BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 12 2012, pages 1663–1664
doi:10.1093/bioinformatics/bts185

Gene expression

Advanced Access publication April 11, 2012

RFMapp: ribosome flow model application

Hadas Zur[1] and Tamir Tuller[2,*]

[1]The Blavatnik School of Computer Science, Faculty of Exact Sciences and [2]Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

Associate Editor: Trey Ideker

ABSTRACT

Summary: The RFMapp is a graphical user interface application based on the RFM (ribosome flow model), enabling the estimation of the translation elongation rates of messenger ribonucleic acids (mRNAs) and the profile of ribosomal densities along the mRNAs, in a computationally efficient way. The RFMapp is based on the approach previously described by Reuveni *et al*., and unlike other traditional approaches in the field, which are mainly related to the genes' mean codon translation efficiency, the RFM additionally considers the codon order, the ribosomes' size and their order. Thus, it has been shown that RFM outperforms traditional predictors when analyzing both heterologous and endogenous genes.

Availability and implementation: Distributable cross-platform application and guideline are available for download at:
http://www.cs.tau.ac.il/~tamirtul/RFM_Installers/install.htm

Contact: tamirtul@post.tau.ac.il

Received on February 15, 2012; revised on February 15, 2012; accepted on April 2, 2012

1 INTRODUCTION

Gene translation is a complex process through which a messenger ribonucleic acid (mRNA) sequence is decoded by the ribosome to produce a specific protein. The elongation step of this process is an iterative procedure in which each codon in the mRNA sequence is recognized by a specific transfer RNA (tRNA), which adds one additional amino acid to the growing peptide. As gene translation is a central process in all living organisms, its understanding has important ramifications to every biomedical field, namely human health (Kimchi-Sarfaty *et al*., 2007), biotechnology (Gustafsson *et al*., 2004) and evolution (Tuller *et al*., 2011).

In recent years, there has been a sharp increase in emerging technologies for measuring different features related to the process of gene translation (Ingolia *et al*., 2009; Schwanhausser *et al*., 2011). However, this process is still enigmatic and contradicting conclusions regarding the essential parameters that determine translation rates appear in different studies (Plotkin and Kudla, 2010); suggesting that the rate-limiting parameters vary across organisms, under diverse conditions and in different genes.

Ribosome flow model (RFM) is a new approach for modeling the process of translation elongation, affording the first large-scale analysis of gene translation elongation, while taking into account the stochastic nature of the translation process. RFM is aimed at capturing the effect of codon order and composition on translation rates, the interactions between ribosomes and the characteristics of the translation elongation process on all its various physical aspects.

The main advantage of the RFM is 2-fold. First, it considers additional aspects of translation elongation, in contrast with the most widely used predictors of elongation that are based on the coding sequence, such as the codon adaptation index (Sharp and Li, 1987) and the tRNA adaptation index (dos Reis *et al*., 2004).

Second, it is orders of magnitude computationally faster than more comprehensive models of translation elongation, such as the totally asymmetric simple exclusion process (TASEP; first suggested by Heinrich and Rapoport, 1980), whereas its predictions are relatively similar to the ones obtained by the TASEP.

RFM is a simple, physically plausible computational model that is solely based on the coding sequence (i.e. a vector of codons in each gene) and on an additional parameter (initiation rate). The model allows for a computationally efficient analysis of the translation process on a genome-wide scale and across all species. Focusing on the coding sequence, we by no means to imply that it is the only factor taking place in the determination of translation rates. Nevertheless, as it has been widely recognized as a prime factor in the translation elongation process, we chose to concentrate on it.

Herein, we present RFMapp, a distributable cross-platform, graphical user interface application based on the approach of Reuveni *et al*. (2011). The new tool will serve the community by enabling the prediction of fundamental features of the gene translation process, including translation rates, ribosomal densities and the relationship between all these variables. Protein abundance levels are proportional to the multiplication of the predicted translation rates and mRNA levels and can be extrapolated if mRNA levels are available (Reuveni *et al*., 2011). We provide below a high-level description of RFMapp, with a small-scale example.

2 APPLICATION DESCRIPTION

The usage of RFMapp is straightforward. The input is genes' coding sequences, codon translation times, initiation rate (which can either be global/identical for all the organism's genes or specific per gene) and number of approximated translation sites (chunk size), for a certain organism. The output is a text file of the predicted translation features and an optional graphical-predicted ribosomal density profile. RFMapp provides the genomic data (coding sequences and codon translation times) for six model organisms (for a list of supported organisms, see the RFMapp

*To whom correspondence should be addressed.

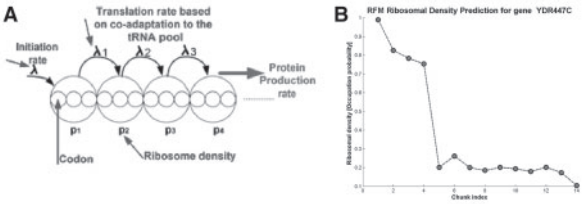

Fig. 1. (**A**) An illustration of the RFM model. (**B**) An example ribosomal density profile of the ribosomal protein RP51B (YDR447C)—an output of RFMapp

User Guide) and supports both organisms across all the domains of life and synthetic simulations.

The RFM has two free parameters: the initiation rate λ and the number of codons C at each 'site' (can be proportional to the ribosome size). Each site has a corresponding transition rate λ, that is estimated based on the co-adaptation between the codons of the site and the tRNA pool of the organism [see the Methods section of Reuveni *et al*. (2011)]. The output of the model consists of the steady-state occupancy probabilities of ribosomes at each site and the steady-state translation rates or ribosome flow through the system (Fig. 1A).

As aforementioned, the input to the RFMapp includes a coding sequence, estimated translation time for each codon, initiation rate (λ) and chunk size (C). The mRNA molecules are coarse grained into sites of codons according to the chunk size and the coding sequence (in Fig. 1A, $C=3$); ribosomes arrive at the first site with an initiation rate λ but are only able to bind if this site is not occupied by another ribosome. The rate of each chunk is computed by the RFMapp based on the translation times of the organism's codons and the codon composition of the chunk. In this application, the transition time of a codon can be determined by the tRNA pool of the organism. Briefly, taking into account the affinity between tRNA species and codons, the translation rate of a codon is proportional to the abundance of the tRNA species that recognize it (Reuveni *et al*., 2011).

By the RFM, a ribosome that occupies the ith site moves, with rate λ, to the consecutive site provided the latter is not occupied by another ribosome. Denoting the probability that the ith site is occupied at time t by $P_i(t)$, it follows that the rate of ribosome flow into/out of the system is given by $\lambda(1-P_1(t))$ and $\lambda_n P_n(t)$ respectively. The rate of ribosome 'flow' from site i to site $i+1$ is given by $\lambda_i P_i(t)(1-P_{i+1}(t))$. The RFM determines the steady-state solution of the occupation probabilities of the equations below and specifically the rate of protein production at steady state.

$$\begin{cases} \frac{dp_1(t)}{dt} = \lambda[1-p_1(t)] - \lambda_1 p_1(t)[1-p_2(t)] \\ \frac{dp_i(t)}{dt} = \lambda_{i-1} p_{i-1}(t)[1-p_i(t)] - \lambda_i p_i(t)[1-p_{i+1}(t)]_1 < i < n \\ \frac{dp_n(t)}{dt} = \lambda_{n-1} p_{n-1}[1-p_n(t)] - \lambda_n p_n(t) \end{cases}$$

3 SMALL-SCALE EXAMPLE

We describe the application of RFMapp on *Saccharomyces cerevisiae* gene.

We select the initiation time to be 0.06 and the chunk size 10 codons (approximate eukaryote ribosome footprint). The gene FASTA file is as follows:

> YDR447C
ATGGGTAGAGTTAGAACCAAGACCGTCAAACGTGCCTCCA
AGGCTTTGATTGAACGTTACTATCCAAAGTTGACCTTGGAT
TTCCAAACTAACAAGAGACTTTGTGATGAAATTGCAACTA
TCCAATCCAAGAGATTGAGAAACAAGATTGCTGGTTACAC
TACTCATTTGATGAAAAGAATCCAAAAGGGTCCAGTTAGA
GGTATTTCTTTCAAATTGCAAGAAGAAGAAAGAGAAAGA
AAGGATCAATACGTCCCAGAAGTCTCTGCTTTGGACTTGT
CTCGTTCTAACGGTGTTTTGAACGTTGACAACCAAACCTC
TGACTTGGTTAAATCTTTGGGTTTGAAGTTGCCATTATCTG
TCATCAACGTTTCCGCTCAAAGAGACAGACGTTACAGAA
AGAGAAACTAA.

A text file containing the organisms' 61 codon translation rates must be provided. The textual output containing the predicted translation rate, and occupation probabilities at steady state (ribosomal density profile), is as follows:

> YDR447C
Translation rate = 0.000599
Occupation probabilities = 0.99002 0.82625 0.78358 0.75364 0.2011 0.26172 0.19952 0.18489 0.20107 0.19337 0.17971 0.20113 0.17339 0.10386

The graphical-predicted ribosome density profile is shown in Figure 1B.

ACKNOWLEDGEMENT

H.Z. is supported by the Edmond J Safra Bioinformatics program at Tel Aviv University.

Conflict of Interest: none declared.

REFERENCES

dos Reis,M., *et al.* (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**, 5036–5044.

Gustafsson,C., *et al*. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.*, **22**, 346–353.

Heinrich,R. and Rapoport,T.A. (1980) Mathematical modelling of translation of mRNA in eucaryotes; steady state, time-dependent processes and application to reticulocytes. *J. Theor. Biol.*, **86**, 279–313.

Ingolia,N.T., *et al*. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218.

Kimchi-Sarfaty,C., *et al*. (2007) A "Silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, **315**, 525–528.

Plotkin,J.B. and Kudla,G. (2010) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.

Reuveni,S., *et al*. (2011) Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput. Biol.*, **7**, e1002127.

Schwanhausser,B., *et al*. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.

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.

Tuller,T., *et al*. (2011) Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res.*, **39**, 4743–4755.