Computational modeling and quantitative analysis of an *in vitro* DNA transcriptional switch system based on experimental data

Kenneth L. Ho

Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125

Introduction

Rationally designed circuits in synthetic biology are difficult to characterize and often exhibit unexpected departures from theoretically predicted behaviors. Although researchers have made significant progress toward a qualitative understanding of such phenomena in recent years, an effective quantitative description, on which the precise engineering of genetic networks must ultimately rely, remained largely lacking. Thus, the natural question arose: Can effective quantitative descriptions of biological systems be derived?

We consider a bistable *in vitro* biochemical feedback circuit of two logical transcriptional inverters constructed in the Winfree Lab. Hypothesized biomolecular interactions from previous qualitative work fueled the development of a quantitative approach via a computational kinetic model of the system dynamics. The timecourse of the modeled system output in conjunction with sample experimental data provided the means for model quality assessment.

Objectives

The primary objective of this project was to produce an effective model that is:

(1) Capable of explaining the experimental data to within a normalized mean squared error (MSE) of 0.01 (roughly 10% of the signal span); and

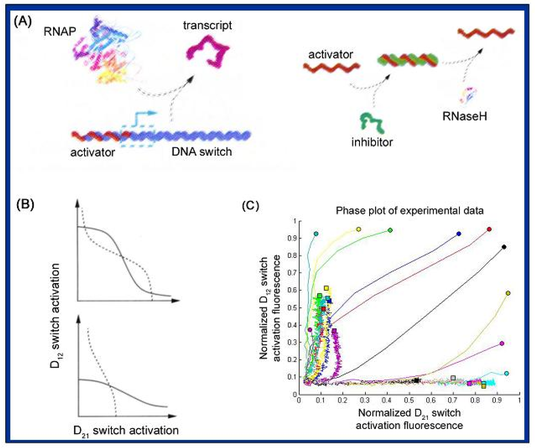
(2) Predictive in the sense that it is capable of generating output which matches experimental observations at the MSE = 0.01 level for out-of-sample system initial conditions.

In addition, we pursued the derivation of a reduced-order model of the system dynamics, which therefore simplifies the governing differential equations, but retains the essential character of the full-order model.

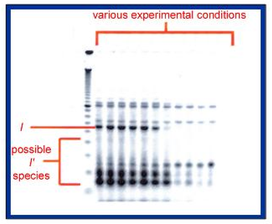
Approach

We began by enumerating a minimal set of system reactions and then used mass action to derive a minimal differential equation model for the system dynamics. Previous qualitative work highlighted the importance of enzyme competition; direct experimental observation suggested the prevalence of abortive transcription and partial degradation; and indirect experimental results inspired the possibility of RNA polymerase (RNAP) initial burst. We successively incorporated each of these three effects into the model and measured their individual contributions to simulation error reduction.

All models considered were deterministic and continuous implementations; the stochastic and discrete nature of biomolecular interactions were not explored. Simulations were implemented in MATLAB.



***(A)** DNA transcriptional inverter. DNA switch requires activator, which can be negated by inhibitor, for transcript synthesis. As a function of inhibitor with the transcript as signal, the switch is thus an inverter. **(B)** Hypothetical steady states of the bistable switch system. Solid and dashed curves trace out nullclines of each switch. Intersections mark steady states. Top/bottom, high/low RNAP concentration. **(C)** Phase plot of experimental data. Circles denote initial conditions; square denote final states. Steady states observed in narrow strips of the dominant switch.*

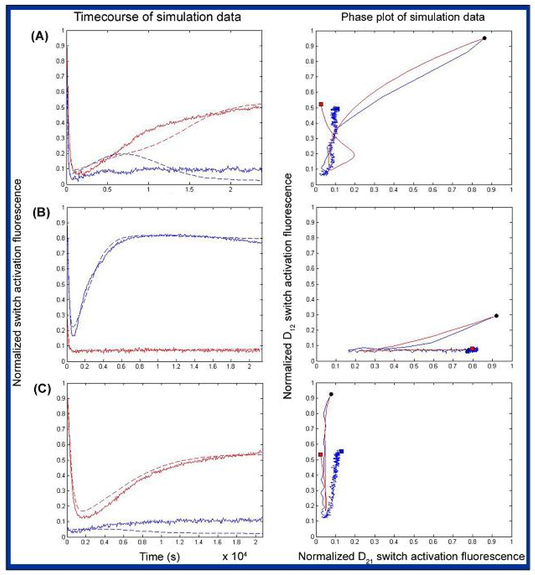


Electrophoresis gel showing the nucleic acids resulting from experimental runs of the bistable switch system. The dark bands near the bottom indicate the existence of significant amounts of partially transcribed or degraded inhibitor, with a global length preference consistent across all lanes.

Results

Least-squares parameter fittings with reasonable bounds showed the potential of each model combination of the three hypothesized effects. The minimum MSE of 0.003843 was achieved by the full model incorporating all three effects. This is well below the established threshold of MSE = 0.01. Thus, this model is capable of explaining the experimental data reasonably well.

Expectedly, enzyme competition was the dominant effect in the system, capturing the general character of the data and by itself accounting for 52% of the error drop



*Sample plots of simulation versus experimental data for the original data set using the full model. On the timecourse plots, solid lines represent experimental data and dashed lines represent simulation data; colors denote different switches. On the phase plots, blue represents experimental data and red represents simulation data. A MSE of 0.003843 was achieved. **(A)**, **(B)**, and **(C)** refer to different experimental runs.*

Enzyme competition	Abortive transcription and partial degradation	RNAP initial burst	MSE
Disabled	Disabled	Disabled	6.166×10^{-3}
Enabled	Disabled	Disabled	4.949×10^{-3}
Disabled	Enabled	Disabled	1.067×10^{-2}
Disabled	Disabled	Enabled	5.644×10^{-3}
Enabled	Enabled	Disabled	4.719×10^{-3}
Enabled	Disabled	Enabled	3.923×10^{-3}
Disabled	Enabled	Enabled	5.732×10^{-3}
Enabled	Enabled	Enabled	3.843×10^{-3}

Errors produced by simulations under models incorporating various combinations of enzyme competition, abortive transcription and partial degradation, and RNAP initial burst. These errors characterize the explanatory capacity of each model as the parameters used were obtained through least-squares parameter fittings against experimental data.

from the minimal to the full model. RNAP initial burst is second in importance, accurately tracing the starting trajectories though failing on intermediate time scales (22% error drop). Together, enzyme competition and RNAP initial burst account for 96% of the final error drop.

When tested against a new data set, however, the model (with parameters trained on the original data set) failed to accurately track the switch behavior at long time scales. Minor adjustments showed that the model was quite sensitive in this region of parameter space.

To establish possible explanatory power, a final parameter fit was run on both data sets, producing MSEs of 0.005748 and 0.003536 for the first and second sets, respectively. These are both below the MSE threshold and thus indicate that both data sets can, in principle, be effectively described by a single model. Simulations of the resulting model showed that it operates on the correct time scales and accurately predicts the general character of both data sets very well, falling under expectations only when the initial states of the system demand a high level of interswitch competition.

Conclusion

Our results show that the effects of enzyme competition, abortive transcription and partial degradation, and RNAP initial burst interact together to significantly influence the behavior of the bistable switch system. The dramatic reduction in simulation error exhibited demonstrates the explanatory power obtained by their incorporation. Given the general contexts in which these effects appear, it is likely that many such synthetic circuits rely on their contribution to produce effective quantitative descriptions of their underlying system dynamics. Thus, we postulate that future engineering of genetic networks will increasingly view such effects as undeniably integral aspects of overall system operation. As synthetic biology becomes increasingly infused with strict engineering principles, the quantitative models thus enabled will prove to be invaluable tools to rational design.

Acknowledgements

We thank Richard Murray, Erik Winfree, and Jongmin Kim for mentorship, and the Caltech SURF program for the opportunity. This work was supported by SURF.