Modeling, Computation, and Characterization to Accelerate the Development of Synthetic Gene Circuits in Cell-Free Extracts

Thesis by

Vipul Singhal

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2019

(Defended June 11, 2018)

Vipul Singhal

ORCID: 0000-0003-1670-1824

Dedicated to my Family

Acknowledgments

I want to start by thanking my advisor, Richard Murray. Richard, among the many things that I am grateful for, the biggest has been your patience while I struggled to develop my scientific voice. I also want to thank you for the incredible generosity you have shown with your time, and for always providing me with any resource or advice I needed.

I am grateful to Professors Lea Goentoro, Erik Winfree and Matthew Thomson for their support, and for being on my committee. Erik and Lea, you provided me with insightful feedback, which my research has benefited from greatly. Matt, though we have just met, I am looking forward to picking your brain about what mathematics can contribute to biology.

I want to express my heartfelt thanks to the faculty of the Computation and Neural Systems program, and in particular Thanos Siapas, Erik Winfree, Pietro Perona and Christof Koch, for giving me the chance to pursue my graduate studies as part of this program. I could not have asked for a better experience. I must also thank Eduardo Sontag for always being willing to discuss theory work at a moment's notice. Your enthusiasm is contagious!

I must thank my lab mates, Sam Clamons, Andy Halleran, James Parkin, Mark Prator, Reed McCardell, Shaobin Guo, Anu Thubagere, Andrey Shur, Yong Wu, Yutaka Hori, Zoltan Tuza and Abel Chiao for the fantastic working environment they created. I also want to thank Victoria Hsiao for being a shining example of how to maintain balance during grad school; Enoch Young for the long mathematical and philosophical digressions we took; Zach Sun for all that work we did on the project that never made it here, yet informed so much of my subsequent thinking; William Poole, Sam Clamons and Wolfgang Halter for extremely useful discussions on the identifiability theory work; and Jongmin Kim for being (much) better than Google at answering all sorts of research questions I had during my

early years. It would probably not be too much of an exaggeration to say that discussions with Anandh Swaminathan have probably helped me avoid about half of the bad ideas I was able to avoid, and adopt half of the good ones I adopted. Last but not least, I would like to thank Miryong Yun, our lab manager, whose efficiency and professionalism have been enabling forces to be reckoned with.

A big part of the summers here were spent mentoring students, and I would be remiss if I did not mention them. Flora Meng, Cody Dunn, Ronnie Rodrigues Pereira , Miroslav Gasparek, Anushka Rau, Enrique Amaya, Tiffany Zhou, Pulkit Malik and Anton Frisk, you were an absolute inspiration to work with. Co-mentors Clare Hayes and Ania Baetica: this task would have been impossible without you. Clare, I must also thank you for teaching me lab work, which I imagine was not easy.

Tristan McKinney, Stephen Perry, Howard Hui and Anandh Swaminathan, you guys have been my support system away from home, and the countably infinite adventures we have had will forever be with me. The ski/snowboard crew, Karthik, Eric, Bassam, Anandh, thanks for abandoning me on the mountain during those early days, and forcing me to up my skills. Your hands-off approach worked, and so did the knee pads. I must also mention all the other people who have helped make grad school fun: Brian Brophy and the storytelling class of 2014, OASIS, and the badminton club.

Most importantly, I want to thank my family. Mom, Dad, Niki, Tashu, Neha Didi, and Sanchit Jijaji. Without you, none of this would be possible.

Abstract

Synthetic biology may be defined as an attempt at using engineering principles to design and build novel biological functionalities. An important class of such functionalities involves the bottom up design of genetic networks (or 'circuits') to control cellular behavior. Performing design iterations on these circuits *in vivo* is often a time consuming process. One approach that has been developed to address these long design times is to use *E. coli* cell extracts as simplified circuit prototyping environments. The analogy with similar approaches in engineering, such as prototyping using wind tunnels and breadboards, may be extended by developing accompanying computer aided design tools. In this thesis, we discuss the development of computational and mathematical tools to accelerate circuit prototyping in the TX-TL cell free prototyping platform, and demonstrate some applications of these tools.

We start by discussing the problem of reducing circuit behavior variability between different batches of TX-TL cell extracts. To this end, we demonstrate a model-based methodology for calibrating extract batches, and for using the calibrations to 'correct' the behavior of genetic circuits between batches. We also look at the interaction of this methodology with the phenomenon of parameter non-identifiability, which occurs when the parameter identification inverse problem has multiple solutions. In particular, we derive conditions under which parameter non-identifiability does not hinder our modeling objectives, and subsequently demonstrate the use of such non-identifiable models in performing data variability reduction.

Next, we describe `txtlsim`, a MATLAB® Simbiology® based toolbox for automatically generating models of genetic circuits in TX-TL, and for using these models for part characterization and circuit behavior prediction. Large genetic circuits can have non-negligible

resource usage needs, leading to unintended interactions between circuit nodes arising due to the loading of cellular machinery, transcription factors or other regulatory elements. The usage of consumable resources like nucleotides and amino acids can also have non-trivial effects on complex genetic circuits. These types of effects are handled by the modeling framework of `txtlsim` in a natural way.

We also highlight mcmc_simbio, a smaller toolbox within `txtlsim` for performing concurrent Bayesian parameter inference on Simbiology® models. Concurrent inference here means that a common set of parameters can be identified using data from an ensemble of different circuits and experiments, with each experiment informing a subset of the parameters. The combination of the concurrence feature with the fact that Markov chain Monte Carlo (MCMC) based Bayesian inference methods allow for the direct visualization of parameter non-identifiability enables the design of ensembles of experiments that reduce such non-identifiability.

Finally, we end with a method for performing model order reduction on transcription and translation elongation models while maintaining the ability of these models to track resource consumption. We show that due to their network topology, our models cannot be brought into the two-timescale form of singular perturbation theory when written in species concentration coordinates. We identify a coordinate system in which singular perturbation theory may be applied to chemical reaction networks more naturally, and use this to achieve the desired model reduction.

Published Content and Contributions

X. F. Meng, A.-A. Baetica, V. Singhal, and R. M. Murray. (2017). Recursively constructing analytic expressions for equilibrium distributions of stochastic biochemical reaction networks. *J. R. Soc. Interface* 2017 14 20170157; *10.1098/rsif.2017.0157*. AB and VS conceived the project. XFM, AB and VS performed the mathematical analysis. XFM did the computational experiments. All authors wrote and edited the manuscript.

Z. Sun, J. Kim, V. Singhal, R. M. Murray. (2015). Protein degradation in a TX-TL cell-free expression system using ClpXP protease. *bioRxiv* 019695; *10.1101/019695*. ZS and JK did the experiments. VS and JK did the mathematical and computational modeling. ZS wrote the manuscript, and all authors edited it.

M. Takahashi, J. Chappell, C. A. Hayes, Z. Z. Sun, J. Kim, V. Singhal, K. Spring, S. Al-Khabouri, C. P. Fall, V. Noireaux, R. M. Murray, J. B. Lucks. (2014). Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription–Translation (TX-TL) Systems. *ACS Synthetic Biology*. 4. *10.1021/sb400206c*. MT and JBL conceived the project, and intitial experiments were done by VS, KS, SAK and CP under the guidance of MT and JBL at the first Cold Spring Harbor Laboratory synthetic Biology summer course. Subsequent experiments were performed by MT with help from JC, CH, ZS and JK. MT wrote the manuscript, with input from all the authors.

Z. Tuza, V. Singhal, J. Kim, and R. M. Murray. (2013). An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular "breadboard" system. *Proceedings of the IEEE Conference on Decision and Control*. 1404-1410. *10.1109/CDC.2013.6760079*. RMM conceived the project. The software was written by ZT and VS. The experiments were performed by ZT and JK. The manuscript was written by VS and ZT, and edited by all the authors.

Contents