

## A novel and versatile computational tool to model translation

Dominique Chu<sup>1,\*</sup>, Nicolae Zabet<sup>2</sup> and Tobias von der Haar<sup>3</sup><sup>1</sup>School of Computing, University of Kent, Kent CT2 7NF, <sup>2</sup>Department of Genetics, University of Cambridge, CB2 1QR and <sup>3</sup>School of Biosciences, University of Kent, Kent CT2 7NJ, UK

Associate Editor: Trey Ideker

### ABSTRACT

**Motivation:** Much is now known about the mechanistic details of gene translation. There are also rapid advances in high-throughput technologies to determine quantitative aspects of the system. As a consequence realistic and system-wide simulation models of translation are now feasible. Such models are also needed as devices to integrate a large volume of highly fragmented data known about translation.

**Software:** In this application note, we present a novel, highly efficient software tool to model translation. The tool represents the main aspects of translation. Features include a representation of exhaustible tRNA pools, ribosome–ribosome interactions and differential initiation rates for different mRNA species. The tool is written in Java, and is hence portable and can be parameterized for any organism.

**Availability:** The model can be obtained from the authors or directly downloaded from the authors' home-page (<http://goo.gl/JUWVl>).

**Contact:** d.f.chu@kent.ac.uk

Received and revised on September 26, 2011; accepted on November 7, 2011

### 1 INTRODUCTION

Progress in biotechnological methods means that we now have a very good working knowledge of the biological components that are essential for gene translation. Moreover, rapid improvement of quantitative methods in high-throughput biology enables us to obtain kinetic rate constants and other parameters of the system. To date there are no available tools that allow the user to develop a quantitative system-wide understanding of translation from the available data. In this application note, we present a generic model of translation that can be used precisely for this purpose. It can be parameterized to describe translation of any eukaryotic or prokaryotic cell. The model can be used to answer fundamental questions about the dynamics of translation, including its limiting factors under various conditions. Moreover, it could also be of significant practical and industrial utility, for example within bioprocessing to understand how recombinant protein impacts on the translation of endogenous genes.

*Prima facie* translation is a driven walk of ribosomes over a sequence of codons contained on the mRNA. A ribosome can only progress to the next codon if an aa-tRNA has bound to it and delivered its amino acid to the growing polypeptide. There are a number of couplings between system variables. Transcripts interact with one another via the competition for aa-tRNA and ribosomes.

Within a polysome, ribosomes may collide with one another potentially resulting in 'traffic jams' that may slow down translation significantly. The relative importance of these interactions is parameter and hence species dependent.

There have been many previous attempts to model translation. None of those provides a system-wide and dynamic representation of the system. The biophysical mechanism of polypeptide elongation as an 18-step process that has first been described by Gromadski and Rodnina (2004). Zouridis and Hatzimanikatis (2007) have performed a sensitivity analysis of the variables involved in the translation process and also modelled translation of a single transcript. Romano *et al.* (2009) modelled translation as a totally asymmetric exclusion process. While they provide analytic insights specifically illuminating the effects of ribosome–ribosome interactions on transcripts, their model assumptions are by necessity too simplified to provide a comprehensive insight into the various factors influencing translation dynamics. Chu *et al.* (2011) presented a stochastic mean-field model of translation where the heterogeneity of the model transcript population is represented by its average sequence. While this model reproduces overall features of the translation system well, its mean-field assumption leaves out crucial aspects relating to transcript–transcript competition for ribosomes as well as translation initiation. Most recently, Reuveni *et al.* (2011) presented a model based on a master equation approach. This model does not represent free ribosomes explicitly, nor does it take into account competition for aa-tRNA or ribosomes.

### 2 RESULTS

The tool described here uses a hybrid approach mixing agent-based techniques with event-driven stochastic simulation algorithms (Gillespie, 1977). The individual mRNAs are explicitly represented as strings of codons; they also contain an 'initial binding site' that summarizes the 5'-UTR of every transcript. Each transcript can be occupied by one or more ribosomes. The spatial arrangement of codons makes it necessary to keep track of the position of individual ribosomes on the mRNA; consequently, ribosomes are represented as agents encapsulating information about the particular transcript they are bound to, the position thereon as well as information about the tRNA they are interacting with. Individual tRNA molecules on the other hand are assumed to be perfectly mixed within the cell compartment and need not be accounted for individually. Therefore, the model merely tracks the number of amino-acylated, uncharged, bound and unbound tRNAs. The total numbers of tRNA, mRNA and ribosomes are assumed to be unchanged over the course of a simulation run.

\*To whom correspondence should be addressed.

The simulation is driven by discrete events that are scheduled using Gillespie's algorithm. These events are as follows:

- (i) Binding of free ribosomes to an initial binding site on a transcript.
- (ii) Binding of ribosomes located on an initial binding site to the first codon (AUG site).
- (iii) Binding of near/non-cognate aa-tRNAs to unoccupied bound ribosomes.
- (iv) Unbinding of near/non-cognate tRNAs from a bound ribosome.
- (v) Translocation of ribosomes to the next codon.
- (vi) Aminoacylation of uncharged tRNA.
- (vii) Re-initiation/dissociation of ribosome once the ribosome has reached the end of the ORF.

In real systems, the rate with which free ribosomes bind to the 5' cap of the transcript is partially determined by structural properties of the 5'-UTR. Furthermore, translation initiation is a multi-step process that involves the assembly of ribosomal subunits, scanning of the 5'-UTR (von der Haar and McCarthy, 2002) but also regulatory interactions at  $\mu$ -ORFs. Simulating these processes within the context of a system-wide model of translation would not be possible. Moreover, it would not be useful because the relevant parameters of initiation are poorly known. Hence, for reasons of parsimony and computational efficiency this complex sequence of events was collapsed into a single initiation step (Event (i)) following second-order kinetics with a mRNA-type specific rate constant. Translation proper starts when the ribosome binds to the first AUG site (Event (ii)). This step follows 0-th order dynamics with a rate that is ORF specific.

Movement of a ribosome from one codon to the next requires an aa-tRNA to bind to the ribosomal binding site, to pass a proof-reading step and to deliver the amino acid. Gromadski and Rodnina (2004) described this in detail as an 18-step process. Within the present model, this is reduced to three essential steps, corresponding to events (iii) to (v). The model assumes that both cognate and near-cognate (but not non-cognate) aa-tRNAs bind to the ribosome. Once an aa-tRNA is bound to a ribosome, it is decided probabilistically whether or not the tRNA is accepted or rejected. The probabilities depend on whether the aa-tRNA is cognate or near-cognate. If it is accepted then an event of type (v) is scheduled. Otherwise the rejected aa-tRNA unbinds during an event of type (iv) and the aa-tRNA is returned to the pool. It is generally thought (Fluitt *et al.*, 2007) that significant amounts of time are taken up by aa-tRNA probing the ribosomal binding site without transferring the amino acid.

A translocation event (v) results in a de-aminoacylation of the tRNA at the ribosomal P-site. The unbinding of the spent tRNA is delayed until the next elongation cycle is completed. Following the exit of the uncharged tRNA, an aminoacylation event (type (vi)) is scheduled using Gillespie's algorithm assuming first-order kinetics and a fixed rate constant that is equal for all tRNA species (Chu *et al.*, 2011).

Once a ribosome has finished translating the mRNA, there are two possibilities. Either it is returned to the pool of free ribosomes

or it re-binds again at the first AUG site to go through another round of translation. Each re-initiation event is a stochastic event. In addition, the program allows the user to set a maximum number of re-initiations per ribosome.

Ribosomes on the mRNA have an excluded volume that can be set by the user. By default, it is assumed that a ribosome occupies six codons to either side of its central codon, preventing other ribosomes to occupy this space. During elongation, and also initiation a situation may occur where a ribosome attempts to bind to excluded codons. If this happens, then the ribosome does not proceed with its movement until the relevant site is freed again.

### 3 DISCUSSION

The software is written in Java (version 1.6) and runs on any computer platform that supports a Java virtual machine. Parameters are entered via parameter files. The user can select the ORFs to be expressed (using FASTA format) and specify mRNA copy numbers and ribosome-mRNA affinities for each ORF. During execution, the program outputs data whenever a ribosome has finished a translation round. This output includes the ORF that has just been translated, the translation time, the length of the ORF and the proportion of free ribosomes. Upon completion, the program also outputs a summary file with comprehensive statistics about each codon and ORF including translation error, usage rates (for codons and tRNA) and total translation numbers and polysome sizes for individual ORFs. The run-time of the model depends on the parameters, the number of transcripts to be simulated and the number of ribosomes. Using realistic parameters for *Saccharomyces cerevisiae*, we found that 15 000 s of translation can be simulated within 5 days using a single core of a 2 GHz Intel Xeon processor.

**Funding:** BBSRC/BRIC grant (BB/I010351/1); Royal Society research grant (RG090115).

**Conflict of Interest:** none declared.

### REFERENCES

- Chu, D. *et al.* (2011) The role of tRNA and ribosome competition in coupling the expression of different mRNAs in *saccharomyces cerevisiae*. *Nucleic Acids Res.*, **39**, 6705–6714.
- Fluitt, A. *et al.* (2007) Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.*, **31**, 335–346.
- Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Gromadski, K. and Rodnina, M. (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell*, **13**, 191–200.
- Reuveni, S. *et al.* (2011) Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput. Biol.*, **7**, e1002127.
- Romano, M. *et al.* (2009) Queueing phase transition: theory of translation. *Phys. Rev. Lett.*, **102**, 198104.
- von der Haar, T. and McCarthy, J. (2002) Intracellular translation initiation factor levels in *saccharomyces cerevisiae* and their role in cap-complex function. *Mol. Microbiol.*, **46**, 531–544.
- Zouridis, H. and Hatzimanikatis, V. (2007) A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. *Biophys. J.*, **92**, 717–730.