Simulating and fitting stochastic models of RNA transcription via the Gillespie algorithm

Gennady Gorin[1], Mengyu Wang[1,2,3], Heng Xu[4,5], and Ido Golding[1,2,6]

***Short Abstract* — Testing mechanisms of RNA transcription requires predictive models and experimental data. We present an implementation of the Gillespie algorithm that simulates the stochastic kinetics of nascent (actively transcribed) and mature RNA, including two- and three-state gene regulation, RNA synthesis initiation and stepwise elongation, release to the cytoplasm, and stepwise degradation. To facilitate comparison with experimental data, the algorithm predicts probe signals measurable by single-cell imaging. By minimizing statistical distance from experimental signal distributions, we can estimate underlying parameters.**

***Keywords* — Transcription kinetics, single-cell imaging, parameter estimation.**

I. Background

THE transcription of RNA is the product of both stochastic and deterministic dynamical processes in the cell [1]. The wealth of hypotheses about these processes motivates the development of a modular framework to test proposed mechanisms and quantify their kinetic parameters. Due to challenges in producing closed-form solutions to arbitrary kinetic models [1], simulations offer an attractive alternative. We present a simulation platform that easily incorporates new mechanisms, offers a graphical user interface (GUI), provides outputs comparable to experimental measurements, and efficiently scales when scanning kinetic parameters.

II. Methods

We modified the Gillespie algorithm to simulate the kinetics of nascent and mature RNA. Stochastic reactions model initiation of transcription, stepwise elongation along the DNA strand, release to the cytoplasm, attachment of RNase and stepwise degradation [2], while stochastic or deterministic schema model gene regulation. The algorithm is implemented in MATLAB and can be scaled using distributed computing, such as the Amazon Web Services (AWS) cloud. Probe signals predicted by the simulation are verified using differential equation solutions and the finite state projection (FSP) algorithm [3]. Parameter estimation uses the genetic algorithm to fit synthetic FSP data and single-molecule fluorescence *in situ* hybridization (smFISH) measurements of nascent and mature RNA in cells [4].

III. Results

The simulation reproduces the average smFISH probe signals of the nascent and mature RNA populations predicted by the differential equation solution and the corresponding copy-number histograms predicted by the FSP algorithm. It has already been modified to incorporate multiple gene copies, deterministic changes in kinetic parameters, variable transcription speeds, co-transcriptional degradation by RNase, and other features. A GUI is available, enabling parameter input and inspection of graphical outputs.

Using cloud computation, the algorithm can generate (within minutes to hours) time-dependent signal predictions for hundreds of kinetic parameter sets, each applied to thousands of cells. For parameter estimation, we bin the predicted measurements to generate time-dependent histograms for each parameter set, calculate their statistical distance away from "target" histograms, and generate new parameter sets by using the genetic algorithm to minimize the statistical distance. We currently fit synthetic data to determine the best methods for fitting to experiments.

IV. Conclusion

The platform facilitates testing and fitting of transcription models, as well as casual use via the GUI. These capabilities make it a useful addition to the toolbox of quantitative biology.

Acknowledgements: This work was partially supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (PHY-1427654). This research was also funded by the National Science Foundation (PHY-1430124 and PHY-1147498), National Institute of Health (R01 GM082837), Welch Foundation (Q-1759) and John S. Dunn Foundation.

[1]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX

[2]Center for Theoretical Biological Physics, Rice University, Houston, TX

[3]Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX

[4]Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, P.R. China

[5]School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, P.R. China

[6]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL

References

[1] Xu H, et al. (2016) Stochastic Kinetics of Nascent RNA. *Phys Rev Lett* ***117***, 128101.

[2] Chen H, et al. (2015) Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Mol Syst Biol* **11**, 781.

[3] Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* **124**, 044104.

[4] Skinner SO, et al. (2013) Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat Protoc* **8**, 1100-1113.