI}{\sigma + 1} \frac{e^{-kx}}{G(n+1)} dx, \quad k = m/\sigma^2, \quad n = (m/\sigma)^2.
$$

Here $m$ and $\sigma$ are the mean and variance of gamma-distribution, respectively, and LCI (local complementarity index) is the measure of self-complementarity of mRNA region before the ribosome (Likhoshvai, Matushkin, 2002).

**Implementation and Results**

The program is implemented in C++. Ribosomes are independent objects with a set of states (current position on mRNA, type of codon in A-site, etc.) and pointers to the neighboring ribosomes. Polysome is implemented as dynamic list of ribosomes.

Calculation of elongation profiles for a sample of *E. coli* K12 genes confirms the previously known effect of frequent usage of non-optimal codons in the beginning of *E. coli* genes (Bulmer 1991) (Fig. 1). This result supports the validity of calculation algorithm for translation rate constants of individual codons (Likhoshvai, Matushkin, 2002).

We have carried out a numerical analysis to investigate how one non-optimal codon influence the evolutional drift of other codons in its 5′-proximal region of coding sequence. We speculated as follows. Let us consider an mRNA consisting of optimal codons and one non-optimal among them. A ribosome will pause on this non-optimal codon, waiting for cognate tRNA. It is obvious that the following ribosomes will form a queue. The A-sites of queuing ribosomes will be separated from each other by a distance approximately equal to the size of ribosome. The codons in these A-sites will be exposed to weaker natural selection pressure by optimality because changing them to
optimal ones will not increase the overall rate of mRNA translation (ribosomes pause on them anyway). Therefore, we can suppose that in the 5′-proximal region of coding sequence from the non-optimal codon there will appear new non-optimal codons due to mutation process. In this case the distance between these non-optimal codons will be approximately equal to the size of ribosome. The results of simulation for hypothetic mRNA are shown in Fig. 2.

To verify our reasons we have analyzed average translation rates of codons in highly expressed genes of *E. coli* K12, and indeed revealed this effect (Fig. 3). For this purpose we have taken 90 genes with the highest frequency of optimal codons usage, and considered the fragments consisting of the first 50 codons (from #1 to #50). In these fragments we have selected non-optimal codons and calculated average translation rates of all codons at distance of \(D=2,\ldots,d_{\text{max}}\) to the left from the selected ones, in their 5′-proximal regions (\(d_{\text{max}}\) is the distance between selected non-optimal codon and the start one). Results are shown in Fig. 3.

---

**Fig. 1.** Averaged elongation profile for mRNAs of highly expressed *E. coli* genes. The horizontal axis shows the number of codon in the ORF ( #1 is the codon next to the start one). The vertical axis (\(V_{\text{rib}}\)) is the average speed of ribosome (codons/sec).

**Fig. 2.** Elongation profile of hypothetic mRNA consisting of optimal codons with high translation rate, and one non-optimal in position 60. \(D_0\) is the putative size of ribosome. Axes are the same as in Fig. 1.

**Fig. 3.** Average translation rates of codons in the 5′-proximal regions of non-optimal codons. Horizontal axis D is the distance from a non-optimal codon, vertical axis is the average translation rate of codons.
The length of period $D_0$ obtained is $11 \pm 1$ codons. The decreases of average translation rates of codons repeat at the distances of $11, 21, 32, 42$ codons ($D_0, 2D_0, 3D_0, 4D_0$, respectively) to the left from non-optimal codons (Fig 3), in accordance with the effect predicted on hypothetic mRNA (Fig. 2). The length of period ($33 \pm 3$ nucleotides) with accuracy to $3$ nucleotides is equal to the steric size of ribosome from experimental studies (Kozak, 1983) – about $30$ nucleotides. This result sustains the validity of our hypotheses and the applicability of the model for solving biological problems related to codon biases.

The model developed may be applied for solving problems of evolution and bioinformatics related to optimization of codon composition of genes.

**Acknowledgements**

The work was supported by the grants No. 03-07-96833, 03-04-48829, 03-04-48506, 03-01-00328, 02-07-90359, 02-04-48802 of the RFBR, by the Interdisciplinary grant No. 119 and the Project No. 10.4 of the RAS Presidium Program “Molecular and Cellular Biology”, Integration project of SB RAS No. 148.

**References**


