

A finite model for the random behavior in the lactose regulation system of *Escherichia coli*

Agung Julius, Ádám Halász, Vijay Kumar and George Pappas, GRASP Lab, University of Pennsylvania, Philadelphia, USA

Introduction

Induction in the lactose uptake network has been described in the 1950's in the pioneering work of Monod et al [1] who demonstrated hysteresis in the synthesis of the enzyme β -galactosidase on the level of a cell population. Novick and Weiner [2] discovered the underlying bistability on the level of individual cells, which they termed the "all-or-none" phenomenon.

Mathematical models of the lactose uptake network also date back to the 1950's. Initially model building focused on the dynamical system arising from the ordinary differential equations describing the rates of relevant molecular processes. In this approach, the rate constants and other kinetic parameters are obtained from traditional "macroscopic" biochemical experiments designed to measure a subset of the model parameters. The work of Yildirim and Mackey [3] is an example of such a model which is also consistent with a number of classic experimental results on induction.

These ordinary differential equation (ODE) models are meant to describe the population of many cells as a single reactor, not addressing the behavior of individual cells. This approach is not necessarily flawed, but it is not at all obvious that the behavior of a large number of small cells is quantitatively, or even qualitatively, similar to that of a single large reactor that occupies the aggregate volume.

Current experimental methods can directly investigate stochastic switching on the level of individual cells. The work of Ozbudak et al [4] and Mettetal et al [5] provides the time evolution of the entire distribution of *lac* expression levels during transitions. These results are fitted with semi-empirical models which do not yet make the full connection with molecular biological mechanisms and independently determined kinetic parameters.

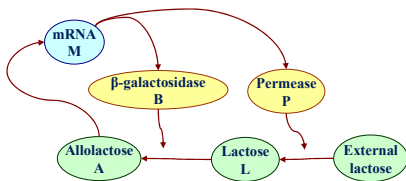
Constructing a biochemically sound model which is simultaneously consistent with the microscopic results of [4, 5] as well as with macroscopic experiments such as [2] is a challenging task. One difficulty is that the macroscopic model prediction which makes the connection with experiments is an aggregate of many individual stochastic single cell simulations. This makes parameter fitting approaches prohibitively costly.

Our aim is to provide a simple mathematical abstraction for the behavior of a microscopic stochastic model of the lactose uptake network. The parameters of this simplified model can serve as an intermediate step between the full stochastic model and the observed macroscopic population-level behavior.

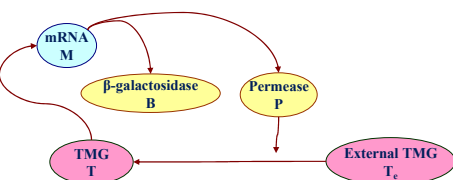
Future work on understanding the dependence of the transition rates and the variance of the induction states on the underlying full model parameters may help in the derivation of a new generation of dynamic stochastic models that can reproduce the observed kinetics from first principles. Another possible use of the simplified model is in the design of engineering applications where the control of aggregate cell behavior is of interest.

Background

The main elements of the lactose uptake network are encoded by two of the genes composing the *lac* operon: *lacY* codes for lactose permease which allows for the uptake of lactose and similar molecules, and *lacZ* codes for β -galactosidase which helps metabolize lactose into allolactose. Allolactose binds to LacI, a negative transcriptional regulator of the *lac* operon, exerting a positive effect on its expression. The positive feedback loop thus formed is the source of the observed bistable behavior.



Since lactose serves as an energy source, its introduction into the cell upon induction increases the growth rate. For this reason, experiments are often carried out using gratuitous inducers such as thiomethyl- β -D-galactoside (TMG). Gratuitous inducers are molecules similar to lactose that can not be metabolized, but which can be taken up by permease and which can induce the *lac* operon. The TMG uptake network is therefore simpler but consists of the same elements.



Model

Our model equations are based on the model of Yildirim and Mackey [3], modified for the use of TMG as inducer.

$$\begin{aligned} \frac{dM}{dt} &= \alpha_M \frac{1 + K_1 (e^{-\mu T_M} T(t - \tau_M))^2}{K + K_1 (e^{-\mu T_M} T(t - \tau_M))^2} + \Gamma_0 - (\gamma_M + \mu)M \\ \frac{dB}{dt} &= \alpha_B e^{-\mu T_B} M(t - \tau_B) - (\gamma_B + \mu)B \\ \frac{dT}{dt} &= \alpha_L P \frac{T_e}{K_T + T_e} - \beta_L P \frac{T}{K_L + T} - (\gamma_T + \mu)T \\ \frac{dP}{dt} &= \alpha_P e^{-\mu(\tau_P + \tau_B)} M(t - \tau_P - \tau_B) - (\gamma_P + \mu)P \end{aligned}$$

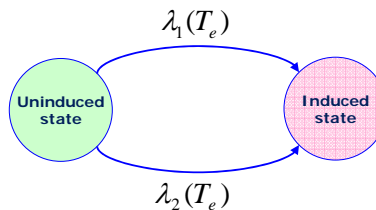
The ODE model variables are the internal TMG concentration T , the concentration of permease P , β -galactosidase B and of mRNA M . The external TMG concentration T_e can be thought of as an input to the system. The constant values are based on those of [3] but have been modified to give consistent behavior to the model in the limit of a large cell population.

To study the effect of stochasticity, we adopt a hybrid stochastic-deterministic model where only two of the model variables, namely the mRNA and β -galactosidase concentrations are replaced by discrete molecule numbers. The conversion from concentrations to molecule numbers per cell is through a constant, C_N . The evolution of the discretized variables is through two Poisson processes for each species, describing the creation and the destruction of the respective molecules.

$$\begin{aligned} dM_i &= d\tilde{M}_i - d\tilde{M}_i^- ; \quad M(t) = M_i / C_N \\ dB_i &= d\tilde{B}_i - d\tilde{B}_i^- ; \quad B(t) = B_i / C_N \\ \lambda_{\tilde{M}_i}(t) &= C_N \left[\alpha_M \frac{1 + K_1 (e^{-\mu T_M} T(t - \tau_M))^2}{K + K_1 (e^{-\mu T_M} T(t - \tau_M))^2} + \Gamma_0 \right] \\ \lambda_{\tilde{M}_i^-}(t) &= (\gamma_M + \mu)M_i \\ \lambda_{\tilde{B}_i}(t) &= \alpha_B e^{-\mu T_B} M_{i-\tau_B} \\ \lambda_{\tilde{B}_i^-}(t) &= (\gamma_B + \mu)B_i \end{aligned}$$

Two state abstraction

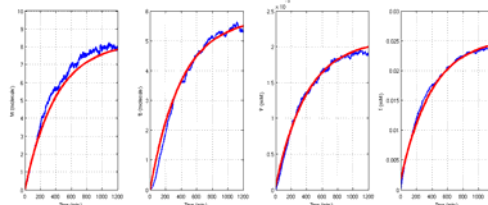
The stochastic model is quite complex and requires considerable computational resources if it is to be simulated for a large number of cells. We constructed a simple abstraction for the stochastic model in the form of **two state continuous time Markov chain**. The states of the Markov chain correspond to the low and high stable equilibria of the system, also known as the uninduced and induced states. The rates of switching between the two states are given as a function of the external TMG concentration.



Using a standard result in the analysis of continuous time Markov chains, we can derive the evolution of the probability densities at the states as:

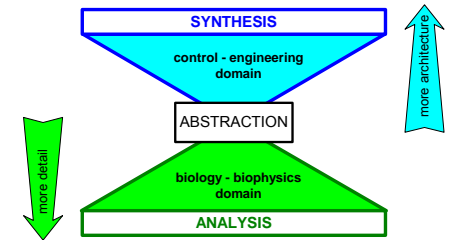
$$\frac{d}{dt} \begin{bmatrix} p_{lo} \\ p_{hi} \end{bmatrix} = \begin{bmatrix} -\lambda_1(T_e) & \lambda_2(T_e) \\ \lambda_1(T_e) & -\lambda_2(T_e) \end{bmatrix} \begin{bmatrix} p_{lo} \\ p_{hi} \end{bmatrix}$$

We use the abstract model to predict the average behavior of a colony of 1000 cells when it is exposed to the inducer TMG, and compare the result with that of the simulation of the full stochastic model of 1000 cells.

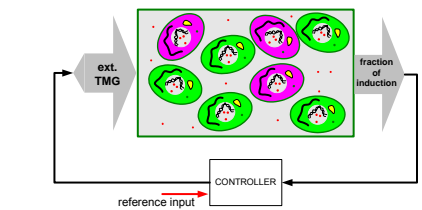


Control application

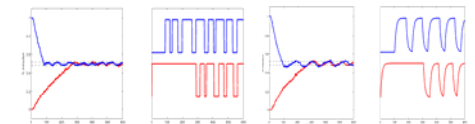
Since the abstract model is a good approximation for modeling the system on the macroscopic scale, we use it as a building block for modeling of a colony with a large number of cells, when we want to design control algorithm for the system. Thus, we can view the abstract model as a separator between the biology/biophysics domain and the engineering domain, where an engineer designing a control algorithm for the system can adopt the abstract model as the standing model of the control plant, without having to know about the biological/biophysical aspects of the system.



The control architecture that we consider is basically a feedback control with global measurement and global actuation, which is also known as broadcast feedback. The globally measured output, which is also the to-be-controlled quantity is the fraction of induced cells in the population. The control actuation is the external concentration of TMG. Thus we influence the system by influencing the transition rates between the induced and uninduced states.



We simulate two broadcast feedback control algorithms, with the goal of attaining and maintaining 50% level of induction.



We aim at formulating more advanced control algorithms, for example, optimal control, that are relevant to the biological system. Also, we are interested in learning the possibility of actually implementing a feedback control algorithm in the same organism.

Acknowledgements

Á. Halász is supported by a research grant from the Commonwealth of Pennsylvania and previously by an NIH-NLM Individual Biomedical Informatics Fellowship.

A. Julius and George Pappas are supported by the NSF Presidential Early CAREER (PECASE) Grant 0132716.

References

- [1] J. Monod, A. M. Pappenheimer, G. Cohen-Bazire, *Biochim. Biophys. Acta*, **9**, 648 (1952).
- [2] A. Novick and M. Weiner, *Proc. Natl. Acad. Sci. USA*, **43**, 553 (1957)
- [3] N. Yildirim and M. C. Mackey, *Biophys. J.* **84**, 2841 (2003)
- [4] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, A. van Oudenaarden, *Nature*, **427**, 737 (2004)
- [5] J. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, A. van Oudenaarden, *Proc. Natl. Acad. Sci. USA*, **103**, 7304 (2006)